
BOLLETTINO UNIONE MATEMATICA ITALIANA

CLAUDIO CANUTO

Adaptive convex optimization in Banach spaces: a multilevel approach

*Bollettino dell'Unione Matematica Italiana, Serie 8, Vol. 6-B (2003),
n.2, p. 263–287.*

Unione Matematica Italiana

http://www.bdim.eu/item?id=BUMI_2003_8_6B_2_263_0

L'utilizzo e la stampa di questo documento digitale è consentito liberamente per motivi di ricerca e studio. Non è consentito l'utilizzo dello stesso per motivi commerciali. Tutte le copie di questo documento devono riportare questo avvertimento.

*Articolo digitalizzato nel quadro del programma
bdim (Biblioteca Digitale Italiana di Matematica)
SIMAI & UMI*

<http://www.bdim.eu/>

Adaptive Convex Optimization in Banach Spaces: a Multilevel Approach (*)(**).

CLAUDIO CANUTO

Sunto. – *In questo articolo, a prevalente carattere di rassegna, si considerano varie applicazioni del concetto di Approssimazione Nonlineare alla minimizzazione convessa adattativa. Dapprima, si ricordano alcuni concetti di base e si confrontano l'approssimazione lineare e quella nonlineare nel caso di tre basi funzionali notevoli: la base di Fourier, le basi degli elementi finiti e le basi di ondine. Successivamente, indichiamo come l'approssimazione nonlineare possa essere usata nella definizione di metodi adattativi per la risoluzione di problemi di minimizzazione astratta in spazi di Banach. Gli algoritmi risultanti, che impiegano sia basi di ondine sia basi di elementi finiti, risultano rigorosamente giustificabili e con proprietà di ottimalità dal punto di vista dell'efficienza. In questo ambito, si descrive con un qualche dettaglio un algoritmo di «steepest-descent» per discretizzazioni in ondine.*

Summary. – *This is mainly a review paper, concerned with some applications of the concept of Nonlinear Approximation to adaptive convex minimization. At first, we recall the basic ideas and we compare linear to nonlinear approximation for three relevant families of bases used in practice: Fourier bases, finite element bases, wavelet bases. Next, we show how nonlinear approximation can be used to design rigorously justified and optimally efficient adaptive methods to solve abstract minimization problems in Banach spaces, using either wavelet or finite element bases. In particular, a wavelet adaptive steepest-descent algorithm is presented and investigated.*

1. – Introduction.

The purpose of this paper is to illustrate some of the ideas around the recently introduced concept of «Nonlinear Approximation» [29], as well as their application to the adaptive numerical discretization of certain variational problems. Nonlinear approximation means that the procedure which leads to ap-

(*) Il testo è una versione ampliata della conferenza tenuta a Pisa il 16 giugno 2002 in occasione del «First AMS-UMI Joint Meeting».

(**) Il lavoro è stato parzialmente finanziato dal Progetto di ricerca MURST Cofin 2000 «Calcolo scientifico: modelli e metodi numerici innovativi».

proximate a given function is intrinsically nonlinear, as it is adapted to the particular structure of the function. This process contrasts with the usual approach of linear approximation, based on the linear projection of the given function upon a linear manifold. Both kinds of approximation, when applied to the numerical treatment of variational problems, suffer from the difficulty that the function to be approximated is not directly accessible as, e.g., a signal or an image, but it is defined only implicitly by the problem.

The application we have in mind concerns the numerical discretization of an abstract minimization problem in a Banach space V . Given a functional $J: V \rightarrow \mathbb{R} \cup \{+\infty\}$, the problem consists of finding an element $u \in V$ such that

$$(1.1) \quad J(u) = \min_{v \in V} J(v).$$

It is well known that this formulation includes several situations of interest. Obviously, many problems in the Calculus of Variations have the form (1.1). In particular, the solution of an operator equation of the form

$$(1.2) \quad A(u) = f \quad \text{in } V'$$

can be reduced to (1.1) if $A: V \rightarrow V'$ is such that $A(v) - f = J'(v)$, the Fréchet (or Gateaux) derivative at v of a convex functional $J: V \rightarrow \mathbb{R}$. The perhaps most classical example is the homogeneous Dirichlet problem for the Laplace operator,

$$(1.3) \quad -\Delta u = f \quad \text{in } \Omega, \quad u = 0 \quad \text{on } \partial\Omega$$

(where Ω is a bounded domain in \mathbb{R}^n with sufficiently smooth boundary $\partial\Omega$ and $f \in L^2(\Omega)$), which corresponds to the minimization of the energy functional

$$(1.4) \quad J: H_0^1(\Omega) \rightarrow \mathbb{R}, \quad J(v) = \frac{1}{2} \int_{\Omega} |\nabla v|^2 - \int_{\Omega} f v.$$

Even if the equation of interest is not the Euler equation of a natural functional, it can be formulated variationally: it is enough to resort to a generalized least-square formulation, in which the functional to be minimized is defined as

$$J(v) = \|A(v) - f\|_{V'}.$$

We also mention that a constrained minimization problem for a functional $J_0: V \rightarrow \mathbb{R}$

$$J_0(u) = \min_{v \in K} J_0(v)$$

(such as a variational inequality) can be reduced to (1.1) by introducing the functional $J(v) = J_0(v) + I_K(v)$, where I_K denotes the indicator of the closed set $K \subseteq V$, defined as $I_K(v) = 0$ if $v \in V$, $I_K(v) = +\infty$ otherwise.

From now on, we will assume that Problem (1.1) has a unique solution (at least locally). This is certainly true if V is reflexive and J is strictly convex, lower-semicontinuous and unbounded as $\|v\| \rightarrow \infty$ (see, e.g., [13]).

A quite ambitious goal of any adaptive approximation strategy for Problem (1.1) is as follows.

GOAL 1. – *Given tolerances $\varepsilon_1 > 0$, $\varepsilon_2 > 0$, compute an approximation $u_\varepsilon \in V$ of u satisfying*

$$\|u - u_\varepsilon\|_V < \varepsilon_1 \quad \text{and} \quad J(u_\varepsilon) < J(u) + \varepsilon_2,$$

with the minimal computational complexity (memory occupancy, floating-point operations).

Unfortunately, often the complexity of the problem at hand is so huge, that such a goal cannot be achieved in practice, unless the specified tolerances are fairly large. This is due to memory and/or CPU time limitations. In these cases, a more realistic goal consists of optimizing the use of the available resources. For instance, if the most severe restrictions come from memory occupancy, one can set the following goal.

GOAL 2. – *Given N memory locations, compute an approximation $u_N \in V$ of u identified by the value of N real parameters (we will write $u_N \sim \mathbf{u} \in \mathbb{R}^N$), so that*

$$\|u - u_N\|_V \quad \text{and} \quad J(u_N) - J(u)$$

are as small as possible, at a cost at most proportional to N .

A similar goal can be set if CPU time is the dominant constraint. In the following, we will be mainly interested in satisfying Goal 2. The N parameters on which the approximation u_N depends are often called the *degrees of freedom* of the approximation.

Given this goal, the following central question in Approximation Theory arises: *how an element $v \in V$ can be accurately approximated by some $v_N \sim \mathbf{v} \in \mathbb{R}^N$?* The classical answer is based on a *linear* process, namely defining v_N as a linear function of v , and leads to a linear approximation theory. However, as we will see, this is not the most appropriate strategy to reach our goal: the best (or near-best) approximation depends *nonlinearly* on v , and a nonlinear approximation theory has to be developed. In the two coming sections, we will discuss linear and nonlinear approximations.

Throughout the paper, we will use the notation $A \lesssim B$ to indicate $A \leq cB$ for a suitable constant c independent of the relevant parameters in the formula. The notation $A \asymp B$ will mean $A \lesssim B$ and $B \lesssim A$.

2. – Linear approximation.

The classical approach consists of choosing a linear subspace $V_\delta \subset V$ of finite dimension N , and defining a linear projection $P_\delta: V \rightarrow V_\delta$ such that

$$\|v - P_\delta v\|_V \asymp \inf_{v_\delta \in V_\delta} \|v - v_\delta\|_V.$$

The question of approximation theory posed above takes here the following form: *how the quantity*

$$(2.1) \quad \|v - P_\delta v\|_V$$

depends on the dimension N of V_δ ? The question is relevant to our problem: indeed, if we define $u_\delta \in V_\delta$ by

$$J(u_\delta) = \min_{v_\delta \in V_\delta} J(v_\delta),$$

in many situations one has

$$(2.2) \quad \|u - u_\delta\|_V \asymp \inf_{v_\delta \in V_\delta} \|u - v_\delta\|_V.$$

A result of this type, sometimes referred to as Céa's Lemma [14], is often obtained by exploiting the *stability* of the numerical approximation, i.e., the fact that the norm of u_δ in V is bounded in terms of the data of the problem, uniformly in V_δ . *Stability* and *consistency*, i.e., the fact that (2.1) tends to 0 as V_δ saturates V , imply the *convergence* of the approximation, according to the well-known Lax-Richtmyer Theorem. Furthermore, (2.2) yields a so-called *a priori* error estimate for the approximation.

We will study the quantity (2.1) for three relevant approximation methods used in the numerical discretization of boundary value problems. We assume that Ω is a bounded domain in \mathbb{R}^d with Lipschitz boundary, and we denote by $H^s(\Omega) = W^{s,2}(\Omega) = B_{22}^s(\Omega)$ the Sobolev space of order s based on $L^2(\Omega)$ and equipped with the natural norm and seminorm. V will be a closed subspace of some $H^{s_0}(\Omega)$, defined by some boundary conditions.

2.1. Fourier approximation.

Set $\Omega = (0, 2\pi)^d$ and, for a given $s_0 \geq 0$, let $V = H_{\text{per}}^{s_0}(\Omega)$ be the closure in $H^{s_0}(\Omega)$ of the 2π -periodic functions of $C^\infty(\overline{\Omega})$. For any $M > 0$, let us define

$$V_M = \text{span} \{e^{ik \cdot x}: \|k\|_\infty \leq M\}, \quad P_M v = \sum_{\|k\|_\infty \leq M} \hat{v}_k e^{ik \cdot x}.$$

Then, if $v \in H_{\text{per}}^s(\Omega)$ for some $s \geq s_0$, one has the Jackson inequality (see, e.g., [9])

$$\|v - P_M v\|_{H^{s_0}(\Omega)} \lesssim M^{s_0 - s} |v|_{H^s(\Omega)}.$$

Let us set $N = \dim V_M$. Observing that $N \asymp M^d$, one obtains

$$\|v - P_M v\|_{H^{s_0}(\Omega)} \lesssim N^{-\Delta s/d} |v|_{H^s(\Omega)}, \quad \text{where } \Delta s = s - s_0.$$

Note that the projection $P_M v$ yields a *uniform* approximation of v , i.e., all Fourier modes satisfying the cut-off condition $\|k\|_\infty \leq M$ are equally represented, independently of which are the most significant for v . Also note that the only obstruction to the rate of decay of the error comes from the smoothness of v .

2.2. Finite element approximation.

Let Ω be a polygonal or polyhedral domain, decomposed into non-overlapping «elements» E (such as triangles, tetrahedrons, ...) of diameter $\asymp h$. For any integer $m \geq 0$, let us denote by $\mathcal{P}_m(E)$ the space of polynomials of degree $\leq m$ on E . Let $V = H_b^{s_0}(\Omega)$ be the closed subspace of some $H^{s_0}(\Omega)$ defined by some boundary conditions. We set

$$V_{h,m} = \{v \in V : v|_E \in \mathcal{P}_m(E), \forall E\}$$

and we denote by $P_{h,m} v$ an approximation of v in $V_{h,m}$, as defined e.g. by an orthogonal projection upon $V_{h,m}$, or by a suitable interpolation procedure. The typical approximation result is as follows: if $v \in H^s(\Omega)$ for some $s \geq s_0$, then

$$\|v - P_{h,m} v\|_{H^{s_0}(\Omega)} \lesssim \frac{h^{\min(s, m+1) - s_0}}{(m+1)^{s-s_0}} |v|_{H^s(\Omega)}$$

(see, e.g., [14, 9, 2, 35]).

In order to express the result in terms of $N = \dim V_{h,m}$, we note that $\text{card}\{E\} \asymp h^{-d}$, whereas $\dim \mathcal{P}_m(E) \asymp (m+1)^d$. Thus, we obtain $N \asymp \left(\frac{m+1}{h}\right)^d$, i.e., if $m+1 \geq s$,

$$\|v - P_{h,m} v\|_{H^{s_0}(\Omega)} \lesssim N^{-\Delta s/d} |v|_{H^s(\Omega)}.$$

We have again a uniform approximation, since Ω is partitioned into elements of comparable size, each one carrying a space of polynomials of the same degree. We now have two obstructions to the rate of decay of the error, one coming as above from the smoothness of v , the other one from the bound $s \leq m+1$.

2.3. Wavelet approximation.

Wavelets have emerged in the last decade as a powerful tool in signal and image processing, and their importance is increasing also in the numerical treatment of various operator equations, such as partial differential and integral equations; we refer to [28, 15] for a comprehensive presentation.

In view of the applications we have in mind, we will consider compactly supported, biorthogonal wavelet bases in $L^2(\Omega)$. Precisely, we assume that we are given two bases in $L^2(\Omega)$, $\{\psi_\lambda\}_{\lambda \in \mathcal{N}}$ and $\{\tilde{\psi}_\lambda\}_{\lambda \in \mathcal{N}}$. The notation λ stands for (j, k) , where j is the scale index and k is the position index: the wavelets ψ_λ and $\tilde{\psi}_\lambda$ have a support of diameter $\asymp 2^{-j}$ located around a point in $\bar{\Omega}$ which depends on k . The index set has the structure $\mathcal{N} = \{\lambda = (j, k) \mid j \geq j_0, k \in \mathcal{X}_j\}$, where \mathcal{X}_j are suitable index sets, which satisfy $\text{card } \mathcal{X}_j \asymp 2^{dj}$. It is convenient to set $|\lambda| = j$.

The bases satisfy the biorthogonality conditions

$$(\psi_\lambda, \tilde{\psi}_\mu)_{L^2(\Omega)} = \delta_{\lambda, \mu}, \quad \forall \lambda, \mu \in \mathcal{N},$$

as well as the representation property

$$(2.3) \quad \|v\|_{L^2(\Omega)}^2 \asymp \sum_{\lambda \in \mathcal{N}} |\hat{v}_\lambda|^2 \quad \text{for all } v = \sum_{\lambda \in \mathcal{N}} \hat{v}_\lambda \psi_\lambda \in L^2(\Omega).$$

Furthermore, we assume that the wavelets form a Riesz basis in a scale of Sobolev spaces; precisely, we assume that the condition

$$(2.4) \quad \|v\|_{H^s(\Omega)}^2 \asymp \sum_{\lambda \in \mathcal{N}} 2^{2|\lambda|s} |\hat{v}_\lambda|^2 \quad \text{iff } v \in H^s(\Omega)$$

holds for all s in an interval $(-s_*, s^*)$ depending on the smoothness and the number of vanishing moments of the generating wavelets.

The prototype of the wavelets of interest is the Haar wavelet [33]. On the real line, the most popular families are the Daubechies orthogonal wavelets [28] and the Cohen, Daubechies and Feauveau biorthogonal spline wavelets [19]. These bases can be restricted to the unit interval and suitably adapted to form biorthogonal bases in $L^2(0, 1)$ [20, 25, 32], with the possibility of enforcing vanishing conditions at the endpoints of the interval (clearly in view of their use for the solution of boundary value problems). Wavelets on cartesian products of intervals are built in an obvious manner by tensor products. Several wavelet families have been constructed on less trivial domains, using domain decomposition and patching [10, 26] or domain decomposition and prolongation [27]. In particular, we highlight the Wavelet Element Method [10, 11], which provides an explicit construction of matched wavelets across subdomains. The corresponding software `WemLib` is available at PoliTo [3].

A natural way to define a wavelet-based linear approximation method is to

introduce, for any integer $J \geq j_0$, the linear subspace of $L^2(0, 1)$

$$V_J = \text{span} \{ \psi_\lambda : |\lambda| \leq J \}$$

and the biorthogonal projection operator $P_J: L^2(0, 1) \rightarrow V_J$ defined by

$$P_J v = \sum_{|\lambda| \leq J} \hat{v}_\lambda \psi_\lambda.$$

Then, for any s_0 and s satisfying $-s_* < s_0 \leq s < s^*$, (2.4) yields the Jackson inequality

$$\|v - P_J v\|_{H^{s_0}(\Omega)} \lesssim 2^{J(s_0 - s)} |v|_{H^s(\Omega)}.$$

If $\Omega \subset \mathbb{R}^d$ is bounded, then $N \asymp 2^{Jd}$, whence again

$$(2.5) \quad \|v - P_J v\|_{H^{s_0}(\Omega)} \lesssim N^{-\Delta s/d} |v|_{H^s(\Omega)}.$$

As for the finite element approximation, the obstructions to the rate of decay of the error come from the smoothness of v and the range of validity of (2.4). The basis functions used in this approximation are again *uniformly distributed* in Ω , independently of where the function to be approximated has stronger or weaker gradients or singularities. It may happen that some (or perhaps many) of the basis functions bring a very small or even a null contribution to the quality of the approximation, since their support is located in a portion of the domain where v is very smooth. The aim of Nonlinear Approximation is to select only the most relevant contributions to the approximation.

3. – Nonlinear approximation.

In this section, we present the basic concepts of Nonlinear Approximation Theory. We follow the beautiful presentations by Ron DeVore [29] and Albert Cohen [15], to which we address the reader for further material. For the sake of definiteness, we confine ourselves to wavelet approximations, as introduced above. Similar concepts can be elaborated for, e.g., free-partition splines or finite elements, although at a less developed stage.

Let us introduce the following notation. Given any finite index set $\mathcal{A} \subset \mathcal{N}$, we define the subspace of V

$$V_{\mathcal{A}} = \text{span} \{ \psi_\lambda \mid \lambda \in \mathcal{A} \}$$

and the (biorthogonal) projection upon $V_{\mathcal{A}}$

$$\mathbf{P}_{\mathcal{A}} v = \sum_{\lambda \in \mathcal{A}} \hat{v}_\lambda \psi_\lambda.$$

If $\mathcal{A} = \{ \lambda \in \mathcal{N} : |\lambda| \leq J \}$, we obtain the linear approximation discussed

above. On the other hand, if \mathcal{A} depends on v , we obtain a nonlinear operator. We focus on three possible choices for \mathcal{A} .

[i] (Near) best N -term approximation. Let us order the wavelet coefficients of v in decreasing order of modulus

$$|\widehat{v}_{\lambda_1}| \geq \dots \geq |\widehat{v}_{\lambda_n}| \geq |\widehat{v}_{\lambda_{n+1}}| \geq \dots$$

For any given integer $N > 0$, set $\mathcal{A}_B(N, v) = \{\lambda_n : 1 \leq n \leq N\}$ and define

$$P_N(v) = \mathbf{P}_{\mathcal{A}_B(N, v)} v = \sum_{n=1}^N \widehat{v}_{\lambda_n} \psi_{\lambda_n}.$$

The approximation $P_N(v)$ is defined via the so-called Pure Greedy Algorithm, which picks the N largest in absolute value wavelet coefficients of the function v . Another way to look at P_N is as follows. Let us define the *manifold*

$$S_N = \left\{ \sum_{\lambda \in \mathcal{A}} \widehat{v}_{\lambda} \psi_{\lambda} : \text{card } \mathcal{A} \leq N \right\}$$

of the linear combinations of wavelets having at most N terms. Obviously, such a set is not a linear space, since $S_N + S_N \not\subseteq S_N$; however, we note that the following properties hold:

$$S_N + S_N \subseteq S_{2N} \quad \text{and} \quad aS_N = S_N, \quad \forall a \in \mathbb{R} \setminus \{0\}.$$

A *best N -term approximation* of v is any element $v_N \in S_N$ satisfying

$$\|v - v_N\|_{L^2(\Omega)} = \inf_{s_N \in S_N} \|v - s_N\|_{L^2(\Omega)} =: \sigma_N(v)$$

(note that there might be several elements v_N satisfying this property). Recalling (2.3), we immediately have

$$\|v - P_N(v)\|_{L^2(\Omega)} \asymp \sigma_N(v),$$

which justifies the term «near best N -term approximation».

[ii] (Absolute) Thresholding. For any given $\varepsilon > 0$, set $\mathcal{A}_T(\varepsilon, v) = \{\lambda \in \mathcal{M} : |\widehat{v}_{\lambda}| \geq \varepsilon\}$ and define the thresholding operator

$$P_{\varepsilon}(v) = \mathbf{P}_{\mathcal{A}_T(\varepsilon, v)} v = \sum_{|\widehat{v}_{\lambda}| \geq \varepsilon} \widehat{v}_{\lambda} \psi_{\lambda}.$$

[iii] Target Accuracy Approximation. For any given $\eta > 0$, denote by \mathcal{A}_{η} any subset of \mathcal{A} such that

$$\sum_{\lambda \notin \mathcal{A}_{\eta}} |\widehat{v}_{\lambda}|^2 \leq \eta^2$$

and let $\mathcal{A}_A(\eta, v)$ be any \mathcal{A}_{η} of minimal cardinality. Let us define the target ac-

curacy operator

$$P_\eta(v) = \mathbf{P}_{A^{A(\eta, v)}} v,$$

for which we have

$$\|v - P_\eta(v)\|_{L^2(\Omega)} \lesssim \eta.$$

The three nonlinear operators defined above are intimately related to each other, as we shall see below. Their theoretical properties have been intensively studied by R. DeVore and his collaborators (see [29] and the references therein). The rest of this section will be mostly devoted to investigate the behaviour of the approximation error in the L^2 -norm, expressed in terms of one of the approximation parameters N , ε or η and of a suitable «norm» of the function v .

3.1. Best N -term approximation.

Let us denote by $\mathbf{v} = \{\widehat{v}_\lambda\}_{\lambda \in \mathcal{J}\mathcal{I}}$ the sequence of wavelet coefficients of v ; it belongs to $l^2(\mathcal{J}\mathcal{I})$. We are going to define subspaces of $l^2(\mathcal{J}\mathcal{I})$ by looking at the decay rate of the sequence components. To this end, let us denote by $\{v_n = \widehat{v}_{\lambda_n}\}_{n \in \mathbb{N}}$ the decreasing rearrangement of the sequence \mathbf{v} , already introduced above, and let us observe that

$$\mathbf{v} \in l^2(\mathcal{J}\mathcal{I}) \Rightarrow n^{1/2} |v_n| \leq \|\mathbf{v}\|_{l^2(\mathcal{J}\mathcal{I})}, \quad \forall n \in \mathbb{N}.$$

Thus, for any real τ such that $0 < \tau < 2$, it is natural to consider the subspace $l_w^\tau(\mathcal{J}\mathcal{I})$ of the sequences such that

$$(3.1) \quad |\mathbf{v}|_{l_w^\tau(\mathcal{J}\mathcal{I})} := \sup_{n \in \mathbb{N}} n^{1/\tau} |v_n| < +\infty.$$

The decreasing rearrangement of a sequence $\mathbf{v} \in l_w^\tau(\mathcal{J}\mathcal{I})$ decays faster and faster as τ approaches 0; this means that \mathbf{v} can be described by a smaller and smaller number of significant components. Thus, we can consider $|\mathbf{v}|_{l_w^\tau(\mathcal{J}\mathcal{I})}$ as a measure of the *sparsity* of \mathbf{v} . We define

$$\|\mathbf{v}\|_{l_w^\tau(\mathcal{J}\mathcal{I})} := \|\mathbf{v}\|_{l^2(\mathcal{J}\mathcal{I})} + |\mathbf{v}|_{l_w^\tau(\mathcal{J}\mathcal{I})},$$

which is a (quasi-) norm for $l_w^\tau(\mathcal{J}\mathcal{I})$. It is easy to derive an estimate of the best N -term approximation error for a function v such that $\mathbf{v} \in l_w^\tau(\mathcal{J}\mathcal{I})$. Indeed,

$$\sum_{n > N} |v_n|^2 = \sum_{n > N} n^{-2/\tau} n^{2/\tau} |v_n|^2 = \left(\sum_{n > N} n^{-2/\tau} \right) |\mathbf{v}|_{l_w^\tau(\mathcal{J}\mathcal{I})}^2 \lesssim N^{1-2/\tau} |\mathbf{v}|_{l_w^\tau(\mathcal{J}\mathcal{I})}^2.$$

Introducing the quantity $s > 0$ by the relation

$$(3.2) \quad \frac{s}{d} = \frac{1}{\tau} - \frac{1}{2}$$

(such a definition will be justified below), we conclude that

$$(3.3) \quad \sigma_N(v) \asymp \|v - P_N(v)\|_{L^2(\Omega)} \lesssim N^{-s/d} |\mathbf{v}|_{l_w^\tau(\mathcal{N})}.$$

It is remarkable that the rate of decay $N^{-s/d}$ of the best N -term approximation error of v is *equivalent* to the condition $\mathbf{v} \in l_w^\tau(\mathcal{N})$. Indeed, for all N , we have

$$N|v_{2N}|^2 \leq \sum_{N < n \leq 2N} |v_n|^2 \leq \sigma_N(v)^2 \leq CN^{-2s/d},$$

which implies

$$|v_{2N}| \leq CN^{-\left(\frac{s}{d} + \frac{1}{2}\right)} = CN^{-\frac{1}{\tau}},$$

from which it is easily derived $\mathbf{v} \in l_w^\tau(\mathcal{N})$. Summarizing, we have

PROPOSITION 1. – *Let s and τ be related by (3.2). Then, $\sigma_N(v) = O(N^{-s/d})$ if and only if $\mathbf{v} \in l_w^\tau(\mathcal{N})$, and*

$$\sigma_N(v) \lesssim N^{-s/d} |\mathbf{v}|_{l_w^\tau(\mathcal{N})}.$$

In other words, the family of the so-called Lorentz spaces $l_w^\tau(\mathcal{N}) = l^{\tau, \infty}(\mathcal{N})$ is the natural scale in which to study the convergence of the best N -term approximation of a function v . We now investigate in more detail the nature of these spaces, and how they are related to standard function spaces for v . At first, note that

$$\|\mathbf{v}\|_{l^{\tau'}(\mathcal{N})} \geq \sum_{n \leq N} |v_n|^{\tau'} \geq N|v_N|^{\tau'},$$

which implies $l^\tau(\mathcal{N}) \subset l_w^\tau(\mathcal{N})$. On the other hand, if $\tau' > \tau$

$$\sum_n |v_n|^{\tau'} = \sum_n n^{-\tau'/\tau} (n^{1/\tau} |v_n|)^{\tau'} \lesssim |\mathbf{v}|_{l_w^{\tau'}(\mathcal{N})}^{\tau'}$$

which implies $l_w^\tau(\mathcal{N}) \subset l^{\tau'}(\mathcal{N})$. It is easily seen that all these inclusions are strict.

Next, we relate the condition $\mathbf{v} \in l^\tau(\mathcal{N})$ to the smoothness of v . To this end, we observe that, since $L^2(\Omega) \subset L^\tau(\Omega)$, the wavelet system $\{\psi_\lambda\}_{\lambda \in \mathcal{N}}$ allows the characterization of the family of Besov spaces $B_{\tau\tau}^s(\Omega)$ constructed on $L^\tau(\Omega)$. Precisely, there exists $s_\tau^* > 0$ depending on the smoothness of the generating wavelets in the scale $B_{\tau\tau}^s(\Omega)$ and on the number of their vanishing moments, such that one has

$$\|v\|_{B_{\tau\tau}^s(\Omega)} \asymp \left(\sum_{\lambda \in \mathcal{N}} 2^{s\tau|\lambda|} |\widehat{v}_\lambda^{(\tau)}|^\tau \right)^{1/\tau}, \quad \forall v \in B_{\tau\tau}^s(\Omega), \quad 0 \leq s < s_\tau^*.$$

Here, the wavelet coefficients are computed with respect to the basis normalized in $L^\tau(\Omega)$, i.e.,

$$\widehat{v}_\lambda^{(\tau)} = \int_\Omega v(x) \widetilde{\psi}_\lambda^{(\tau)}(x) dx$$

with

$$\tilde{\psi}_\lambda^{(r)}(x) = 2^{-\frac{d}{r}|\lambda|} \tilde{\psi}(2^{|\lambda|}x - k) = 2^{d(\frac{1}{2} - \frac{1}{r})|\lambda|} \tilde{\psi}_\lambda(x).$$

Hence, $\widehat{v}_\lambda^{(r)} = 2^{d(\frac{1}{2} - \frac{1}{r})|\lambda|} \widehat{v}_\lambda$ so that $\|v\|_{B_{rr}^s(\Omega)} \asymp \left(\sum_{\lambda \in \mathcal{M}} 2^{(sr + \frac{rd}{2} - d)|\lambda|} |\widehat{v}_\lambda|^\tau \right)^{1/\tau}$. Choosing s as in (3.2), we obtain

$$\|v\|_{B_{rr}^s(\Omega)} \asymp \left(\sum_{\lambda \in \mathcal{M}} |\widehat{v}_\lambda|^\tau \right)^{1/\tau} = \|v\|_{l^\tau(\mathcal{M})}.$$

The present result, combined with Proposition 1, yields the following Jackson estimate for a function $v \in B_{rr}^s(\Omega)$:

$$(3.4) \quad \|v - P_N(v)\|_{L^2(\Omega)} \lesssim N^{-s/d} \|v\|_{B_{rr}^s(\Omega)}.$$

Actually, one can prove the following analogue of Proposition 1: a function $v \in L^2(\Omega)$ belongs to $B_{rr}^s(\Omega)$ if and only if

$$\sum_N (N^{s/d} \sigma_N(v))^r N^{-1} < +\infty.$$

In order to grasp the intimate difference between linear and nonlinear approximation, one should compare the Jackson inequalities (2.5) (with $s_0 = 0$) and (3.4). The rate of decay of the approximation error is the same in both estimates, but the required smoothness of the function to be approximated is not. The function spaces $B_{pp}^s(\Omega)$ involved in the estimates can be pictorially represented as points $(1/p, s)$ in the plane (see Figure 1); moving right means decreasing summability, moving up means increasing differentiability. The fami-

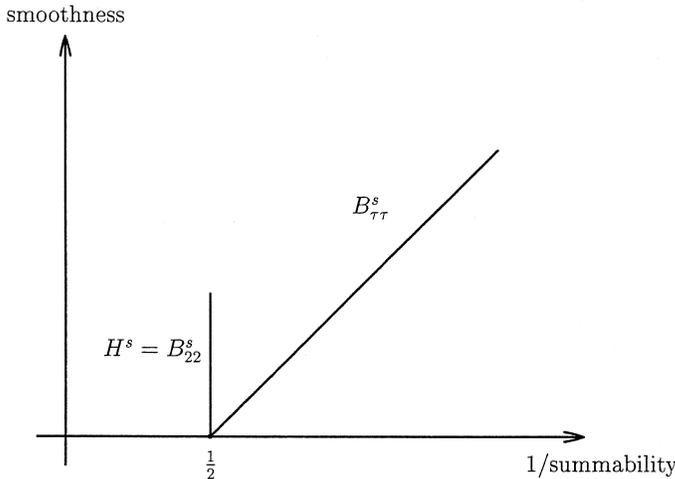


Figure 1. – DeVore diagram.

ly $H^s(\Omega) = B_{22}^s(\Omega)$ which naturally comes into play in linear approximation lies on the vertical line issuing from the point $(\frac{1}{2}, 0)$, whereas the family $B_{\tau\tau}^s(\Omega)$ which comes into play in nonlinear approximation lies on the oblique line, issuing from the same point with slope d . Note that this line corresponds to the critical exponent of the embedding of a Besov space $B_{pp}^s(\Omega)$ into $L^2(\Omega)$: spaces represented by points above this line are compactly embedded into $L^2(\Omega)$, spaces represented by points below this line are not contained in $L^2(\Omega)$. Also note that the spaces $B_{\tau\tau}^s(\Omega)$ are not contained in any Sobolev space $H^s(\Omega)$ for $s > 0$. The pictorial representation of the families $H^s(\Omega)$ and $B_{\tau\tau}^s(\Omega)$ in the $(1/p, s)$ -plane is known as the *DeVore diagram*.

For a fixed $s > 0$, the space $B_{\tau\tau}^s(\Omega)$ is larger than the space $H^s(\Omega) = B_{22}^s(\Omega)$; indeed, when we move on the oblique line, we require less and less summability to the «derivatives» of the functions. In terms of approximation, this means that the same rate of decay of the error can be achieved by nonlinear approximation for a wider set of functions than by linear approximation. Equivalently, for a fixed function to be approximated, it is likely that the rate of decay of the error will be higher with nonlinear approximation than with linear approximation, since the upper bound for the smoothness index s of the function will be higher along the oblique line than along the vertical line. In particular, no convergence rate can be expected for the linear approximation of functions which do not belong to a better space than $B_{\tau\tau}^s(\Omega)$.

As an example, set $\Omega = (-1, 1)$ and define

$$v_\alpha(x) = \begin{cases} 0 & x \leq 0, \\ x^\alpha & x > 0, \end{cases} \quad \alpha \in \mathbb{R}.$$

Then

$$v_\alpha \in B_{pp}^s(\Omega) \quad \text{iff} \quad \begin{cases} \frac{1}{p} > \max(-\alpha, 0), \\ s < \alpha + \frac{1}{p}. \end{cases}$$

Thus, $v_\alpha \in L^2(\Omega)$ iff $\alpha > -\frac{1}{2}$; in this case, we have $v_\alpha \in H^s(\Omega)$ for all $s < \alpha + \frac{1}{2}$, whereas $v_\alpha \in B_{\tau\tau}^s(\Omega)$ for all $s > 0$. Thus, the rate of decay of the best N -term approximation error of such a function is solely determined by the number of vanishing moments of the generating wavelets used in the approximation.

For a general domain in \mathbb{R}^d , the spaces $B_{\tau\tau}^s(\Omega)$ contain discontinuous functions for arbitrarily large values of s , whereas functions in $H^s(\Omega)$ are necessarily continuous for $s > \frac{d}{2}$.

Finally, it is of interest to relate the number N of retained wavelets to the threshold given by the minimal value $\varepsilon = \varepsilon_N$ of the absolute value of the retained wavelet coefficients, as well as to the accuracy $\eta = \eta_N$ guaranteed by the near best N -term approximation. To this end, we observe that by definition $\varepsilon_N = |v_N|$, whence $N^{1/\tau} \varepsilon_N = N^{1/\tau} |v_N| \leq |\mathbf{v}|_{l_w^\tau(\mathcal{N})}$, so that

$$\varepsilon_N \leq N^{-1/\tau} |\mathbf{v}|_{l_w^\tau(\mathcal{N})}.$$

On the other hand, estimate (3.3) yields

$$\eta_N \leq N^{-s/d} |\mathbf{v}|_{l_w^\tau(\mathcal{N})}.$$

3.2. Thresholding.

The Lorentz space $l_w^\tau(\mathcal{N})$, defined by (3.1) in terms of the decreasing rearrangement of a sequence, can be equivalently defined from the point of view of thresholding. Indeed, the following result holds.

PROPOSITION 2. – *A sequence $\mathbf{v} \in l^2(\mathcal{N})$ belongs to $l_w^\tau(\mathcal{N})$ if and only if, for all $\varepsilon > 0$,*

$$(3.5) \quad \text{card} \{ \lambda : |\hat{v}_\lambda| \geq \varepsilon \} \leq C\varepsilon^{-\tau}.$$

Moreover, the smallest value of C for which the previous estimate holds coincides with $|\mathbf{v}|_{l_w^\tau(\mathcal{N})}^\tau$.

PROOF. – Assume that \mathbf{v} belongs to $l_w^\tau(\mathcal{N})$ and set $N_\varepsilon = \text{card} \{ \lambda : |\hat{v}_\lambda| \geq \varepsilon \}$. We have $|v_{N_\varepsilon}| \geq \varepsilon$, so that

$$N_\varepsilon^{1/\tau} \varepsilon \leq N_\varepsilon^{1/\tau} |v_{N_\varepsilon}| \leq |\mathbf{v}|_{l_w^\tau(\mathcal{N})},$$

from which we obtain $N_\varepsilon \leq |\mathbf{v}|_{l_w^\tau(\mathcal{N})}^\tau \varepsilon^{-\tau}$. Conversely, assume that (3.5) hold. For any fixed $n \in \mathbb{N}$, set $\varepsilon := |v_n|$ and define $n_\varepsilon = \text{card} \{ \lambda : |\hat{v}_\lambda| \geq \varepsilon \}$. We have

$$n \leq n_\varepsilon \leq C\varepsilon^{-\tau},$$

which yields $n^{1/\tau} |v_n| \leq C^{1/\tau}$, i.e., \mathbf{v} belongs to $l_w^\tau(\mathcal{N})$.

Let us go back to the thresholding operator $P_\varepsilon(v)$ defined above. Denote again by $N_\varepsilon = \text{card} \mathcal{A}_T(\varepsilon, v)$ the number of retained wavelet components of v . Then, Proposition 2 tells us that

$$N_\varepsilon \leq \varepsilon^{-\tau} |\mathbf{v}|_{l_w^\tau(\mathcal{N})}^\tau.$$

Concerning the approximation error, we have

$$\begin{aligned} \|v - P_\varepsilon(v)\|_{L^2(\Omega)}^2 &\leq \sum_{k=0}^{\infty} 2^{-k-1} \varepsilon < |\hat{v}_\lambda| \leq 2^{-k} \varepsilon \quad |\hat{v}_\lambda|^2 \\ &\leq \sum_{k=0}^{\infty} (2^{-k} \varepsilon)^2 (2^{-k-1} \varepsilon)^{-\tau} |\mathbf{v}|_{l_w^\tau(\mathcal{D}\mathbb{R})}^\tau \leq \varepsilon^{2-\tau} |\mathbf{v}|_{l_w^\tau(\mathcal{D}\mathbb{R})}^\tau, \end{aligned}$$

which yields

$$(3.6) \quad \|v - P_\varepsilon(v)\|_{L^2(\Omega)} \leq \varepsilon^{1-\tau/2} |\mathbf{v}|_{l_w^\tau(\mathcal{D}\mathbb{R})}^{\tau/2}.$$

Equivalently, if η_ε denotes the accuracy guaranteed by the thresholding approximation, we have

$$\eta_\varepsilon \leq \varepsilon^{1-\tau/2} |\mathbf{v}|_{l_w^\tau(\mathcal{D}\mathbb{R})}^{\tau/2}.$$

3.3. Target accuracy.

Finally, we consider the target accuracy operator $P_\eta(v)$ which yields an approximation error of order η with the minimal number of wavelet components, say N_η . This means that $P_\eta(v) = P_{N_\eta}(v)$. According to the definition, we have

$$\sum_{n > N_\eta} |v_n|^2 \leq \eta^2, \quad \text{whereas} \quad \sum_{n > N_{\eta-1}} |v_n|^2 > \eta^2.$$

Thus, using (3.3), we get

$$\eta \leq \|v - P_{N_\eta-1}(v)\|_{L^2(\Omega)} \leq (N_\eta - 1)^{-s/d} |\mathbf{v}|_{l_w^\tau(\mathcal{D}\mathbb{R})}$$

whence we obtain

$$N_\eta \leq \eta^{-d/s} |\mathbf{v}|_{l_w^\tau(\mathcal{D}\mathbb{R})}^{d/s}.$$

Let us now denote by ε_η the minimal value of the absolute value of the retained wavelet coefficients of v . We have

$$\eta^2 < \sum_{n \geq N_\eta} |v_n|^2 = \sum_{n \geq N_\eta} |v_n|^\tau |v_n|^{2-\tau} \leq \sum_{n \geq N_\eta} |v_n|^\tau \varepsilon_\eta^{2-\tau} \leq \|\mathbf{v}\|_{l_w^\tau(\mathcal{D}\mathbb{R})}^\tau \varepsilon_\eta^{2-\tau},$$

whence we obtain

$$(3.7) \quad \varepsilon_\eta \geq \eta^{\frac{2}{2-\tau}} \|\mathbf{v}\|_{l_w^\tau(\mathcal{D}\mathbb{R})}^{-\frac{\tau}{2-\tau}}.$$

The last result means that an approximation error below a fixed accuracy η is guaranteed if we apply a thresholding procedure with a threshold given by the right-hand side of (3.7).

3.4. *Extensions.*

So far, we have investigated the nonlinear approximation error measured in the $L^2(\Omega)$ -norm. In the applications to data compression or to the numerical discretization of operator equations (boundary value problems for partial differential equations, integral equations, and so on), one may be interested in measuring the approximation error in some other norms, e.g., varying either the smoothness index or the summability index.

Suppose that we want to measure the approximation error in some Sobolev space $W^{s_0,p}(\Omega)$, where either $p=2$ and $s_0>0$ is arbitrary, or $p\in(1, +\infty)\setminus\{2\}$ and $s_0\notin\mathbb{N}$. In these cases, the space $W^{s_0,p}(\Omega)$ coincides with the Besov space $B_{pp}^{s_0}(\Omega)$. If we assume that our wavelet system allows the characterization of such a space, we have

$$\|v\|_{W^{s_0,p}(\Omega)}^p \asymp \sum_{\lambda} 2^{(s_0 p + \frac{pd}{2} - d)|\lambda|} |\widehat{v}_{\lambda}|^p.$$

If we set $\tilde{v}_{\lambda} := 2^{(s_0 + \frac{d}{2} - \frac{d}{p})|\lambda|} \widehat{v}_{\lambda}$ and $\tilde{v} = \{\tilde{v}_{\lambda}\}_{\lambda \in \mathcal{J}\mathbb{N}}$, we have the norm equivalence

$$\|v\|_{W^{s_0,p}(\Omega)} \asymp \|\tilde{v}\|_{l^p(\mathcal{J}\mathbb{N})},$$

so that we are back to the situation discussed above, i.e., we apply the nonlinear operators to the sequence \tilde{v} . For instance, if we denote by $\tilde{P}_N(v)$ the near best N -term approximation operator in the $W^{s_0,p}(\Omega)$ -norm, we have as in (3.3)

$$(3.8) \quad \inf_{s_N \in \mathcal{S}_N} \|v - s_N\|_{W^{s_0,p}(\Omega)} \asymp \|v - \tilde{P}_N(v)\|_{W^{s_0,p}(\Omega)} \leq N^{-\Delta s/d} |\tilde{v}|_{l_w^s(\mathcal{J}\mathbb{N})},$$

where now the quantity Δs is defined as

$$(3.9) \quad \frac{\Delta s}{d} = \frac{1}{\tau} - \frac{1}{p}.$$

Thus, setting $s = s_0 + \Delta s$, we obtain

$$(3.10) \quad \|v - \tilde{P}_N(v)\|_{W^{s_0,p}(\Omega)} \leq N^{-\Delta s/d} \|v\|_{B_{\tau\tau}^s(\Omega)}.$$

In other words, the rate of decay of the error $- \frac{\Delta s}{d}$ is achieved if we move the Besov space $B_{\tau\tau}^s(\Omega)$ along the oblique line issueing from the point $(\frac{1}{p}, s_0)$ with slope d . Similar results hold for the other nonlinear operators considered in the previous sections.

Finally, we consider the case in which we want to measure the approximation error in $L^p(\Omega)$ or in some Sobolev space $W^{m,p}(\Omega)$ with $m \in \mathbb{N}$, which for $p \neq 2$ is known not to coincide with $B_{pp}^m(\Omega)$. In this case, the analysis is by far more complex than above, since one cannot exploit any norm equivalence simi-

lar to (2.4). One has to resort to the Littlewood-Paley theory, using the so-called square function Sv defined as

$$Sv(x) = \left(\sum_{\lambda \in \mathfrak{M}} |\widehat{v}_\lambda|^2 |\psi_\lambda(x)|^2 \right)^{1/2},$$

for which the result $\|v\|_{L^p(\Omega)} \asymp \|Sv\|_{L^p(\Omega)}$ holds. However, the conclusions of the analysis are remarkably the same as before. Indeed, a result by Temlyakov [36] states that a near best N -term approximation in $L^p(\Omega)$ can be achieved by a simple thresholding procedure: precisely, let $\mathcal{A}(N, v)$ denote any set of indices λ such that $\|\widehat{v}_\lambda \psi_\lambda\|_{L^p(\Omega)}$ take the N largest values, and set $P_N(v) = P_{\mathcal{A}(N, v)}v$. Then, one has

$$\inf_{s_N \in S_N} \|v - s_N\|_{L^p(\Omega)} \asymp \|v - P_N(v)\|_{L^p(\Omega)},$$

from which an estimate like (3.10) can be obtained.

4. – Adaptivity.

In the two previous sections, we have discussed the problem of approximating a given function in a domain Ω . We have assumed that the function is *explicitly accessible*, in the sense that the complete knowledge of, e.g., the set of its values in Ω , or the sequence of its Fourier or wavelet coefficients, is available. Now, let us consider the more difficult situation, in which the function u to be approximated by a N -term expansion is only *implicitly accessible*, since it is defined as the solution of the minimization problem (1.1), or the operator equation (1.2). In this case, very often we can only generate approximations of u , by some discrete mechanism which mimics (1.1) or (1.2) in finite dimension. Adaptivity is accomplished by means of an iterative procedure, which incorporates best N -term approximation ideas as described in the previous section.

Let us denote by u_N^{discr} the computed N -term approximation of u produced by some discrete mechanism. We call the discretization procedure *asymptotically optimal* if, for $N \rightarrow \infty$, the function u_N^{discr} behaves as the best N -term approximation of u in the same basis, i.e.,

$$\|u - u_N^{\text{discr}}\|_V \asymp \sigma_N(u), \quad N \rightarrow \infty,$$

and the computational cost to obtain u_N^{discr} is at most proportional to N . If we use a wavelet basis, then by (3.3) such a behaviour of the error is equivalent to the bound

$$\|u - u_N^{\text{discr}}\|_V \lesssim N^{-\Delta s/d} |\mathbf{u}|_{L_w^r(\mathfrak{M})},$$

provided V is a space of smoothness index s_0 . In particular, optimality is

achieved if the error behaviour is described by the estimate (see (3.4))

$$\|u - u_N^{\text{discr}}\|_V \lesssim N^{-\Delta s/d} \|u\|_{B_{rr}^s(\Omega)}.$$

The key ingredients of the design of an optimal adaptive algorithm include:

1. The knowledge of regularity results for the solution of minimization problems or operator equations in the *diagonal* scale of Besov spaces $B_{rr}^s(\Omega)$. Existing results concern conservation laws [30], elliptic equations in Lipschitz and polygonal domains [24, 21], the Stokes equations [23].

2. The a posteriori error control, based on explicit relations between the current error $\|u - u_\delta\|_V$ and some dual norm of the residual, such as $\|J'(u_\delta)\|_{V'}$ in smooth minimization problems, or $\inf_{s \in \partial J(u_\delta)} \|s\|_{V'}$ (where $\partial J(u_\delta)$ denotes the subdifferential of J at u_δ) in convex, non-differentiable minimization problems. We note that dual norms are difficult to compute in practice, unless one uses multilevel or hierarchical bases; the difficulty may be circumvented by resorting to equivalent expressions, based on weighted local L^2 -norms.

3. A judiciously designed adaptation strategy, which incorporates in the iterative minimization process (or in the iterative solution of linear/nonlinear algebraic systems) some actions of mesh refinement/coarsening or Riesz basis enrichment/thresholding, which are grounded on and controlled by the underlying theoretical analysis.

4.1. Wavelet approximations.

Starting from the pioneering works [4, 22] (we also mention [8] for an early implementation of wavelet-based adaptivity for elliptic problems), theoretically sound adaptive wavelet strategies have been developed in the last few years for a wide class of differential and integral operators, including elliptic operators, mixed variational formulations, Calderon-Zygmund operators, and certain nonlinear operators. We mention the results by A. Cohen, W. Dahmen, R. DeVore and collaborators [16, 17, 23, 18] and those by S. Bertoluzza and collaborators [5, 6]. For the study of a class of convex minimization problems, see [12]. The analysis contained in these and other contributions proves the asymptotical optimality of the adaptive wavelet discretizations considered therein (possibly up to an extra logarithmic term in the computational complexity, and assuming an $O(1)$ -cost in the computation of each entry of the stiffness and mass matrices). Numerical results which provide a quantitative support to the theory can be found, e.g., in [1].

In order to illustrate the methodology, let us consider the minimization problem (1.1) for the Dirichlet functional (1.4). Following [12] and using the approach of [17], we study an adaptive steepest-descent algorithm. We assume

that the wavelet basis $\{\psi_\lambda\}_{\lambda \in \mathcal{N}}$ is a Riesz basis in $V = H_0^1(\Omega)$; consequently, the biorthogonal basis $\{\tilde{\psi}_\lambda\}_{\lambda \in \mathcal{N}}$ is a Riesz basis in $V' = H^{-1}(\Omega)$. This means that (2.4) holds with $s = 1$ and $H^s(\Omega)$ replaced by $H_0^1(\Omega)$, whereas the dual space $H^{-1}(\Omega)$ is characterized as follows:

$$\|F\|_{H^{-1}(\Omega)}^2 \asymp \sum_{\lambda \in \mathcal{N}} 2^{-2|\lambda|s} |\widehat{F}_\lambda|^2 \quad \text{iff } F \in H^{-1}(\Omega),$$

with $\widehat{F}_\lambda = \langle F, \psi_\lambda \rangle$, the duality pairing between F and ψ_λ in $H_0^1(\Omega)$. In the present situation, it is customary to normalize primal wavelets in the $H_0^1(\Omega)$ -norm and dual wavelets in the $H^{-1}(\Omega)$ -norm, i.e., replace each ψ_λ by $2^{-|\lambda|} \psi_\lambda$ and each $\tilde{\psi}_\lambda$ by $2^{|\lambda|} \tilde{\psi}_\lambda$. If we still denote the wavelet coefficients of a function $v \in H_0^1(\Omega)$ in the normalized basis by $\{\widehat{v}_\lambda\}$, and those of a form $F \in H^{-1}(\Omega)$ by $\{\widehat{F}_\lambda\}$, then we can identify v to the vector $\mathbf{v} = \{\widehat{v}_\lambda\} \in l^2(\mathcal{N})$, and F to the vector $\mathbf{F} = \{\widehat{F}_\lambda\} \in l^2(\mathcal{N})$. Furthermore, we have $\|v\|_{H_0^1(\Omega)} \asymp \|\mathbf{v}\|_{l^2(\mathcal{N})}$ and $\|F\|_{H^{-1}(\Omega)} \asymp \|\mathbf{F}\|_{l^2(\mathcal{N})}$. From now on, we set $(\mathbf{u}, \mathbf{v}) = (\mathbf{u}, \mathbf{v})_{l^2(\mathcal{N})}$ and $\|\mathbf{v}\| = \|\mathbf{v}\|_{l^2(\mathcal{N})}$.

Using these representations, the minimization problem (1.1) and (1.4) can be formulated as an infinite dimensional problem in $l^2(\mathcal{N})$. Indeed, we have

$$(4.1) \quad J(u) = \min_{v \in H_0^1(\Omega)} J(v) \quad \text{equivalent to} \quad \mathbf{J}(\mathbf{u}) = \min_{\mathbf{v} \in l^2(\mathcal{N})} \mathbf{J}(\mathbf{v}),$$

where

$$\mathbf{J}(\mathbf{v}) = J\left(\sum_{\lambda \in \mathcal{N}} \widehat{v}_\lambda \psi_\lambda\right) = \frac{1}{2}(\mathbf{A}\mathbf{v}, \mathbf{v}) - (\mathbf{f}, \mathbf{v})$$

with

$$\mathbf{A} = \{a_{\lambda\mu}\}_{\lambda, \mu \in \mathcal{N}}, \quad a_{\lambda\mu} = \int_{\Omega} \nabla \psi_\mu \cdot \nabla \psi_\lambda, \quad \text{and} \quad \mathbf{f} = \{\widehat{f}_\lambda\}_{\lambda \in \mathcal{N}}, \quad \widehat{f}_\lambda = \int_{\Omega} f \psi_\lambda.$$

We have $\mathbf{J}'(\mathbf{v}) = \mathbf{A}\mathbf{v} - \mathbf{f}$ and $\mathbf{J}''(\mathbf{v}) = \mathbf{A}$, for all $\mathbf{v} \in l^2(\mathcal{N})$. Furthermore, let u_δ be any approximation of u in $H_0^1(\Omega)$, and let \mathbf{u}_δ be the vector of its coefficients. The associated residual is $-\mathbf{J}'(u_\delta) = \Delta u_\delta + f \in H^{-1}(\Omega)$ and one has the error representation

$$\|u - u_\delta\|_{H_0^1(\Omega)}^2 \asymp \|\mathbf{J}'(u_\delta)\|_{H^{-1}(\Omega)}^2 \asymp \|\mathbf{J}'(\mathbf{u}_\delta)\|^2 = \|\mathbf{A}\mathbf{u}_\delta - \mathbf{f}\|^2.$$

It will be important for the sequel to observe that the previous equivalences between the $H_0^1(\Omega)$ -norm of a function and the $l^2(\mathcal{N})$ -norm of the vector of its coefficients can be rephrased as

$$(4.2) \quad C_1 \|\mathbf{v}\|^2 \leq (\mathbf{A}\mathbf{v}, \mathbf{v}) \leq C_2 \|\mathbf{v}\|^2, \quad \forall \mathbf{v} \in l^2(\mathcal{N});$$

in other words, the condition number $\kappa_2(\mathbf{A})$ of the infinite matrix \mathbf{A} in the Euclidean norm is bounded by C_2/C_1 . We will set $\|\mathbf{v}\|_{\mathbf{A}} = (\mathbf{A}\mathbf{v}, \mathbf{v})^{1/2}$.

The infinite dimensional steepest-descent algorithm to solve (4.1) is as follows. Let $\mathbf{u}^{(0)} \in l^2(\mathcal{D}\mathcal{N})$ be arbitrary; then, for $k \geq 0$, given $\mathbf{u}^{(k)}$, define

$$\begin{aligned} \mathbf{s}^{(k)} &= -\mathbf{J}'(\mathbf{u}^{(k)}), \\ \beta^{(k)} &= \arg \min_{\beta \geq 0} \mathbf{J}(\mathbf{u}^{(k)} + \beta \mathbf{s}^{(k)}) = \frac{(\mathbf{s}^{(k)}, \mathbf{s}^{(k)})}{(\mathbf{A}\mathbf{s}^{(k)}, \mathbf{s}^{(k)})}, \\ \mathbf{u}^{(k+1)} &= \mathbf{u}^{(k)} + \beta^{(k)} \mathbf{s}^{(k)}. \end{aligned}$$

It is well known that the sequence $\{\mathbf{u}^{(k)}\}$ converges to \mathbf{u} . Precisely, one has

$$\|\mathbf{u}^{(k+1)} - \mathbf{u}\|_A \leq \varrho_E \|\mathbf{u}^{(k)} - \mathbf{u}\|_A \quad \text{with} \quad \varrho_E = \left(\frac{\kappa_2(\mathbf{A}) - 1}{\kappa_2(\mathbf{A}) + 1} \right)^{1/2}.$$

Let us now introduce a finite dimensional, adaptive version of this algorithm. We assume that the best N -term approximation of u in the wavelet basis has, for some $s > 1$, the decay rate

$$\inf_{s_N \in S_N} \|u - v_N\|_{H_0^1(\Omega)} \lesssim N^{-s^*} \|\mathbf{u}\|_{l_w^s(\mathcal{D}\mathcal{N})}, \quad \text{with} \quad s^* = \frac{s-1}{d}, \quad \frac{1}{\tau} = \frac{s-1}{d} + \frac{1}{2},$$

and we recall some fact from [17], which are crucial in the design of the algorithm. From now on, the symbol \mathbf{v}_A will denote a vector of finite support, i.e., with finitely many non-zero components \hat{v}_λ . We will set $\text{card}(\mathbf{v}_A) = \text{card}\{\lambda \in \mathcal{D}\mathcal{N} : \hat{v}_\lambda \neq 0\}$; the symbol $\text{cost}(\mathbf{v}_A)$ will indicate the computational cost (floating point operations and/or sorting) to get \mathbf{v}_A .

PROPERTY 1. – Given $\eta > 0$ and \mathbf{v}_A , a vector $\mathbf{z}_A = A(\mathbf{v}_A, \eta)$ can be computed, such that $\|\mathbf{A}\mathbf{v}_A - \mathbf{z}_A\| \leq \eta$, $\|\mathbf{z}_A\|_{l_w^\tau(\mathcal{D}\mathcal{N})} \lesssim \|\mathbf{v}_A\|_{l_w^\tau(\mathcal{D}\mathcal{N})}$, $\text{card}(\mathbf{z}_A) \lesssim \eta^{-1/s^*} \|\mathbf{v}_A\|_{l_w^\tau(\mathcal{D}\mathcal{N})}^{1/s^*}$ and $\text{cost}(\mathbf{z}_A) \asymp \text{card}(\mathbf{v}_A) + \eta^{-1/s^*} \|\mathbf{v}_A\|_{l_w^\tau(\mathcal{D}\mathcal{N})}^{1/s^*}$.

PROPERTY 2. – Given $\eta > 0$ and the vector $\mathbf{f} = \mathbf{A}\mathbf{u}$, a vector $\mathbf{f}_A = f(\mathbf{f}, \eta)$ can be computed by thresholding, such that $\|\mathbf{f} - \mathbf{f}_A\| \leq \eta$, $\|\mathbf{f}_A\|_{l_w^\tau(\mathcal{D}\mathcal{N})} \lesssim \|\mathbf{u}\|_{l_w^\tau(\mathcal{D}\mathcal{N})}$ and $\text{card}(\mathbf{f}_A) \asymp \text{cost}(\mathbf{f}_A) \lesssim \eta^{-1/s^*} \|\mathbf{u}\|_{l_w^\tau(\mathcal{D}\mathcal{N})}^{1/s^*}$.

PROPERTY 3. – Given $\eta > 0$ and a vector \mathbf{v}_A such that $\|\mathbf{v}_A - \mathbf{u}\| \leq \eta$, a vector $\mathbf{w}_A = T(\mathbf{v}_A, \eta)$ can be computed by thresholding, which satisfies $\|\mathbf{w}_A - \mathbf{u}\| \leq 5\eta$, $\|\mathbf{w}_A\|_{l_w^\tau(\mathcal{D}\mathcal{N})} \lesssim \|\mathbf{u}\|_{l_w^\tau(\mathcal{D}\mathcal{N})}$, $\text{card}(\mathbf{w}_A) \lesssim \eta^{-1/s^*} \|\mathbf{u}\|_{l_w^\tau(\mathcal{D}\mathcal{N})}^{1/s^*}$ and $\text{cost}(\mathbf{w}_A) \asymp \text{card}(\mathbf{v}_A)$.

We are now in a position to define the single descent step of our adaptive algorithm.

PROPOSITION 3. – Let $\varrho \in (\varrho_E, 1)$ be fixed. Given $\eta > 0$ and a vector \mathbf{v}_A such that $\|\mathbf{v}_A - \mathbf{u}\|_A \leq \eta$ and $\text{card}(\mathbf{v}_A) \leq \eta^{-1/s^*} \|\mathbf{u}\|_{l_w^s(\mathcal{D}\mathcal{C})}^{1/s^*}$, let us set, for some $\vartheta \in (0, 1)$,

$$\mathbf{s}_A = -A(\mathbf{v}_A, \vartheta\eta) + f(\mathbf{f}, \vartheta\eta),$$

$$\mathbf{z}_A = A(\mathbf{s}_A, \vartheta\|\mathbf{s}_A\|),$$

$$\tilde{\beta} = \frac{(\mathbf{s}_A, \mathbf{s}_A)}{(\mathbf{z}_A, \mathbf{s}_A)}$$

and finally

$$\mathbf{w}_A = \mathbf{v}_A + \tilde{\beta}\mathbf{s}_A =: D(\mathbf{v}_A, \eta).$$

Then, there exists ϑ such that

$$\|\mathbf{w}_A - \mathbf{u}\|_A \leq \varrho\eta, \quad \text{card}(\mathbf{w}_A) \asymp \text{cost}(\mathbf{w}_A) \leq (\varrho\eta)^{-1/s^*} \|\mathbf{u}\|_{l_w^s(\mathcal{D}\mathcal{C})}^{1/s^*}.$$

The constant involved in the last symbol \asymp is independent of η and \mathbf{v}_A .

PROOF. – At first, we check that $\tilde{\beta}$ is well defined. We have $(\mathbf{z}_A, \mathbf{s}_A) = (\mathbf{A}\mathbf{s}_A, \mathbf{s}_A) + (\boldsymbol{\mu}_z, \mathbf{s}_A)$ for some $\boldsymbol{\mu}_z$ satisfying $\|\boldsymbol{\mu}_z\| \leq \vartheta\|\mathbf{s}_A\|$. Thus, recalling (4.2), we get $|(\boldsymbol{\mu}_z, \mathbf{s}_A)| \leq \vartheta\|\mathbf{s}_A\|^2 \leq \vartheta C_1^{-1}(\mathbf{A}\mathbf{s}_A, \mathbf{s}_A)$. Choosing $\vartheta < C_1$ and setting

$$\bar{\beta} = \frac{(\mathbf{s}_A, \mathbf{s}_A)}{(\mathbf{A}\mathbf{s}_A, \mathbf{s}_A)}, \quad \text{we obtain } \tilde{\beta} = \bar{\beta} + \vartheta O(1).$$

Now, let us set $\mathbf{e}_0 = \mathbf{v}_A - \mathbf{u}$ and $\mathbf{e}_n = \mathbf{w}_A - \mathbf{u}$, so that $\mathbf{e}_n = \mathbf{e}_0 + \tilde{\beta}\mathbf{s}_A$ and $(\mathbf{A}\mathbf{e}_n, \mathbf{e}_n) = (\mathbf{A}\mathbf{e}_0, \mathbf{e}_0) + 2\tilde{\beta}(\mathbf{A}\mathbf{e}_0, \mathbf{s}_A) + \tilde{\beta}^2(\mathbf{A}\mathbf{s}_A, \mathbf{s}_A)$. Set $\mathbf{s} = -\mathbf{J}'(\mathbf{v}_A) = -\mathbf{A}\mathbf{v}_A + \mathbf{f}$ and note that, by construction, we have $\|\mathbf{s} - \mathbf{s}_A\| = \vartheta O(\eta)$. Next, observe that $\mathbf{A}\mathbf{e}_0 = -\mathbf{s}$; thus, keeping into account (4.2) again, we have $(\mathbf{A}\mathbf{e}_0, \mathbf{e}_0) = (\mathbf{s}, \mathbf{A}^{-1}\mathbf{s}) = (\mathbf{s}_A, \mathbf{A}^{-1}\mathbf{s}_A) + \vartheta O(\eta^2)$ as well as $(\mathbf{A}\mathbf{e}_0, \mathbf{s}_A) = -(\mathbf{s}, \mathbf{s}_A) = -(\mathbf{s}_A, \mathbf{s}_A) + \vartheta O(\eta^2)$. Consequently, after some algebra, we obtain

$$(\mathbf{A}\mathbf{e}_n, \mathbf{e}_n) = \left[1 - \frac{(\mathbf{s}_A, \mathbf{s}_A)^2}{(\mathbf{A}\mathbf{s}_A, \mathbf{s}_A)(\mathbf{s}_A, \mathbf{A}^{-1}\mathbf{s}_A)} \right] (\mathbf{s}_A, \mathbf{A}^{-1}\mathbf{s}_A) + \vartheta O(\eta^2).$$

By the Kronecker inequality $1 \leq (\mathbf{A}\mathbf{x}, \mathbf{x})(\mathbf{A}^{-1}\mathbf{x}, \mathbf{x})/\|\mathbf{x}\|^4 \leq (\kappa_2(\mathbf{A})^{1/2} + \kappa_2(\mathbf{A})^{-1/2})^2/4$, which holds for all $\mathbf{x} \in l^2(\mathcal{D}\mathcal{C})$, the term in square brackets is bounded by ϱ_E^2 . Thus,

$$(\mathbf{A}\mathbf{e}_n, \mathbf{e}_n) \leq \varrho_E^2(\mathbf{s}_A, \mathbf{A}^{-1}\mathbf{s}_A) + \vartheta O(\eta^2) = \varrho_E^2(\mathbf{A}\mathbf{e}_0, \mathbf{e}_0) + \vartheta O(\eta^2) \leq (\varrho_E^2 + \vartheta O(1))\eta^2.$$

Choosing ϑ so that $\varrho_E^2 + \vartheta O(1) \leq \varrho^2$, we obtain the estimate $\|\mathbf{w}_A - \mathbf{u}\|_A \leq \varrho\eta$. The remaining statements follow from Properties 1 and 2. ■

A fixed number of descent steps guarantees a prescribed error reduction.

PROPOSITION 4. – Let $\alpha \in (0, 1)$ be fixed. Given $\eta > 0$ and a vector \mathbf{v}_A such that $\|\mathbf{v}_A - \mathbf{u}\| \leq \eta$ and $\text{card}(\mathbf{v}_A) \leq \eta^{-1/s^*} |\mathbf{u}|_{L_w^s(\mathcal{R})}^{1/s^*}$, set $\mathbf{w}_A^{(0)} = \mathbf{v}_A$ and, for $k \geq 0$, $\mathbf{w}_A^{(k+1)} = D(\mathbf{w}_A^{(k)}, \varrho^k \eta)$. Then, there exists an integer K depending only on α and ϱ such that the vector $\mathbf{w}_A := \mathbf{w}_A^{(K)} =: D(\mathbf{v}_A, \eta, \alpha)$ satisfies

$$\|\mathbf{w}_A - \mathbf{u}\| \leq \alpha \eta, \quad \text{card}(\mathbf{w}_A) \asymp \text{cost}(\mathbf{w}_A) \leq (\alpha \eta)^{-1/s^*} |\mathbf{u}|_{L_w^s(\mathcal{R})}^{1/s^*}.$$

The constant involved in the symbol \asymp depends on K but is independent of η and \mathbf{v}_A .

PROOF. – Recalling (4.2), we have $\|\mathbf{w}_A - \mathbf{u}\| \leq C_1^{-1/2} \varrho^K C_2^{1/2} \eta$. Thus, it is enough to choose K such that $C_1^{-1/2} \varrho^K C_2^{1/2} \leq \alpha$. ■

The complete adaptive algorithm is as follows. Let $\gamma \in (0, 1)$ be fixed. For $m = 0$, set $\mathbf{u}_A^{(0)} = \mathbf{0}$. For $m \geq 0$, given $\mathbf{u}_A^{(m)}$, define

$$(4.3) \quad \tilde{\mathbf{u}}_A^{(m+1)} = D(\mathbf{u}_A^{(m)}, \gamma^m, \gamma/5),$$

$$(4.4) \quad \mathbf{u}_A^{(m+1)} = T(\tilde{\mathbf{u}}_A^{(m+1)}, \gamma^{m+1}/5),$$

where the operators D and T are defined in Proposition 4 and Property 3, respectively. Collecting all previous estimates, one has the following final result.

THEOREM 3. – The sequence $\{\mathbf{u}_A^{(m)}\}$ defined above satisfies, for all $m \geq 0$,

$$\|\mathbf{u}_A^{(m)} - \mathbf{u}\| \leq \eta_m := \gamma^m, \quad \text{card}(\mathbf{u}_A^{(m)}) \asymp \text{cost}(\mathbf{u}_A^{(m)}) \leq \eta_m^{-1/s^*} |\mathbf{u}|_{L_w^s(\mathcal{R})}^{1/s^*}. \quad \blacksquare$$

Obviously, the result can be rephrased in terms of the quantity $N_m := \text{card}(\mathbf{u}_A^{(m)})$. Setting $u_{N_m} = \sum_{\lambda \in \mathcal{R}} \hat{u}_\lambda^{(m)} \psi_\lambda$, we obtain

$$\|u - u_{N_m}\|_{H_0^1(\Omega)} \leq N_m^{-s^*} |\mathbf{u}|_{L_w^s(\mathcal{R})}, \quad m \rightarrow \infty.$$

We conclude that the adaptive minimization algorithm (4.3)-(4.4) is asymptotically optimal.

4.2. Finite element approximations.

Adaptive finite element methods are by now widely and successfully used in practice. Mesh adaptation is based on the inspection of local error indicators, derived by a careful a posteriori error analysis (see, e.g., [37]).

As an example, consider again the model problem (1.3). Denote by $\mathcal{E} = \{E\}$ a geometrically conforming «triangulation» of $\bar{\Omega}$ by elements E of diameter

h_E , having «sides» $e \subset \partial E$ of diameter h_e . Let V_h be a subspace of $H_0^1(\Omega)$ made of piecewise polynomial functions on \mathcal{T} . If u_h denotes the Galerkin approximation of u in V_h , i.e., the solution of the problem

$$\int_{\Omega} \nabla u_h \cdot \nabla v_h \, dx = \int_{\Omega} f v \, dx, \quad \forall v_h \in V_h,$$

which corresponds to minimizing the energy functional J over V_h , then one can prove that

$$\|u - u_h\|_{H_0^1(\Omega)}^2 \asymp \|J'(u_h)\|_{H^{-1}(\Omega)}^2 = \|\Delta u_h + f\|_{H^{-1}(\Omega)}^2 \asymp \sum_{E \in \mathcal{T}} \eta_E^2,$$

where the local error indicators η_E are given by

$$\eta_E^2 = h_E^2 \|\Delta u_h + f\|_{L^2(E)}^2 + \sum_{e \subset \partial E} h_e \left\| \left[\frac{\partial u_h}{\partial n} \right] \right\|_{L^2(e)}^2$$

($[\cdot]$ being the jump across e). Popular mesh adaptation strategies mark each element E for refinement or coarsening depending on the relative size of the indicators η_E .

Despite the easiness of implementation of this and similar strategies, as well as the extremely good performances observed in practice, the theoretical foundations of adaptive finite element methods are still in their infancy. The first convergence proof of an adaptive algorithm (in the multidimensional case) was given in [31] and refined in [34]. It is only recently [7] that the rate of decay of the error, in terms of the dimension N of the subspace V_h and the Besov regularity of the solution, has been established for Problem (1.3). The analysis is based on concepts of nonlinear approximation theory, as described in Section 3, originally developed for wavelets. The conclusions are similar to those reported in Section 4.1. Precisely, a mesh adaptation algorithm is devised, which is shown to have the following properties: given an integer N , the algorithm generates an admissible mesh with $\asymp N$ elements, such that the piecewise linear Galerkin approximation built on that mesh approximates the exact solution to an error, in the energy norm, which behaves as the best approximation error among all the piecewise linear functions defined on admissible meshes with N elements; furthermore, the computational cost is at most proportional to N .

Remarkably, a simple refinement strategy, which starts from a coarse grid and recursively attempts to equidistribute the local error, often works well in practice, producing asymptotically optimal results. For instance, let us consider the Dirichlet problem for the Laplace equation

$$(4.5) \quad \Delta u = 0 \quad \text{in } \Omega = (0, 1)^2, \quad u = g \quad \text{on } \partial\Omega,$$

which can be easily reduced to Problem (1.3). Let us choose g to be identically

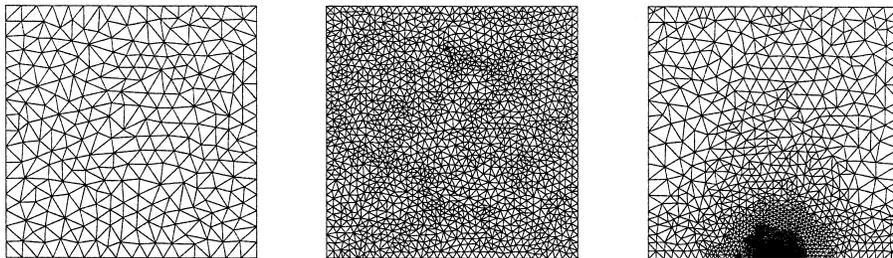


Figure 2. – Finite element meshes for Problem (4.5): initial mesh (left), uniformly refined mesh (center), adapted mesh (right).

0 on $\partial\Omega$, except for a neighborhood of the midpoint $(0.5, 0)$ of the bottom side, where it behaves like $|x-0.5|^{1/2}$ and then decays smoothly to 0. The corresponding solution u belongs to $H^s(\Omega)$ for all $s < \frac{3}{2}$, whereas it belongs to $B_{rr}^s(\Omega)$ for all $s \geq 0$. Then, one observes that the error $\|u - u_N^{\text{discr}}\|_{H_0^1(\Omega)}$ behaves like $O(N^{-1/4})$ if u_N^{discr} is the piecewise linear Galerkin solution defined on a uniformly refined mesh with N triangles, whereas it behaves like $O(N^{-1/2})$ if u_N^{discr} is the solution defined on a grid, again with N triangles, which is adaptively refined around the singularity point (see Figures 2). These results match the theoretical predictions, based on the linear and nonlinear approximation theories.

Acknowledgement. The author wishes to thank S. Berrone, K. Urban and A. Tabacco for providing some material used in the presentation.

REFERENCES

- [1] A. BARINKA - T. BARSCH - PH. CHARTON - A. COHEN - S. DAHLKE - W. DAHMEN - K. URBAN, *Adaptive wavelet schemes for elliptic problems - implementation and numerical experiments*, SIAM J. Sci. Comput., **23** (2001), 910-939.
- [2] C. BERNARDI - Y. MADAY, Spectral Methods, pp. 209-486 in *Handbook of Numerical Analysis, Vol. V* (Ph.G. Ciarlet and J.L. Lions, eds.), North Holland, Amsterdam, 1997.
- [3] S. BERRONE - L. EMMEL, *A realization of a wavelet Galerkin method on non-trivial domains*, Math. Models Meth. Appl. Sci., **12** (2002), 1525-1554.
- [4] S. BERTOLUZZA, *A posteriori error estimates for the wavelet Galerkin method*, Appl. Math. Lett., **8** (1995), 1-6.
- [5] S. BERTOLUZZA - S. MAZET - M. VERANI, *A nonlinear Richardson algorithm for the solution of elliptic PDE's*, Pubbl. IAN-CNR n. 1227, Pavia (2001).

- [6] S. BERTOLUZZA - M. VERANI, *Convergence of a non-linear wavelet algorithm for the solution of PDE's*, Pubbl. IAN-CNR n. 1205, Pavia (2001), to appear in Appl. Math. Lett.
- [7] S. BINEV - W. DAHMEN - R. A. DEVORE, *Adaptive finite element methods with convergence rates*, IGPM Report No. 219, RWTH Aachen, June 2002.
- [8] C. CANUTO - I. CRAVERO, *A wavelet-based adaptive finite element method for the advection-diffusion equations*, Math. Models Meths. Appl. Sci., **7** (1997), 265-289.
- [9] C. CANUTO - M. Y. HUSSAINI - A. QUARTERONI - T. A. ZANG, *Spectral Methods in Fluid Dynamics*, Springer, New York, 1990.
- [10] C. CANUTO - A. TABACCO - K. URBAN, *The Wavelet Element Method Part I: construction and analysis*, Appl. Comput. Harm. Anal., **6** (1999), 1-52.
- [11] C. CANUTO - A. TABACCO - K. URBAN, *The Wavelet Element Method Part II: realization and additional features in 2D and 3D*, Appl. Comp. Harm. Anal., **8** (2000), 123-165.
- [12] C. CANUTO - K. URBAN, *Adaptive optimization of convex functionals in Banach spaces*, in preparation.
- [13] J. CÉA, *Optimisation, théorie et algorithmes*, Dunod, Paris, 1971.
- [14] PH. G. CIARLET, Basic Error Estimates for Elliptic Problems, pp. 17-352 in *Handbook of Numerical Analysis, Vol. II* (Ph. G. Ciarlet and J. L. Lions, eds.), North Holland, Amsterdam, 1991.
- [15] A. COHEN, Wavelet Methods in Numerical Analysis, pp. 417-711 in *Handbook of Numerical Analysis, Vol. VII* (Ph.G. Ciarlet and J. L. Lions, eds.), North Holland, Amsterdam, 2000.
- [16] A. COHEN - W. DAHMEN - R. A. DEVORE, *Adaptive wavelet methods for elliptic operator equations - convergence rates*, Math. Comput., **70** (2001), 27-75.
- [17] A. COHEN - W. DAHMEN - R. A. DEVORE, *Adaptive wavelet methods II - beyond the elliptic case*, Found. Comput. Math., **2** (2002), 203-245.
- [18] A. COHEN - W. DAHMEN - R. A. DEVORE, *Adaptive Wavelet Schemes for Nonlinear Variational Problems*, IGPM Report No. 221, RWTH Aachen, June 2002.
- [19] A. COHEN - I. DAUBECHIES - J. FEAUVEAU, *Biorthogonal bases of compactly supported wavelet*, Comm. Pure Appl. Math., **45** (1992), 485-560.
- [20] A. COHEN - I. DAUBECHIES - P. VIAL, *Wavelets on the interval and fast wavelet transform*, Appl. Comp. Harm. Anal., **1** (1993), 54-81.
- [21] S. DAHLKE, *Besov regularity for elliptic boundary value problems on polygonal domains*, Appl. Math. Lett., **12** (1999), 31-36.
- [22] S. DAHLKE - W. DAHMEN - R. HOCHMUTH - R. SCHNEIDER, *Stable multiscale bases and local error estimation for elliptic problems*, Appl. Numer. Math., **23** (1997), 21-48.
- [23] S. DAHLKE - W. DAHMEN - K. URBAN, *Adaptive wavelet methods for saddle point problems - optimal convergence rates*, IGPM Report No. 204, RWTH Aachen, 2001, to appear in SIAM J. Numer. Anal.
- [24] S. DAHLKE - R. A. DEVORE, *Besov regularity for elliptic boundary value problems*, Commun. Partial Diff. Eqns., **22** (1997), 1-16.
- [25] W. DAHMEN - A. KUNOTH - K. URBAN, *Biorthogonal spline-wavelets on the interval - Stability and moment conditions*, Appl. Comp. Harm. Anal., **6** (1999), 132-196.
- [26] W. DAHMEN - R. SCHNEIDER, *Wavelets on manifolds I. Construction and domain decompositions*, SIAM J. Math. Anal., **31** (1999), 184-230.

- [27] W. DAHMEN - R. SCHNEIDER, *Composite wavelet bases for operator equations*, Math. Comp., **68** (1999), 1533-1567.
- [28] I. DAUBECHIES, *Ten Lectures on Wavelets*, CBMS-NSF Series in Applied Mathematics 61, SIAM, Philadelphia, 1992.
- [29] R. A. DEVORE, *Nonlinear Approximation*, pp. 51-150 in *Acta Numerica 1998*, Cambridge University Press, Cambridge, 1998.
- [30] R. A. DEVORE - B. J. LUCIER, *Wavelets*, pp. 1-56 in *Acta Numerica 1992*, Cambridge University Press, Cambridge, 1992.
- [31] W. DÖRFLER, *A convergent adaptive algorithm for Poisson's equation*, SIAM J. Numer. Anal., **33** (1996), 1106-1124.
- [32] S. GRIVET TALOCIA - A. TABACCO, *Wavelets on the interval with optimal localization*, Math. Models Meth. Appl. Sci., **10** (2000), 441-462.
- [33] A. HAAR, *Zur Theorie der orthogonalen Funktionen-Systeme*, Math. Ann., **69** (1910), 331-371.
- [34] P. MORIN - R. NOCHETTO - K. SIEBERT, *Data oscillation and convergence of adaptive FEM*, SIAM J. Numer. Anal., **38** (2000), 466-488.
- [35] C. SCHWAB, *p- and hp-Finite Element Methods*, Clarendon Press, Oxford, 1998.
- [36] V. N. TEMLYAKOV, *The best m-term approximation and greedy algorithms*, Adv. Comput. Math., **8** (1998), pp. 249-265.
- [37] R. VERFÜRTH, *A Review of A Posteriori Error Estimation and Adaptive Mesh Refinement Techniques*, Wiley-Teubner, Chichester, 1996.

Dipartimento di Matematica, Politecnico di Torino
Corso Duca degli Abruzzi 24, 10129 Torino, Italy
e-mail: ccanuto@polito.it