MOX–Report No. 36/2012

On the Numerical Analysis of Adaptive Spectral/hp Methods for Elliptic Problems

Canuto, C.; Verani, M.

MOX, Dipartimento di Matematica "F. Brioschi"
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox@mate.polimi.it http://mox.polimi.it

On the Numerical Analysis of Adaptive Spectral/*hp* Methods for Elliptic Problems

Claudio Canuto[a] and Marco Verani[b]

September 14, 2012

[a] Dipartimento di Scienze Matematiche, Politecnico di Torino
Corso Duca degli Abruzzi 24, 10129 Torino, Italy
E-mail: claudio.canuto@polito.it

[b] MOX, Dipartimento di Matematica, Politecnico di Milano
Piazza Leonardo da Vinci 32, I-20133 Milano, Italy
E-mail: marco.verani@polimi.it

The transcendental is not infinite and unattainable tasks,
but the neighbor who is within reach in any given situation.
(D. Bonhoeffer)

Abstract

We provide an overview of the state of the art of adaptive strategies for high-order *hp* discretizations of partial differential equations; at the same time, we draw attention on some recent results of ours concerning the convergence and complexity analysis of adaptive algorithm of spectral and spectral-element type. Complexity is studied under the assumption that the solution belongs to a sparsity class of exponential type, which means that its best N-term approximation error in the chosen piecewise polynomial basis decays at an exponential rate with respect to N.

1 Introduction

The present authors are some generations apart, yet both of them have been deeply influenced by the gigantic human and professional figure of Enrico Magenes; this paper is a modest yet heartfelt tribute to his memory.

On the scientific ground, professor Magenes not only gave outstanding contributions to the mathematical theory of partial differential equations, but he was a pioneer in fostering

the development of a Computational Mathematics at the same time soundly tied to functional analysis and strongly immersed in real-life applications. The "Laboratorio di Analisi Numerica", that he founded in Pavia in the Seventies and directed for several decades, had a paramount impact on the Italian Applied Mathematics community, and soon become a recognized reference for the whole international community.

Enrico Magenes was able to create, and maintain over the years, a scientific environment extremely open and favorable to the development and exchange of new ideas. Many young researchers have been attracted by that atmosphere; some of them, such as the senior author (CC), had the chance to become members of that institution, and bring everlasting gratitude for the opportunity they had of being exposed day by day to the charismatic personality of the founder; some others, such as the junior author (MV), have remained fascinated and deeply influenced by his human and scientific legacy.

Enrico Magenes was always very careful in granting freedom of research to the members of his group, the only conditions to respect being the quality and the interest of the undertaken investigations. Good projects pushed forward by younger collaborators had the chance of being supported as more mature lines of research. The onset of interest for spectral methods, which are the object of the present contribution, at the "Laboratorio" in the early Eighties is precisely an example of this favorable environment. Influenced by a stimulating visit at the "Laboratoire d'Analyse Numérique" (created in Paris by Magenes' coworker and friend Jacques-Louis Lions), Alfio Quarteroni initiated a fruitful and long-lasting collaboration with the present senior author on the numerical analysis of spectral and high-order methods for boundary-value problems; while in the early stage the scientific guidance was provided by Franco Brezzi, the full and continuous strategic and logistic support granted by professor Magenes was certainly a key ingredient for the success of that research.

Since then, high-order methods such as spectral(-element) methods or the *hp*-version of finite element methods (the two categories being often hardly distinguishable from each other) have reached their maturity, both in the full understanding of their theoretical properties and in the penetration into the scientific computing practices, in various applicative environments (see, e.g., [50, 12, 13]). Yet, some relevant aspects of these methods are still far from being in a satisfactory shape, and deserve further investigations. An example is given by the so-called *hp*-adaptivity; indeed, even for steady problems, a full and rigorous understanding of the selection strategies between h-refinement and p-enrichment, and their influence on the complexity and optimality of the related algorithms, is still lacking.

The purpose of the present paper is to provide a soft overview of the state of the art of adaptive strategies for high-order discretizations of partial differential equations; at the same time, we aim at drawing attention on some recent results of ours concerning the convergence and complexity analysis of adaptive algorithm of spectral and spectral-element type.

We begin by recalling various approximation results which show that a proper choice of the mesh and the polynomial degree distribution over the mesh guarantees an exponential decay of the error even if the solution of the equation exhibits singularities inside the domain or at the boundary, provided their position is known. The free parameter is the cardinality of the set of active degrees of freedom. An elementary derivation is detailed, in the case of algebraic singularity or piecewise analyticity. Obviously, such results on optimal "a-priori adaptivity" constitute a benchmark for the "a-posteriori" adaptive strategies, which need to detect the singularities and properly allocate the degrees of freedom around them. Thus, we are led to review the main error estimators proposed in the literature for *hp* methods, and the various adaptation strategies which exploit their information. While the number of different strategies

is fairly large, with different mathematical sources and different practical performances, very few algorithms can be rigorously proven to be convergent, with a precise estimate of their rate of convergence. The situation becomes even worse if complexity or cost issues are to be taken into account.

In this respect, we devote the second part of the paper to illustrate some recent results we have obtained in collaboration with Ricardo H. Nochetto; we consider adaptive spectral methods of Legendre type, and actually we extend them to cover the case of spectral-element discretizations (or p-type finite elements). A representative algorithm (out of several possible variants) is described, and its convergence properties are discussed. Furthermore, we investigate its complexity, by comparing the output of the algorithm with the best possible approximation of the exact solution in the chosen piecewise polynomial basis, for the same accuracy – this point of view being related to the so-called "best N-term approximation" of a function. The novelty of the analysis, compared to the available results in the literature, is that optimality is discussed with respect to an assumed exponential (or sub-exponential) decay of the best N-approximation error; this assumption appears indeed to be coherent with the use of spectral-type methods in the discretization of the boundary-value problem.

Notation. Throughout the paper, by $A \lesssim B$ we mean that the quantity A can be bounded by a multiple of B, the multiplicator being independent of those parameters A and B may depend on. Likewise $A \simeq B$ means $A \lesssim B$ and $B \lesssim A$, whereas $A \sim B$ means $A = B + o(B)$, with $o(B)$ negligible with respect to B.

2 From approximation theory to a-priori adaptive hp methods

We begin by recalling some classical results concerning the approximation of a univariate function having Sobolev or analytical regularity, by means of algebraic polynomials. These estimates will be useful in the subsequent analysis of piecewise polynomial approximation.

Let I denote the reference interval $(-1, 1)$. If v belongs to the Sobolev space $H^m(I)$, $m \geq 0$, then

$$\inf_{w \in \mathbb{P}_p(I)} \|v - w\|_{L^2(I)} \leq C(v, m) p^{-m} , \tag{2.1}$$

where the positive constant $C(v, m)$ can be bounded by the norm of v in $H^m(I)$ (actually, the bound holds under weaker assumptions, see, e.g., [50, 12]). On the other hand, if v can be extended to an analytic function on the closed ellipse $E(-1, 1; \sigma)$ in the complex plane having foci at $z = \pm 1$ and semiaxes' sum $\sigma > 1$, then setting $\eta = \log \sigma$ one has

$$\inf_{w \in \mathbb{P}_p(I)} \|v - w\|_{L^2(I)} \leq C(v, \eta) p^{-1/2} \mathrm{e}^{-\eta p} \tag{2.2}$$

(see, e.g., [30, 50]). A different estimate involves the maximum modulus of v over the ellipse $E(-1, 1; \sigma)$ and reads as follows:

$$\inf_{w \in \mathbb{P}_p(I)} \|v - w\|_{L^2(I)} \leq C \, \frac{1}{\sinh \eta} \mathrm{e}^{-\eta p} \max_{z \in E(-1,1;\sigma)} |v(z)| \tag{2.3}$$

(see, e.g., [12]).

The previous results can be easily combined to provide bounds of the approximation error for piecewise smooth functions on a finite partition of the domain by piecewise polynomials; in such cases, the estimates involve not only the polynomial degrees but also the sizes of the

subdomains. Less trivial is the problem of expressing the error in terms of the total number, say N, of employed degrees of freedom, and, even more, of optimizing the allocation of degrees of freedom for a given target accuracy. This is precisely the crucial problem of *hp* adaptivity.

The earliest attempts to study the adaptive approximation of a univariate function, having a finite number of singularities and otherwise smooth, by means of piecewise polynomials of variable degree dates back to the late Seventies, with the pioneering works [20] and [24] (see also [21] and the references therein).

In [20], the best N-term approximation of a univariate function in the maximum norm by piecewise polynomials of variable degree is studied, and in particular it is proven that for certain classes of analytic functions the best N-term approximation is achieved by a single polynomial over the entire domain. On the other hand, [24] deals with functions with singularities of the type x^α near the origin, and proves that a proper combination of graded mesh and linear increase of polynomial degrees (see below) yields exponential decay of the best N-term approximation error, with exponent proportional to $\sqrt{N}$.

This result influenced Gui and Babuška [30] in their study of the convergence rate of the *hp* approximation to a model elliptic problem in 1D. As usual, a Céa Lemma argument reduces the problem to estimate the best N-term approximation error in the energy norm. Let us give some detail.

Let $\Omega = (0,1)$. Suppose that the solution u of the underlying elliptic PDE is real analytic in $(0,1]$ and behaves like $u(x) = x^\alpha$ for some $\alpha > 1/2$. Consider a partition of $\overline{\Omega}$ into contiguous intervals K_j $(0 \le j \le J)$ and a corresponding distribution of polynomial degrees $p_j \ge 1$. Define the subspace

$$V_\delta = \{v \in H^1(\Omega) \ : \ v_{|K_j} \in \mathbb{P}_{p_j}(K_j) \ , \ 0 \le j \le J\} \ . \tag{2.4}$$

Let us assume that the mesh is geometrically graded towards 0, i.e., there exists $0 < \rho < 1$ such that

$$K_j = [\rho^{j+1}, \rho^j] \quad \text{for} \quad 0 \le j < J \ , \qquad K_J = [0, \rho^J] \ , \tag{2.5}$$

whereas the polynomial degrees grow linearly away from 0, i.e.,

$$p_j = \max(1, [\lambda(J-j)]) \tag{2.6}$$

for some $\lambda > 0$. Under these assumptions, there exists λ such that one has for $N = \dim V_\delta$

$$\inf_{v_\delta \in V_\delta} \|u - v_\delta\|_{H^1(\Omega)} \le C\mathrm{e}^{-b\sqrt{N}} \ , \tag{2.7}$$

where the constants $C > 0$ and $b > 0$ are independent of N. In particular, the choice $\rho^* = (\sqrt{2}-1)^2$ for the grading factor is optimal for any α.

The result can be extended to more general functions; in particular, to those belonging to the class $B^\ell_\beta(\Omega)$ for $\ell = 1, 2$ and some $\beta \in (0,1)$; these are the functions $u \in H^{\ell-1}(\Omega)$ such that, setting $\Phi_\gamma(x) = |x|^\gamma$, the functions $\Phi_{\beta+k-\ell} D^k u$ belong to $L^2(\Omega)$ for any $k \ge \ell$, and there exist constants $C > 0$ and $d \ge 1$ for which

$$\|\Phi_{\beta+k-\ell} D^k u\|_{L^2(\Omega)} \le C d^{k-l} (k-l)! \ . \tag{2.8}$$

Based on the previously described a-priori analysis, Gui and Babuška ([31]) proposed what is probably the first *hp* adaptive algorithm (see also Subsection 3.2 below). Given an elemental error estimator satisfying suitable assumptions, the elements of the partition on which the

error estimator is larger than a fixed fraction of the largest estimator are marked for refinement/enrichment; by inspecting the ratio between two error estimators on the marked interval with two consecutive polynomial degrees, it is decided whether to divide the interval into two parts carrying the same polynomial degree as before, or to keep the interval unchanged but increase the polynomial degree by one. The algorithm is proven to be convergent, with a predicted rate. However, the assumptions on the admissible error estimators appear to be overly restrictive, essentially they are taylored on the x^α-type singularity, for which indeed the algorithm produces a nearly optimal discretization.

In 2D, the counterpart of the previous a-priori analysis is as follows. Consider a bounded polygon Ω, having the origin 0 as a vertex. Let u be the solution of an elliptic problem in Ω, which is real analytic in $\overline{\Omega} \setminus \{0\}$ and behaves like $|x|^\alpha$ as $|x| \to 0$. Consider a conforming and regular partition of $\overline{\Omega}$ by J layers of elements around the origin; all the elements in the j-th layer have diameter of the order of ρ^j for some fixed $0 < \rho < 1$. Assume that all elements in the j-th layer carry polynomial degrees of the order of p_j, with $p_j = \max(1, [\lambda(J-j)])$ for some $\lambda > 0$. Let the subspace $V_\delta \subset H^1(\Omega)$ be defined in the standard way, and let again $N = \dim V_\delta$. Then, the following error bound is proven in [32, 33]:

$$\inf_{v_\delta \in V_\delta} \|u - v_\delta\|_{H^1(\Omega)} \leq C \mathrm{e}^{-b\sqrt[3]{N}} , \tag{2.9}$$

with $C > 0$ and $b > 0$ independent of N. The result extends to solutions in the class $B^2_\beta(\Omega)$, locally defined in a neighborhood of each vertex in a manner similar as above. This is relevant, since the solution of elliptic problems in Ω with data having suitable piecewise-analytic regularity can be shown to belong to such a class, see [5].

The situation in 3D is more complex, since in polyhedra singularities occur not only at vertices, but also along edges. Thus, an adapted mesh has a different structure in different regions of the domain, to accomodate for the local structure of the solution: it is quasi-uniform away from the boundary, it is isotropically graded towards a vertex, it is anisotropically graded towards the central part of an edge (being quasi-uniform in the tangential direction), and finally it has a transitional nature near the portion of an edge that gets close to a vertex. Then, a proper distribution of the polynomial degrees over such a mesh guarantees the following behavior of the best-approximation error vs the dimension N of the corresponding subspace V_δ, for the solution u of an elliptic problem in a polyhedron, with piecewise analytic data:

$$\inf_{v_\delta \in V_\delta} \|u - v_\delta\|_{H^1(\Omega)} \leq C \mathrm{e}^{-b\sqrt[5]{N}} , \tag{2.10}$$

again with $C > 0$ and $b > 0$ independent of N. The result (that should be compared to (2.7) in 1D and (2.9) in 2D) was first asserted by Guo and Babuška [6]; for the proof, we refer to [48, 49], where both Continuous- and Discontinuous-Galerkin *hp* discretizations are considered. The analysis relies on very accurate estimates of suitable weighted norms of the solution, of the type (2.8); we refer to [34] and to the more recent and comprehensive result [19].

2.1 An elementary analysis of *hp* approximations over dyadic partitions

Herafter, we use elementary arguments based on the repeated application of the error estimate (2.3) in order to establish the exponential convergence of suitable *hp* approximations over dyadic partitions to singular functions or piecewise-analytic functions in one space dimension.

To this end, given real numbers $r < s$ and $\sigma > h := s - r$, let $E(r, s; \sigma)$ denote the closed ellipse in the complex plane having foci at $z = r, s$ and semiaxes' sum σ; let us set $c = (r + s)/2$ and $f = h/2$. Let v be a function defined on the interval (r, s) of the real line, that can be extended to an analytic function on the closed ellipse $E(r, s; \sigma)$. Then, if we apply the change of variable

$$\hat{x} = \frac{x - c}{f} , \qquad \hat{y} = \frac{y}{f} ,$$

it is easily seen that the function $\hat{v}$ such that $\hat{v}(\hat{x}) = v(x) = v(c + f\hat{x})$ is defined on the reference interval $I = (-1, 1)$ and can be extended to an analytic function on the closed ellipse $E(-1, 1; \hat{\sigma})$ with $\hat{\sigma} = \sigma/f$. Thus, we can apply the bound (2.3) with $\eta = \log \hat{\sigma} = \log \sigma - \log f$, to obtain

$$\begin{aligned} \inf_{w \in \mathbb{P}_p((r,s))} \|v - w\|_{L^2((r,s))} &= h^{1/2} \inf_{\hat{w} \in \mathbb{P}_p(I)} \|\hat{v} - \hat{w}\|_{L^2(I)} \\ &\leq C h^{1/2} \frac{1}{\sinh \eta} \mathrm{e}^{-\eta p} \max_{z \in E(-1,1;\hat{\sigma})} |\hat{v}(z)| \\ &= C h^{1/2} \frac{1}{\sinh \eta} \mathrm{e}^{-\eta p} \max_{z \in E(r,s;\sigma)} |v(z)| . \end{aligned} \tag{2.11}$$

2.1.1 The case of an algebraic singularity

Let $\Omega = (0, 1)$. Suppose again that the solution u is real analytic in $(0, 1]$ and behaves like $u(x) = x^\alpha$ for some $\alpha > 1/2$. Let us consider the subspace of $L^2(\Omega)$ defined similarly to (2.4), i.e.,

$$V_\delta = \{v \in L^2(\Omega) \ : \ v_{|K_j} \in \mathbb{P}_{p_j}(K_j) \ , \ 0 \leq j \leq J\} , \tag{2.12}$$

where the K_j are defined in (2.5) with $\rho = \frac{1}{2}$ and the p_j are to be determined in the sequel. If v_δ denotes any function in V_δ, we split the approximation error as

$$\|u - v_\delta\|^2_{L^2(\Omega)} = \sum_{j=0}^{J} \|u_j - v_j\|^2_{L^2(K_j)}$$

with $u_j = u_{|K_j}$ and $v_j = v_{\delta|K_j}$. For $j = J$, we take as v_J the linear interpolant of u_J. This yields, with $h_J = 2^{-J}$,

$$\|u_J - v_J\|^2_{L^2(K_J)} \simeq \int_0^{h_J} (s^\alpha - h_J^{\alpha-1} s)^2 ds \simeq h_J^{2\alpha+1} \simeq 2^{-(2\alpha+1)J} .$$

Consider now any interval K_j with $0 \leq j < J$ and set $h_j = 2^{-(j+1)}$. We can think u_j as a real analytic function in K_j which can be extended to an analytic function in any closed ellipse $E(\rho^{j+1}\, \rho^j; \sigma)$ with $\sigma < \sigma_j = \frac{3}{2} h_j$. Hence, setting $\hat{\sigma}_j = \sigma_j / f_j = (\frac{3}{2} h_j)/(\frac{1}{2} h_j) = 3$ and $\eta_j = \log \hat{\sigma}_j = \log 3$, we can apply (2.11) in K_j and find $v_j \in \mathbb{P}_{p_j}(K_j)$ such that

$$\begin{aligned} \|u_j - v_j\|^2_{L^2(K_j)} &\leq C h_j \, \mathrm{e}^{-2\eta p_j} \max_{z \in E(\rho^{j+1}\, \rho^j; \sigma_j)} |u_j(z)|^2 \\ &\leq C(u) h_j \, \mathrm{e}^{-2\eta p_j} = C(u) 2^{-(j+1+\eta^* p_j)} , \end{aligned}$$

with $\eta^* = 2(\log_2 \mathrm{e})\, \eta$.

Let λ, μ be fixed constants ≥ 0 such that $\lambda + \mu = 1$. Then,

$$\|u - v_\delta\|^2_{L^2(\Omega)} \leq C \sum_{j=0}^{J-1} 2^{-\lambda j} 2^{-(\mu j + 1 + \eta^* p_j)} + 2^{-(2\alpha+1)J} .$$

Let us enforce that

$$\|u - v_\delta\|^2_{L^2(\Omega)} \leq C 2^{-2M}$$

for any given M. The bound on the error given above suggests to choose $J \sim 2M/(2\alpha+1)$ as well as $\mu j + \eta^* p_j \sim 2M$, i.e., $p_j \sim (2M - \mu j)/\eta^*$. With such choices, it is readily seen that the total number N of activated degrees of freedom satisfies

$$N = \sum_{j=0}^{J} p_j \simeq M^2 \,,$$

i.e., $M \simeq \sqrt{N}$. We conclude that the best approximation error satisfies

$$\inf_{v_\delta \in V_\delta} \|u - v_\delta\|_{L^2(\Omega)} \leq C \mathrm{e}^{-b\sqrt{N}} \tag{2.13}$$

for some $b > 0$, i.e., a bound of the same type as (2.7). Note that the definition of p_j given above is of the same type as (2.6).

The best approximation error in the $H^1(\Omega)$-norm can be estimated in a similar manner.

2.1.2 The case of a piecewise-analytic function

Assume now that u is a piecewise analytic function in $\overline{\Omega}$. It is not restrictive to assume the existence of just one singular point, say $x_s \in \Omega$. Thus, both $u_l := u_{|[0,x_s]}$ and $u_r := u_{|[x_s,1]}$ can be extended to analytic functions in a neighborhood of their intervals of definition in the complex plane.

With the aim of mimiking an adaptive algorithm which detects the position of the singularity by some error indicator, we consider the approximation procedure that generates a dyadic partition of $\overline{\Omega}$ by recursively halving the subinterval which contains the singular point x_s. Obviously, if x_s itself is a dyadic point, the procedure stops after a finite number of subdivisions, and we are just required to approximate by polynomials a finite number of analytic functions over a partitions of $\overline{\Omega}$; then, it is enough to apply (2.11) to each of them. If x_s is not a dyadic point, then at iteration $J \geq 0$ of the recursive algorithm we have a partition of the domain into $J+1$ subintervals K_j, such that $h_j := |K_j| = 2^{-j}$ for $0 \leq j \leq J$, and such that K_J is the only interval containing x_s. Let us set $u_j := u_{|K_j}$.

If u_J is of class C^ℓ In the interval K_J, for some $\ell \geq -1$ ($\ell = -1$ meaning that u_J has a jump at x_s), then we can find a polynomial v_J of degree $\ell + 1$ such that

$$\|u_J - v_J\|^2_{L^2(K_J)} \simeq h_J^{2\ell+3} \simeq 2^{-(2\ell+3)J} \,.$$

On the other hand, in any interval K_j with $0 \leq j < J$, u_j is a real analytic function which can be extended to an analytic function in some closed ellipse $E(r_j, s_j; \sigma_j)$ where $K_j = [r_j, s_j]$ and $\sigma_j \simeq 1$ depending on the size of the ellipse of analyticity of either u_l or u_r. In view of applying (2.11) in K_j we observe that $f_j = h_j/2 = 2^{-(j+1)}$, so that $\eta_j = \log(\sigma_j/f_j) \simeq a + bj$ and $\sinh \eta_j \sim \frac{1}{2}\mathrm{e}^{\eta_j} \simeq 2^j$. Hence,

$$\begin{aligned} \|u_j - v_j\|^2_{L^2(K_j)} &\leq C h_j \, \mathrm{e}^{-2\eta_j p_j} \max_{z \in E(r_j, s_j; \sigma_j)} |u_j(z)|^2 \\ &\leq C(u) 2^{-3j} \, \mathrm{e}^{-2(a+bj)p_j} = C(u) 2^{-(3j + (a^* + b^* j) p_j)} \end{aligned}$$

with $a^* = a \log_2 \mathrm{e}$ and $b^* = b \log_2 \mathrm{e}$. Summing-up, we obtain

$$\|u - v_\delta\|^2_{L^2(\Omega)} \leq C \sum_{j=0}^{J-1} 2^{-j} 2^{-(2j+(a^*+b^*j)p_j)} + 2^{-(2\ell+3)J} \ .$$

Let us enforce that

$$\|u - v_\delta\|^2_{L^2(\Omega)} \leq C 2^{-2M}$$

for any given M. The bound on the error given above suggests to choose $J \sim 2M/(2\ell+3)$ as well as $2j + (a^* + b^* j)p_j \sim 2M$, i.e., $p_j \sim 2(M-j)/(a^* + b^* j)$. With such choices, it is readily seen that the total number N of activated degrees of freedom satisfies

$$N = \sum_{j=0}^{J} p_j \simeq \sum_{j=0}^{J-1} \frac{M-j}{a^* + b^* j} \simeq M \log J \simeq M \log M \ ,$$

i.e., $M \simeq \phi(N)$, where $x = \phi(y)$ is the inverse function of $y = x \log x$ for $x \geq 1$. We conclude that the best approximation error satisfies

$$\inf_{v_\delta \in V_\delta} \|u - v_\delta\|_{L^2(\Omega)} \leq C \mathrm{e}^{-b\phi(N)} \tag{2.14}$$

for some $b > 0$. The result indicates that the behavior of the best approximation error in the presence of piecewise analyticity is only marginally worse than the one in the case of full analiticity, see (2.7).

Again, the best approximation error in the $H^1(\Omega)$-norm can be estimated in a similar manner.

3 *hp* adaptivity

Over the last few decades, adaptive algorithms have become a standard technique for solving partial differential equations via the finite element method. The general form of an adaptive algorithm can be stated as follows:

$$\ldots \rightarrow \texttt{SOLVE} \rightarrow \texttt{ESTIMATE} \rightarrow \texttt{MARK} \rightarrow \texttt{ENRICH} \rightarrow \texttt{SOLVE} \rightarrow \ldots$$

Generally speaking, the algorithm starts computing the discrete solution (`SOLVE`) employing a low-dimensional approximation space. Thereafter, in order to improve the accuracy of the approximation, an error indicator is employed (`ESTIMATE`) to obtain information about the error distribution. Based on this error distribution, a set of elements are flagged (`MARK`) to be enriched and a suitable enrichment of the approximation space is chosen (`ENRICH`). A new approximation of higher accuracy is computed and a new adaptive iteration is performed in case the approximation is not sufficiently accurate. In the adaptive h-FEM, the enrichment of the finite element space is simply done by subdividing into smaller elements all those elements with a large error indicator. However, in the hp-FEM one has the option to split an element or to increase its approximation order. Thus, as already pointed out, a main difficulty in hp-adaptivity is to decide whether to increase the approximation order p or to split an element whose error is large. The importance of making the correct decisions is highlighted by the a priori results mentioned in Section 2, from which it is evident that for a large class of problems an exponential rate of convergence can be achieved if the mesh and the polynomial degree distribution are chosen suitably.

Although considerable progress has been made in the context of adaptive h-FEM on both the a posteriori error analysis and the theoretical and computational assessment of the convergence properties of the adaptive refinement strategies (see, e.g., [43] for a comprehensive introduction), in contrast the theory of adaptive hp-FEM is far less advanced. Below we provide a brief review of existing a-posteriori hp error estimates (Section 3.1) and hp adaptive methods (Section 3.2).

3.1 A-posteriori hp error estimates

In the hp framework, similarly to the case of h-FEM, error indicators can be subdivided into the following categories:

- *Estimators based on the (approximate) solution of suitably defined local problems.* This includes [1, 2, 3, 4, 44]. The estimators of [1, 2, 3, 4] are based on solving local problems with Neumann type boundary conditions; a forerunner of this approach is [44]. Additionally, [44] discusses in detail other techniques known from h-FEM that can be extended to the hp-context such as solving local Dirichlet problems on patches, employing duality theory from convex optimization to derive upper and lower bounds of the local errors and employing various interpolation/postprocessing techniques to obtain more accurate approximations.

  At this point, we also mention the equilibrated residual estimators introduced in [10]. Although the method of [10] uses equilibrated fluxes, it differs from estimators via local Neumann problems as the estimators are obtained by the hypercircle method.

- *Residual based a-posteriori error estimators.* In the pioneering work [8], a posteriori error error indicators of residual type have been considered. However, the two-dimensional analysis of [8] is restricted to meshes consisting of axiparallel rectangles. In [40] the results of [8] are extended to meshes containing quadrilaterals and triangles and a family η_α , $\alpha \in [0, 1]$, of error indicators given by weighted residuals on the elements and on the edges is introduced. It is shown that that η_0 is reliable. As in [8], the reason for considering a family of indicators is that simultaneous reliability and efficiency cannot be proved for any fixed $\alpha \in [0, 1]$ due to the poor p-dependence of polynomial inverse estimates. For a related residual based a posteriori error estimate in one dimension, see also [47].

- *Estimators based on more accurate approximate solutions of the global problem.* This approach is based on the following steps: (1) a reference (finer) solution is computed by performing a global hp-refinement, i.e., breaking each element isotropically and enriching the polynomial order of approximation by one; (2) an error indicator is built by computing (and localizing) a suitable projection-based interpolation error of the reference solution. Roughly speaking, the indicator is computed by projecting the reference solution onto a finite element space employing the original mesh, but with a local polynomial degree incremented by one, as well as on a sequence of finite element spaces corresponding to a local h-refinement of the element that results in the same increase in the number of degrees of freedom as the p-enrichment. This approach has been introduced in [23] and further developed in [38].

Finally, we refer to [35] for goal-oriented hp-type error estimators.

3.2 Adaptive hp methods

Classical h adaptive finite element methods simply subdivide elements where the local error indicator is large, while keeping the polynomial degree fixed (at some low value). In general, this may not be the most efficient strategy in terms of error reduction per unit cost. For example, if the analytical solution to the underlying partial differential equation is smooth, or at least locally smooth, an enrichment of the polynomial degree (p-refinement) may be much more effective in reducing the local error per unit cost than a simple element subdivision (h refinement). Generally speaking, a local p-refinement is expected to be more efficient on elements where the solution is smooth, while local h-refinement is preferable for regions where the solution is not smooth.

In the following we will briefly review existing hp adaptive strategies. In particular, we will highlight the mechanism driving the choice between h or p refinement.

1. *Optimization strategy based on reference solution.* In this strategy a reference solution is computed on a finer finite element space, which is obtained by uniformly refining all the elements and globally incrementing the polynomial degree by one. Then, on each element in the coarser finite element mesh, the projection-based interpolation error of the reference solution is computed (see in the previous subsection the description of the estimators based on more accurate approximate solutions of the global problem). The optimal refinement of each element is then chosen to be the one which leads to the smallest projection-based interpolation error; elements in the mesh are then refined based on those that will lead to the greatest decrease in the projection error per degree of freedom. This strategy was first introduced by [46, 44, 22]; see also [23] for more recent work.

2. *Relative size of the error estimators.* In this strategy (originally introduced in [31]) it is assumed the existence of a local error indicator $\eta_K(u_{h,p}, h_K, p_K)$ which depends on the element K, the approximate solution $u_{h,p}$, the local mesh-size h_K and the local polynomial degree p_K. Then, the choice between h and p refinement is based on the ratio $r_K = \eta_K(u_{h,p_K}, h_K, p_K)/\eta_K(u_{h,p_K-1}, h_K, p_K - 1)$. In particular, if $c_k \leq \gamma$, $0 < \gamma < 1$ then p-enrichment should be performed as the error decreases when the polynomial degree is raised. On the other hand, if $c_K > \gamma$ then the element K is subdivided.

3. *Comparison of estimated and predicted error.* This strategy has been proposed in [40] (see also [27]) where the decision whether to subdivide an element or to increase its polynomial degree depends on the refinement history of the element. In particular, it is introduced a predicted local error indicator η_K^{pred} which can be viewed as a simple extrapolation of the error indicators computed during the previous refinement steps under the assumption that the solution is (locally) smooth. If the computed error indicator (which reflects the actual error) is larger than the predicted one, then an h-refinement is performed since the assumption of (local) smoothness, under which the computation of the predicted indicator is performed, is false. Conversely, if the indicated error is smaller than the predicted one, then p-refinement is performed.

4. *Analyticity check by estimating the decay rate of expansion coefficients.* In [21], the authors propose to determine whether the solution is locally smooth or non-smooth by calculating the decay rate of the Legendre expansion coefficients of the solution; this is performed by a least-squares best fit. More recently, a strategy has been developed in [37] for estimating the size of the Bernstein ellipse of the solution, thereby determining whether the solution

is analytic. In the case when it is not analytic, a second strategy, based on the work developed in [36], seeks to directly compute the local Sobolev index of the solution.

5. *Local regularity estimation.* This strategy, first proposed in [4], relies on estimating in each element K the local Sobolev regularity index m_k, by using a local error indicator which is computed by solving a series of local problems with different polynomial degrees. The local Sobolev regularity is then employed to perform h or p refinement. In particular, if $p_K + 1 \leq m_K$, where p_K denotes the current local polynomial order, then p-refinement is performed in K, otherwise h-refinement is selected. This latter criterium relies on two ingredients: i) the following hp a-priori error estimate on quasi-uniform mesh of size h and elements of uniform polynomial degree p:

$$\|u - u_{hp}\|_{H^1(\Omega)} \leq Ch^{\mu} p^{-(m-1)} \|u\|_{H^m(\Omega)}$$

where $\mu = \min(p, m-1)$ and $u \in H^m(\Omega)$, and ii) the idea that if the regularity of the solution is such that the rate of convergence of the h-type finite element method with elements of fixed degree p turns out to be sub-optimal, then an h refinement is needed. Recently, related ideas have been exploited in [47, 26] where the residual, instead of the error, has been employed to choose between p enrichment or h refinement. More recently, an approach based on the Sobolev embedding of H^1 into an appropriate L^p space has been proposed for the one-dimension case in [54].

6. *"Texas Three step".* This strategy was first introduced [45] and is based on a three-step scheme where only three solutions of the problem are needed. First, the initial mesh details as well as intermediate and final error tolerances are specified and the problem is solved. Then, the h-refinements take place in order to guarantee that the (intermediate) error (measured in some appropriate norm) is less than the intermediate tolerance. In the final third step, the mesh is kept fixed and the p-refinements are carried out to achieve the final error tolerance. For related work, we refer, e.g., to [53] and the references cited therein.

At last, we mention that a thorough comparison among various hp adaptive strategies has been recently accomplished in [41]. Algorithms have been tested on different kinds of representative solutions (analytic solution, corner singularity, peak, boundary layer, wavefront, and so on) and their preformance has been evaluated according to different measures of efficiency, such as the number of activated degrees of freedom or the computational time. None of the considered strategies has emerged as the best one in all situations, although some strategies perform better than the others for specific kinds of solutions.

3.3 Convergence of adaptive spectral/hp methods

The theory of h adaptive finite element (AFEM) schemes for elliptic problems is quite satisfactory: it started with the convergence results of [25] and [42]; the first optimality result was derived in [9] for $d = 2$ and extended by [51] to any d. The most comprehensive results for AFEM are contained in [16] for any d and L^2 data, and [18] for $d = 2$ and H^{-1} data; we refer to the survey [43]. In contrast, very little is known on convergence and optimality properties of adaptive hp methods. The first pioneering result goes back to [31] where an adaptive hp algorithm (see Subsection 3.2 for the description) is proven to be convergent, with a predicted rate. However, due to the assumptions on the admissible error estimators which appear to be

overly restrictive, the results in [31] cannot be considered completely satisfactory. Only after two decades, in [26] a contraction result of the form

$$\|u - u_{n+1}\|_{H^1} \leq \mu \|u - u_n\|_{H^1} \ , \quad \mu < 1 \ ,$$

has been proven, where u_n and u_{n+1} are the coarse and the enriched discrete solutions built by an adaptive *hp* algorithm approximating a one dimensional elliptic problem. More recently, the convergence result of [26] has been extended to higher dimensions in [11]. In this respect, it is also worth mentioning the result in [47] where an estimation (from above and below) of the error between the actual discrete solution and its (h or p) enriched version is presented. However, to the best of the authors' knowledge, there are no optimality results for *hp* adaptive algorithms: this is still a completely open issue.

4 Spectral adaptive algorithms with optimality properties

Inspired by the analysis performed in the wavelet framework by [17, 29, 52] and in the finite element framework by [25, 42, 9, 51, 16, 43], the present authors in collaboration with Ricardo H. Nochetto have recently initiated the study of the convergence and complexity properties of adaptive Fourier Galerkin methods in arbitrary d-dimension [14], and of adaptive Legendre-Galerkin methods in one-dimension [15]. Hereafter, we present a short account of the latter results, which incorporates their extension to the case of discretizations by spectral elements.

We consider the elliptic problem in $\Omega = (a, b)$

$$\begin{cases} Au = -D \cdot (\nu Du) + \sigma u = f \quad \text{in } \Omega \ , \\ u(a) = u(b) = 0 \ , \end{cases} \tag{4.1}$$

where ν and σ are sufficiently smooth real coefficients satisfying $0 < \nu_* \leq \nu(x) \leq \nu^* < \infty$ and $0 < \sigma_* \leq \sigma(x) \leq \sigma^* < \infty$ in Ω; let us set

$$\alpha_* = \min(\nu_*, \sigma_*) \qquad \text{and} \qquad \alpha^* = \max(\nu^*, \sigma^*) \ .$$

We formulate this problem variationally as

$$u \in V := H^1_0(\Omega) \quad : \quad a(u, v) = \langle f, v \rangle \qquad \forall v \in H^1_0(\Omega) \ , \tag{4.2}$$

where $a(u, v) = \int_\Omega \nu Du Dv + \int_\Omega \sigma uv$. We denote by $|\!|\!| v |\!|\!| = \sqrt{a(v, v)}$ the energy norm of any $v \in H^1_0(\Omega)$, which satisfies

$$\sqrt{\alpha_*} \|v\| \leq |\!|\!| v |\!|\!| \leq \sqrt{\alpha^*} \|v\| \ . \tag{4.3}$$

Our error estimators will be of residual type. Therefore, for any w belonging to some finite dimensional subspace V_Λ of $H^1_0(\Omega)$, we define the residual $r(w) = f - Aw \in H^{-1}(\Omega)$. Then, by the continuity and coercivity of the bilinear form a, one has

$$\frac{1}{\alpha^*} \|r(w)\| \leq \|u - w\| \leq \frac{1}{\alpha_*} \|r(w)\| \ , \tag{4.4}$$

or, equivalently,

$$\frac{1}{\sqrt{\alpha^*}} \|r(w)\| \leq |\!|\!| u - w |\!|\!| \leq \frac{1}{\sqrt{\alpha_*}} \|r(w)\| \ . \tag{4.5}$$

4.1 Bases and norm representations

Form now on, we assume that the coefficients and data of the problem are piecewise analytic on a finite partition $\mathcal{T} = \{K\}$ of $\overline{\Omega}$. Let us introduce the subspace of $H_0^1(\Omega)$ of the piecewise linear functions on $\mathcal{T}$, i.e.,

$$V_L(\mathcal{T}) = \{v \in H_0^1(\Omega) \mid v_{|K} \in \mathbb{P}_1(K) \quad \forall K \in \mathcal{T}\} ;$$

then,

$$V := H_0^1(\Omega) = V_L(\mathcal{T}) \oplus \bigoplus_{K\in\mathcal{T}} H_0^1(K) ,$$

where, for convenience, we assume that functions in $H_0^1(K)$ are extended by 0 outside the interval K; indeed, for any $v \in V$, we have

$$v = v_L + \sum_{K\in\mathcal{T}} v_K ,$$

where $v_L \in V_L(\mathcal{T})$ is the piecewise linear interpolant of v on $\mathcal{T}$ and $v_K = (v - v_L)_{|K} \in H_0^1(K)$. Since

$$(v_L, v_K)_{H_0^1(\Omega)} = \int_K v_L'(v-v_L)' = v_{L|K}' \int_K (v-v_L)' = 0 \qquad \text{and} \qquad (v_K, v_{K'})_{H_0^1(\Omega)} = 0 \quad \text{if } K \neq K' ,$$

we have

$$\|v\|^2_{H_0^1(\Omega)} = \|v_L\|^2_{H_0^1(\Omega)} + \sum_{K\in\mathcal{T}} \|v_K\|^2_{H_0^1(K)} .$$

Given any $F \in V' = H^{-1}(\Omega)$, let $F_L \in V_L(\mathcal{T})'$ denote the restriction of F to $V_L(\mathcal{T})$; similarly, for each $K \in \mathcal{T}$, let $F_K \in H^{-1}(K)$ denote the restriction of F to $H_0^1(K)$. Then,

$$\langle F, v\rangle = \langle F_L, v_L\rangle + \sum_{K\in\mathcal{T}} \langle F_K, v_K\rangle \qquad \forall v \in V ,$$

which easily implies

$$\|F\|^2_{H^{-1}(\Omega)} = \|F_L\|^2_{V_L(\mathcal{T})'} + \sum_{K\in\mathcal{T}} \|F_K\|^2_{H^{-1}(K)} . \tag{4.6}$$

Let us now introduce the Lagrangean basis functions ψ_q in $V_L(\mathcal{T})$ associated with the internal nodes of the partition, say x_q for $1 \le q \le Q$, so that

$$V_L(\mathcal{T}) = \operatorname{span}\{\psi_q \mid 1 \le q \le Q\} .$$

On the other hand, on the reference element $\hat{I} = (-1, 1)$, we consider the Babuška-Shen basis, made of polynomials of strictly increasing degree,

$$\hat{\phi}_k(\hat{x}) = \sqrt{\frac{2k-1}{2}} \int_{\hat{x}}^1 L_{k-1}(s)\, ds = \frac{1}{\sqrt{4k-2}}\big(L_{k-2}(\hat{x}) - L_k(\hat{x})\big) \qquad k \ge 2 , \tag{4.7}$$

where $L_k(\hat{x})$, $k \ge 0$, stands for the k-th Legendre polynomial, which satisfies $\deg L_k = k$, $L_k(1) = 1$ and

$$\int_{-1}^1 L_k(\hat{x}) L_m(\hat{x})\, d\hat{x} = \frac{2}{2k+1}\, \delta_{km} , \qquad m \ge 0 .$$

It is easily seen that the basis functions satisfy

$$(\hat{\phi}_k, \hat{\phi}_m)_{H_0^1(\hat{I})} = \int_{-1}^{1} \hat{\phi}_k'(\hat{x})\hat{\phi}_m'(\hat{x})\, d\hat{x} = \delta_{km} \ , \qquad k, m \geq 2 \ ,$$

i.e., they form an orthonormal system with respect to the $H_0^1(\hat{I})$-inner product. Going back to our partition, for any element $K \in \mathcal{T}$ of size h_K, we can map the reference element to the element K via an affine transformation $x = h_K\hat{x} + c$, yielding the functions

$$\phi_{K,k}(x) = \sqrt{\frac{h_K}{2}}\hat{\phi}_k(\hat{x}) \ , \qquad k \geq 2 \ , \tag{4.8}$$

which form an orthonormal system with respect to the $H_0^1(K)$-inner product.

At this point, we are ready to give a representation of the norms in $H_0^1(\Omega)$ and $H^{-1}(\Omega)$. Precisely, any $v \in H_0^1(\Omega)$ is expanded as

$$v = \sum_{q=1}^{Q} \hat{v}_q\psi_q + \sum_{K\in\mathcal{T}} \sum_{k=2}^{\infty} \hat{v}_{K,k}\phi_{K,k} =: \sum_{\lambda\in\mathbb{L}} \hat{v}_\lambda\varphi_\lambda \ ,$$

so that

$$\|v\|^2_{H_0^1(\Omega)} = \sum_{q=1}^{Q} |\hat{v}_q|^2 + \sum_{K\in\mathcal{T}} \sum_{k=2}^{\infty} |\hat{v}_{K,k}|^2 =: \sum_{\lambda\in\mathbb{L}} |\hat{v}_\lambda|^2 \ , \tag{4.9}$$

where the new notation on the right-most side has been introduced for subsequent convenience. Similarly, for any $F \in H^{-1}(\Omega)$, we set

$$\hat{F}_q = \langle F, \psi_q\rangle \qquad \text{and} \qquad \hat{F}_{K,k} = \langle F, \phi_{K,k}\rangle \ ,$$

and we obtain

$$\|F\|^2_{H^{-1}(\Omega)} = \sum_{q=1}^{Q} |\hat{F}_q|^2 + \sum_{K\in\mathcal{T}} \sum_{k=2}^{\infty} |\hat{F}_{K,k}|^2 =: \sum_{\lambda\in\mathbb{L}} |\hat{F}_\lambda|^2 \ . \tag{4.10}$$

The formal analogy between (4.9) and (4.10) suggests us to use the notation $\| \ . \ \|$ to indicate both the $H_0^1(\Omega)$-norm of a function v, or the $H^{-1}(\Omega)$-norm of a linear form F; the specific meaning will be clear from the context.

Moreover, given any finite index set $\Lambda \subset \mathbb{L}$, we define the subspace of $H_0^1(\Omega)$

$$V_\Lambda := \operatorname{span}\{\varphi_\lambda \,|\, \lambda \in \Lambda\} \ ;$$

we set $|\Lambda| = \operatorname{card}\Lambda$, so that $\dim \mathrm{V}_\Lambda = |\Lambda|$. If g admits an expansion $g = \sum_{\lambda\in\mathbb{L}} \hat{g}_\lambda\varphi_\lambda$ (converging in an appropriate norm), then we define its projection $P_\Lambda g$ onto V_Λ by setting

$$P_\Lambda g = \sum_{\lambda\in\Lambda} \hat{g}_\lambda\varphi_\lambda \ .$$

Also note that if $r = r(v_\Lambda)$ is a residual, then its norm is given by

$$\|r\|^2 = \sum_{\lambda\in\mathbb{L}} |\hat{r}_\lambda|^2 \ , \tag{4.11}$$

with $\hat{r}_\lambda = \langle f, \varphi_\lambda\rangle - a(v_\Lambda, \varphi_\lambda)$.

4.1.1 Algebraic representation and properties of the stiffness matrix

Let us introduce the semi-infinite, symmetric and positive-definite matrix

$$\mathbf{A} = (a_{\lambda,\mu})_{\lambda,\mu\in\mathbb{L}} \qquad \text{with} \qquad a_{\lambda,\mu} = a(\varphi_\mu, \varphi_\lambda) \ . \tag{4.12}$$

Then, Problem (4.2) can be equivalently written as

$$\mathbf{A}\mathbf{u} = \mathbf{f} \ , \tag{4.13}$$

where the vectors $\mathbf{u} = (\hat{u}_\mu)$ and $\mathbf{f} = (\hat{f}_\lambda)$ collect, respectively, the coefficients of the solution $u = \sum_\mu \hat{u}_\mu \varphi_\mu$ of Problem (4.2), and of the right-hand side f.

For any element $K \in \mathcal{T}$, let us denote by $\mathbf{A}_K$ the square block of $\mathbf{A}$ associated with the basis functions $\{\phi_{K,n}\}$ in K, i.e.,

$$\mathbf{A}_K = (a_{\lambda,\mu})_{\lambda,\mu\in\mathbb{L}(K)} \qquad \text{where} \quad \mathbb{L}(K) = \{\lambda \in \mathbb{L} \ : \ \lambda = (K,k) \text{ for some } k \geq 2\} \ .$$

For convenience, let us write $k = k(\lambda)$.

Let us assume that the restrictions to any $K \in \mathcal{T}$ of the coefficients ν and σ of the differential operator in (4.2) are analytic functions, so that their (scaled) Legendre coefficients in K decay at an exponential rate. Then, one can prove the existence of strictly positive constants η_A and c_A such that

$$|a_{\lambda,\mu}| \leq c_A \mathrm{e}^{-\eta_A |k(\lambda)-k(\mu)|} \qquad \forall \lambda, \mu \in \mathbb{L}(K) \ . \tag{4.14}$$

We say that any $\mathbf{A}_K$ belongs to the exponential class $\mathcal{D}_e(\eta_A, c_A)$. Note that it is not restrictive to assume η_A and c_A independent of K, since the partition $\mathcal{T}$ has been fixed once and for all.

The following properties hold (see [14, 15]).

Proposition 4.1. *Assume that the constant c_A satisfying (4.14) is such that for each $K \in \mathcal{T}$*

$$c_A < \frac{1}{2}(\mathrm{e}^{\eta_A} - 1) \min_{\lambda\in\mathbb{L}(K)} a_{\lambda,\lambda} \ . \tag{4.15}$$

Then each $\mathbf{A}_K$ is invertible and $\mathbf{A}_K^{-1} \in \mathcal{D}_e(\bar{\eta}_A, \bar{c}_A)$ for some $\bar{\eta}_A \in (0, \eta_A]$ and $\bar{c}_A > 0$. □

Proposition 4.2. *For any $K \in \mathcal{T}$ and any integer $J \geq 0$, the truncated matrices $(\mathbf{A}_K)_J$ such that*

$$((\mathbf{A}_K)_J)_{\lambda,\mu} = \begin{cases} a_{\lambda,\mu} & \text{if } |k(\lambda) - k(\mu)| \leq J \ , \\ 0 & \text{elsewhere} \ , \end{cases} \tag{4.16}$$

satisfy the inequalities

$$\|\mathbf{A}_K - (\mathbf{A}_K)_J\| \leq C_{\mathbf{A}} \mathrm{e}^{-\eta_A J}$$

for some constant $C_{\mathbf{A}} > 0$. Furthermore, under the assumptions of Proposition 4.1, there exists a constant $\bar{C}_{\mathbf{A}} > 0$ such that

$$\|\mathbf{A}_K^{-1} - (\mathbf{A}_K^{-1})_J\| \leq \bar{C}_{\mathbf{A}} \mathrm{e}^{-\bar{\eta}_A J} \ . \tag{4.17}$$

4.2 The constitutive elements of an adaptive algorithm

We are going to present an adaptive algorithm which implements the following recursion: (i) compute a Galerkin approximation of the exact solution; (ii) compute the corresponding residual, actually a feasible (finite-dimensional) version of it, so that its norm can be taken as an error estimator; (iii) apply Dörfler's marking, also known as bulk-chasing, to the components of the residual in order to identify a set of new basis functions to be activated for the next Galerkin solve; (iv) expand this set using properties of the stiffness matrix of the problem; (v) compute the new Galerkin solution on the enriched finite-dimensional subspace; (vi) get rid of the negligible components of this solution by applying a coarsening procedure.

We anticipate that step (iv) guarantees an arbitrarily large error reduction, whereas step (vi) assures a quasi-optimal complexity count.

We us now introduce the specific procedures, which will enter the definition of our adaptive algorithm.

- $u_\Lambda := \mathbf{GAL}(\Lambda)$
  Given a finite subset $\Lambda \subset \mathbb{L}$, the output $u_\Lambda \in V_\Lambda$ is the solution of the Galerkin problem .

$$u_\Lambda \in V_\Lambda \quad : \quad a(u_\Lambda, v_\Lambda) = \langle f, v_\Lambda \rangle \qquad \forall v_\Lambda \in V_\Lambda \, . \tag{4.18}$$

- $r := \mathbf{RES}(v_\Lambda)$
  Given a function $v_\Lambda \in V_\Lambda$ for some finite index set Λ, the output r is, in an ideal algorithm, the residual $r(v_\Lambda) = f - Av_\Lambda$. In a feasible version, the output, say $\tilde{r}$ is a function with a finite expansion along the chosen basis, obtained by suitably approximating the right-hand side f and the image Av_Λ; it satisfies the inequality

$$\|r - \tilde{r}\| \leq \gamma \|\tilde{r}\|$$

  for some fixed constant $\gamma \in (0, 1)$. In the following, we restrict ourselves to the ideal case where the residual is assumed to be computed exactly.

- $\Lambda^* := \mathbf{DÖRFLER}(r, \theta)$
  Given $\theta \in (0, 1)$ and an element $r \in H^{-1}(I)$, the ouput $\Lambda^* \subset \mathbb{L}$ is a finite set of minimal cardinality such that the inequality

$$\|P_{\Lambda^*} r\| \geq \theta \|r\| \, , \tag{4.19}$$

  or equivalently

$$\|r - P_{\Lambda^*} r\| \leq \sqrt{1 - \theta^2} \|r\| \, , \tag{4.20}$$

  is satisfied. In terms of expansion coefficients, condition (4.19) can be equivalently stated as

$$\sum_{\lambda \in \Lambda^*} |\hat{r}_\lambda|^2 \geq \theta^2 \sum_{\lambda \in \mathbb{L}} |\hat{r}_\lambda|^2 \, . \tag{4.21}$$

  Thus, the output set Λ^* of minimal cardinality can be immediately determined by a greedy algorithm, i.e., by rearranging the coefficients $\hat{r}_\lambda$ in non-increasing order of modulus and retaining the largest ones until (4.21) is fulfilled.

- $\Lambda^* := \mathbf{ENRICH}(\Lambda, J)$
  Given an integer $J \geq 0$ and a finite set $\Lambda \subset \mathbb{L}$, the output is the set

$$\Lambda^* := \{\mu = (K,k) \in \mathbb{L} \ : \ \text{ there exists } \lambda = (K,k') \in \Lambda \text{ such that } |k-k'| \leq J\} \ .$$

  Note that Λ is enriched element-by-element with respect to the fixed partition $\mathcal{T} = \{K\}$ of Ω.

- $\Lambda^* := \mathbf{E\text{-}DÖRFLER}(r, \theta)$
  The two previous procedures are combined as follows. Given $\theta \in (0,1)$ and an element $r \in H^{-1}(I)$, the ouput $\Lambda^* \subset \mathbb{L}$ is defined by the sequence

$$\begin{aligned} \widetilde{\Lambda} :=& \mathbf{DÖRFLER}(r, \theta) \\ \Lambda^* :=& \mathbf{ENRICH}(\widetilde{\Lambda}, J_\theta) \ , \end{aligned} \tag{4.22}$$

  where, based on Proposition 4.2, J_θ is chosen as the smallest integer which satisfies

$$\bar{C}_{\mathbf{A}} \mathrm{e}^{-\bar{\eta}_A J} \leq \sqrt{\frac{1-\theta^2}{\alpha_* \alpha^*}} \tag{4.23}$$

  (see [15] for more details).

- $\Lambda := \mathbf{COARSE}(w, \epsilon)$
  Given a function $w \in V_{\Lambda^*}$ for some finite index set Λ^*, and an accuracy $\epsilon > 0$ which is known to satisfy $\|u - w\| \leq \epsilon$, the output $\Lambda \subseteq \Lambda^*$ is a set of minimal cardinality such that

$$\|w - P_\Lambda w\| \leq 2\epsilon \ , \tag{4.24}$$

  which obviously implies $\|u - P_\Lambda w\| \leq 3\epsilon$.

4.3 An adaptive algorithm with convergence rate

We are ready to present our adaptive algorithm. Each iteration can be viewed as a prediction step, based on the inspection of the current residual and the application of (enriched) Dörfler marking, followed by a correction step, based on coarsening. For this reason, we call it **PC-ADLEG** – Predictor-Corrector ADaptive LEGendre algorithm.

Given two parameters $\theta \in (0,1)$ and $tol \in [0,1)$, let us define

Algorithm PC-ADLEG$(\theta, \ tol)$

Set $r_0 := f$, $\Lambda_0 := \emptyset$, $n = -1$

do

$n \leftarrow n+1$
$\widehat{\partial\Lambda}_n := \mathbf{E\text{-}DÖRFLER}(r_n, \theta)$
$\widehat{\Lambda}_{n+1} := \Lambda_n \cup \widehat{\partial\Lambda}_n$

$\widehat{u}_{n+1} := \mathbf{GAL}(\widehat{\Lambda}_n)$

$\Lambda_{n+1} := \mathbf{COARSE}\left(\widehat{u}_{n+1}, \frac{2}{\alpha_*}\sqrt{1-\theta^2}\|r_n\|\right)$

$u_{n+1} := \mathbf{GAL}(\Lambda_{n+1})$

$r_{n+1} := \mathbf{RES}(u_{n+1})$

while $\|r_{n+1}\| > tol$

The following convergence result can be proven, by adapting the arguments given in [15] for the single-element case.

Theorem 4.1. *Let $0 < \theta < 1$ be chosen so that*

$$\rho = \rho(\theta) = 6\frac{\alpha^*}{\alpha_*}\sqrt{1-\theta^2} < 1 \,. \tag{4.25}$$

If the assumptions of Proposition 4.1 are fulfilled, the sequence of errors $u - u_n$ generated for $n \geq 0$ by the algorithm satisfies the inequality

$$|\!|\!| u - u_{n+1} |\!|\!| \leq \rho |\!|\!| u - u_n |\!|\!| \,.$$

Thus, for any $tol > 0$ the algorithm terminates in a finite number of iterations, whereas for $tol = 0$ the sequence u_n converges to u in $H^1(I)$ as $n \to \infty$.

Note that the rate of decay of the error can be brought as close to 0 as desired by choosing θ close enough to 1. This is a feature stemming from the Enrichment procedure, keeping into account the decay properties of the inverse of the stiffness matrices $\mathbf{A}_K$, $K \in \mathcal{T}$.

4.4 Nonlinear approximation in Gevrey spaces

In order to estimate the complexity of our algorithm, and evaluate its optimality, we have to make assumptions on the structure of the solution u. Precisely, we have to make assumptions on the minimal number of degrees of freedom (i.e., active basis functions) needed to build an approximation of u within a given tolerance. This is usually expressed as the condition that u belongs to a suitable *sparsity class*. Once this is done, we can compare the number of degrees of freedom activated by our algorithm at a certain iteration (actually, an estimate of this number) to the minimal number of degrees of freedom needed to obtain the same accuracy; optimality usually means that the two numbers are within a constant independent of the solution and the current iteration.

Sparsity classes typically involved in finite-order approximations such as wavelets or h-type finite elements describe an *algebraic* decay of the best approximation error vs the number of activated degrees of freedom. Hereafter, we will rather consider sparsity classes describing an *exponential* decay of that error; this choice is coherent with the nature of our discretization approach, which uses an infinite-order spectral-element method, or p-type finite element method, hence providing faster-than-algebraic decay of the error whenever the solution is piecewise smooth on the partition $\mathcal{T}$ of the domain.

The definition of sparsity class is based on the concept of *best N-term approximation error*, that we now recall. Given any nonempty finite index set $\Lambda \subset \mathbb{L}$ and the corresponding subspace

$V_\Lambda \subset V$ of dimension $|\Lambda| = \text{card}\,\Lambda$, the best approximation of v in V_Λ is the orthogonal projection of v upon V_Λ, i.e. the function $P_\Lambda v = \sum_{\lambda\in\Lambda} \hat{v}_\lambda \varphi_\lambda$, which satisfies

$$\|v - P_\Lambda v\| = \left(\sum_{\lambda \notin \Lambda} |\hat{v}_\lambda|^2 \right)^{1/2} .$$

For any integer $N \geq 1$, we minimize this error over all possible choices of Λ with cardinality N, thereby leading to the best N-term approximation error

$$E_N(v) = \inf_{\Lambda\subset\mathbb{L},\ |\Lambda|=N} \|v - P_\Lambda v\| .$$

A way to construct a best N-term approximation v_N of v consists of rearranging the coefficients of v in decreasing order of modulus

$$|\hat{v}_{\lambda_1}| \geq \ldots \geq |\hat{v}_{\lambda_n}| \geq |\hat{v}_{\lambda_{n+1}}| \geq \ldots$$

and setting $v_N = P_{\Lambda_N} v$ with $\Lambda_N = \{\lambda_n \ : \ 1 \leq n \leq N\}$.

We are ready to give the following fundamental definition.

Definition 4.1. *Given two real numbers $\eta > 0$ and $t \in (0, 1]$, we denote by $\mathcal{A}_G^{\eta,t}(\Omega, \mathcal{T})$ the set defined as*

$$\mathcal{A}_G^{\eta,t}(\Omega, \mathcal{T}) := \left\{ v \in V = H_0^1(\Omega) \ : \ \|v\|_{\mathcal{A}_G^{\eta,t}(\Omega,\mathcal{T})} := \sup_{N\geq 0} \ E_N(v)\, \mathrm{e}^{\eta N^t} < +\infty \right\} .$$

As shown in [14], the set $\mathcal{A}_G^{\eta,t}(\Omega, \mathcal{T})$ is not a vector space, since it may happen that $u,\ v$ belong to this set, whereas $u + v$ does not; however, one can show that $u + v \in \mathcal{A}_G^{\bar{\eta},t}(\Omega, \mathcal{T})$ with $\bar{\eta} = 2^{-t}\eta$.

The quantity $\|v\|_{\mathcal{A}_G^{\eta,t}(\Omega,\mathcal{T})}$ dictates the minimal number N_ε of basis functions needed to approximate v with accuracy ε. In fact, from the relations

$$E_{N_\varepsilon}(v) \leq \varepsilon < E_{N_\varepsilon - 1}(v) \leq \mathrm{e}^{-\eta(N_\varepsilon - 1)^t} \|v\|_{\mathcal{A}_G^{\eta,t}(\Omega,\mathcal{T})} , \tag{4.26}$$

we obtain

$$N_\varepsilon \leq \frac{1}{\eta^{1/t}} \left(\log \frac{\|v\|_{\mathcal{A}_G^{\eta,t}(\Omega,\mathcal{T})}}{\varepsilon} \right)^{1/t} + 1 . \tag{4.27}$$

In order to motivate our definition, let us first assume that $\mathcal{T} = \{\Omega\}$, i.e., let us concentrate on a single element. Then, inspired by [7], one can introduce the following family of spaces of Gevrey type: given any $v \in V$, let $v = \sum_{k\geq 2} \hat{v}_k \phi_k$ be its expansion along the Babuška-Shen basis defined as in (4.8) relative to the interval Ω. Then, we set

$$A^{\eta,t}(\Omega) = \{ v \in V \ : \ \text{there exists a constant } C > 0 \text{ such that } \ |\hat{v}_k| \leq C\mathrm{e}^{-\eta k^t} \ \forall k \geq 2 \} .$$

It is well-known that for $t = 1$ we get analytic functions in a neighborhood of $\overline{\Omega}$. A slightly stronger family of spaces is represented by the Sobolev-Gevrey spaces (see [28]; see also [39]) defined as

$$G^{\eta,t}(\Omega) = \{ v \in V : \quad \|v\|^2_{G^{\eta,t}(\Omega)} := \sum_{k=2}^{\infty} e^{2\eta k^t} |\hat{v}_k|^2 < +\infty \} . \tag{4.28}$$

We immediately observe that $G^{\eta,t}(\Omega) \subset A^{\eta,t}(\Omega)$. Furthermore, given any $v \in G^{\eta,t}(\Omega)$ and approximating it by the linear projection

$$P_N v = \sum_{k=2}^{N} \hat{v}_k \phi_k \ ,$$

we immediately get

$$E_N(v) \leq \|v - P_N v\| \leq \mathrm{e}^{-\eta N^t} \|v\|_{G^{\eta,t}(\Omega)} \ ,$$

which implies $G^{\eta,t}(\Omega) \subset \mathcal{A}_G^{\eta,t}(\Omega, \mathcal{T})$. Thus, the latter space contains, in particular, analytic functions and Gevrey functions.

Let us now consider an arbitrary finite partition $\mathcal{T}$ of Ω. For any $K \in \mathcal{T}$, let $v \in H_0^1(K)$; its best N-term approximation error in K is defined as

$$E_{K,N}(v) := \inf_{\Lambda \subset \mathbb{L}(K),\ |\Lambda| = N} \|v - P_\Lambda v\|_{H_0^1(K)} \ .$$

Consequently, we can define the class $\mathcal{A}_G^{\eta,t}(K)$ by setting

$$\mathcal{A}_G^{\eta,t}(K) := \left\{ v \in H_0^1(K) \ : \ \|v\|_{\mathcal{A}_G^{\eta,t}(K)} := \sup_{N \geq 0} E_{K,N}(v)\, e^{\eta N^t} < +\infty \right\} .$$

Given any $v \in H_0^1(\Omega)$, denote by $v_L \in V_L(\mathcal{T})$ its piecewise linear interpolant, and set $\tilde{v} = v - v_L$, so that $\tilde{v}_{|K} \in H_0^1(K)$ for all $K \in \mathcal{T}$. Now, assume that $v \in \mathcal{A}_G^{\eta,t}(\Omega, \mathcal{T})$, and let w be a best N-term approximation of v, i.e., a linear combination of at most N basis functions (we will write $|\operatorname{supp} w| \leq N$) such that

$$\|v - w\| \leq \mathrm{e}^{-\eta N^t} \|v\|_{\mathcal{A}_G^{\eta,t}(\Omega,\mathcal{T})} \ .$$

Writing $v - w = (v - v_L) - (w - w_L) + (v_L - w_L) = \tilde{v} - \tilde{w} + z_L$ and using the orthogonality of the basis functions, we have

$$\|v - w\|^2 = \|\tilde{v} - \tilde{w} + z_L\|^2 \geq \|\tilde{v} - \tilde{w}\|^2 = \sum_{K \in \mathcal{T}} \|\tilde{v}_K - \tilde{w}_K\|^2_{H_0^1(K)} \ ,$$

where the appended K denotes restriction of a function to K; thus,

$$\|\tilde{v}_K - \tilde{w}_K\|_{H_0^1(K)} \leq \mathrm{e}^{-\eta N^t} \|v\|_{\mathcal{A}_G^{\eta,t}(\Omega,\mathcal{T})} \qquad \forall K \in \mathcal{T}$$

and since $|\operatorname{supp} w_K| \leq |\operatorname{supp} w| \leq N$, we deduce that $\tilde{v}_K \in \mathcal{A}_G^{\eta,t}(K)$ for all $K \in \mathcal{T}$.

On the other hand, let $v \in H_0^1(\Omega)$ be such that $\tilde{v}_{|K} \in \mathcal{A}_G^{\tilde{\eta},t}(K)$ for all $K \in \mathcal{T}$, for some $\tilde{\eta} > 0$ to be determined later on. Then, there exist a constant $C > 0$ and functions $\tilde{w}_K \in H_0^1(K)$ with $|\operatorname{supp} \tilde{w}_K| \leq N$ such that

$$\|\tilde{v}_K - \tilde{w}_K\|_{H_0^1(K)} \leq C\mathrm{e}^{-\eta N^t} \qquad \forall K \in \mathcal{T} \ .$$

Denoting by $\tilde{w}$ the function in Ω which coincides with $\tilde{w}_K$ in each K, and setting $w = v_L + \tilde{w}$, we have

$$\|v - w\|^2 = \|\tilde{v} - \tilde{w}\|^2 = \sum_{K \in \mathcal{T}} \|\tilde{v}_K - \tilde{w}_K\|^2_{H_0^1(K)} \leq (Q+1) C^2 \mathrm{e}^{-2\tilde{\eta} N^t} \ .$$

Now, observe that $|\operatorname{supp} w| = |\operatorname{supp} v_I| + \sum_{K \in \mathcal{T}} |\operatorname{supp} \tilde{w}_K| \leq Q + (Q+1)N \leq (Q+2)N$. Choosing $\tilde{\eta} = (Q+2)^t \eta$ and letting $N \to \infty$, we conclude that $v \in \mathcal{A}_G^{\eta,t}(\Omega, \mathcal{T})$.

4.5 Complexity analysis of the algorithm

We are now ready to investigate complexity issues for the sequence of approximations $u_n = u_{\Lambda_n}$ generated by **PC-ADLEG**, under the assumption that the solution u belongs to a class $\mathcal{A}_G^{\eta,t}(\Omega,\mathcal{T})$ for some $\eta > 0$ and $t \in (0,1]$.

At first we note that each set Λ_{n+1} of the active degrees of freedom produced by the algorithm is generated by the procedure **COARSE** with a suitable tolerance ε_n. A general result about coarsening (see, e.g., [52]) allows us to estimate its cardinality $|\Lambda_{n+1}| = N_{\varepsilon_n}$ according to (4.27). On the other hand, one can prove that $\|u - u_{n+1}\| \lesssim \varepsilon_n$. Thus, we obtain the following optimal result.

Theorem 4.2. *Suppose that $u \in \mathcal{A}_G^{\eta,t}$, for some $\eta > 0$ and $t \in (0,1]$. Then, there exists a constant $C > 1$ such that the cardinality of the set Λ_{n+1} of the active degrees of freedom produced by* **PC-ADLEG** *satisfies the bound*

$$|\Lambda_{n+1}| \leq \frac{1}{\eta^{1/t}} \left(\log \frac{\|u\|_{\mathcal{A}_G^{\eta,t}}}{\|u - u_{n+1}\|} + \log C \right)^{1/t} + 1 \,, \qquad \forall\ n \geq 0.$$

Next, we focus on the cardinality of the intermediate set $\widehat{\Lambda}_{n+1}$, which depends on that of the incremental set $\widehat{\partial\Lambda}_{n+1}$; in turns, this can be bounded by $2J_\theta$ times the cardinality of the incremental set $\widetilde{\partial\Lambda}_{n+1}$ generated by **DÖRFLER** with residual r_n. Although under certain assumptions on θ it is possible to estimate such cardinality in terms of the sparsity class of the solution (see [52]), in the most general situation as the one we want to consider here, it is the sparsity class of the residual that influences the growth of degrees of freedom. Indeed, we recall that the step

$$\partial\Lambda := \mathbf{DÖRFLER}(r,\theta)$$

selects a set $\partial\Lambda$ of minimal cardinality in $\mathbb{L} \setminus \Lambda$ for which $\|r - P_{\partial\Lambda} r\| \leq \sqrt{1-\theta^2}\|r\|$. In other words, it performs a best approximation of the residual for the accuracy $\varepsilon = \sqrt{1-\theta^2}\|r\|$. Thus, if r belongs to a certain sparsity class $\mathcal{A}_G^{\bar\eta,\bar t}(\Omega,\mathcal{T})$ for some $\bar\eta > 0$ and $\bar t > 0$, we have by (4.27)

$$|\partial\Lambda| \leq \frac{1}{\bar\eta^{1/\bar t}} \left(\log \frac{\|r\|_{\mathcal{A}_G^{\bar\eta,\bar t}}}{\sqrt{1-\theta^2}\ \|r\|} \right)^{1/\bar t} + 1 \,. \tag{4.29}$$

Thus, it make sense to investigate the sparsity class of the residual. In a sparsity class of algebraic type, this is the same as the class of the solution (see again [52]). Unfortunately, in a sparsity class of exponential type such a property does not hold [14], and we have to expect the generic residual to be less sparse than the exact solution.

The best result we can expect is as follows.

Proposition 4.3. *Let $v \in \mathcal{A}_G^{\eta,t}(\Omega,\mathcal{T})$ for some $\eta > 0$ and $t \in (0,1]$. Assume that $\eta < \eta_A$, where η_A is the constant for which (4.14) holds. Let us set*

$$\bar\eta = \zeta(t)\eta \,, \qquad \bar t = \frac{t}{1+t} \,,$$

where we define

$$\zeta(t) := \left(\frac{1+t}{2}\right)^{\frac{t}{1+t}} \qquad \forall\ 0 < t \leq 1. \tag{4.30}$$

Then, one has $Av \in \mathcal{A}_G^{\bar{\eta},\bar{t}}(\Omega,\mathcal{T})$*, with*

$$\|Av\|_{\mathcal{A}_G^{\bar{\eta},\bar{t}}(\Omega,\mathcal{T})} \lesssim \|v\|_{\mathcal{A}_G^{\eta,t}(\Omega,\mathcal{T})} . \tag{4.31}$$

Under the sparsity assumption on the solution u made in the previous theorem, this implies that $f = Au \in \mathcal{A}_G^{\bar{\eta},\bar{t}}(\Omega,\mathcal{T})$. On the other hand, it is possible to prove that any Galerkin solution produced by **PC-ADLEG** satisfies $\|u_n\|_{\mathcal{A}_G^{\eta,t}(\Omega,\mathcal{T})} \lesssim \|u\|_{\mathcal{A}_G^{\eta,t}(\Omega,\mathcal{T})}$, so that $Au_n \in \mathcal{A}_G^{\bar{\eta},\bar{t}}(\Omega,\mathcal{T})$. Keeping into account the remark after Definition 4.1, we obtain the following result.

Proposition 4.4. *Let* $u \in \mathcal{A}_G^{\eta,t}(\Omega,\mathcal{T})$ *for some* $\eta > 0$ *and* $t \in (0,1]$*. There exists* $\tilde{\eta} \leq \eta$ *such that* $r_n = r(u_n) \in \mathcal{A}_G^{\tilde{\eta},\bar{t}}(\Omega,\mathcal{T})$ *with*

$$\|r_n\|_{\mathcal{A}_G^{\tilde{\eta},\bar{t}}(\Omega,\mathcal{T})} \lesssim \|u\|_{\mathcal{A}_G^{\eta,t}(\Omega,\mathcal{T})} .$$

Using (4.29), we arrive at the following final estimate.

Theorem 4.3. *Suppose that* $u \in \mathcal{A}_G^{\eta,t}(\Omega,\mathcal{T})$ *for some* $\eta > 0$ *and* $t \in (0,1]$ *and that the assumptions of Proposition 4.1 are satisfied. Then, there exist positive constants* $\tilde{\eta} \leq \eta$*,* $\bar{t} \leq t$ *and* C *such that the cardinality of the intermediate sets* $\widehat{\Lambda}_{n+1}$ *activated in the predictor step of* **PC-ADLEG** *can be estimated as*

$$|\widehat{\Lambda}_{n+1}| \leq |\Lambda_n| + \frac{2J_\theta}{\tilde{\eta}^{1/\bar{t}}} \left(\log \frac{\|u\|_{\mathcal{A}_G^{\eta,t}}}{\|u - u_{n+1}\|} + \log C \right)^{1/\bar{t}} + 2J_\theta , \qquad \forall\ n \geq 0 .$$

Keeping into account the conditions on $\tilde{\eta}$ and $\bar{t}$, we expect the cardinality of $|\widehat{\Lambda}_{n+1}|$ to be asymptotically larger than the optimal one of $|\Lambda_{n+1}|$, estimated in Theorem 4.2. Precisely for this reason, a coarsening step has been added at the end of each adaptive iteration: coarsening brings complexity from the one dictated by the sparsity class of the residual back to the one associated with the exact solution. On the other hand, we consider such intermediate loss of optimality to be worth of being accepted, since it should be compensated by the fast convergence of our algorithm, guaranteed by the allowed aggressive policy of degree of freedom enrichment.

We mention that the sparsity class of the residual influences complexity even in other instances of the algorithm, not discussed here. For instance, this is the case when a feasible computation of the residual-based error estimator is considered: to avoid degradation of the contraction property of the algorithm, approximate finite-dimensional residuals should be sufficiently close to the exact ones, which can be obtained with a complexity related to the sparseness of the residuals themselves. We refer to [14] for more details.

References

[1] M. Ainsworth and J. T. Oden. A procedure for a posteriori error estimation for *h-p* finite element methods. *Comput. Methods Appl. Mech. Engrg.*, 101(1-3):73–96, 1992.

[2] M. Ainsworth and J. T. Oden. A unified approach to a posteriori error estimation using element residual methods. *Numer. Math.*, 65(1):23–50, 1993.

[3] M. Ainsworth and B. Senior. Aspects of an adaptive *hp*-finite element method: adaptive strategy, conforming approximation and efficient solvers. *Comput. Methods Appl. Mech. Engrg.*, 150(1-4):65–87, 1997. Symposium on Advances in Computational Mechanics, Vol. 2 (Austin, TX, 1997).

[4] M. Ainsworth and B. Senior. An adaptive refinement strategy for *hp*-finite element computations. *Appl. Numer. Math.*, 26(1-2):165–178, 1998.

[5] I. Babuška and B. Q. Guo. Regularity of the solutions of elliptic problems with piecewise analytic data. *SIAM J. Numer. Anal.*, 20:763–781, 1989.

[6] I. Babuška and B. Q. Guo. Approximation properties of the *h-p* version of the finite element method. *Comput. Methods Appl. Mech. Engrg.*, 133(3-4):319–346, 1996.

[7] M. S. Baouendi and C. Goulaouic. Régularité analytique et itérés d'opérateurs elliptiques dégénérés; applications. *J. Functional Analysis*, 9:208–248, 1972.

[8] C. Bernardi. Indicateurs d'erreur en *h-N* version des éléments spectraux. *RAIRO Modél. Math. Anal. Numér.*, 30(1):1–38, 1996.

[9] P. Binev, W. Dahmen, and R. DeVore. Adaptive finite element methods with convergence rates. *Numer. Math.*, 97(2):219–268, 2004.

[10] D. Braess, V. Pillwein, and J. Schöberl. Equilibrated residual error estimates are *p*-robust. *Comput. Methods Appl. Mech. Engrg.*, 198(13-14):1189–1197, 2009.

[11] M. Bürg and W. Dörfler. Convergence of an adaptive *hp* finite element strategy in higher space-dimensions. *Appl. Numer. Math.*, 61(11):1132–1146, 2011.

[12] C. Canuto, M.Y. Hussaini, A. Quarteroni, and T.A. Zang. *Spectral Methods. Fundamentals in Single Domains*. Scientific Computation. Springer-Verlag, Berlin, 2006.

[13] C. Canuto, M.Y. Hussaini, A. Quarteroni, and T.A. Zang. *Spectral methods. Evolution to Complex Geometries*. Scientific Computation. Springer-Verlag, Berlin, 2007.

[14] C. Canuto, R.H. Nochetto, and M. Verani. Adaptive Fourier-Galerkin Methods. *arXiv:1201.5648v1, submitted*, pages 1–48, 2011.

[15] C. Canuto, R.H. Nochetto, and M. Verani. Contraction and optimality properties of adaptive legendre-galerkin methods: the 1-dimensional case. *arXiv:1206.5524v1, submitted*, pages 1–26, 2011.

[16] J. M. Cascon, C. Kreuzer, R. H. Nochetto, and K. G. Siebert. Quasi-optimal convergence rate for an adaptive finite element method. *SIAM J. Numer. Anal.*, 46(5):2524–2550, 2008.

[17] A. Cohen, W. Dahmen, and R. DeVore. Adaptive wavelet methods for elliptic operator equations – convergence rates. *Math. Comp*, 70:27–75, 1998.

[18] A. Cohen, R. DeVore, and R.H. Nochetto. Convergence rates for afem with H^{-1} data. *Found. Comput. Math., to appear*, 2011.

[19] M. Costabel, M. Dauge, and S. Nicaise. Analytic regularity for linear elliptic systems in polygons and polyhedra. *Math. Models Methods Appl. Sci.*, 22(8), 2012.

[20] W. Dahmen and K. Scherer. Best approximation by piecewise polynomials with variable knots and degrees. *J. Approx. Theory*, 26(1):1–13, 1979.

[21] W. Dahmen and K. Scherer. On optimal global error bounds obtained by scaled local error estimates. *Numer. Math.*, 36:151–176, 1981.

[22] L. Demkowicz, J. T. Oden, W. Rachowicz, and O. Hardy. Toward a universal *h-p* adaptive finite element strategy. I. Constrained approximation and data structure. *Comput. Methods Appl. Mech. Engrg.*, 77(1-2):79–112, 1989.

[23] L. Demkowicz, W. Rachowicz, and Ph. Devloo. A fully automatic *hp*-adaptivity. *J. Sci. Comput.*, 17(1-4):127–155, 2002.

[24] R. DeVore and K. Scherer. Variable knot, variable degree spline approximation to x^β. In *Quantitative approximation (Proc. Internat. Sympos., Bonn, 1979)*, pages 121–131. Academic Press, New York, 1980.

[25] W. Dörfler. A convergent adaptive algorithm for Poisson's equation. *SIAM J. Numer. Anal.*, 33(3):1106–1124, 1996.

[26] W. Dörfler and V. Heuveline. Convergence of an adaptive *hp* finite element strategy in one space dimension. *Appl. Numer. Math.*, 57(10):1108–1124, 2007.

[27] T. Eibner and J. M. Melenk. An adaptive strategy for *hp*-FEM based on testing for analyticity. *Comput. Mech.*, 39(5):575–595, 2007.

[28] C. Foias and R. Temam. Gevrey class regularity for the solutions of the Navier-Stokes equations. *J. Funct. Anal.*, 87(2):359–369, 1989.

[29] T. Gantumur, H. Harbrecht, and R. Stevenson. An optimal adaptive wavelet method without coarsening of the iterands. *Math. Comp.*, 76(258):615–629, 2007.

[30] W. Gui and I. Babuška. The *h*, *p* and *h-p* versions of the finite element method in 1 dimension. II. The error analysis of the *h*- and *h-p* versions. *Numer. Math.*, 49(6):613–657, 1986.

[31] W. Gui and I. Babuška. The *h*, *p* and *h-p* versions of the finite element method in 1 dimension. III. The adaptive *h-p* version. *Numer. Math.*, 49(6):659–683, 1986.

[32] B. Guo and I. Babuška. The *hp*-version of the finite element method i: the basic approximation results. *Comp. Mech.*, 1:21–41, 1986.

[33] B. Guo and I. Babuška. The *hp*-version of the finite element method ii: general results and applications. *Comp. Mech.*, 1:203–226, 1986.

[34] B. Guo and I. Babuška. Regularity of the solutions for elliptic problems on nonsmooth domains in $\mathbf{R^3}$. II. Regularity in neighbourhoods of edges. *Proc. Roy. Soc. Edinburgh Sect. A*, 127(3):517–545, 1997.

[35] V. Heuveline and R. Rannacher. Duality-based adaptivity in the *hp*-finite element method. *J. Numer. Math.*, 11(2):95–113, 2003.

[36] P. Houston, B. Senior, and E. Süli. *hp*-discontinuous Galerkin finite element methods for hyperbolic problems: error analysis and adaptivity. *Internat. J. Numer. Methods Fluids*, 40(1-2):153–169, 2002. ICFD Conference on Numerical Methods for Fluid Dynamics (Oxford, 2001).

[37] P. Houston and E. Süli. A note on the design of *hp*-adaptive finite element methods for elliptic partial differential equations. *Comput. Methods Appl. Mech. Engrg.*, 194(2-5):229–243, 2005.

[38] J. Kurtz and L. Demkowicz. A fully automatic *hp*-adaptivity for elliptic PDEs in three dimensions. *Comput. Methods Appl. Mech. Engrg.*, 196(37-40):3534–3545, 2007.

[39] J.-L. Lions and E. Magenes. *Problèmes aux Limites Non Homogènes et Applications - Vol. III.* Dunod, Paris, 1970.

[40] J. M. Melenk and B. I. Wohlmuth. On residual-based a posteriori error estimation in *hp*-FEM. *Adv. Comput. Math.*, 15(1-4):311–331 (2002), 2001.

[41] W.F. Mitchell and M. McCain. A comparison of *hp*-adaptive strategies for elliptic partial differential equations. *NIST Report, submitted*, pages 1–39, 2011.

[42] P. Morin, R. H. Nochetto, and K. G. Siebert. Data oscillation and convergence of adaptive FEM. *SIAM J. Numer. Anal.*, 38(2):466–488 (electronic), 2000.

[43] R. H. Nochetto, K. G. Siebert, and A. Veeser. Theory of adaptive finite element methods: an introduction. In *Multiscale, nonlinear and adaptive approximation*, pages 409–542. Springer, Berlin, 2009.

[44] J. T. Oden, L. Demkowicz, W. Rachowicz, and T. A. Westermann. Toward a universal *h-p* adaptive finite element strategy. II. A posteriori error estimation. *Comput. Methods Appl. Mech. Engrg.*, 77(1-2):113–180, 1989.

[45] J.T. Oden, A. Patra, and Y. Feng. *An hp adaptive strategy*, volume 157, pages 23–46. ASME Publication, 1992.

[46] W. Rachowicz, J. T. Oden, and L. Demkowicz. Toward a universal *h-p* adaptive finite element strategy. III. Design of *h-p* meshes. *Comput. Methods Appl. Mech. Engrg.*, 77(1-2):181–212, 1989.

[47] A. Schmidt and K. G. Siebert. A posteriori estimators for the *h-p* version of the finite element method in 1D. *Appl. Numer. Math.*, 35(1):43–66, 2000.

[48] D. Schötzau, C. Schwab, and T. Wihler. *hp*-dgfem for elliptic problems in polyhedra i: Stability and quasi-optimality on geometric meshes. *SAM Report ETHZ*, 2012.

[49] D. Schötzau, C. Schwab, and T. Wihler. *hp*-dgfem for elliptic problems in polyhedra ii: Exponential convergence. *SAM Report ETHZ*, 2012.

[50] Ch. Schwab. *p- and hp-finite element methods*. Numerical Mathematics and Scientific Computation. The Clarendon Press Oxford University Press, New York, 1998. Theory and applications in solid and fluid mechanics.

[51] R. Stevenson. Optimality of a standard adaptive finite element method. *Found. Comput. Math.*, 7(2):245–269, 2007.

[52] R. Stevenson. Adaptive wavelet methods for solving operator equations: an overview. In *Multiscale, nonlinear and adaptive approximation*, pages 543–597. Springer, Berlin, 2009.

[53] J. Valenciano and R. G. Owens. A new adaptive modification strategy for numerical solutions to elliptic boundary value problems. *Appl. Numer. Math.*, 32(3):305–329, 2000.

[54] T. P. Wihler. An *hp*-adaptive strategy based on continuous Sobolev embeddings. *J. Comput. Appl. Math.*, 235(8):2731–2739, 2011.

MOX Technical Reports, last issues

Dipartimento di Matematica "F. Brioschi",
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

36/2012 Canuto, C.; Verani, M.
On the Numerical Analysis of Adaptive Spectral/hp Methods for Elliptic Problems

35/2012 Pigoli, D.; Aston, J.A.D.; Dryden, I.L.; Secchi, P.
Distances and Inference for Covariance Functions

34/2012 Menafoglio, A.; Dalla Rosa, M.; Secchi, P.
A Universal Kriging predictor for spatially dependent functional data of a Hilbert Space

33/2012 Motamed, M.; Nobile, F.; Tempone, R.
Analysis and computation of the elastic wave equation with random coefficients

32/2012 Formaggia, L.; Fumagalli, A.; Scotti A.; Ruffo, P
A reduced model for Darcy s problem in networks of fractures

31/2012 Bonizzoni, F.; Buffa, A; Nobile, F:
Moment equations for the mixed formulation of the Hodge Laplacian with stochastic data

30/2012 Beck, J.; Nobile, F.; Tamellini, L.; Tempone, R.;
Convergence of quasi-optimal Stochastic Galerkin Methods for a class of PDES with random coefficients

29/2012 Chen, P.; Quarteroni, A.; Rozza, G.
Stochastic Optimal Robin Boundary Control Problems of Advection-Dominated Elliptic Equations

28/2012 Canuto, C.; Nochetto, R.H.; Verani, M.
Contraction and optimality properties of adaptive Legendre-Galerkin methods: the 1-dimensional case

27/2012 Pigoli, D.; Secchi,P.
Estimation of the mean for spatially dependent data belonging to a Riemannian manifold