APPROXIMATION CLASSES FOR ADAPTIVE HIGHER ORDER FINITE ELEMENT APPROXIMATION

FERNANDO D. GASPOZ AND PEDRO MORIN

Abstract. We provide an almost characterization of the approximation classes appearing when using adaptive finite elements of Lagrange type of any fixed polynomial degree. The characterization is stated in terms of Besov regularity, and requires the approximation within spaces with integrability indices below one. This article generalizes to higher order finite elements the results presented for linear finite elements by Binev *et. al.* [BDDP 2002].

1. Introduction

Adaptive finite element methods (AFEM) for the numerical solution of partial differential equations are now widely used in scientific computation, with the purpose of reducing the computational cost by yielding the automatic construction of a sequence of meshes that would equidistribute, eventually, the approximation errors, producing (quasi-)optimal meshes.

Recent articles [BDD 2004, S 2006, CKNS 2007, DX 2008, GM 2010] tackle the issue of optimality of AFEM by proving results of the following type for different stationary problems:

> Let u denote the exact solution being approximated in the norm $\|\cdot\|$. Assume that there exists $s > 0$ and a constant C such that, for each $N \in \mathbb{N}$, there is a triangulation $\mathcal{T}$ obtained with at most N bisections from the initial triangulation $\mathcal{T}_0$ such that
> $$\inf \|u - v_{\mathcal{T}}\| \leq CN^{-s} = C\big(\#\mathcal{T} - \#\mathcal{T}_0\big)^{-s},$$
> where the infimum is taken over all finite element functions $v_{\mathcal{T}}$ over the mesh $\mathcal{T}$. Then, the adaptive cycle generates a sequence $\{(\mathcal{T}_k, u_k)\}_{k\in\mathbb{N}}$ of meshes, and corresponding finite element approximations such that
> $$\|u - u_k\| \leq \hat{C}(\#\mathcal{T}_k - \#\mathcal{T}_0)^{-s}, \qquad k = 1, 2, \dots.$$

In other words, if the exact solution can be ideally approximated with complexity $O(N^{-s})$, then the AFEM generates a sequence of discrete solutions converging with the same order.

The goal of this paper is to shed some light into the understanding of the approximation classes $\mathbb{A}_s$ of functions that can be approximated with complexity of

---

2010 *Mathematics Subject Classification.* Primary 41A25, 65D05; Secondary 65N30, 65N50.

Key words and phrases. Adaptive finite elements, Besov spaces, convergence rates, approximation classes.

Partially supported by CONICET through grant PIP 112-200801-02182, Universidad Nacional del Litoral through grants CAI+D 062-312, 062-309, and Agencia Nacional de Promoción Científica y Tecnológica, through grant PICT-2008-0622 (Argentina).

order N^{-s}, with Lagrange finite elements of arbitrary (but fixed) order $r \in \mathbb{N}$ on domains of arbitrary dimension d. In [GM 2008] it is proven that the so-called optimal convergence $O(N^{-r/d})$ when measuring the error in H^1 is obtained whenever the function to approximate can be decomposed as a sum of a regular part plus a singular part with singularities around a finite number of points. This decomposition is usual in regularity results of Partial Differential Equations in terms of Sobolev norms. Nevertheless, Sobolev spaces are not sufficient to *characterize* these approximation classes.

In [BDDP 2002] an almost characterization of these classes is obtained, for the case of *linear* finite elements, in terms of Besov regularity. In this work we state and prove analogous results for higher order finite elements. For the proofs we follow the steps of [BDDP 2002], filling in all the details necessary for studying the higher order case. In particular, it is necessary to work with Besov spaces with integrability index below one (see Remark 2.4), requiring the definition of a local polynomial approximation operator $\Pi_{p,G}$ and the study of best approximation properties in L^p for $0 < p < \infty$, among others (see Definition 3.7 and Theorem 3.8); this is a key novel result of this article.

We end this introduction noticing that some of the results presented in this article are rather technical and others are known to researchers from approximation theory. The former were necessary in order to obtain a rigorous proof of the main results, the latter were included for two reasons. First, we could not find proofs of those results under the precise assumptions necessary for our argument, they were proved either for one-dimensional domains or for smooth or convex domains; we compare our results with the existing ones along their presentation throughout the article. Secondly, those results are not so familiar to the finite element community, and including them here makes this article more self-contained and easier to read for a wider audience.

In the next section we state precisely our main results, and at its end we outline the organization of the rest of this article.

2. Main Results

In this section we state the main results of this article, and start by setting the finite element framework. We consider an initial triangulation $\mathcal{T}_0$ of the polyhedral domain Ω into simplices, and we let the admissible triangulations be those belonging to the family $\mathbb{T}$ of all conforming partitions of Ω obtained from $\mathcal{T}_0$ by refinement using the bisection rules from [M 1995, Tr 1997], considered in [S 2007]. These rules coincide (after some re-labeling) with the *newest-vertex* bisection procedure in two dimensions and the bisection procedure of Kossaczký in three dimensions [K 1994], and yield a shape regular family $\mathbb{T}$:

$$\sup_{\mathcal{T}\in\mathbb{T}} \sup_{T\in\mathcal{T}} \frac{\operatorname{diam}(T)}{\rho_T} =: \kappa_{\mathbb{T}} < \infty,$$

where $\operatorname{diam}(T)$ is the diameter of T and ρ_T is the radius of the largest ball contained in it. Throughout this article, we only consider meshes $\mathcal{T}$ that belong to the family $\mathbb{T}$, so the shape regularity of all of them is bounded by the uniform constant $\kappa_{\mathbb{T}}$ which only depends on the initial triangulation $\mathcal{T}_0$ [SS 2005, S 2007]. Besides, these bisection rules ensure that the extra refinement needed to keep conformity of the meshes is controlled by the number of elements marked for refinement (see Theorem 6.1 below).

From now on, for any admissible triangulation $\mathcal{T}$, we let $\mathbb{V}_{\mathcal{T}}$ denote the finite element space of continuous piecewise polynomials of degree at most r, where r is a fixed positive integer, i.e.,

$$\mathbb{V}_{\mathcal{T}} = \left\{ v \in C(\overline{\Omega}) : v_{|T} \in \mathcal{P}^r \text{ for all } T \in \mathcal{T} \right\}.$$

Throughout the paper we fix $B_0 = B^{\alpha}_{p,p}(\Omega)$, $0 < p < \infty$, $0 < \alpha < 1 + \frac{1}{p}$ or $B_0 = L^p(\Omega)$ if $\alpha = 0$ and we recall that the Besov space $B = B^{\alpha+s}_{\tau,\tau}(\Omega)$ $(s > 0)$ is compactly embedded in B_0 if and only if [BDDP 2002]

$$\frac{1}{\tau} < \frac{s}{d} + \frac{1}{p} \qquad \text{or} \quad \delta := \frac{s}{d} + \frac{1}{p} - \frac{1}{\tau} > 0.$$

The positive number δ is called the discrepancy for B relative to B_0. The constraint $\alpha < 1 + \frac{1}{p}$ is required to guarantee that $\mathbb{V}_{\mathcal{T}} \subset B_0$, because we are considering C^0 finite elements (see Proposition 4.7).

One of the main results of this article is the existence of a quasi-interpolant $Q_{\mathcal{T}}$ with properties that are fundamental for the construction of quasi-optimal meshes. These properties involve local and global bounds for the interpolation error in terms of Besov and broken Besov norms and are explicitly stated in the following.

Proposition 2.1. *Let $B_0 = B^{\alpha}_{p,p}(\Omega)$, $0 < p < \infty$, $0 < \alpha < \min\{r+1, 1 + \frac{1}{p}\}$ or $B_0 = L^p(\Omega)$ if $\alpha = 0$. If $f \in B = B^{\alpha+s}_{\tau,\tau}(\Omega)$ with $\frac{1}{\tau} < \frac{s}{d} + \frac{1}{p}$, $s > 0$, and $s + \alpha \le r + \frac{1}{\tau_*}$, where $\tau_* = \min\{1, \tau\}$. Then, for any mesh $\mathcal{T} \in \mathbb{T}$ there exists an interpolant $Q_{\mathcal{T}} : B_0 \to \mathbb{V}_{\mathcal{T}}$, and constants C_1, C_2, C_3 such that the following inequalities hold:*

$$\|f - Q_{\mathcal{T}}(f)\|_{L^p(T)} \le C_1 |T|^{\delta} |f|_{B(\omega_{\mathcal{T}}(T))}, \tag{2.1}$$

$$|f|_{B_0(\omega_{\mathcal{T}}(T))} \le C_1 |T|^{\delta} |f|_{B(\omega_{\mathcal{T}}(T))} \qquad (\alpha > 0), \tag{2.2}$$

$$\|f - Q_{\mathcal{T}}(f)\|^p_{L^p(\Omega)} \le C_2 \sum_{T \in \mathcal{T}} |f|^p_{B_0(\omega_{\mathcal{T}}(T))}, \tag{2.3}$$

$$|f - Q_{\mathcal{T}}(f)|^p_{B_0(\Omega)} \le C_3 \sum_{T \in \mathcal{T}} |f|^p_{B_0(\omega_{\mathcal{T}}(T))} \qquad (\alpha > 0). \tag{2.4}$$

where $\delta = \frac{s}{d} + \frac{1}{p} - \frac{1}{\tau} > 0$ and $C_1 = C_1(p, s, \tau, d, r, \kappa_{\mathbb{T}})$, $C_2 = C_2(p, \rho, \alpha, d, r, \kappa_{\mathbb{T}})$ and $C_3 = C_3(p, \rho, \alpha, d, r, \mathrm{diam}(\Omega), \kappa_{\mathbb{T}})$; hereafter $\omega_{\mathcal{T}}(T)$ denotes the patch of elements of $\mathcal{T}$ that have nonempty intersection with T, $|\cdot|_{B_0(\omega)}$ is the Besov seminorm (4.7) if $\alpha > 0$ and $|\cdot|_{B_0(\omega)} = \|\cdot\|_{L^p(\omega)}$ when $\alpha = 0$.

The construction of the operator $Q_{\mathcal{T}}$ is presented in Section 3 for a generic function $f \in L^p(\Omega)$, without assuming any extra regularity. The proof of Proposition 2.1 is presented in Section 5, after the Besov spaces are introduced and its properties are discussed in Section 4.

In the rest of this article, we will use the notation $C = C(\square)$ to emphasize the dependence of the constant C upon the parameters $\square$. For the sake of simplicity, the notation $\lesssim$ will be used to indicate that $a \le Cb$ with $C > 0$ a constant depending on the parameters defined in the corresponding theorems, lemmas or propositions, also $a \simeq b$ will indicate that $a \lesssim b$ and $b \lesssim a$.

One of the main consequences of this interpolation operator and its properties is an embedding theorem between the Besov space B and an approximation class, through the construction of optimal meshes using a *greedy algorithm*. In order to define the approximation classes we introduce the notion of best approximation error with complexity $N \in \mathbb{N}$ on a quasi-Banach space B_0 of functions defined on Ω, as follows:

$$\sigma_N(u)_{B_0} = \min_{\mathcal{T}\in\mathbb{T}_N} \inf_{v\in\mathbb{V}_\mathcal{T}} \|u-v\|_{B_0},$$

where $\mathbb{T}_N := \{\mathcal{T} \in \mathbb{T} : (\#\mathcal{T} - \#\mathcal{T}_0) \le N\}$ that is, the minimum over $\mathcal{T}$ is taken over all admissible triangulations obtained with at most N bisections.

Next, we define, for $s > 0$ the approximation classes $\mathbb{A}_s(B_0)$ containing the functions that can be approximated with best approximation error of order N^{-s}. More precisely,

$$\mathbb{A}_s(B_0) = \left\{v \in B_0 : \exists C \text{ such that } \sigma_N(v)_{B_0} \le CN^{-s}, \forall N \in \mathbb{N}\right\}.$$

Equivalently, we can define $\mathbb{A}_s(B_0)$ through a semi-(quasi)norm as follows:

$$\mathbb{A}_s(B_0) = \left\{v \in B_0 : |v|_{\mathbb{A}_s(B_0)} < \infty\right\} \quad \text{with} \quad |v|_{\mathbb{A}_s(B_0)} := \sup_{N\in\mathbb{N}} N^s \sigma_N(v)_{B_0}.$$

This definition is also equivalent to saying that $v \in \mathbb{A}_s(B_0)$ if there is a constant C such that for all $\epsilon > 0$, there exists a mesh $\mathcal{T}$ that satisfies:

$$\inf_{v_\mathcal{T}\in\mathbb{V}_\mathcal{T}} \|v - v_\mathcal{T}\|_{B_0} \le \epsilon \quad \text{and} \quad (\#\mathcal{T} - \#\mathcal{T}_0) \le C\epsilon^{-\frac{1}{s}}, \tag{2.5}$$

and $|v|_{\mathbb{A}_s(B_0)}$ is equivalent to the infimum of all constants C that satisfy (2.5).

This scale of spaces can be extended adding a parameter $0 < q < \infty$ in the following way:

$$\mathbb{A}_{s,q}(B_0) := \left\{v \in B_0 : |v|_{\mathbb{A}_{s,q}(B_0)} < \infty\right\}$$
$$\text{with} \qquad |v|_{\mathbb{A}_{s,q}(B_0)} := \Big(\sum_{n\in\mathbb{N}} [\sigma_{2^n}(v)_{B_0} 2^{ns}]^q\Big)^{\frac{1}{q}}.$$

This more general class will be useful for proving the inverse estimates. We identify $\mathbb{A}_{s,\infty}(B_0) = \mathbb{A}_s(B_0)$.

An important consequence of Proposition 2.1 is that the optimal error with complexity N decays as $O(N^{-s/d})$ when a function being approximated in B_0-norm belongs to B, and the discrepancy $\delta = \frac{s}{d} + \frac{1}{p} - \frac{1}{\tau}$ is positive. More precisely,

Theorem 2.2 (Direct Theorem)**.** *Let $B_0 = B^\alpha_{p,p}(\Omega)$, $0 < p < \infty$, $0 < \alpha < \min\{r+1, 1+\frac{1}{p}\}$ or $B_0 = L^p(\Omega)$ if $\alpha = 0$. If $f \in B = B^{\alpha+s}_{\tau,\tau}(\Omega)$ with $\frac{1}{\tau} < \frac{s}{d} + \frac{1}{p}$, $s > 0$, and $s + \alpha \le r + \frac{1}{\tau_*}$, where $\tau_* = \min\{1, \tau\}$, then*

$$\sigma_N(f)_{B_0} \le CN^{-s/d}|f|_B, \qquad N \ge 1. \tag{2.6}$$

where $C = C(p, \alpha, s, r, \tau, d, \Omega, \kappa_\mathbb{T}, \mathcal{T}_0)$.

In terms of approximation classes, Theorem 2.2 can be stated as follows:

Corollary 2.3. *Under the assumptions of Theorem 2.2 we have:*

$$\begin{aligned} B^{\alpha+s}_{\tau,\tau}(\Omega) &\subset \mathbb{A}_{s/d}(B^\alpha_{p,p}(\Omega)) \qquad && (\alpha > 0),\\ B^{s}_{\tau,\tau}(\Omega) &\subset \mathbb{A}_{s/d}(L^p(\Omega)) \qquad && (\alpha = 0). \end{aligned} \tag{2.7}$$

Remark 2.4. It is important to notice that the decay $O(N^{-s/d})$ of the optimal error with complexity N can be achieved if $s + \alpha \leq r + \frac{1}{\tau}$. Thus, in order to take full advantage of the polynomial degree r we need τ to be strictly less than 1; when $r \geq 2$, in fact we need to set $\tau = 1/r$. In the case of linear polynomials, taking $\tau = 1$ was sufficient [BDDP 2002]. Now we need to cope with Besov spaces with integration power less than one, and this poses an additional difficulty when defining the best approximation, due to the lack of uniqueness implied by the non-convexity of the balls. This is dealt with in Section 3, precisely in Definition 3.7 and in Theorem 3.8.

When approximating the most studied case of second order elliptic problems, the usual energy norm is equivalent to the Sobolev $H^1(\Omega)$ norm. It is generally said that an adaptive method with Lagrange finite elements of degree r for this class of problems is quasi-optimal when the method converges with order $N^{-r/d}$. This is the order obtained when the solutions belong to $H^{r+1}(\Omega)$ and the meshes are refined uniformly. The same order is often observed in practice when the solutions are not so regular, and adaptive refinement is used. By inspection of the assumptions of Theorem 2.2, and using the fact that $H^1(\Omega) = B^1_{2,2}(\Omega)$ we conclude that it is possible to approximate optimally a solution, whenever it belongs to $B^{r+1}_{\tau,\tau}(\Omega)$ with $0 < \frac{1}{\tau} < \frac{r}{d} + \frac{1}{2}$, i.e., $\tau > \frac{2d}{2r+d}$. In order to illustrate better we provide some specific examples of spaces included in $\mathbb{A}_{r/d}(H^1(\Omega))$ in Table 1. These few examples in the

	$d = 2$	$d = 3$
$r = 1$	$B^2_{1+\varepsilon,1+\varepsilon}(\Omega)$	$B^2_{\frac{6}{5}+\varepsilon,\frac{6}{5}+\varepsilon}(\Omega)$
$r = 2$	$B^3_{\frac{2}{3}+\varepsilon,\frac{2}{3}+\varepsilon}(\Omega)$	$B^3_{\frac{6}{7}+\varepsilon,\frac{6}{7}+\varepsilon}(\Omega)$
$r = 3$	$B^4_{\frac{1}{2}+\varepsilon,\frac{1}{2}+\varepsilon}(\Omega)$	$B^4_{\frac{6}{9}+\varepsilon,\frac{6}{9}+\varepsilon}(\Omega)$
$r = 4$	$B^5_{\frac{2}{5}+\varepsilon,\frac{2}{5}+\varepsilon}(\Omega)$	$B^5_{\frac{6}{11}+\varepsilon,\frac{6}{11}+\varepsilon}(\Omega)$

TABLE 1. Besov spaces contained in $\mathbb{A}_{r/d}(H^1(\Omega))$ for $d = 2, 3$ and $r = 1, 2, 3, 4$. These few examples show the need of using Besov spaces with integrability index below one.

simplest and most widely studied case of second order elliptic problems show that, in order to find the largest class of functions which can be approximated at optimal rates, it is unavoidable to get involved with function spaces with integrability indices below one.

The last main result is a kind of inverse result from the previous one and states which generalized approximation classes are embedded into which generalized Besov spaces (these generalized spaces are defined in Section 7):

Theorem 2.5 (Inverse Theorem)**.** *Let $0 < p < \infty$, $\alpha \geq 0$, $s > 0$ and $\frac{1}{\tau} = \frac{s}{d} + \frac{1}{p}$. Then*

$$\mathbb{A}_{\frac{s}{d},\tau}(\widehat{B}^{\alpha}_{p,p}(\Omega)) \subset \widehat{B}^{\alpha+s}_{\tau,\tau}(\Omega) \qquad (\alpha > 0),$$
$$\mathbb{A}_{\frac{s}{d},\tau}(L^p(\Omega)) \subset \widehat{B}^{s}_{\tau,\tau}(\Omega) \qquad (\alpha = 0).$$

Remark 2.6. The generalized Besov spaces $\widehat{B}^{\alpha}_{p,p}(\Omega)$ are an extension of the classical Besov spaces using a multiscale norm that coincides with the norm of the classical

Besov spaces $B^\alpha_{p,p}(\Omega)$ when $\alpha < 1 + \frac{1}{p}$. $\widehat{B}^\alpha_{p,p}$ *strictly* contains the classical Besov space $B^\alpha_{p,p}$ when $\alpha \geq 1 + \frac{1}{p}$ (see Section 7).

The rest of this article is organized as follows. In Section 3 we present the construction of an interpolant that will satisfy the estimates of Proposition 2.1. This construction is a key step of this article, and not obvious, since we are approximating in spaces L^p with $p < 1$, which implies that the projections are not well defined in a natural way. In Section 5 we prove that such an interpolant satisfies the properties stated in Proposition 2.1. In order to do so, we use properties of Besov spaces which are presented in Section 4, where we describe the Besov spaces in their classical and multilevel form. In Section 6, we construct a mesh that allows us to prove Theorem 2.2, whereas in Section 7 we prove the inverse estimates of Theorem 2.5. Sections 6 and 7 are generalizations of the results from [BDDP 2002] and the proofs follow the same lines; they are included here for the sake of completeness.

3. Finite Element Basis and Quasi-interpolant

In this section we will construct our interpolant $Q_{\mathcal{T}}$ and in Section 5 we will prove Proposition 2.1, after having defined the Besov spaces and stated some of their properties in Section 4. As usual, the construction does not make use of the regularity of the function being approximated. We will follow the steps highlighted in [BDDP 2002] generalizing them to the case of functions belonging to $L^p(\Omega)$ with $0 < p < 1$ and Lagrange C^0 finite elements of arbitrary polynomial degree. The main contribution of this section are the constructive Definition 3.7 of a quasi-best local polynomial approximation, and the proof of its properties in Theorem 3.8.

Let $\mathcal{T} \in \mathbb{T}$, recall that $\mathbb{V}_{\mathcal{T}} = \left\{ v \in C(\overline{\Omega}) : v_{|T} \in \mathcal{P}^r \quad \forall T \in \mathcal{T} \right\}$, and let

$$\Xi_{\mathcal{T}} = \{\nu : \nu \text{ is a node of } \mathbb{V}_{\mathcal{T}}\}$$

denote the set of nodes, or location of the degrees of freedom of the finite element space. We define, for each $\nu \in \Xi_{\mathcal{T}}$, the basis function ϕ_ν as the only function of $\mathbb{V}_{\mathcal{T}}$ with value 1 at the node ν and zero at the rest of the nodes. Then $\{\phi_\nu\}_{\nu \in \Xi_{\mathcal{T}}}$ is the nodal or canonical basis of $\mathbb{V}_{\mathcal{T}}$.

By standard scaling arguments, if $0 < p < \infty$ and $g = \sum_{\nu \in \Xi_{\mathcal{T}}} a_\nu \phi_\nu$,

$$\|g\|_{L^p(\Omega)} \simeq \left(\sum_{\nu \in \Xi_{\mathcal{T}}} \|a_\nu \phi_\nu\|^p_{L^p(\Omega)} \right)^{\frac{1}{p}} \tag{3.1}$$

where the equivalence constants depend only on p, the polynomial degree r and the mesh regularity.

To construct a dual basis we restrict ourselves to an element T and there we generate the biorthonormal dual basis $\{\varsigma_{T,\nu}\}_{\nu \in T}$ to $\{\phi_\nu\}_{\nu \in T}$, where $\varsigma_{T,\nu}$ is the polynomial of $\mathcal{P}^r$ defined in T such that

$$\langle \varsigma_{T,\nu}, \phi_{\nu'} \rangle_T := \int_T \varsigma_{T,\nu}\, \phi_{\nu'}|_T = \delta_{\nu,\nu'}, \quad \text{for all } \nu' \in T. \tag{3.2}$$

Then, we define the dual basis $\left\{ \widetilde{\phi}_\nu \right\}_{\nu \in \Xi_{\mathcal{T}}}$ as follows:

$$\widetilde{\phi}_\nu = \frac{1}{m_\nu} \sum_{T \in \mathcal{T} : \nu \in T} \varsigma_{T,\nu}\, \chi_T, \tag{3.3}$$

where $m_\nu = \#\{T \in \mathcal{T} : \nu \in T\} = \#\{T \in \mathcal{T} : T \subset \operatorname{supp}(\phi_\nu)\}$. It is straightforward to see that for this dual basis the following properties are valid:

$$\langle \phi_\nu, \widetilde{\phi}_{\nu'} \rangle = \int_\Omega \phi_\nu \widetilde{\phi}_{\nu'} = \delta_{\nu,\nu'}, \tag{3.4}$$

$$\theta_\nu := \operatorname{supp}(\phi_\nu) = \operatorname{supp}(\widetilde{\phi}_\nu) = \bigcup_{T \ni \nu} T, \tag{3.5}$$

and $\widetilde{\phi}_\nu|_T \in \mathcal{P}^r$ for all $T \in \mathcal{T}$, albeit in general $\widetilde{\phi}_\nu$ is discontinuous.

Remark 3.1. The equivalence (3.1) is also true when we change $\{\phi_\nu\}_{\Xi_\mathcal{T}}$ by $\{\widetilde{\phi}_\nu\}_{\Xi_\mathcal{T}}$, and $g = \sum_{\nu \in \Xi_\mathcal{T}} a_\nu \widetilde{\phi}_\nu$.

It is easy to see that the linear operator $\mathcal{Q}_\mathcal{T} : L^1(\Omega) \to \mathbb{V}_\mathcal{T}$ defined as

$$\mathcal{Q}_\mathcal{T}(f) := \sum_{\nu \in \Xi_\mathcal{T}} \langle f, \widetilde{\phi}_\nu \rangle \phi_\nu, \tag{3.6}$$

is a projection mapping $L^1(\Omega)$ into $\mathbb{V}_\mathcal{T}$, i.e., if $f \in \mathbb{V}_\mathcal{T}$, then $\mathcal{Q}_\mathcal{T}(f) = f$. And moreover, if $f = g\, \chi_{\omega_\mathcal{T}(T)}$, with $g \in \mathcal{P}^r$, then $\mathcal{Q}_\mathcal{T}(f)|_T = g|_T$.

Note that this quasi-interpolant is not well defined for $f \in L^p(\Omega)$ when $0 < p < 1$, since $f\widetilde{\phi}_\nu$ could be non-integrable in some cases. To overcome this difficulty we first define (see Definition 3.11 below) a quasi-best polinomial approximation operator $\Pi_{p,\mathcal{T}}$, which yields, for each $f \in L^p(\Omega)$, a (possibly discontinuous) piecewise polynomial function $\Pi_{p,\mathcal{T}}(f)$, and then apply $\mathcal{Q}_\mathcal{T}$ to $\Pi_{p,\mathcal{T}}(f)$.

Lemma 3.2. *The linear projector $\mathcal{Q}_\mathcal{T}$ has the following local stability properties:*

(1) If $1 \le p < \infty$ and $f \in L^p(\Omega)$, then:

$$\|\mathcal{Q}_\mathcal{T}(f)\|_{L^p(T)} \lesssim \|f\|_{L^p(\omega_\mathcal{T}(T))} \qquad \text{for all } T \in \mathcal{T}.$$

(2) If $0 < p < \infty$ and $g = \sum_{T \in \mathcal{T}} \chi_T g_T$ with $g_T \in \mathcal{P}^r$ for each $T \in \mathcal{T}$, then:

$$\|\mathcal{Q}_\mathcal{T}(g)\|_{L^p(T)} \lesssim \|g\|_{L^p(\omega_\mathcal{T}(T))} \qquad \text{for all } T \in \mathcal{T}.$$

The constants involved in the previous bounds depend on p, r, and $\kappa_\mathbb{T}$.

Proof. Let $1 \le p < \infty$, and notice that using (3.3) and (3.4), and scaling to a reference element $\left\|\widetilde{\phi}_\nu\right\|_{L^{p'}(\Omega)} \simeq |\theta_\nu|^{1/p'-1}$, $\|\phi_\nu\|_{L^p(\Omega)} \simeq |\theta_\nu|^{1/p}$ with constants depending only on p, r and $\kappa_\mathbb{T}$. Then, for $f \in L^p(\Omega)$, recalling the definition (3.6) of $\mathcal{Q}_\mathcal{T}$ we obtain

$$\|\mathcal{Q}_\mathcal{T}(f)\|_{L^p(T)} \le \sum_{\nu \in T} |\langle f, \widetilde{\phi}_\nu \rangle| \, \|\phi_\nu\|_{L^p(T)} \lesssim \sum_{\nu \in T} \|f\|_{L^p(\theta_\nu)} \lesssim \|f\|_{L^p(\omega_\mathcal{T}(T))},$$

where the last inequality follows from the finite overlapping of the patches $\theta_\nu \subset \omega_\mathcal{T}(T)$, for $\nu \in T$. Therefore, (1) is proved.

Notice that (2) for $1 \le p < \infty$ is a particular case of (1).

Let $0 < p < 1$ and $g = \sum_{T \in \mathcal{T}} \chi_T g_T$, with $g_T \in \mathcal{P}^r$. Recalling the definition (3.6) of $\mathcal{Q}_\mathcal{T}$ and the fact that for $0 < p < 1$ the triangle inequality holds for $\|\cdot\|^p_{L^p(T)}$ we have

$$\|\mathcal{Q}_\mathcal{T}(g)\|^p_{L^p(T)} \le \sum_{\nu \in T} |\langle g, \widetilde{\phi}_\nu \rangle|^p \, \|\phi_\nu\|^p_{L^p(T)}.$$

On the other hand, by (3.2)

$$\langle g, \widetilde{\phi}_\nu \rangle = \sum_{T' \subset \theta_\nu} \langle g, \widetilde{\phi}_\nu \rangle_{T'} = \frac{1}{m_\nu} \sum_{T' \subset \theta_\nu} \langle g, \varsigma_{T',\nu} \rangle_{T'} = \frac{1}{m_\nu} \sum_{T' \subset \theta_\nu} g_{T'}(\nu).$$

Therefore, using Lemma 3.3 below

$$|\langle g, \widetilde{\phi}_\nu \rangle| \lesssim \sum_{T' \subset \theta_\nu} \|g_{T'}\|_{L^\infty(T')} \lesssim \sum_{T' \subset \theta_\nu} |T'|^{-1/p} \|g\|_{L^p(T')} \lesssim |T|^{-1/p} \|g\|_{L^p(\omega_{\mathcal{T}}(T))}.$$

Since by scaling $\|\phi_\nu\|^p_{L^p(T)} \simeq |T|$, we obtain the desired estimate for (2). □

From now on, G will denote a subdomain of Ω composed of elements of an admissible triangulation $\mathcal{T} \in \mathbb{T}$. More precisely, G will be assumed to be either a patch formed by one element $T \in \mathcal{T}$ or a patch of neighboring elements $\omega_{\mathcal{T}}(T) = \{T' \in \mathcal{T} : T \cap T' \neq \emptyset\}$. Standard scaling arguments can be used to prove the following.

Lemma 3.3. *Let $\mathcal{T} \in \mathbb{T}$, $T \in \mathcal{T}$ and $G = T$ or $G = \omega_{\mathcal{T}}(T)$. If $0 < p, q \leq \infty$ and $r \geq 1$, then*

$$\|g\|_{L^q(G)} \simeq |G|^{\frac{1}{q} - \frac{1}{p}} \|g\|_{L^p(G)} \qquad \forall g \in \mathcal{P}^r, \tag{3.7}$$

with equivalence constants depending only on p, q, r, $\kappa_{\mathbb{T}}$, but independent of G.

Definition 3.4 (Best and near best approximation)**.** Let G be any domain of $\mathbb{R}^d$, let $0 < p < \infty$, let r be a fixed polynomial degree, and let $f \in L^p(G)$.

- The *best approximation error in $L^p(G)$* is defined as
$$E(f, G)_p = \inf_{g \in \mathcal{P}^r} \|f - g\|_{L^p(G)}.$$
- We say that $g \in \mathcal{P}^r$ is a *near best $L^p(G)$-approximation* of f with constant $A > 1$ if
$$\|f - g\|_{L^p(G)} \leq A E(f, G)_p.$$

It is easy to see that

Lemma 3.5. *Let $\mathcal{T} \in \mathbb{T}$, $T \in \mathcal{T}$ and $G = T$ or $G = \omega_{\mathcal{T}}(T)$. If $0 < \rho \leq p \leq \infty$ and $g \in \mathcal{P}^r$ is a near best $L^\rho(G)$-approximation of f with constant A, then g a near best $L^p(G)$-approximation of f with constant cA, where c depends on r, p, ρ and $\kappa_{\mathbb{T}}$, but is independent of f, g and the size of G.*

Given these preliminary definitions and observations we are now in position to construct a quasi-best local approximation operator $\Pi_{p,G} : L^p(G) \to \mathcal{P}^r$ for $0 < p < \infty$. This approximation will have some properties desired for proving Proposition 2.1. In particular, the *linearity* property (3.9) stated in Theorem 3.8 below cannot be guaranteed by the mere existence of a quasi-best approximation; an explicit construction is needed. The case $p > 1$ is simple, because the projection of $L^p(G)$ onto $\mathcal{P}^r$ is linear, and thus (3.9) holds trivially. The main difficulty in the case $0 < p < 1$ is the lack of convexity of the balls in $L^p(G)$, that implies the possible existence of multiple best approximations. We propose the following constructive definition, which requires the definition of barycenter of some class of subsets of the space $\mathcal{P}^r$.

Definition 3.6. Let $k = \dim(\mathcal{P}^r)$ and let $W : \mathcal{P}^r \longmapsto \mathbb{R}^k$ be the canonical linear transformation that maps each polynomial in $\mathcal{P}^r$ to the vector of its coefficients in a given fixed basis. For a nonempty open set $S \subset \mathcal{P}^r$, let $\bar{S} = W(S)$, which is also nonempty and open. Utilizing the Lebesgue measure in $\mathbb{R}^k$, we let Λ be the barycenter of $\bar{S}$, i.e.

$$\Lambda_i = \frac{1}{|\bar{S}|} \int_{\bar{S}} x_i \, dx, \qquad i = 1, 2, \ldots, k,$$

where x_i denotes the i-th component of $x \in \mathbb{R}^k$. We define the barycenter of S as $W^{-1}(\Lambda)$. Notice that since $\bar{S}$ is a nonempty open set of $\mathbb{R}^k$, $|S| > 0$ and the barycenter is well defined.

For a unitary set $S = \{g\} \subset \mathcal{P}^r$ we define the barycenter of S as g.

Definition 3.7. Let G be a domain in $\mathbb{R}^d$ and let $0 < p < \infty$, we define the *quasi-best local polynomial approximation operator* $\Pi_{p,G} : L^p(G) \to \mathcal{P}^r$ as follows. Given $f \in L^p(G)$, we let $S_f = \{f\}$ if $E(f, G)_p = 0$, which happens only when $f \in \mathcal{P}^r$. Otherwise, let $S_f = \{g \in \mathcal{P}^r : \|f - g\|_{L^p(G)} < \frac{3}{2} E(f, G)_p\}$, which is open and nonempty. We define $\Pi_{p,G}(f)$ as the barycenter of S_f.

The following theorem is the main new result of this section, and essentially states that the operator $\Pi_{p,G}$ is a quasi-best approximation in $L^p(G)$ and it is linear with respect to functions in $\mathcal{P}^r$ (but not necessarily linear in general for $p \leq 1$).

Theorem 3.8. *The operator $\Pi_{p,G}$ has the following properties for $0 < p < \infty$:*

(1) $\Pi_{p,G}$ is well defined in $L^p(G)$, and for any $f \in L^p(G)$, $\Pi_{p,G}(f) \in \operatorname{co}(S_f)$, the convex hull of S_f, i.e.,

$$\Pi_{p,G}(f) \in \operatorname{co}(S_f) = \left\{ \sum_{i=1}^{n} \lambda_i g_i : g_i \in S_f,\ \sum_{i=1}^{n} \lambda_i = 1,\ \lambda_i \geq 0, i = 1, 2, \ldots, n,\ n \in \mathbb{N} \right\}.$$

In particular, $\Pi_{p,G} f = f$ if $f \in \mathcal{P}^r$.

(2) For $A_{p,r} = \frac{3}{2}(\dim(\mathcal{P}^r) + 1)^{\frac{1-p}{p}}$ if $0 < p \leq 1$, and $A_{p,r} = \frac{3}{2}$ if $1 < p < \infty$,

$$\|f - \Pi_{p,G}(f)\|_{L^p(G)} \leq A_{p,r} E(f, G)_p, \qquad \forall f \in L^p(G), \tag{3.8}$$

and then $\Pi_{p,G}$ is a quasi-best approximation operator in $L^p(G)$.

(3) $\Pi_{p,G}$ is a bounded operator in $L^p(G)$ in the sense that

$$\|\Pi_{p,G}(f)\|_{L^p(G)} \leq (A_{p,r}^{p_*} + 1)^{1/p_*} \|f\|_{L^p(G)}, \qquad \forall f \in L^p(G),$$

with $p_* = \min\{p, 1\}$, but it is not necessarily linear if $0 < p < 1$.

(4) $\Pi_{p,G}$ is linear for functions in $\mathcal{P}^r$, that is:

$$\Pi_{p,G}(f + g) = \Pi_{p,G}(f) + g, \qquad \forall f \in L^p(G),\ \forall g \in \mathcal{P}^r. \tag{3.9}$$

Proof. $\boxed{1.}$ The first assertion is trivial when $E(f, G)_p = 0$ since in this case $\operatorname{co}(S_f) = S_f = \{f\} = \{\Pi_{p,G}(f)\}$. When $E(f, G)_p > 0$, S_f is open and nonempty, and by virtue of the separation theorem for convex sets [vT 1984, Ch. 3], the barycenter Λ of $\bar{S}_f$ belongs to $\operatorname{co}(\bar{S}_f)$. Since W is linear, $\Pi_{p,G}(f) = W^{-1}(\Lambda) \in W^{-1}(\operatorname{co}(\bar{S}_f)) = \operatorname{co}(S_f)$.

$\boxed{2.}$ The second assertion is trivial for $1 < p < \infty$. To prove it in the case $0 < p \leq 1$ we use the Carathéodory Fundamental Theorem [vT 1984, Ch. 4], which says that each element of the convex hull $\operatorname{co}(S)$ of a set S on a space of finite dimension k, is a convex combination of at most $k+1$ elements of S. Then, since $\Pi_{p,G}(f) \in \operatorname{co}(S_f)$ there exists a set $\{g_i\}_{i=0}^k \subset S_f$, with $k = \dim(\mathcal{P}^r)$ and a set of nonnegative real numbers $\{\lambda_i\}_{i=0}^k$ with $\sum_{i=0}^k \lambda_i = 1$, such that $\Pi_{p,G}(f) = \sum_{i=0}^k \lambda_i g_i$, and then

$$\begin{aligned}
\|f - \Pi_{p,G}(f)\|_{L^p(G)}^p &= \left\| f - \sum_{i=0}^k \lambda_i g_i \right\|_{L^p(G)}^p = \left\| \sum_{i=0}^k \lambda_i (f - g_i) \right\|_{L^p(G)}^p \\
&\le \sum_{i=0}^k \lambda_i^p \|f - g_i\|_{L^p(G)}^p < \left(\frac{3}{2}\right)^p \sum_{i=0}^k \lambda_i^p E(f,G)_p^p \\
&= \left(\frac{3}{2}\right)^p \left(\sum_{i=0}^k \lambda_i^p\right) E(f,G)_p^p \le \left(\frac{3}{2}\right)^p (k+1)^{1-p} E(f,G)_p^p
\end{aligned}$$

where we have used that $g_i \in S_f$ yields $\|f - g_i\| < \frac{3}{2} E(f,G)_p$ for $0 \le i \le k$.

$\boxed{3.}$ The boundedness of the operator in $L^p(G)$ follows from the previous item and the triangle inequality:

$$\begin{aligned}
\|\Pi_{p,G}(f)\|_{L^p(G)}^{p_*} &\le \|f\|_{L^p(G)}^{p_*} + \|f - \Pi_{p,G}(f)\|_{L^p(G)}^{p_*} \\
&\le \|f\|_{L^p(G)}^{p_*} + A_{p,r}^{p_*} E(f,G)_p^{p_*} \le (A_{p,r}^{p_*} + 1) \|f\|_{L^p(G)}^{p_*},
\end{aligned}$$

since $0 \in \mathcal{P}^r$ yields $E(f,G)_p \le \|f\|_{L^p(G)}$.

$\boxed{4}$ If $f \in L^p(G)$ and $g \in \mathcal{P}^r$, then clearly $E(f+g,G)_p = E(f,G)_p$ and also $S_{f+g} = S_f + g$. Since W is linear, $\overline{S_f + g} = \bar{S}_f + g$, which readily implies that $\Pi_{p,G}(f+g) = \Pi_{p,G}(f) + g$ because the Lebesgue measure is translation invariant. $\square$

Remark 3.9. In order to understand the idea of the definition, some remarks are in order:

- The minimum $\min_{g \in \mathcal{P}^r} \|f - g\|_{L^p(G)} = E(f,G)_p$ is always realized because the space $\mathcal{P}^r$ is finite dimensional. It is realized by exactly one polynomial when $p > 1$ and it may be realized by multiple polynomials if $0 < p < 1$.
- When $f \in \mathcal{P}^r$, $S_f = \{f\}$ and $\Pi_{p,G}(f) = f$.
- When $f \notin \mathcal{P}^r$, S_f is the (nonempty) set of all elements of $\mathcal{P}^r$ at distance $< \frac{3}{2} E(f,G)_p$ of f; i.e., S_f contains quasi-best approximants with a uniform constant $(\frac{3}{2})$. The factor $\frac{3}{2}$ can be replaced by any number greater than one,
- Finally, $\Pi_{f,G}(f)$ is defined as the barycenter of S_f, which is also a quasi-best approximation because it is the convex combination of at most $\dim(P^r)+1$ polynomials.

Remark 3.10. It is not clear to us if the operator $\Pi_{p,G}$ is continuous in $L^p(G)$. Even though it would be interesting to know more properties of $\Pi_{p,G}$, we notice that the bound (3.8) is sufficient for our purposes.

Definition 3.11. Given a triangulation $\mathcal{T} \in \mathbb{T}$, $0 < p < \infty$ and $f \in L^p(\Omega)$ we define

$$\Pi_{p,\mathcal{T}}(f) := \sum_{T \in \mathcal{T}} \chi_T \Pi_{p,T}(f),$$

where χ_T is the characteristic function of T. Notice that $\Pi_{p,\mathcal{T}}(f)$ is a piecewise polynomial function over $\mathcal{T}$, not necessarily continuous.

The *quasi-interpolant* $Q_{p,\mathcal{T}} : L^p(\Omega) \to \mathbb{V}_{\mathcal{T}}$ is then defined as:

$$Q_{p,\mathcal{T}}(f) := \mathcal{Q}_{\mathcal{T}}(\Pi_{p,\mathcal{T}}(f)) \tag{3.10}$$

for all $f \in L^p(\Omega)$, $0 < p < \infty$, where $\mathcal{Q}_{\mathcal{T}}$ was defined in (3.6).

Remark 3.12. The *quasi-best approximation* estimate (3.8) of $\Pi_{p,T}$ implies that $Q_{p,\mathcal{T}}(g) = g$ whenever $g \in \mathbb{V}_{\mathcal{T}}$, and the *linearity* property (3.9) implies that

$$Q_{p,\mathcal{T}}(f - Q_{p,\mathcal{T}}(f)) = Q_{p,\mathcal{T}}(f) - Q_{p,\mathcal{T}}(Q_{p,\mathcal{T}}(f)) = Q_{p,\mathcal{T}}(f) - Q_{p,\mathcal{T}}(f) = 0, \tag{3.11}$$

for any $f \in L^p(\Omega)$.

Notice that even though $\mathcal{Q}_{\mathcal{T}}(f)$ may be undefined for $f \in L^p(\Omega)$ if $0 < p < 1$, (3.10) is always well defined, and moreover, we have the following best approximation bound.

Lemma 3.13. *If $f \in L^p(\Omega)$ and $0 < \rho \leq p < \infty$, then there exists a constant c such that:*

$$\|f - Q_{\rho,\mathcal{T}}(f)\|_{L^p(T)} \leq cE(f, \omega_{\mathcal{T}}(T))_p \tag{3.12}$$

where c depends only on p, ρ and the mesh regularity, that is, the approximation interpolant in L^ρ is a quasi-best aproximation locally in $L^p(\Omega)$ for $p \geq \rho$.

Proof. Let $P \in \mathcal{P}^r$ such that $E(f, \omega_{\mathcal{T}})_p = \|f - P\|_{L^p(\omega_{\mathcal{T}}(T))}$, then:

$$\begin{aligned}
\|f - Q_{\rho,\mathcal{T}}(f)\|_{L^p(T)} &\lesssim \|f - P\|_{L^p(T)} + \|P - Q_{\rho,\mathcal{T}}(f)\|_{L^p(T)} \\
&= \|f - P\|_{L^p(T)} + \|\mathcal{Q}_{\mathcal{T}}(P - \Pi_{\rho,\mathcal{T}}(f))\|_{L^p(T)} \\
&\lesssim \|f - P\|_{L^p(T)} + \|P - \Pi_{\rho,\mathcal{T}}(f)\|_{L^p(\omega_{\mathcal{T}}(T))} \\
&= \|f - P\|_{L^p(T)} + \|P - f + f - \Pi_{\rho,\mathcal{T}}(f)\|_{L^p(\omega_{\mathcal{T}}(T))} \\
&\lesssim \|f - P\|_{L^p(\omega_{\mathcal{T}}(T))} + \|f - \Pi_{\rho,\mathcal{T}}(f)\|_{L^p(\omega_{\mathcal{T}}(T))} \\
&\lesssim E(f, \omega_{\mathcal{T}})_p
\end{aligned}$$

where we have used Lemmas 3.2 and 3.5. ☐

This quasi-interpolant, for an appriopriate choice of ρ is the quasi-interpolant $Q_{\mathcal{T}}$ referred to in Proposition 2.1 (see (5.1) below).

4. Besov Spaces and Polynomial Approximation

In this section we define the Besov Spaces and give an equivalent multiscale norm using a finite element multiscale decomposition of the functions, then we present some classical results on polynomial and finite element approximation on Besov Spaces.

4.1. **Moduli of Smoothness and Polynomial Approximation.**

In the following we will give different definitions of *moduli of smoothness*, based on *difference operators*, and we will show some equivalences between the different definitions. Some proofs are rather technical, but necessary to be able to prove Proposition 2.1 in Section 5.

Let G be an open domain in $\mathbb{R}^d$, and let $f : G \to \mathbb{R}$ be a measurable function. Let $h \in \mathbb{R}^d$ and $k \in \mathbb{N}$, we will call *h-difference of order k* of f in G to the function $\Delta_h^k(f, \cdot, G) : G \to \mathbb{R}$ given by

$$\Delta_h^k(f, x, G) := \begin{cases} \displaystyle\sum_{j=0}^{k} (-1)^{k+j} \binom{k}{j} f(x + jh), & \text{if } [x, x + k\,h] \subset G, \\ 0, & \text{otherwise,} \end{cases}$$

where $[x, x + k\,h]$ is the prism $x + [0, k\,h_1] \times \cdots \times [0, k\,h_d]$. It is easy to see that if $1 \le \ell \le k - 1$, then

$$\Delta_h^k = \Delta_h^1(\Delta_h^{k-1}) = \Delta_h^\ell(\Delta_h^{k-\ell}),$$

and also

$$\Delta_h^k(P, \cdot) \equiv 0, \quad \text{if } P \in \mathcal{P}_{k-1}. \tag{4.1}$$

Using these difference operators we define the *modulus of smootheness* of order k in $L^p(G)$ as:

$$\omega_k(f, t)_p = \omega_k(f, t, G)_p := \sup_{|h| \le t} \left\| \Delta_h^k(f, \cdot\,, G) \right\|_{L^p(G)}, \qquad t > 0,$$

which satisfies the basic property [DeL 1993]

$$\omega_k(f, at)_p \le (a + 1)^k\, \omega_k(f, t)_p, \quad \text{for all } a,\, t > 0. \tag{4.2}$$

For $0 < q < \infty$ the *q-modulus of smootheness* of order k in $L^p(G)$ is defined as

$$\mathbf{w}_k(f, t)_{p,q} = \mathbf{w}_k(f, t, G)_{p,q} := \left[\frac{1}{(2t)^d} \int_{[-t,t]^d} \left(\int_G |\Delta_h^k(f, x, G)|^p dx \right)^{\frac{q}{p}} dh \right]^{\frac{1}{q}}.$$

Note that if we extended the definition of $\mathbf{w}_k(f, t)_{p,q}$ to the case $q = \infty$ in a standard manner, we would obtain $\mathbf{w}_k(f, t)_{p,\infty} = \omega_k(f, t)_p$.

The following lemma states that these moduli of smoothness are equivalent if the domain satisfies an *interior cone* condition. Its proof for an interval (1d situation) can be found in [DeL 1993]. We present here a proof valid in any dimension for domains satisfying a cone condition.

Lemma 4.1. *Let $G \subset \mathbb{R}^d$ be a bounded domain such that there exists a family $\{B_j\}_{j=0}^N = \{B(x_j, r_j)\}_{j=0}^N$ and a family of bounded cones $\{\mathcal{C}_j\}_{j=0}^N$ (all congruent to a fixed bounded cone $\mathcal{C}$) satisfying:*

$$G \subset \cup_{j=0}^N B_j \text{ and } \forall x \in B(x_j, 2r_j) \cap G,\ x + \mathcal{C}_j \subset G.$$

Then for each $0 < p, q < \infty$ there exist three constants C_i, $i = 1, 2, 3$ depending only on k, the aspect ratio of $\mathcal{C}$, p, q, N and $\min_{1 \le j \le N}\{r_j\}$, such that:

$$C_1 \mathbf{w}_k(f, t)_{p,q} \le \omega_k(f, t)_p \le C_2 \mathbf{w}_k(f, t)_{p,q}, \quad \text{if } 0 < t \le C_3 \operatorname{diam}(\mathcal{C}),$$

for all measurable functions f defined on G.

Proof. $\boxed{1}$ The left-hand inequality is trivial using (4.2) and that

$$\left\| \Delta_h^k(f, \cdot\,, G) \right\|_{L^p} \le \omega_k(f, t)_p,$$

for all $t > 0$ and all $|h| \le t$.

$\boxed{2}$ To prove the right hand inequality suppose that $\operatorname{diam}(G) = 1$, the other cases follow from this one by standard scaling arguments.

$\boxed{3}$ Consider first the case $q = 1$. For a given $v \in \mathbb{R}^d$ we define $G_v = \{x \in G : x + v \in G\}$. Note that if $h, s \in \mathbb{R}^d$:

$$\begin{aligned}
&\sum_{\ell=0}^{k}(-1)^{k+\ell}\binom{k}{\ell}\Delta^k_{h+\ell s}(f,x)\\
&\qquad\qquad = \sum_{\ell=0}^{k}(-1)^{k+\ell}\binom{k}{\ell}\sum_{j=0}^{k}(-1)^{k+j}\binom{k}{j}f(x+j(h+\ell s))\\
&\qquad\qquad = \sum_{j=0}^{k}(-1)^{k+j}\binom{k}{j}\sum_{\ell=0}^{k}(-1)^{k+\ell}\binom{k}{\ell}f(x+jh+j\ell s)\\
&\qquad\qquad = \sum_{j=1}^{k}(-1)^{k+j}\binom{k}{j}\Delta^k_{js}(f,x+jh).
\end{aligned}$$

whenever all the arguments of f appearing in the last expression belong to G, i.e., if $x \in G_{kh} \cap G_{kh+k^2 s}$. The term corresponding to $\ell = 0$ in the left hand side is $(-1)^k\Delta^k_h(f,x)$, and then we obtain:

$$\Delta^k_h(f,x) = \sum_{\ell=1}^{k}(-1)^{\ell}\binom{k}{\ell}[\Delta^k_{\ell s}(f,x+\ell h) - \Delta^k_{h+\ell s}(f,x)], \tag{4.3}$$

for all $x \in G_{kh} \cap G_{kh+k^2 s}$.

We now consider the family $\{G^j\}_{j=1}^N = \{G \cap B_j\}_{j=1}^N$, which is by assumption a covering of G.

Let $\varrho = \min\{\min_{1\le j\le N}\{r_j\}, \operatorname{diam}(\mathcal{C})\}$, note that if $t \le \frac{\varrho}{k^2}$, $h \in \mathbb{R}^d$, $0 \le |h| \le t$ and $s \in \mathcal{C}^t_j = \{x : \frac{x}{t} \in \mathcal{C}_j\}$, then $x \in G^j \cap G_{kh}$ implies that $x + \ell h \in G_{k\ell s}$ and thus $x \in G_{\ell(h+\ell s)}$, for all $1 \le \ell \le k$. Then taking $L^p(G^j)$-norm in (4.3) we obtain, for all $s \in \mathcal{C}^t_j$:

$$\begin{aligned}
&\left\|\Delta^k_h(f,\cdot)\right\|_{L^p(G_j\cap G_{kh})} = \left\|\Delta^k_h(f,\cdot)\right\|_{L^p(G_j)}\\
&\qquad \le 2^{\frac{1}{p}}\sum_{\ell=1}^{k}\binom{k}{\ell}\left[\left\|\Delta^k_{\ell s}(f,\cdot+\ell h)\right\|_{L^p(G_{k\ell s})} + \left\|\Delta^k_{h+\ell s}(f,\cdot)\right\|_{L^p(G_{k(h+\ell s)})}\right].
\end{aligned}$$

Averaging on $s \in \mathcal{C}^t_j$, a change of variables gives us:

$$\begin{aligned}
&\left\|\Delta^k_h(f,\cdot)\right\|_{L^p(G_j)}\\
&\quad \lesssim \frac{1}{|\mathcal{C}^t_j|}\sum_{\ell=1}^{k}\left[\int_{\mathcal{C}^{\ell t}_j}\left\|\Delta^k_u(f,\cdot)\right\|_{L^p(G)}du + \int_{h+\mathcal{C}^{\ell t}_j}\left\|\Delta^k_u(f,\cdot)\right\|_{L^p(G)}du\right]\\
&\quad \lesssim \frac{1}{t^d}\int_{[-(k+1)t,(k+1)t]^d}\left\|\Delta^k_u(f)\right\|_{L^p(G)}du = (2k+2)^d\mathbf{w}_k(f,(k+1)t)_{p,1},
\end{aligned}$$

Where the constants involved depend on k and the aspect ratio of $\mathcal{C}$. Adding over $j = 1, 2, \ldots, N$ we obtain

$$\left\|\Delta^k_h(f)\right\|_{L^p(G)} \lesssim \mathbf{w}_k(f,(k+1)t)_{p,1}, \qquad \text{for all } t \le \frac{\varrho}{k^2} \text{ and } |h| \le t,$$

where the constant involved depends also on N. Taking supremum over $|h| \le t$ we obtain

$$\omega_k(f,t)_p \lesssim \mathbf{w}_k(f,(k+1)t)_{p,1},$$

and using the basic property (4.2), we obtain that for $0 < t \leq \frac{\varrho}{k^2}$:

$$\omega_k(f,(k+1)t)_p \leq (k+2)^k \omega_k(f,t)_p \lesssim \mathbf{w}_k(f,(k+1)t)_{p,1} \lesssim \omega_k(f,(k+1)t)_p.$$

The result is thus proved for $q = 1$. The cases $0 < q < 1$ and $1 < q < \infty$ are consequences of this first case $q = 1$.

$\boxed{4}$ When $1 < q < \infty$ we use Hölder's inequality to obtain:

$$\begin{aligned}
\mathbf{w}_k(f,t)_{p,1} &= \frac{1}{(2t)^d} \int_{[-t,t]^d} \left\| \Delta_u^k(f) \right\|_{L^p(G_{ku})} du \\
&\leq \frac{1}{(2t)^d} \left(\int_{[-t,t]^d} \left\| \Delta_u^k(f) \right\|_{L^p(G_{ku})}^q du \right)^{\frac{1}{q}} \left(\int_{[-t,t]^d} du \right)^{1-\frac{1}{q}} \\
&= \left(\frac{1}{(2t)^d} \int_{[-t,t]^d} \left\| \Delta_u^k(f) \right\|_{L^p(G_{ku})}^q du \right)^{\frac{1}{q}} = \mathbf{w}_k(f,t)_{p,q}.
\end{aligned}$$

When $0 < q < 1$ we bound $\left\| \Delta_u^k(f) \right\|_{L^p(G_{ku})} \leq \omega_k(f, \sqrt{d}\, t)_p$ and obtain, using again (4.2) that

$$\begin{aligned}
\mathbf{w}_k(f,t)_{p,1} &= \frac{1}{(2t)^d} \int_{[-t,t]^d} \left\| \Delta_u^k(f) \right\|_{L^p(G_{ku})} du \\
&\leq \frac{1}{(2t)^d} \int_{[-t,t]^d} \left\| \Delta_u^k(f) \right\|_{L^p(G_{ku})}^q \omega_k(f, \sqrt{d}\, t)_p^{1-q} du \\
&\leq (\sqrt{d}+1)^{k(1-q)} \omega_k(f,t)_p^{1-q} \frac{1}{(2t)^d} \int_{[-t,t]^d} \left\| \Delta_u^k(f) \right\|_{L^p(G_{ku})}^q du \\
&\lesssim \mathbf{w}_k(f,t)_{p,1}^{1-q} \left(\frac{1}{(2t)^d} \int_{[-t,t]^d} \left\| \Delta_u^k(f) \right\|_{L^p(G_{ku})}^q du \right) \\
&= \mathbf{w}_k(f,t)_{p,1}^{1-q} \, \mathbf{w}_k(f,t)_{p,q}^q,
\end{aligned}$$

where in the last inequality we have used the proved the case $q = 1$, whence:

$$\mathbf{w}_k(f,t)_{p,1} \lesssim \mathbf{w}_k(f,t)_{p,q}, \tag{4.4}$$

and the result is proved for $0 < q < 1$. □

Thanks to the characterization given in [Sh 1983, Theorem 1], any Lipschitz domain satisfies the assumptions of Lemma 4.1, and we thus obtain the same equivalence of moduli of smoothness on Lipschitz domains.

Corollary 4.2. *If G is a Lipschitz domain, for any $0 < p, q < \infty$ there exist constants $C_i = C_i(p,q,k,G)$, $i = 1,2,3$ such that:*

$$C_1 \mathbf{w}_k(f,t)_{p,q} \leq \omega_k(f,t)_p \leq C_2 \mathbf{w}_k(f,t)_{p,q}$$

for all measurable functions f defined on G and all $t \leq C_3$.

The next corollary states that given a triangulation $\mathcal{T}$ of Ω and $T \in \mathcal{T}$, if $G = T$ or $G = \omega_{\mathcal{T}}(T) = \{T' \in \mathcal{T} : T \cap T' \neq \emptyset\}$, then the equivalence constants of the moduli of smoothness from Lemma 4.1 do not depend on the domain G directly, but through the regularity constant $\kappa_{\mathcal{T}}$.

Corollary 4.3. *Given an admissible triangulation $\mathcal{T}$ of a Lipschitz domain Ω, and $0 < p, q < \infty$, there exist three constants $C_i = C_i(p, q, k, \kappa_{\mathcal{T}}, \Omega)$, $i = 1, 2, 3$ such that: If $T \in \mathcal{T}$ and $G = T$ or $G = \omega_{\mathcal{T}}(T)$ then*

$$C_1 \mathbf{w}_k(f, t)_{p,q} \leq \omega_k(f, t)_p \leq C_2 \mathbf{w}_k(f, t)_{p,q}, \quad \textit{if } 0 < t \leq C_3 \operatorname{diam}(T) \tag{4.5}$$

for all measurable functions f defined on G.

Proof. The proof consists in defining a covering $\{B_j\}_{j=1}^N$ and corresponding cones $\{\mathcal{C}_j\}_{j=1}^N$, all congruent to a fixed bounded cone $\mathcal{C}$ independent of $T \in \mathcal{T}$, satisfying the assumptions of Lemma 4.1. Let $\{x_j\}_{j=1}^{\ell}$ be the set of vertices of G, let

$$R = \sup\{r \in \mathbb{R} : B(x_j, s) \cap G \text{ is connected for all } j = 1, 2, \dots \ell, \text{ and all } 0 < s < r\}$$

and $\alpha = \min\{\alpha_1, \alpha_2\}$, where $\alpha_1 = \frac{1}{2}\arctan(\frac{1}{\kappa_{\mathcal{T}}})$ and α_2 is the maximum angle for which the cone condition holds for all $x \in \partial\Omega$. Now take $h = \min\{\min_{E \subset G}\{h_E\}, R\}$, where the minimum is taken over all the edges of $\mathcal{T}$ included in G, and let us define $\mathcal{C}$ a cone with angle α and height $\frac{h \sin(\alpha)}{8}$, from the choice of α and h it follows that for each $x \in G$ there exists a rotation T_x such that $T_x(\mathcal{C}) + x \subset G$.

Now, for each vertex $\{x_j\}_{j=1}^{\ell}$, take $B_j = B(x_j, \frac{h}{4})$ and assign to each B_j the cone $\mathcal{C}_j = T_{x_j}(\mathcal{C})$. To complete the covering let $n = \left\lfloor \dfrac{16H}{h \sin(\alpha)} \right\rfloor$ with $H = \max_{E \subset G} h_E$ and let $\{y_i\}_{i=1}^m \subset G$ be the set of all Lagrange nodes of degree n such that $B(y_i, \frac{h \sin \alpha}{16}) \cap (G \setminus \cup_{j=1}^N B_j) \neq \emptyset$. Now, define $B_{\ell+i} = B(y_i, \frac{h \sin \alpha}{16})$, $1 \leq i \leq m$ and assign to each $B_{\ell+i}$ a cone in the following way:

- if $\exists j,\ 1 \leq j \leq \ell$ such that $B(y_i, \frac{h \sin \alpha}{4}) \cap B_j \neq \emptyset$ let $\mathcal{C}_{\ell+i} = \mathcal{C}_j$;
- if $B(y_i, \frac{h \sin \alpha}{4}) \cap B_j = \emptyset$ for all $j = 1, 2, \dots, \ell$, but $B(y_i, \frac{h \sin \alpha}{4})$ intersects some faces then let $\mathcal{C}_{\ell+i} = \mathcal{C}_j$ for any x_j that belongs to the intersection of those faces (by the definition of R and α, it is clear that the set of such vertices is non empty);
- for any other case assign $\mathcal{C}_{\ell+i} = \mathcal{C}$.

By construction $\{B_j\}_{j=1}^{\ell+m}$ is a covering of G, which satisfies the conditions of Lemma 4.1 with $N = \ell + m \leq c_1(\kappa_{\mathcal{T}}) + c_2(\kappa_{\mathcal{T}}, \Omega)\frac{1}{(\sin \alpha)^d}$, and the aspect ratio of $\mathcal{C}$ depending on $\kappa_{\mathcal{T}}$ and the Lipschitz constant of Ω. Using Lemma 4.1 we conclude that (4.5) holds for all $t \leq C_3 \operatorname{diam}(G)$ and C_1, C_2, C_3 depending only on p, q, k, $\kappa_{\mathcal{T}}$ and Ω. $\square$

The following lemma, known as Whitney's Lemma, is proved in [DL 2004] for convex domains G. It thus holds directly for $G = T$, we sketch the proof for $G = \omega_{\mathcal{T}}(T)$, which follows the steps of [DL 2004].

Lemma 4.4 (Whitney's Lemma, [DL 2004])**.** *Let $\mathcal{T}$ be an admissible mesh, $T \in \mathcal{T}$ and $G = T$ or $G = \omega_{\mathcal{T}}(T) = \{T' \in \mathcal{T} : T \cap T' \neq \emptyset\}$. If $0 < p < \infty$ and $r \geq 1$, then*

$$E(f, G)_p \leq C\omega_{r+1}(f, G)_p, \qquad \textit{for all } f \in L^p(G), \tag{4.6}$$

where $C = C(p, r, d)$ if $G = T$, and $C = C(p, r, d, \kappa_{\mathbb{T}})$ if $G = \omega_{\mathcal{T}}(T)$, but is independent of f and the size of G.

Remark 4.5. It is worth noticing that in general, the constant C from the previous lemma depends on the minimal head-angle condition of the cones associated with the boundary, and it can be difficult to control in general. Moreover, it can be shown that the constant *blows up* for some sequence of domains (see [DL 2004]). The result that we prove holds true with a uniform constant because we are considering only a small family of Lipschitz sets, either elements T of $\mathcal{T}$ or *stars* $\omega_{\mathcal{T}}(T)$, for $\mathcal{T} \in \mathbb{T}$.

Sketch of the proof of Lemma 4.4. The result for $G = T$ is proved in [DL 2004], because T is convex.

Let $G = \omega_{\mathcal{T}}(T)$, and notice that there exist a finite number of *reference stars* such that for every star $\omega_{\mathcal{T}}(T)$, with $\mathcal{T} \in \mathbb{T}$, there exists an affine map $A_{\mathcal{T},T}$ (more precisely the composition of a rotation, a translation and a dilation such that $A_{\mathcal{T},T}(\omega_{\mathcal{T}}(T))$ is a *reference star*. By dilation and translation of the reference stars we may ask that the maximal interior ball of $A_{\mathcal{T},T}(T)$ is $B(0,1)$. Observe also that for $f \in L^p(G)$

$$E(f \circ A_{\mathcal{T},T}^{-1}, A_{\mathcal{T},T}(G))_p = |G|^{1/p} E(f, G)_p,$$
$$\omega_{r+1}(f \circ A_{\mathcal{T},T}^{-1}, A_{\mathcal{T},T}(G)) = |G|^{1/p} \omega_{r+1}(f, G)_p,$$

and it is thus sufficient to prove the result in each of the reference stars, which are all Lipschitz domains.

The case $p \geq 1$ is proved in [JS 1977]. For the case $0 < p < 1$ we should follow the steps in [DL 2004]:

$\boxed{1}$ Let G^{ref} be one of the reference stars, and let $R > 0$ be such that $B(0,1) \subset G^{\mathrm{ref}} \subset B(0,R)$.

$\boxed{2}$ The same proof of [DL 2004, Corollary 2.7] yields the existence of a constant C, depending only on mesh regularity, d, R and p, such that, if $f \in L^p(G^{\mathrm{ref}})$,

$$\|f - c\|_{L^p(G^{\mathrm{ref}})} \leq C\omega_1(f, G^{\mathrm{ref}})_p, \quad \text{for some } c \in \mathbb{R}.$$

$\boxed{3}$ The construction of the step function φ given in [DL 2004, Lemma 2.8] still works in our case, since $B(0,1) \subset G^{\mathrm{ref}} \subset B(0,R)$. The constant $C(d,p)$ from that lemma will now depend on R and the Lipschitz constant of G^{ref}.

$\boxed{4}$ The Proof of Theorem 1.4 for the case $0 < p < 1$ from [DL 2004, pag. 359] now holds because it is based on the previously mentioned results. Moreover, in our case the domain is fixed so there is no need to consider a sequence of domains in the contradiction argument. □

4.2. **Besov Spaces and Classical Results.** Given $s > 0$ and $0 < q, p \leq \infty$, for any $r \in \mathbb{N}$ such that $s < r + \max\{1, \frac{1}{p}\} = r + \frac{1}{p_*}$ for $p_* = \min\{1, p\}$, the Besov space $B^s_{p,q}(\Omega)$, is the set of all functions $f \in L^p(\Omega)$ such that the semi-(quasi)norm $|f|_{B^s_{p,q}(\Omega)}$ is finite, with

$$|f|_{B^s_{p,q}(\Omega)} := \begin{cases} \left(\displaystyle\int_0^\infty [t^{-s}\omega_{r+1}(f,t)_p]^q \frac{dt}{t} \right)^{\frac{1}{q}}, & \text{if } 0 < q < \infty \\ \displaystyle\sup_{t>0} t^{-s}\omega_{r+1}(f,t)_p, & \text{if } q = \infty. \end{cases} \tag{4.7}$$

The (quasi)norm of $B^s_{p,q}(\Omega)$ is defined by:

$$\|f\|_{B^s_{p,q}(\Omega)} = \|f\|_{L^p(\Omega)} + |f|_{B^s_{p,q}(\Omega)}. \tag{4.8}$$

Remark 4.6. The condition $s < r + \frac{1}{p_*}$ with $p_* = \min\{1, p\}$ follows from the fact that if $s \geq r + \frac{1}{p_*}$ and $|f|_{B^s_{p,q}}(\Omega) < \infty$, then $\omega_{r+1}(f,t) = 0$, for all $0 < t < \infty$, which in turn implies that $f \in \mathcal{P}^r$. Besides, if $s < r + \frac{1}{p_*}$, there exist nontrivial functions in $L^p(\Omega)$ with $|f|_{B^s_{p,q}}(\Omega) < \infty$. This is a consequence of the following inequality which holds for $0 < p < \infty$ and $f \in L^p(\Omega)$:

$$\omega_{r+1}(f,t)_p \geq C_r \omega_{r+1}(f,1)_p \, t^{r+\max\{1,\frac{1}{p}\}}, \qquad 0 < t < 1. \tag{4.9}$$

This inequality is proved in [DeL 1993, pag. 370] as a consequence of the operator identities in [DeL 1993, Ch. 12, Lemma 5.1], and the triangle inequality for $\|\cdot\|_{L^p(\Omega)}^{\min\{1,p\}}$. It is interesting to notice that the operator identities are independent on the dimension d, and as a consequence, also (4.9) is independent of the dimension d of the underlying space. In the case $p > 1$ the condition reads $s < r+1$. Inequality (4.9) guarantees that $\omega_{r+1}(f,t)_p = O(t^{r+\max\{1,\frac{1}{p}\}})$ for nontrivial functions $f \in L^p(\Omega)$. Notice that the constraint $s < r + \frac{1}{p_*}$ should not be seen as a restriction on s but rather as a way to choose r for the correct definition of $B^s_{p,q}(\Omega)$.

In order to shed some light into the understanding of Besov spaces, we compute $\omega_{r+1}(\phi_\nu, t)$ for a basis function ϕ_ν of $\mathbb{V}_\mathcal{T}$ and determine for which values of α and p it is true that $\mathbb{V}_\mathcal{T} \subset B^\alpha_{p,p}(\Omega)$.

Proposition 4.7. *Let ϕ_ν be a basis function of $\mathbb{V}_\mathcal{T}$ for some triangulation $\mathcal{T} \in \mathbb{T}$ with support θ_ν. Then,*

$$\omega_{r+1}(\phi_\nu, t) \lesssim \begin{cases} |\theta_\nu|^{\frac{d-1-p}{dp}} t^{1+\frac{1}{p}}, & \text{if } 0 < t \leq |\theta_\nu|^{\frac{1}{d}}, \\ |\theta_\nu|, & \text{if } t > |\theta_\nu|^{\frac{1}{d}}. \end{cases} \tag{4.10}$$

As an immediate consequence, $\mathbb{V}_\mathcal{T} \subset B^\alpha_{p,p}(\Omega)$ if $0 < \alpha < 1 + \frac{1}{p}$.

Proof. Since $\|\phi_\nu\|_\infty = 1$ and ϕ_ν is Lipschitz continuous with Lipschitz constant $\simeq |\theta_\nu|^{-\frac{1}{d}}$, we have that, for all $t > 0$,

$$|\Delta_h^{r+1}\phi_\nu| \lesssim \max\{1, \frac{t}{|\theta_\nu|^{\frac{1}{d}}}\}, \qquad \text{for all } |h| \leq t.$$

Due to the fact that $\Delta_h^{r+1} P = 0$ for all $P \in \mathcal{P}^r$, we have that $\Delta_h^{r+1}\phi_\nu(x) = 0$ if x is at a distance bigger than $(r+1)|h|$ of the *skeleton* of θ_ν, that is, the union of the sides of the mesh that touch θ_ν. Therefore, for $|h| \leq t$, the support of $\Delta_h^{r+1}\phi_\nu$ is contained in a neighbourhood of radius $(r+1)t$ of the *skeleton* of θ_ν, and

$$|\operatorname{supp}(\Delta_h^{r+1}\phi_\nu)| \lesssim \max\{|\theta_\nu|, |\theta_\nu|^{\frac{d-1}{d}} t\},$$

Using the two previous inequalities and integrating, we obtain the following bound for the moduli of smoothness of ϕ_ν

$$\omega_{r+1}(\phi_\nu, t)_p^p = \sup_{|h| \leq t} \|\Delta_h^{r+1}\phi_\nu\|^p_{L^p(\Omega)} \lesssim |\theta_\nu|^{\frac{d-1-p}{d}} t^{1+p}$$

if $0 < t \leq |\theta_\nu|^{\frac{1}{d}}$. □

Remark 4.8. The definition of $B^s_{p,q}(\Omega)$ is independent of r in the sense that if r is replaced by $r' \in \mathbb{N}$ with $s < r' + \max\{1, \frac{1}{p}\}$, then the resulting space is the same with equivalent (quasi)norms. This is a consequence of the fact that $\omega_{r+1}(f,t)_p \leq 2\omega_r(f,t)$ and the Marchaud inequality [D 1988]:

$$\omega_r(f,t)_p \leq C t^r \left(\|f\|_p^{p_*} + \int_t^\infty (u^{-r}\omega_{r+1}(f,u)_p)^{p_*} \frac{du}{u} \right)^{\frac{1}{p_*}}$$

where $p_* := \min\{1, p\}$ and the constant C depends on the Lipschitz constant of the domain and r.

Remark 4.9. It is important to observe that the full (quasi)norms (4.8) of $B^s_{p,q}$ are equivalent for different values of $r > s - \max\{1, \frac{1}{p}\}$, but the corresponding semi-(quasi)norms $|\cdot|_{B^s_{p,q}}$ are not. A simple example is the following. Denote momentarily with $|f|_{B^{s,r+1}_{p,q}(G)}$ the seminorm (4.7) with r replaced by $r+1$, consider $s = r+1$ and let $0 < p < 1$. Then, for $f \in \mathcal{P}^{r+1}$ we have $|f|_{B^{r+1,r+1}_{p,q}(G)} = 0$ because $\omega_{r+2}(f,t)_p = 0$ for all $0 < t < \infty$, whereas $|f|_{B^{r+1}_{p,p}(G)} \neq 0$ unless $f \in \mathcal{P}^r \subsetneq \mathcal{P}^{r+1}$. Throughout this article, we consider $r \geq 1$ fixed as the polynomial degree of the finite element spaces $\mathbb{V}_{\mathcal{T}}$, and always use the semi-norm (4.7), since we only consider Besov spaces $B^s_{p,q}$ with differentiability indices $s < r + \max\{1, \frac{1}{p}\}$.

If we split the integral in (4.7) in diadic intervals, we obtain a very helpful equivalent semi-(quasi)norms when $0 < q < \infty$. In fact

$$|f|^q_{B^s_{p,q}(\Omega)} = \sum_{m\in\mathbb{Z}} \int_{2^{-m-1}}^{2^{-m}} t^{-sq}\omega_{r+1}(f,t)^q_p \,\frac{dt}{t} \simeq \sum_{m\in\mathbb{Z}} 2^{msq}\omega_{r+1}(f,2^{-m})^q_p, \tag{4.11}$$

where we have used that $\omega_{r+1}(f,t)_p$ and t^{-s} are both monotone functions; the constants involved in the equivalence depend on s and q but are otherwise independent of f, r and p.

Another equivalent semi-(quasi)norm in $B^s_{p,p}(\Omega)$ $(0 < p = q < \infty)$ can be established using Lemma 4.1, from where we have:

$$|f|_{B^s_{p,p}(\Omega)} \simeq |f|^w_{B^s_{p,p}(\Omega)} := \left(\int_\Omega \int_0^\infty \int_{[0,t]^d} |\Delta^{r+1}_h(f,x,\Omega)|^p t^{-sp-d-1}\, dh\, dt\, dx \right)^{\frac{1}{p}} \tag{4.12}$$

with equivalence constants that depend only on p, r and the Lipschitz constant of the domain.

As a consequence of the equivalence (4.12) the following sub-additivity result for Besov norms follows.

Lemma 4.10. *Let $0 < p < \infty$, $\alpha > 0$, $f \in B^\alpha_{p,p}(\Omega)$, and let $\{T_i\}_{i=1}^N$ be a finite collection of non-overlapping elements such that $T_i \in \mathcal{T}_i \in \mathbb{T}$, $i = 1, 2, \dots, N$. Then*

$$\sum_{i=1}^N |f|^p_{B^\alpha_{p,p}(T_i)} \lesssim |f|^p_{B^\alpha_{p,p}(\cup_{i=1}^N T_i)}$$

and

$$\sum_{i=1}^N |f|^p_{B^\alpha_{p,p}(\omega_{\mathcal{T}_i}(T_i))} \lesssim |f|^p_{B^\alpha_{p,p}(\cup_{i=1}^N \omega_{\mathcal{T}_i}(T_i))} \lesssim |f|^p_{B^\alpha_{p,p}(\Omega)}$$

where the constants involved depend on p, α, d and $\kappa_{\mathbb{T}}$.

The following embedding results for Besov spaces, can be found in [Pe 1976, T 1978, T 2002].

Theorem 4.11. *Let Ω be a Lipschitz domain, $0 < \beta < \alpha < \infty$, $0 < p, q, \tau, t \leq \infty$. Then the following embedding*

$$B^\alpha_{p,q}(\Omega) \subset B^\beta_{\tau,t}(\Omega) \tag{4.13}$$

is true with continuity if one of the following cases occur:

(1) $p > \tau$;
(2) $p \le \tau$ *and* $\alpha - \beta > d(\frac{1}{p} - \frac{1}{\tau})$;
(3) $p \le \tau$, $\alpha - \beta = d(\frac{1}{p} - \frac{1}{\tau})$ *and* $q \ge t$.

Another important embedding result is the following [Pe 1976, T 1978].

Lemma 4.12. *Let* Ω *be a Lipschitz domain,* $0 < \alpha < \infty$, $0 < p, \tau \le \infty$. *Then the following embedding*

$$B^{\alpha}_{\tau,\tau}(\Omega) \subset L^p(\Omega)$$

is true with continuity if

$$\frac{1}{\tau} \le \frac{s}{d} + \frac{1}{p}.$$

We end this section recalling a few classical results in order to clarify the "location" of the Besov spaces in comparison with Sobolev spaces:

$$B^s_{p,q}(\Omega) \subset W^s_p(\Omega), \qquad \text{if } q > 2,\ p \ge 1 \text{ and } s > 0,$$

$$B^s_{p,q}(\Omega) \supset W^s_p(\Omega), \qquad \text{if } q < 2,\ p \ge 1 \text{ and } s > 0,$$

but it is not true in general that $W^s_p(\Omega) = B^s_{p,2}(\Omega)$. This result holds only when $p = 2$, which means that the only equalities between Besov spaces and Sobolev spaces are:

$$B^s_{2,2}(\Omega) = W^s_2(\Omega), \qquad s > 0. \tag{4.14}$$

4.3. **Besov Spaces and Multiscale Decomposition.** Let $\mathcal{T}_0$ be an initial triangulation of a polygonal (polyhedral) Lipschitz domain Ω and assume that $\mathcal{T}_0$ satisfies condition (b) of Section 4 in [S 2007][1]. We define inductively the sequence $\{\mathcal{T}_m\}_{m=0}^{\infty}$ of nested triangulations of $\mathcal{T}_0$, by letting $\mathcal{T}_{m+1}$ be the mesh obtained by d uniform refinements of $\mathcal{T}_m$, using the algorithm from [S 2007]. A uniform refinement of a triangulation $\mathcal{T}$ is obtained by bisecting all the elements of $\mathcal{T}$, and according to [S 2007, Theorem 4.3] any uniform refinement of $\mathcal{T}_0$ is conforming. Therefore $\mathcal{T}_m \in \mathbb{T}$ and if $T \in \mathcal{T}_{m+1}$, $T' \in \mathcal{T}_m$ with $T \subset T'$, then $|T| = |T'|/2^d$. We let

$$\mathbb{V}_{\mathcal{T}_m} = \left\{ v \in C(\overline{\Omega}) : v_{|T} \in \mathcal{P}^r \quad \forall T \in \mathcal{T}_m \right\},$$

and denote with $\Xi_m := \Xi_{\mathcal{T}_m}$ the set of all the nodes of the space $\mathbb{V}_{\mathcal{T}_m}$ over the mesh $\mathcal{T}_m$ and let $\Xi := \{(\nu, m) : \nu \in \Xi_m, m = 0, 1, 2, \dots\}$ be the set of all the nodes of the sequence of meshes $\{\mathcal{T}_m\}_{m=0}^{\infty}$ with their corresponding level, this definition takes into account the fact that a node ν can belong to many levels (each time that ν is a vertex of a mesh it will be also a vertex of all the following meshes). For the rest of this work we will not mention the level and we will write ν to indicate (ν, m), and the corresponding basis functions will be denoted by ϕ_ν.

In [O 1994, Theorem 6 on p. 38], it is proved that, whenever $1 \le p, q \le \infty$, $0 < s < \min\{r + 1, 1 + \frac{1}{p}\}$, the norm of $B^s_{p,q}$ is equivalent to the norm

$$\|f\|^{***}_{B^s_{p,q}(\Omega)} := \left(\sum_{m=0}^{\infty} 2^{msq} \left\|(R_m - R_{m-1})f\right\|^q_{L^p(\Omega)} \right)^{\frac{1}{q}},$$

[1] Notice that if the initial triangulation cannot be labeled to satisfy this condition, it can be refined globally with a simple procedure so that a labeling of vertices and edges satisfying this condition is easily obtained [S 2007, Appendix A].

if $\{R_m\}$ is a sequence of uniformly $L^p(\Omega)$-bounded linear projectors onto $\mathbb{V}_{\mathcal{T}_m}$ and $R_{-1} = 0$. We prove a weaker result for $0 < p, q < \infty$ which is sufficient for our purposes.

Taking $0 < \rho \le p < \infty$ we denote the operator $Q_{\rho,\mathcal{T}_m}$ from Definition 3.11 by Q_m for $m = 0, 1, \dots$ and define $q_m = Q_m - Q_{m-1}$, with $Q_{-1} = 0$. Since $\cup_{m=0}^{\infty}\mathbb{V}_{\mathcal{T}_m}$ is dense in $L^p(\Omega)$, $\|f - Q_m(f)\|_{L^p(\Omega)} \to 0$ when $m \to \infty$ and then:

$$f = \sum_{m=0}^{\infty} q_m(f) \quad \text{with convergence in } L^p(\Omega). \tag{4.15}$$

Moreover, we have the following result.

Theorem 4.13. *Let $0 < \rho \le p < \infty$ be given, and let Ω, $\{\mathcal{T}_m\}$, $\{Q_m\}$, $\{q_m = Q_m - Q_{m-1}\}$ be as in the previous paragraph. For $0 < \alpha < \min\{r+1, 1+\frac{1}{p}\}$ consider the following norm*

$$\|f\|^Q_{B^\alpha_{p,q}(\Omega)} := \left(\sum_{m=0}^{\infty} 2^{m\alpha q}\, \|q_m f\|^q_{L^p(\Omega)}\right)^{\frac{1}{q}}.$$

Then

$$\|f\|^Q_{B^\alpha_{p,q}(\Omega)} \lesssim \|f\|_{B^\alpha_{p,q}(\Omega)}.$$

with a constant depending only on ρ, p, q, α and the regularity of $\mathcal{T}_0$.

Using (3.1) and the fact that $|\theta_\nu| \simeq 2^{-md}$ for $\nu \in \Xi_m$, we have the following corollary for the particular case $p = q$.

Corollary 4.14. *Under the assumptions of Theorem 4.13, for $f \in L^p(\Omega)$ let $\{b_\nu(f)\}_{\nu\in\Xi_m}$ be the set of coefficients of $q_m(f) \in \mathbb{V}_{\mathcal{T}_m}$ in the canonical basis of $\mathbb{V}_{\mathcal{T}_m}$, that is*

$$b_\nu(f) = \langle q_m, \widetilde{\phi}_\nu\rangle, \quad \nu \in \Xi_m, \qquad \text{and} \qquad q_m(f) = \sum_{\nu\in\Xi_m} b_\nu(f)\phi_\nu. \tag{4.16}$$

Then

$$\|f\|_{\widehat{B}^\alpha_{p,p}(\Omega)} := \left(\sum_{\nu\in\Xi} |\theta_\nu|^{-\frac{\alpha p}{d}}\, \|b_\nu(f)\phi_\nu\|^p_{L^p(\Omega)}\right)^{\frac{1}{p}} \simeq \|f\|^Q_{B^\alpha_{p,p}(\Omega)} \lesssim \|f\|_{B^\alpha_{p,p}(\Omega)}. \tag{4.17}$$

Proof of Theorem 4.13. Recall that $q_m = Q_m - Q_{m-1}$ and $Q_m = Q_{\rho,\mathcal{T}_m}$ for $0 < \rho \le p$, so that Lemma 3.13 yields $\|f - Q_m f\|_{L^p(T)} \lesssim E(f, \omega_{\mathcal{T}_m}(T))_p$ for any $T \in \mathcal{T}_m$. Therefore, for $T \in \mathcal{T}_{m-1}$,

$$\|q_m f\|^p_{L^p(T)} \le \|f - Q_m f\|^p_{L^p(T)} + \|f - Q_{m-1} f\|^p_{L^p(T)} \lesssim E(f, \omega_{\mathcal{T}_{m-1}}(T)).$$

By Whitney's Lemma 4.4 $E(f, \omega_{\mathcal{T}_{m-1}}(T)) \lesssim \omega_{r+1}(f, 2^{-m}, \omega_{\mathcal{T}_{m-1}}(T))_p$, and thanks to Corollary 4.3

$$\|q_m f\|^p_{L^p(\Omega)} \lesssim \sum_{T\in\mathcal{T}_{m-1}} \omega_{r+1}(f, 2^{-m}, \omega_{\mathcal{T}_{m-1}}(T))^p_p \lesssim \omega_{r+1}(f, 2^{-m}, \Omega)^p_p.$$

Finally, (4.11) implies that

$$\left(\sum_{m=0}^{\infty} 2^{m\alpha q}\, \|q_m f\|^q_{L^p(\Omega)}\right)^{\frac{1}{q}} \lesssim \left(\sum_{m=0}^{\infty} 2^{m\alpha q} \omega_{r+1}(f, 2^{-m}, \Omega)^q_p\right) \lesssim |f|_{B^\alpha_{p,q}(\Omega)}.$$

which is the desired assertion. $\square$

4.4. **Besov Spaces and Approximation Results.** If we consider all the transformations $G^{\text{ref}} = |G|^{-\frac{1}{d}}G = \left\{|G|^{-\frac{1}{d}}x : x \in G\right\}$ for sets $G = T$ and $G = \omega_{\mathcal{T}}(T)$, with $T \in \mathcal{T}$ and for all $\mathcal{T} \in \mathbb{T}$, we obtain a finite family (up to translations) of reference sets G^{ref} with $|G^{\text{ref}}| \simeq 1$. The scaling of the Besov seminorms for $\beta > 0$, $0 < p, q < \infty$ and $f \in B^{\beta}_{p,q}(G)$ is given by:

$$\begin{aligned}
|f|_{B^{\beta}_{p,q}(G)} &= \left(\int_0^\infty \left(t^{-\beta}\omega_{r+1}(f,t,G)_p\right)^q \frac{dt}{t}\right)^{\frac{1}{q}} \\
&\simeq |G|^{\frac{1}{p}}\left(\int_0^\infty \left(t^{-\beta}\omega_{r+1}(f^{\text{ref}}, \frac{t}{|G|^{\frac{1}{d}}}, G^{\text{ref}})_p\right)^q \frac{dt}{t}\right)^{\frac{1}{q}} \\
&= |G|^{\frac{1}{p}-\frac{\beta}{d}}\left(\int_0^\infty \left(s^{-\beta}\omega_{r+1}(f^{\text{ref}}, s, G^{\text{ref}})_p\right)^q \frac{ds}{s}\right)^{\frac{1}{q}} \\
&= |G|^{\frac{1}{p}-\frac{\beta}{d}}|f^{\text{ref}}|_{B^{\beta}_{p,q}(G^{\text{ref}})} = (|G|^{\frac{1}{d}})^{\frac{d}{p}-\beta}|f^{\text{ref}}|_{B^{\beta}_{p,q}(G^{\text{ref}})},
\end{aligned} \tag{4.18}$$

where $f^{\text{ref}} : G^{\text{ref}} \to \mathbb{R}$ is defined by $f^{\text{ref}}(x) = f(|G|^{\frac{1}{d}}x)$, for $x \in G^{\text{ref}}$.

Lemma 4.15. *Let $\mathcal{T} \in \mathbb{T}$ and $T \in \mathcal{T}$. Let $0 < p < \infty$, $s > 0$, $0 < \frac{1}{\tau} \le \frac{s}{d} + \frac{1}{p}$, $\delta = \frac{s}{d} + \frac{1}{p} - \frac{1}{\tau}$. If $G = T$ or $G = \omega_{\mathcal{T}}(T)$, then for $s < r + \frac{1}{\tau_*}$, with $\tau_* = \min\{1, \tau\}$, we have that:*

$$E(f,G)_p \le C|G|^{\delta}|f|_{B^s_{\tau,\tau}(G)}, \qquad \text{for all } f \in B^s_{\tau,\tau}(G), \tag{4.19}$$

where $C = C(p, s, \tau, d, r, \kappa_{\mathbb{T}})$. Alternatively, if $h_T = \operatorname{diam}(T)$, we have

$$E(f,G)_p \le C\, h_T^{s+\frac{d}{p}-\frac{d}{\tau}}|f|_{B^s_{\tau,\tau}(G)}, \qquad \text{for all } f \in B^s_{\tau,\tau}(G).$$

Remark 4.16. At first sight, it looks surprising that inequality (4.19) holds for $s > r + 1$, when $\tau < 1$. It is important to mention here that in the definition of $|f|_{B^s_{\tau,\tau}(G)}$ given by (4.7) we are considering r fixed, equal to the polynomial degree that takes place in the definition of $E(f,G)_p = \inf_{g \in \mathcal{P}^r} \|f - g\|_{L^p(G)}$, so that $|f|_{B^s_{\tau,\tau}(G)} = 0$ only if $f \in \mathcal{P}^r$. See also Remark 4.9.

Proof of Lemma 4.15. Suppose first that $G = G^{\text{ref}}$ and thus $|G| \simeq 1$. Let $f \in B^s_{\tau,\tau}(G)$, by (4.6) and (4.11) we have:

$$E(f,G)_\tau \lesssim \omega_{r+1}(f,1,G)_\tau \le \left(\sum_{k\in\mathbb{Z}} 2^{ks\tau}\omega_{r+1}(f,2^{-k},G)_\tau^\tau\right)^{\frac{1}{\tau}} \simeq |f|_{B^s_{\tau,\tau}(G)}\,.$$

For any polynomial $P \in \mathcal{P}^r$, Theorem 4.11 yields

$$\begin{aligned}
E(f,G)_p \le \|f-P\|_{L^p(G)} &\lesssim \|f-P\|_{L^\tau(G)} + |f-P|_{B^s_{\tau,\tau}(G)} \\
&= \|f-P\|_{L^\tau(G)} + |f|_{B^s_{\tau,\tau}(G)}\,,
\end{aligned}$$

due to (4.1). Choosing $P \in \mathcal{P}^r$ a best approximation for f in $L^\tau(G)$ we thus obtain

$$E(f,G)_p \lesssim E(f,G)_\tau + |f|_{B^s_{\tau,\tau}(G)} \lesssim |f|_{B^s_{\tau,\tau}(G)}\,,$$

which is the desired result for reference domains G.

The general case follows using (4.18) in the proved case, more precisely:

$$\begin{aligned} E(f,G)_p &\lesssim |G|^{\frac{1}{p}} E(f^{\mathrm{ref}}, G^{\mathrm{ref}})_p \\ &\lesssim |G|^{\frac{1}{p}} \left|f^{\mathrm{ref}}\right|_{B^s_{\tau,\tau}(G^{\mathrm{ref}})} \\ &\lesssim |G|^{\frac{1}{p}} |G|^{-\frac{1}{\tau}+\frac{s}{d}} \left|f\right|_{B^s_{\tau,\tau}(G)} = |G|^{\delta} \left|f\right|_{B^s_{\tau,\tau}(G)}, \end{aligned}$$

and the result is proved. □

Lemma 4.17. *Let $\mathcal{T} \in \mathbb{T}$ and $T \in \mathcal{T}$. Let $0 < p < \infty$, $0 < \alpha < \min\{r+1, 1+\frac{1}{p}\}$, $0 < \frac{1}{\tau} < \frac{s}{d} + \frac{1}{p}$, $s > 0$, and $\alpha + s < r + \frac{1}{\tau_*}$, with $\tau_* = \min\{1, \tau\}$. If $G = T$ or $G = \omega_{\mathcal{T}}(T)$ we have that:*

$$|f|_{B^{\alpha}_{p,p}(G)} \le C|T|^{\delta} |f|_{B^{\alpha+s}_{\tau,\tau}(G)}, \qquad \text{for all } f \in B^{\alpha+s}_{\tau,\tau}(\Omega), \tag{4.20}$$

where $C = C(p, s, \tau, d, r, \rho, \kappa_{\mathbb{T}})$ and $\delta = \frac{s}{d} + \frac{1}{p} - \frac{1}{\tau} > 0$.

Proof. We first prove the following bound assuming G is a reference patch, i.e., $|G| \simeq 1$:

$$|f|_{B^{\alpha}_{p,p}(G)} \lesssim |f|_{B^{\alpha+s}_{\tau,\tau}(G)}, \quad \text{for all } f \in B^{\alpha+s}_{\tau,\tau}(G). \tag{4.21}$$

In order to prove this bound we observe that Theorem 4.13 implies that for all $P \in \mathcal{P}^r(G)$ due to Theorem 4.11, we have that:

$$|f|_{B^{\alpha}_{p,p}(G)} = |f-P|_{B^{\alpha}_{p,p}(G)} \le \|f-P\|_{B^{\alpha}_{p,p}(G)} \lesssim \|f-P\|_{L^{\tau}(G)} + |f-P|_{B^{\alpha+s}_{\tau,\tau}(G)},$$

and choosing $P \in \mathcal{P}^r$ such that $\|f-P\|_{L^{\tau}(G)} = E(f,G)_{\tau}$ we obtain by Lemma 4.15,

$$|f|_{B^{\alpha}_{p,p}(G)} \lesssim |f-P|_{B^{\alpha+s}_{\tau,\tau}(G)} = |f|_{B^{\alpha+s}_{\tau,\tau}(G)}, \quad \text{for all } f \in B^{\alpha+s}_{\tau,\tau}(G),$$

whence the bound (4.21) holds on reference patches. Scaling (4.21) and making use of (4.18) we obtain the desired assertion. □

Remark 4.18. Notice that we use the definition of the Besov seminorms with r fixed, equal to the polynomial degree. This lemma is not true if we use different values of r at the left and right-hand side of (4.20).

5. Proof of Proposition 2.1

In this section we present the proof of Proposition 2.1 making use of the interpolant defined in (3.10).

Proof of Proposition 2.1. $\boxed{1}$ Let ρ be any number satisfying $0 < \rho < \min\{p, \tau\}$, and let

$$Q_{\mathcal{T}} = Q_{\rho,\mathcal{T}}, \tag{5.1}$$

with $Q_{\rho,\mathcal{T}}$ from Definition 3.11.

$\boxed{2}$ Let us consider first the case $\alpha = 0$, with $B_0 = L^p$. The first bound (2.1) follows directly from Lemma 4.15 and (3.12). The other two bounds (2.3), (2.4) coincide and are a consequence of the finite overlapping of elements in $\mathcal{T}$.

$\boxed{3}$ Consider now $\alpha > 0$ and $B_0 = B^{\alpha}_{p,p}$. Inequality (2.2) coincides with (4.20), the assertion of Lemma 4.17.

The estimate (2.3) follows from (3.13) and (4.19) using that

$$\|f - Q_{\mathcal{T}}(f)\|^p_{L^p(\Omega)} \le \sum_{T\in\mathcal{T}} \|f - Q_{\mathcal{T}}(f)\|^p_{L^p(T)} \lesssim \sum_{T\in\mathcal{T}} E(f,\omega_{\mathcal{T}}(T))^p_p$$
$$\lesssim \sum_{T\in\mathcal{T}} |T|^{\frac{\alpha p}{d}} |f|^p_{B^\alpha_{p,p}(\omega_{\mathcal{T}}(T))} \lesssim \sum_{T\in\mathcal{T}} |f|^p_{B^\alpha_{p,p}(\omega_{\mathcal{T}}(T))}.$$

$\boxed{4}$ In order to prove (2.4) we will make use of the multilevel decomposition and the norm defined in Theorem 4.13. We consider $\mathcal{T}$ as the initial mesh $\mathcal{T}_0$ of a sequence $\{\mathcal{T}_m\}_{m=0}^\infty$ of uniform refinements as described at the beginning of Section 4.3. Then, with convergence in $L^p(\Omega)$,

$$f - Q_{\mathcal{T}}(f) = f - q_0(f) = \sum_{m=1}^\infty q_m(f) = \sum_{m=1}^\infty \Big(\sum_{\nu\in\Xi_m} b_\nu(f)\phi_\nu \Big),$$

where $q_m = Q_{\mathcal{T}_m} - Q_{\mathcal{T}_{m-1}}$, $m = 1,2,\dots$, $q_0 = Q_{\mathcal{T}_0} = Q_{\mathcal{T}}$, and $\{b_\nu(f)\}_\nu$ are the coefficients defined in (4.16).

Recalling from Section 4.3 that Ξ denotes the set of all the nodes of the sequence of triangulations $\{\mathcal{T}_m\}_{m=0}^\infty$, and θ_ν is the support of the canonical basis function ϕ_ν corresponding to $\nu \in \Xi$, we define $\Psi_j = \{\nu \in \Xi : 2^{-j-1} < |\theta_\nu|^{\frac{1}{d}} \le 2^{-j}\}$ and $g_j = \sum_{\nu\in\Psi_j} b_\nu(f)\phi_\nu$. Then $f - Q_{\mathcal{T}}(f) = \sum_{j\in\mathbb{Z}} g_j$ is a multiscale decomposition of $f - Q_{\mathcal{T}}(f)$. Recall that we want to estimate $|f - Q_{\mathcal{T}}(f)|_{B_0(\Omega)}$ which satisfies

$$|f - Q_{\mathcal{T}}(f)|_{B_0(\Omega)} \simeq \left(\sum_{m\in\mathbb{Z}} 2^{msp}\omega_{r+1}(f - Q_{\mathcal{T}}(f), 2^{-m})^p_p \right)^{\frac{1}{p}}, \tag{5.2}$$

due to (4.11). Taking $p_* = \min\{p,1\}$, the triangle inequality yields,

$$\omega_{r+1}(f - Q_{\mathcal{T}}(f), 2^{-m})^{p_*}_p \le \sum_{j\in\mathbb{Z}} \omega_{r+1}(g_j, 2^{-m})^{p_*}_p, \qquad \text{for each } m \in \mathbb{Z}, \tag{5.3}$$

and we are thus lead to estimating $\omega_{r+1}(g_j,t)_p$, for $t > 0$ and $j \in \mathbb{Z}$.

$\boxed{5}$ On the one hand, since the mesh is generated by bisection, there exists a constant c, that depends only on d and mesh regularity, such that at most c patches θ_ν with $\nu \in \Psi_j$ overlap, which leads us to:

$$\omega_{r+1}(g_j,t)^p_p \lesssim \|g_j\|^p_{L^p(\Omega)} \lesssim \sum_{\nu\in\Psi_j} \|b_\nu(f)\phi_\nu\|^p_{L^p(\Omega)}, \qquad \text{for all } t > 0. \tag{5.4}$$

On the other hand, using the finite overlapping of θ_ν for $\nu \in \Psi_j$, we obtain:

$$\omega_{r+1}(g_j, 2^{-m})^p_p \lesssim \sum_{\nu\in\Psi_j} \omega_{r+1}(b_\nu(f)\phi_\nu, 2^{-m})^p_p \le |b_\nu(f)|\omega_{r+1}(\phi_\nu, 2^{-m})^p_p,$$

and if $j < m$, Proposition 4.7 implies that

$$\begin{aligned} \omega_{r+1}(g_j, 2^{-m})^p_p &\lesssim 2^{-m(1+p)}2^{-j(d-1-p)} \sum_{\nu\in\Psi_j} |b_\nu(f)|^p \\ &\simeq 2^{-(m-j)(1+p)} \sum_{\nu\in\Psi_j} \|b_\nu(f)\phi_\nu\|^p_{L^p(\Omega)}, \end{aligned} \tag{5.5}$$

where we have used that $\|\phi_\nu\|^p_{L^p(\Omega)} \simeq |\theta_\nu| \simeq 2^{-jd}$ if $\nu \in \Psi_j$.

The bounds (5.4) y (5.5) can be summarized as

$$\omega_{r+1}(g_j,2^{-m})_p^p \lesssim \begin{cases} \sum_{\nu\in\Psi_j} \|b_\nu(f)\phi_\nu\|^p_{L^p(\Omega)}, & \text{if } j>m,\\ 2^{-(m-j)(1+p)} \sum_{\nu\in\Psi_j} \|b_\nu(f)\phi_\nu\|^p_{L^p(\Omega)}, & \text{if } j\le m. \end{cases} \tag{5.6}$$

$\boxed{6}$ Inserting this bound into (5.3) we obtain that

$$\begin{aligned}
\omega_{r+1}(f-Q_{\mathcal{T}}(f),2^{-m})_p^{p_*} &\le \sum_{j=-\infty}^{\infty} \omega_{r+1}(g_j,2^{-m})_p^{p_*}\\
&\lesssim \sum_{j=m+1}^{\infty}\left(\sum_{\nu\in\Psi_j}\|b_\nu(f)\phi_\nu\|^p_{L^p(\Omega)}\right)^{\frac{p_*}{p}}\\
&\quad+\sum_{j=-\infty}^{m} 2^{-(m-j)(1+p)\frac{p_*}{p}}\left(\sum_{\nu\in\Psi_j}\|b_\nu(f)\phi_\nu\|^p_{L^p(\Omega)}\right)^{\frac{p_*}{p}}\\
&\simeq \sum_{j=m+1}^{\infty} 2^{-j\alpha p_*}\left(\sum_{\nu\in\Psi_j}|\theta_\nu|^{-\frac{\alpha p}{d}}\|b_\nu(f)\phi_\nu\|^p_{L^p(\Omega)}\right)^{\frac{p_*}{p}}\\
&\quad+\sum_{j=-\infty}^{m} 2^{-(m-j)(1+p)\frac{p_*}{p}}\,2^{-j\alpha p_*}\left(\sum_{\nu\in\Psi_j}|\theta_\nu|^{-\frac{\alpha p}{d}}\|b_\nu(f)\phi_\nu\|^p_{L^p(\Omega)}\right)^{\frac{p_*}{p}}.
\end{aligned}$$

Defining $K_\nu = |\theta_\nu|^{-\frac{\alpha p}{d}}\|b_\nu(f)\phi_\nu\|^p_{L^p(\Omega)}$, by (5.2) the previous bound yields:

$$\begin{aligned}
|f-Q_{\mathcal{T}}(f)|^p_{B^\alpha_{p,p}(\Omega)} &\lesssim \sum_{m\in\mathbb{Z}} 2^{m\alpha p}\left[\sum_{j=m+1}^{\infty}2^{-j\alpha p_*}\left(\sum_{\nu\in\Psi_j}K_\nu\right)^{\frac{p_*}{p}}\right]^{\frac{p}{p_*}}\\
&\quad+\sum_{m\in\mathbb{Z}} 2^{m\alpha p}\left[\sum_{j=-\infty}^{m}2^{-(m-j)(1+p)\frac{p_*}{p}}2^{-j\alpha p_*}\left(\sum_{\nu\in\Psi_j}K_\nu\right)^{\frac{p_*}{p}}\right]^{\frac{p}{p_*}}\\
&\lesssim \sum_{m\in\mathbb{Z}}\left[\sum_{j=m+1}^{\infty}2^{-(j-m)\alpha p_*}\left(\sum_{\nu\in\Psi_j}K_\nu\right)^{\frac{p_*}{p}}\right]^{\frac{p}{p_*}}\\
&\quad+\sum_{m\in\mathbb{Z}}\left[\sum_{j=-\infty}^{m}2^{-(m-j)(1+p-\alpha p)\frac{p_*}{p}}\left(\sum_{\nu\in\Psi_j}K_\nu\right)^{\frac{p_*}{p}}\right]^{\frac{p}{p_*}}.
\end{aligned}$$

$\boxed{7}$ By assumption, $0<\alpha<1+\frac{1}{p}$, so that $1+p-\alpha p>0$. Using Hardy's inequality (see Lemma 5.1 below) with

$$a_j = 2^{-j\alpha}\Big(\sum_{\nu\in\Psi_j}K_\nu\Big)^{\frac{1}{p}} \qquad \text{and} \qquad z_m = \left(\sum_{j=m+1}^{\infty}\Big(\underbrace{2^{-j\alpha}\Big(\sum_{\nu\in\Psi_j}K_\nu\Big)^{\frac{1}{p}}}_{a_j}\Big)^{p_*}\right)^{\frac{1}{p_*}}$$

in the first summation, and

$$a_j = 2^{-j\frac{(1+p-\alpha p)}{p}} \Big(\sum_{\nu \in \Psi_{-j}} K_\nu \Big)^{\frac{1}{p}} \text{ and } z_m = \Big(\sum_{j=-\infty}^{-m} \underbrace{\Big(2^{-j\frac{(1+p-\alpha p)}{p}} \big(\sum_{\nu \in \Psi_{-j}} K_\nu \big)^{\frac{1}{p}} \Big)}_{a_j}{}^{p_*} \Big)^{\frac{1}{p_*}}$$

in the second summation, we obtain:

$$|f - Q_{\mathcal{T}}(f)|^p_{B^\alpha_{p,p}(\Omega)} \lesssim \sum_{j \in \mathbb{Z}} \sum_{\nu \in \Psi_j} K_\nu = \sum_{\nu \in \Xi} |\theta_\nu|^{-\frac{\alpha p}{d}} \|b_\nu(f)\phi_\nu\|^p_{L^p(\Omega)} . \tag{5.7}$$

$\boxed{8}$ It remains to bound $\|b_\nu(f)\phi_\nu\|^p_{L^p(\Omega)}$. Let $m \in \mathbb{N}$, $\nu \in \Xi_m$ and let $T \in \mathcal{T}_m$ be such that $T \subset \theta_\nu$ (or $\nu \in T$). We will denote with T' the ancestor of T in $\mathcal{T}_{m-1}$ (i.e., T' is the only element in $\mathcal{T}_{m-1}$ such that $T \subset T'$). Due to (3.12) we then obtain

$$\begin{aligned}
\|b_\nu(f)\phi_\nu\|_{L^p(\Omega)} &\lesssim \|q_m(f)\|_{L^p(T)} \\
&\lesssim \|f - Q_{\mathcal{T}_m}(f)\|_{L^p(T)} + \left\|f - Q_{\mathcal{T}_{m-1}}(f)\right\|_{L^p(T')} \\
&\lesssim E(f, \widehat{T})_p + E(f, \widehat{T'})_p,
\end{aligned}$$

with $\widehat{T}$, $\widehat{T'}$ denoting, respectively $\omega_{\mathcal{T}_m}(T)$, $\omega_{\mathcal{T}_{m-1}}(T')$. This bound, and (5.7) yield

$$|f - Q_{\mathcal{T}}(f)|^p_{B^\alpha_{p,p}(\Omega)} \lesssim \sum_{T \in \mathbb{H}} |T|^{-\frac{\alpha p}{d}} E(f, \widehat{T})^p_p$$

where $\mathbb{H} = \cup_{m=0}^\infty \mathcal{T}_m$. Recalling that $\mathcal{T}_0 = \mathcal{T}$, the previous bound reads

$$\begin{aligned}
|f - Q_{\mathcal{T}}(f)|^p_{B^\alpha_{p,p}(\Omega)} &\lesssim \sum_{T^* \in \mathcal{T}} \sum_{T \in \mathbb{H} : \widehat{T} \subset \widehat{T^*}} |T|^{-\frac{\alpha p}{d}} E(f, \widehat{T})^p_p \\
&\lesssim \sum_{T^* \in \mathcal{T}} \sum_{T \in \mathbb{H} : \widehat{T} \subset \widehat{T^*}} |T|^{-\frac{\alpha p}{d}} \omega_{r+1}(f, \widehat{T})^p_p
\end{aligned}$$

due to (4.6).

For each $T^* \in \mathcal{T}$ we call $\mathcal{I}_j^{T^*} = \{T \in \mathbb{H} : \widehat{T} \subset \widehat{T^*} \text{ and } 2^{-j-1} < |T|^{\frac{1}{d}} \leq 2^{-j}\}$, and

$$|f - Q_{\mathcal{T}}(f)|^p_{B^\alpha_{p,p}(\Omega)} \lesssim \sum_{T^* \in \mathcal{T}} \sum_{j \in \mathbb{Z}} 2^{j\alpha p} \sum_{T \in \mathcal{I}_j^{T^*}} \omega_{r+1}(f, \widehat{T})^p_p. \tag{5.8}$$

$\boxed{9}$ Thanks to Corollary 4.3, which shows the equivalence between $\omega_{r+1}(f, \widehat{T})_p$ and $\mathbf{w}_{r+1}(f, 2^{-j}, \widehat{T})_{p,p}$ we obtain the estimate:

$$\begin{aligned}
\sum_{T \in \mathcal{I}_j^{T^*}} \omega_{r+1}(f, \widehat{T})^p_p &\lesssim \sum_{T \in \mathcal{I}_j^{T^*}} \mathbf{w}_{r+1}(f, 2^{-j}, \widehat{T})^p_{p,p} \\
&\lesssim \sum_{T \in \mathcal{I}_j^{T^*}} \left(\frac{1}{2^{-jd}} \int_{[-2^{-j}, 2^{-j}]^d} \int_{\widehat{T}} |\Delta_h^k(f, x, \widehat{T})|^p dx dh \right) \\
&\lesssim \frac{1}{2^{-jd}} \int_{[-2^{-j}, 2^{-j}]^d} \int_{\widehat{T^*}} |\Delta_h^k(f, x, \widehat{T})|^p dx dh \\
&\lesssim \mathbf{w}_{r+1}(f, 2^{-j}, \widehat{T^*})^p_{p,p} \lesssim \omega_{r+1}(f, 2^{-j}, \widehat{T^*})^p_p,
\end{aligned}$$

where we have used that at most c patches $\widehat{T}$ for $T \in \mathcal{I}_j^{T^*}$ can overlap. Inserting this last bound into (5.8) we obtain

$$|f - Q_{\mathcal{T}}(f)|^p_{B^\alpha_{p,p}(\Omega)} \lesssim \sum_{T^* \in \mathcal{T}} \sum_{j \in \mathbb{Z}} 2^{j\alpha p} \omega_{r+1}(f, 2^{-j}, \widehat{T^*})^p_p \lesssim \sum_{T^* \in \mathcal{T}} |f|^p_{B^\alpha_{p,p}(\widehat{T^*})}, \tag{5.9}$$

where we have used the equivalent definition (4.11) one last time. □

In the proof we have used the following well known inequality, whose proof can be found in [DeL 1993].

Lemma 5.1 (Discrete Hardy's Inequality, [DeL 1993])**.** *Let $\{a_m\}_{m\in\mathbb{Z}}$ and $\{z_m\}_{m\in\mathbb{Z}}$ be two sequences of positive real numbers such that for some $C_0 > 0$ and $\mu > 0$, we have the inequality:*

$$z_m \le C_0 \Big(\sum_{j=m}^{\infty} a_j^\mu \Big)^{\frac{1}{\mu}}, \qquad \textit{for all } m \in \mathbb{Z}. \tag{5.10}$$

Then for all $\alpha > 0$ and $p > 0$, there exists $C = C(\alpha, p)$ such that:

$$\Big(\sum_{m\in\mathbb{Z}} \big(2^{m\alpha} z_m\big)^p \Big)^{\frac{1}{p}} \le C C_0 \Big(\sum_{m\in\mathbb{Z}} \big(2^{m\alpha} a_m\big)^p \Big)^{\frac{1}{p}}.$$

6. Proof of Direct Theorem

In this section we present the main ideas for proving the desired rate of convergence necessary to obtain the direct theorem (Theorem 2.2). The main step of the proof is the construction of optimal meshes that will imply (2.7). This construction is based on the interpolation estimates of Proposition 2.1.

6.1. Construction of Optimal Meshes.

We start this section presenting a complexity result for the bisection rules considered here. The following theorem was proved for $d = 2$ in [BDD 2004] and for $d \ge 2$ in [S 2007], and it is crucial for controlling the extra refinements necesary to keep the partitions admissible and shape-regular. The theorem is based on the existence of an algorithm

$$\mathcal{T}_* \;\leftarrow\; \mathsf{REFINE}(\mathcal{T}, \mathcal{M})$$

which, given an admissible mesh $\mathcal{T}$, and a set $\mathcal{M} \subset \mathcal{T}$ of *marked* elements, bisects all elements in $\mathcal{M}$ least once, and outputs the smallest conforming refinement $\mathcal{T}_* \in \mathbb{T}$ of $\mathcal{T}$ with $\mathcal{T}_* \cap \mathcal{M} = \emptyset$. Such an algorithm exists provided $\mathcal{T}_0$ satisfies condition (b) of Section 4 in [S 2007], or Assumption 1 in [NSV 2009, Ch. 4]. A further discussion of the results by Stevenson [S 2007] can be found in [NSV 2009, Ch. 4], including practical recursive and iterative implementations.

Theorem 6.1. *Let $\mathcal{T}_0$ be an initial admissible partition of a polygonal (polyhedral) domain Ω in $\mathbb{R}^2$ ($\mathbb{R}^d$) and assume that $\mathcal{T}_0$ satisfies condition (b) of Section 4 in* [S 2007]*. If the sequence $\{\mathcal{T}_\ell\}_{\ell\ge1}$ is obtained by repeating the step:*

$$\mathcal{T}_{\ell+1} \;\leftarrow\; \mathsf{REFINE}(\mathcal{T}_\ell, \mathcal{M}_\ell),$$

with $\mathcal{M}_\ell$ any subset of $\mathcal{T}_\ell$, then for $k \ge 1$ we have that

$$\#\mathcal{T}_k - \#\mathcal{T}_0 \le \mathcal{C} \Big(\sum_{\ell=1}^{k} \mathcal{M}_\ell \Big),$$

where $\mathcal{C}$ only depends on $\mathcal{T}_0$.

The proof of Theorem 2.2 is based mostly in the following constructive result, called *greedy algorithm* that was presented initially in [BDDP 2002, Proposition 5.2], making use of a result analogous to Proposition 2.1 for linear finite elements. In [NSV 2009, Section 5.4] this construction is used to prove that $W^{2,p}(\Omega) \subset \mathbb{A}_{1/2}\big(H^1(\Omega)\big)$ for any $p > 1$ when Ω is a two dimensional domain. We include this construction in order to illustrate the use of the bounds of Proposition 2.1 and thus make the article more self-contained:

Proposition 6.2. *Suppose that the assumptions of Theorem 2.2 hold, then for all $\epsilon > 0$ there exists an admissible mesh $\mathcal{T} \in \mathbb{T}$, such that:*

$$\|f - Q_{\mathcal{T}}(f)\|_{B_0} \le C_4\, (\#\mathcal{T})^{\frac{1}{p}}\, \epsilon \tag{6.1}$$

with

$$\#\mathcal{T} - \#\mathcal{T}_0 \le C_5\, (\epsilon^{-1}|\Omega|^{\delta}|f|_B)^{\frac{\tau}{1+\delta\tau}} \tag{6.2}$$

where $\delta = \frac{s}{d} + \frac{1}{p} - \frac{1}{\tau}$ and $C_i = C_i(p, s, \tau, d, \mathrm{diam}(\Omega), \kappa_{\mathbb{T}})$, $i = 1, 2$.

Proof. Given an admissible mesh $\mathcal{T} \in \mathbb{T}$ and $T \in \mathcal{T}$ we define the *local error* as:

$$e(T, \mathcal{T}) := |T|^{\delta}|f|_{B(\omega_{\mathcal{T}}(T))}, \tag{6.3}$$

where $\omega_{\mathcal{T}}(T)$ is the patch of elements of $\mathcal{T}$ having nonempty intersection with T.

To construct the desired mesh we fix the tolerance $\epsilon > 0$ and generate recursively the sequence of meshes $\{\mathcal{T}_k\}_{k\ge 0}$ with the following algorithm:

$k = 0$
$\mathcal{M}_k = \{T \in \mathcal{T}_k : e(T, \mathcal{T}_k) > \epsilon\}$
`while` $\mathcal{M}_k \ne \emptyset$
  $\mathcal{T}_{k+1} \leftarrow \mathsf{REFINE}(\mathcal{T}_k, \mathcal{M}_k)$
  $k \leftarrow k + 1$
  $\mathcal{M}_k = \{T \in \mathcal{T}_k : e(T, \mathcal{T}_k) > \epsilon\}$
`end while`

The procedure ends in a finite number of steps because Proposition 2.1 implies that

$$e(T, \mathcal{T}) = |T|^{\delta}|f|_{B(\omega_{\mathcal{T}}(T))} \le |T|^{\delta}|f|_B$$

and $|T|$ is halved when T is refined.

Let $\mathcal{T}$ be the final mesh $\mathcal{T}_k$. By the definition of $\mathcal{M}_k$ in each case, we have that $T \in \mathcal{T}$ satisfies $e(T, \mathcal{T}) \le \epsilon$ and then, again by Proposition 2.1

$$\|f - Q_{\mathcal{T}}(f)\|^p_{B_0(\Omega)} \lesssim \sum_{T\in\mathcal{T}} |f|^p_{B_0(\omega_{\mathcal{T}}(T))} \lesssim \sum_{T\in\mathcal{T}} |T|^{\delta}|f|^p_{B(\omega_{\mathcal{T}}(T))} \lesssim (\#\mathcal{T})\, \epsilon^p$$

where the constants involved depend on p, s, d, $\mathrm{diam}(\Omega)$ and $\kappa_{\mathbb{T}}$. This is the first assertion of the proposition.

To prove the second inequality, we will bound the cardinality of the set $\mathcal{M}$ of all the elements marked in the process to obtain $\mathcal{T}$. More precisely, let $\mathcal{M} = \cup_{\ell=0}^{k}\mathcal{M}_\ell$, noticing that the sets $\mathcal{M}_\ell$, $\ell = 0, 1, \ldots, k$ are pairwise disjoint, and for each $j \in \mathbb{Z}$ define $\Gamma_j = \{T \in \mathcal{M} : 2^{-j-1} \le |T| < 2^{-j}\}$. If j_0 is the smallest integer such that $2^{j_0} > |\Omega|$, then $\mathcal{M} = \cup_{j=-j_0}^{\infty}\Gamma_j$.

We now obtain two different upper bounds for $\#\Gamma_j$. The first bound comes from observing that the elements in Γ_j do not overlap, and thus $\#\Gamma_j\, 2^{-j-1} \le \sum_{T\in\Gamma_j} |T| \le |\Omega|$ which leads to

$$\#\Gamma_j \le 2^{j+1}|\Omega|, \quad \text{for each } j \ge -j_0. \tag{6.4}$$

To obtain the second bound for $\#\Gamma_j$ we observe that each $T \in \Gamma_j$ belongs to $\mathcal{M}_\ell$ for only one $\ell \in \{0, 1, 2, \dots, k\}$ and setting $\widehat{T} = \omega_{\mathcal{T}_\ell}(T)$ we see that

$$\epsilon < e(T, \mathcal{T}_\ell) = |T|^\delta |f|_{B(\widehat{T})} \leq 2^{-j\delta} |f|_{B(\widehat{T})}.$$

Since the elements of Γ_j are pairwise disjoint and of comparable size, each $x \in \Omega$ belongs to at most c sets $\widehat{T}$ with $T \in \Gamma_j$ (with c depending only on mesh regularity), and thus

$$\#\Gamma_j \epsilon^\tau \leq C_1^\tau 2^{-j\delta\tau} \sum_{T \in \Gamma_j} |f|^\tau_{B(\widehat{T})} \lesssim 2^{-j\delta\tau} |f|_B^\tau, \quad \text{for each } j \geq -j_0,$$

where in the last inequality we have used Lemma 4.10.

Therefore, for each $j \geq -j_0$

$$\#\Gamma_j \lesssim \min\{2^j |\Omega|, \epsilon^{-\tau} 2^{-j\delta\tau} |f|_B^\tau\}$$

and

$$\#\mathcal{M} = \sum_{j=-j_0}^{\infty} \#\Gamma_j \lesssim \sum_{j=-j_0}^{\infty} \min\{2^j |\Omega|, \epsilon^{-\tau} 2^{-j\delta\tau} |f|_B^\tau\}.$$

Notice that the two terms inside $\min\{\cdot, \cdot\}$ correspond to geometric series, one increasing and the other decreasing. The sum is then bounded (up to a constant) by the size of the terms when they are of comparable size. More precisely, if we let k be the biggest integer such that $2^k |\Omega| \leq \epsilon^{-\tau} 2^{-k\delta\tau} |f|_B^\tau$, then

$$\#\mathcal{M} \lesssim \epsilon^{-\tau} 2^{-k\delta\tau} |f|_B^\tau,$$

and also $2^k \simeq \left(|\Omega|^{-1} \epsilon^{-\tau} |f|_B^\tau\right)^{\frac{1}{1+\delta\tau}}$.

Finally,

$$\#\mathcal{M} \lesssim \epsilon^{-\tau} 2^{-k\delta\tau} |f|_B^\tau \lesssim \epsilon^{-\tau} |f|_B^\tau \left[\left(|\Omega|^{-1} \epsilon^{-\tau} |f|_B^\tau\right)^{\frac{1}{1+\delta\tau}}\right]^{-\delta\tau} = \left(\epsilon^{-1} |f|_B |\Omega|^\delta\right)^{\frac{\tau}{1+\delta\tau}},$$

and using Theorem 6.1 we obtain:

$$\#\mathcal{T} - \#\mathcal{T}_0 \lesssim (\epsilon^{-1} |f|_B |\Omega|^\delta)^{\frac{\tau}{1+\delta\tau}}, \tag{6.5}$$

which is the second and last assertion of the proposition. □

6.2. **Proof of Direct Theorem.** The proof of the Direct Theorem 2.2 is now an almost immediate consequence of the construction by the greedy algorithm given in Proposition 6.2.

Proof of Theorem 2.2. Given $N \geq \#\mathcal{T}_0$, we let $\mathcal{T}$ be the mesh given by Proposition 6.2, with

$$\epsilon = |\Omega|^\delta |f|_B N^{-\frac{1+\delta\tau}{\tau}} C_5^{\frac{1+\delta\tau}{\tau}} = |\Omega|^\delta |f|_B N^{-\frac{1}{\tau}-\delta} C_5^{\frac{1}{\tau}+\delta}, \tag{6.6}$$

and C_5 from (6.2). Then (6.2) implies that $\#\mathcal{T} - \#\mathcal{T}_0 \leq N$ and (6.1) now yields

$$\sigma_N(f)_{B_0} \leq \|f - Q_\mathcal{T}(f)\|_{B_0} \lesssim (\#\mathcal{T})^{1/p} \epsilon \leq (N + \#\mathcal{T}_0)^{\frac{1}{p}} \epsilon \leq (2N)^{1/p} \epsilon, \tag{6.7}$$

where we have used that $N \geq \#\mathcal{T}_0$. Using (6.6) and the fact that $\delta = \frac{s}{d} + \frac{1}{p} - \frac{1}{\tau}$, we have that

$$\sigma_N(f)_{B_0} \lesssim N^{-s/d} |f|_B.$$

The result for all $N > 0$ follows easily from the case $N \geq \#\mathcal{T}_0$. □

7. Proof of Inverse Theorem

The main goal of this section is to prove Theorem 2.5, which is a kind of inverse result of Theorem 2.2 in the sense that it states that if a function can be approximated with rate $N^{-s/d}$ then it belongs to certain smoothness spaces. It is important to note that if $u \in \mathbb{V}_{\mathcal{T}}$ for an admissible mesh $\mathcal{T}$, then $u \in \mathbb{A}_{s,q}$ for all $s > 0$ and $q > 0$, since $\sigma_N(u) = 0$ for all $N \geq \#\mathcal{T}$. On the other hand, the spaces $\mathbb{V}_{\mathcal{T}}$ only guarantee C^0 regularity (continuity), and are not contained in Besov spaces with smooth index $s \geq 1 + \frac{1}{p}$ (see Proposition 4.7). Thus an exact recyprocal to Theorem 2.2 is not possible. For this reason we define the generalized Besov space using the multiscale norm defined in (4.17) in the following way

Definition 7.1. The generalized Besov space $\widehat{B}^s_{p,p}(\Omega)$ for $0 < p < \infty$ and $s > 0$, is defined as the set of functions $f \in L^p(\Omega)$ such that the norm

$$\|f\|_{\widehat{B}^s_{p,p}(\Omega)} := \left(\sum_{\nu \in \Xi} |\theta_\nu|^{-\frac{sq}{d}} \|b_\nu(f)\phi_\nu\|^p_{L^p(\Omega)} \right)^{\frac{1}{p}} \tag{7.1}$$

is finite, where the coefficients $b_\nu(f)$ are defined in (4.16).

Remark 7.2. It is important to note that due to Theorem 4.13 each space $\widehat{B}^s_{p,p}(\Omega)$ contains the corresponding Besov space $B^s_{p,p}(\Omega)$, and also all the functions of $\mathbb{V}_{\mathcal{T}}$ for any admissible mesh $\mathcal{T}$. If the parameter s is big, this implies necessarily that $B^s_{p,p}(\Omega) \subsetneq \widehat{B}^s_{p,p}(\Omega)$, since we only consider C^0 finite elements.

The following theorem gives us an inverse inequality where the strongest norm of discrete functions is bounded by a weaker norm, a fundamental tool for proving Theorem 2.5.

Theorem 7.3 (Inverse Inequalities). *Let $0 < p < \infty$, $\alpha > 0$, $s > 0$, $\frac{1}{\tau} = \frac{s}{d} + \frac{1}{p}$. Then for all $V \in \mathbb{V}_{\mathcal{T}}$, $\mathcal{T} \in \mathbb{T}_N$, $N \in \mathbb{N}$*

$$\|V\|_{\widehat{B}^{s+\alpha}_{\tau,\tau}(\Omega)} \leq C N^{\frac{s}{d}} \|V\|_{\widehat{B}^{\alpha}_{p,p}(\Omega)}, \qquad (\alpha > 0) \tag{7.2}$$

$$\|V\|_{\widehat{B}^{s}_{\tau,\tau}(\Omega)} \leq C N^{\frac{s}{d}} \|V\|_{L^p(\Omega)}, \qquad (\alpha = 0) \tag{7.3}$$

where $C = C(p, \alpha, s, d, \kappa_{\mathbb{T}}, \#\mathcal{T}_0)$.

Proof. It is sufficient to prove the assertions for $N \in \mathbb{N}$, $N \geq \#\mathcal{T}_0$. Let $\mathcal{T} \in \mathbb{T}_N$ and $V \in \mathbb{V}_{\mathcal{T}}$, using the multiscale decomposition from the initial mesh $\mathcal{T}_0$ we obtain:

$$V = \sum_{m \geq 0} \sum_{\nu \in \Xi_m} b_\nu(V)\phi_\nu = \sum_{\nu \in \mathcal{M}} b_\nu(V)\phi_\nu$$

where $\mathcal{M} \subset \Xi$ is the set of all nodes ν such that the coefficient $b_\nu(V) \neq 0$.

The next step consists in counting how many coefficients are not zero in the previous representation, that is, finding $\#\mathcal{M}$. Consider the tree $\mathbb{T}_{\mathcal{T}}$ from the construction of $\mathcal{T}$ from $\mathcal{T}_0$ by bisection. Since the polynomial degree of the finite element space is fixed, the number of nodes $\nu \in \Xi$ for which $b_\nu \neq 0$ is bounded by the total number of nodes in all the elements $T \in \mathbb{T}_{\mathcal{T}}$, whence $\#\mathcal{M} \lesssim \#\mathbb{T}_{\mathcal{T}}$. Since each bisection increases in one the number of leaf elements ($\#\mathcal{T}$) and in two the total number of elements ($\#\mathbb{T}_{\mathcal{T}}$) we thus obtain

$$\#\mathcal{M} \lesssim \#\mathbb{T}_{\mathcal{T}} \lesssim N + \#\mathcal{T}_0 \lesssim N.$$

Since $|\theta_\nu|^{-\frac{s}{d}}\|\phi_\nu\|_{L^\tau(\Omega)} \simeq \|\phi_\nu\|_{L^p(\Omega)}$ we have

$$\begin{aligned}
\|V\|_{\widehat{B}^{s+\alpha}_{\tau,\tau}(\Omega)} &= \left(\sum_{\nu\in\mathcal{M}} |\theta_\nu|^{-\frac{(\alpha+s)\tau}{d}} \|b_\nu(V)\phi_\nu\|^\tau_{L^\tau(\Omega)}\right)^{\frac{1}{\tau}} \\
&\simeq \left(\sum_{\nu\in\mathcal{M}} \left(|\theta_\nu|^{-\frac{\alpha}{d}} \|b_\nu(V)\phi_\nu\|_{L^p(\Omega)}\right)^\tau\right)^{\frac{1}{\tau}} \\
&\lesssim (\#\mathcal{M})^{\frac{1-\frac{\tau}{p}}{\tau}} \left(\sum_{\nu\in\mathcal{M}} \left(|\theta_\nu|^{-\frac{\alpha}{d}} \|b_\nu(V)\phi_\nu\|_{L^p(\Omega)}\right)^p\right)^{\frac{1}{p}} \\
&\lesssim N^{\frac{s}{d}} \|V\|_{\widehat{B}^{\alpha}_{p,p}(\Omega)},
\end{aligned}$$

due to Hölder's inequality. This concludes the proof of the first assertion of the theorem; the second assertion is analogous. □

In order to state the last tool that we will use for proving Theorem 2.5 we need to introduce the K-functional.

Definition 7.4. For two function spaces $\mathbb{F}_2 \subset \mathbb{F}_1$, the K-functional of $\mathbb{F}_1$, $\mathbb{F}_2$, for $f \in \mathbb{F}_1$ is defined as:

$$K(f,t,\mathbb{F}_1,\mathbb{F}_2) := \inf_{g\in\mathbb{F}_2} \{\|f-g\|_{\mathbb{F}_1} + t\|g\|_{\mathbb{F}_2}\}, \qquad t>0.$$

Then, the K-method of real interpolation consists in defining, for $0<\theta<1$ and $0<q<\infty$, the interpolation space $[\mathbb{F}_1,\mathbb{F}_2]_{\theta,q}$ as the set of all $f\in\mathbb{F}_1$ such that $\|f\|_{[\mathbb{F}_1,\mathbb{F}_2]_{\theta,q}}<\infty$, where

$$\|f\|^q_{[\mathbb{F}_1,\mathbb{F}_2]_{\theta,q}} := \int_0^\infty t^{-\theta q} K(f,t,\mathbb{F}_1,\mathbb{F}_2)^q \,\frac{dt}{t},$$

and due to the fact that $K(f,t,\mathbb{F}_1,\mathbb{F}_2)$ is increasing as a function of t,

$$\|f\|^q_{[\mathbb{F}_1,\mathbb{F}_2]_{\theta,q}} \simeq \sum_{n=0}^\infty \left[a^{n\theta} K(f,a^{-n},\mathbb{F}_1,\mathbb{F}_2)\right]^q, \tag{7.4}$$

with equivalence constant depending on $a>1$.

The following bound for the K-functional is the last tool that we need for proving Theorem 2.5.

Lemma 7.5. *Let $0<p<\infty$, $\alpha>0$, $s>0$ and $\frac{1}{\tau}=\frac{s}{d}+\frac{1}{p}$. There exists a constant $C=C(p,\alpha,d,s,\tau,\Omega)$ such that for $\tau_*=\min\{\tau,1\}$:*

(1) If $\alpha>0$ for $f\in\widehat{B}^\alpha_{p,p}(\Omega)$ and all $n\in\mathbb{N}$ we have:

$$\begin{aligned}
&K\big(f,2^{-\frac{sn}{d}},\widehat{B}^\alpha_{p,p}(\Omega),\widehat{B}^{\alpha+s}_{\tau,\tau}(\Omega)\big) \\
&\qquad\le C2^{-\frac{sn}{d}}\left[\left(\sum_{k=0}^n (2^{\frac{sk}{d}}\sigma_{2^k}(f)_{\widehat{B}^\alpha_{p,p}(\Omega)})^{\tau_*}\right)^{\frac{1}{\tau_*}} + \|f\|_{\widehat{B}^\alpha_{p,p}(\Omega)}\right].
\end{aligned}$$

(2) If $\alpha = 0$ for $f \in L^p(\Omega)$ and all $n \in \mathbb{N}$ we have:

$$K\big(f, 2^{-\frac{sn}{d}}, L^p(\Omega), \widehat{B}^s_{\tau,\tau}(\Omega)\big) \leq C 2^{-\frac{sn}{d}} \left[\left(\sum_{k=0}^{n} (2^{\frac{sk}{d}} \sigma_{2^k}(f)_{L^p(\Omega)})^{\tau_*} \right)^{\frac{1}{\tau_*}} + \|f\|_{L^p(\Omega)} \right].$$

Proof. We prove here the first assertion of the lemma. The second one is analogous and we thus omit it. Let $f \in \widehat{B}^\alpha_{p,p}(\Omega)$, and for each $k \in \mathbb{N}$, let $\widehat{\mathcal{T}}_k \in \mathbb{T}_{2^k}$ and $f_k \in \mathbb{V}_{\widehat{\mathcal{T}}_k}$ be such that $\|f - f_k\|_{\widehat{B}^\alpha_{p,p}(\Omega)} = \sigma_{2^k}(f)_{\widehat{B}^\alpha_{p,p}(\Omega)}$. Then if $g_k = f_k - f_{k-1}$ for $k \in \mathbb{N}$ with $f_{-1} = 0$, we obtain:

$$\begin{aligned} 2^{\frac{sn}{d}} K(f, 2^{-\frac{sn}{d}}, \widehat{B}^\alpha_{p,p}(\Omega), \widehat{B}^{\alpha+s}_{\tau,\tau}(\Omega)) &\leq \|f_n\|_{\widehat{B}^{\alpha+s}_{\tau,\tau}} + 2^{\frac{sn}{d}} \|f - f_n\|_{\widehat{B}^\alpha_{p,p}(\Omega)} \\ &= \left\| \sum_{k=0}^{n} g_k \right\|_{\widehat{B}^{\alpha+s}_{\tau,\tau}(\Omega)} + 2^{\frac{sn}{d}} \|f - f_n\|_{\widehat{B}^\alpha_{p,p}(\Omega)} \\ &\leq \left(\sum_{k=0}^{n} \|g_k\|^{\tau_*}_{\widehat{B}^{\alpha+s}_{\tau,\tau}(\Omega)} + 2^{\frac{sn\tau_*}{d}} \|f - f_n\|^{\tau_*}_{\widehat{B}^\alpha_{p,p}(\Omega)} \right)^{\frac{1}{\tau_*}}, \end{aligned}$$

for $\tau_* = \min\{\tau, 1\}$. Using (7.2) we have

$$\begin{aligned} &2^{\frac{sn}{d}} K\big(f, 2^{-\frac{sn}{d}}, \widehat{B}^\alpha_{p,p}(\Omega), \widehat{B}^{\alpha+s}_{\tau,\tau}(\Omega)\big) \\ &\quad \lesssim \left(\sum_{k=0}^{n} 2^{\frac{sk\tau_*}{d}} \|g_k\|^{\tau_*}_{\widehat{B}^\alpha_{p,p}(\Omega)} + 2^{\frac{sn\tau_*}{d}} \|f - f_n\|^{\tau_*}_{\widehat{B}^\alpha_{p,p}(\Omega)} \right)^{\frac{1}{\tau_*}} \\ &\quad \lesssim \left(\sum_{k=0}^{n} 2^{\frac{sk\tau_*}{d}} \left[\sigma_{2^k}(f)^{\tau_*}_{\widehat{B}^\alpha_{p,p}(\Omega)} + \sigma_{2^{k-1}}(f)^{\tau_*}_{\widehat{B}^\alpha_{p,p}(\Omega)} \right] + 2^{\frac{sn\tau_*}{d}} \sigma_{2^n}(f)^{\tau_*}_{\widehat{B}^\alpha_{p,p}(\Omega)} \right)^{\frac{1}{\tau_*}} \\ &\quad \lesssim \left(\sum_{k=0}^{n} 2^{\frac{sk\tau_*}{d}} \sigma_{2^k}(f)^{\tau_*}_{\widehat{B}^\alpha_{p,p}(\Omega)} \right)^{\frac{1}{\tau_*}} + \|f\|_{\widehat{B}^\alpha_{p,p}(\Omega)}, \end{aligned}$$

and the lemma is proved. $\square$

We proceed now with the proof of the final result:

Proof of Theorem 2.5. As before, we prove the first assertion, since the second one is analogous. Suppose that the assumptions of Theorem 2.5 hold with $\alpha > 0$. Let s_1 and τ_1 be such that $s_1 > s$ and $\frac{1}{\tau_1} = \frac{s_1}{d} + \frac{1}{p}$, and $\tau_1 \geq 1$ if $\tau > 1$. We claim that

$$\widehat{B}^{\alpha+s}_{\tau,\tau}(\Omega) = \left[\widehat{B}^\alpha_{p,p}(\Omega), \widehat{B}^{\alpha+s_1}_{\tau_1,\tau_1}(\Omega) \right]_{\frac{s}{s_1},\tau}$$

and

$$\|f\|_{\widehat{B}^{\alpha+s}_{\tau,\tau}(\Omega)} \simeq \|f\|_{\left[\widehat{B}^\alpha_{p,p}(\Omega), \widehat{B}^{\alpha+s_1}_{\tau_1,\tau_1}(\Omega)\right]_{\frac{s}{s_1},\tau}}. \tag{7.5}$$

This last observation is due to the fact that

$$S : \widehat{B}^\gamma_{\zeta,\zeta}(\Omega) \to \ell^\gamma_\zeta(L^\zeta(\Omega)) \qquad \text{defined by} \quad S(f) = \{q_m(f)\}_{m=0}^\infty$$

is a corretraction operator and

$$R : \ell^{\gamma}_{\zeta}(L^{\zeta}(\Omega)) \to \widehat{B}^{\gamma}_{\zeta,\zeta}(\Omega) \qquad \text{defined by} \quad R(\{f_j\}_{j=0}^{\infty}) = \sum_{j=0}^{\infty} f_j$$

is its correspondent retraction operator, and using that

$$\left[\ell^p_{\alpha}(L^p(\Omega)), \ell^{\tau_1}_{\alpha+s_1}(L^{\tau_1}(\Omega))\right]_{\frac{s}{s_1},\tau} = \ell^{\tau}_{\alpha+s}(L^{\tau}(\Omega))$$

(see [Pe 1976]) and Theorem (1.18.2) of [T 1978].

Therefore, using (7.5) and (7.4) with $a = 2^{\frac{s_1}{d}}$,

$$\|f\|_{\widehat{B}^{\alpha+s}_{\tau,\tau}(\Omega)} \simeq \left(\sum_{n=0}^{\infty} \left[2^{\frac{sn}{d}} K(f, 2^{-\frac{s_1 n}{d}}, \widehat{B}^{\alpha}_{p,p}(\Omega), \widehat{B}^{\alpha+s_1}_{\tau_1,\tau_1}(\Omega)) \right]^{\tau} \right)^{\frac{1}{\tau}}, \tag{7.6}$$

and Lemma 7.5 implies that

$$\begin{aligned} K(f, 2^{-\frac{s_1 n}{d}}, \widehat{B}^{\alpha}_{p,p}(\Omega), \widehat{B}^{\alpha+s_1}_{\tau_1,\tau_1}(\Omega)) \\ \lesssim 2^{-\frac{s_1 n}{d}} \left(\sum_{k=0}^{n} \left(2^{\frac{s_1 k}{d}} \sigma_{2^k}(f)_{\widehat{B}^{\alpha}_{p,p}(\Omega)} \right)^{\tau_*} \right)^{\frac{1}{\tau_*}} + 2^{-\frac{s_1 n}{d}} \|f\|_{\widehat{B}^{\alpha}_{p,p}(\Omega)} . \end{aligned} \tag{7.7}$$

Relations (7.6) and (7.7) yield

$$\begin{aligned} \|f\|^{\tau}_{\widehat{B}^{\alpha+s}_{\tau,\tau}(\Omega)} \lesssim \sum_{n=0}^{\infty} 2^{\frac{(s-s_1)n\tau}{d}} \left(\sum_{k=0}^{n} \left(2^{\frac{s_1 k}{d}} \sigma_{2^k}(f)_{\widehat{B}^{\alpha}_{p,p}(\Omega)} \right)^{\tau_*} \right)^{\frac{\tau}{\tau_*}} \\ + \sum_{n=0}^{\infty} (2^{\frac{(s-s_1)n}{d}} \|f\|_{\widehat{B}^{\alpha}_{p,p}(\Omega)})^{\tau} \end{aligned}$$

Using that $\|f\|_{\widehat{B}^{\alpha}_{p,p}(\Omega)} \leq \|f\|_{\mathbb{A}^r_{\frac{s}{d},\tau}(\widehat{B}^{\alpha}_{p,p}(\Omega))}$ and Hardy's inequality (Lemma 5.1) on the first sum with $\mu = \tau_*$,

$$a_j := \begin{cases} 2^{-\frac{s_1 j}{d}} \sigma_{2^{-j}}(f)_{\widehat{B}^{\alpha}_{p,p}(\Omega)}, & \text{if } j \leq 0, \\ 0, & \text{if } j > 0, \end{cases}$$

and

$$z_m := \left(\sum_{j=m}^{\infty} a_j^{\tau_*} \right)^{\frac{1}{\tau_*}} = \begin{cases} \left(\sum_{k=0}^{-m} \left(2^{\frac{s_1 k}{d}} \sigma_{2^k}(f)_{\widehat{B}^{\alpha}_{p,p}(\Omega)} \right)^{\tau_*} \right)^{\frac{1}{\tau_*}}, & \text{if } m \leq 0, \\ 0, & \text{if } m > 0, \end{cases}$$

we obtain:

$$\|f\|_{\widehat{B}^{\alpha+s}_{\tau,\tau}(\Omega)} \lesssim \|f\|_{\mathbb{A}^r_{\frac{s}{d},\tau}(\widehat{B}^{\alpha}_{p,p}(\Omega))},$$

and the proof is thus complete. □

Acknowledgements. We want to thank an anonymous referee for carefully reading the manuscript and for pointing out many valuable suggestions, which have helped us improve substantially the quality of this article. We would also like to thank Christoph Hartmann for pointing out some inconsistencies in a previous version of this article.

References

[BDD 2004] P. Binev, W. Dahmen, R.A. DeVore, *Adaptive Finite Element Methods with Convergence Rates*, Numer. Math. 97 (2004), 219–268.

[BDDP 2002] P. Binev, W. Dahmen, R.A. DeVore, P. Petrushev, *Approximation Classes for Adaptive Methods*, Serdica Math. J. 28 (2002), pp. 391-416.

[CKNS 2007] J.M. Cascón, Ch. Kreuzer, R.H. Nochetto, K. Siebert, *Quasi-optimal Convergence Rate for an Adaptive Finite Element Method*, SIAM J. Numer. Anal. 46 (2008), 2524-2550.

[DX 2008] X. Dai, J. Xu, A. Zhou, *Convergence and optimal complexity of adaptive finite element eigenvalue computations*, Numer. Math. 110 (2008), 313–355.

[DL 2004] S. Dekel, D. Leviatan, *Whitney estimates for convex domains with applications to multivariate piecewise polynomial approximation*, Found. Comput. Math. 4 (2004), 345–368.

[DeL 1993] R.A. DeVore, G.G. Lorentz, *Constructive Approximation*, Springer-Verlag, 1993.

[D 1988] Z. Ditzian, *On the Marchaud-type inequality*, Proc. Amer. Math. Soc. 103 (1988), 198–202.

[GM 2010] E.M. Garau and P. Morin, *Convergence and quasi-optimality of adaptive FEM for Steklov eigenvalue problems*, IMA Journal of Numerical Analysis 2010; doi: 10.1093/imanum/drp055

[GM 2008] F.D. Gaspoz, P. Morin, *Convergence rates for adaptive finite elements*, IMA J. Numer. Anal. 29 (2009), 917–936.

[JS 1977] H. Johnen, K. Scherer, *On the equivalence of the K-functional and moduli of continuity and some applications*, Constructive theory of functions of several variables (Proc. Conf. , Math. Res. Inst. , Oberwolfach, 1976), pp. 119–140. Lecture Notes in Math., Vol. 571, Springer, Berlin, 1977.

[K 1994] I. Kossaczký, *A recursive approach to local mesh refinement in two and three dimensions*, J. Comput. Appl. Math. 55 (1994), 275–288.

[M 1995] J. Maubach, *Local bisection refinement for n-simplicial grids generated by reflection*, SIAM J. Sci. Comput. 16 (1995), 210–227.

[NSV 2009] R.H. Nochetto, K.G. Siebert, A. Veeser, *Theory of Adaptive Finite Elements Methods: An Introduction*, in: Multiscale, Nonlinear and Adaptive Approximation: Dedicated to Wolfgang Dahmen on the Occasion of his 60th Birthday, R. DeVore, A. Kunoth, eds., Springer (2009), 409–542.

[O 1990] P. Oswald, *On the Degree of Nonlinear Spline Approximation in Besov-Sobolev Spaces*, J. Approx. Theory. 61, No. 2, (1990), 131–157.

[O 1994] P. Oswald, *Multilevel finite element approximation.* Teubner Scripts on Numerical Mathematics. B. G. Teubner, Stuttgart, 1994.

[Pe 1976] J. Peetre, *New thoughts on Besov spaces*, Duke University, 1976.

[Sh 1983] R.C. Sharpley, *Cone conditions and the modulus of continuity*, Second Edmonton conference on approximation theory, (1983), 341–351.

[S 2006] R. Stevenson, *Optimality of a standard adaptive finite element method*, Found. Comput. Math. 7 (2007), 245–269.

[S 2007] R. Stevenson, *The completion of locally refined simplicial particions created by bisection*, Math. Comp. 77 (2008), 227–241.

[SS 2005] A. Schmidt, K.G. Siebert, *Design of Adaptive Finite Element Software. The Finite Element Toolbox ALBERTA*, Springer-Verlag, New York, 2005.

[Tr 1997] C.T. Traxler, *An algorithm for adaptive mesh refinement in n dimensions*, Computing 59, (1978), 115–137.

[T 1978] H. Triebel, *Interpolation theory, function spaces, differential operators*, North-Holland Publishing Co., Amsterdam-New York, 1978.

[T 2002] H. Triebel, *Function spaces in Lipschitz domains and on Lipschitz manifolds. Characteristic functions as pointwise multipliers*, Rev. Mat. Complut. 15 (2002), no. 2, 475–524.

[vT 1984] J. van Tiel *Convex Analysis. An introductory text*, John Wiley & Sons, Inc., New York, 1984.

Institut für Angewandte Mathematik und Numerische Simulation. Universität Stuttgart. *Address:* Pfaffenwaldring 57, D-70569 Stuttgart, Germany
E-mail address: fernando.gaspoz@ians.uni-stuttgart.de

Facultad de Ingeniería Química and Instituto de Matemática Aplicada del Litoral. Universidad Nacional del Litoral and Consejo Nacional de Investigaciones Científicas y Técnicas. *Address:* IMAL, Güemes 3450, S3000GLN Santa Fe, Argentina
E-mail address: pmorin@santafe-conicet.gov.ar