Industrial Mathematics Institute

2002:19

Approximation classes for adaptive methods

P. Binev, W. Dahmen, R. DeVore and P. Petrushev

IMI

Preprint Series

Department of Mathematics
University of South Carolina

Approximation Classes for Adaptive Methods *

Peter Binev, Wolfgang Dahmen, Ronald DeVore,
and Pencho Petrushev

December 9, 2002

Abstract

Adaptive Finite Element Methods (AFEM) are numerical procedures that approximate the solution to a partial differential equation (PDE) by piecewise polynomials on adaptively generated triangulations. Only recently has any analysis of the convergence of these methods [10, 13] or their rates of convergence [2] become available. In the latter paper it is shown that a certain AFEM for solving Laplace's equation on a polygonal domain $\Omega \subset \mathbb{R}^2$ based on newest vertex bisection has an optimal rate of convergence in the following sense. If, for some $s > 0$ and for each $n = 1, 2, \ldots$, the solution u can be approximated in the energy norm to order $O(n^{-s})$ by piecewise linear functions on a partition P obtained from n newest vertex bisections, then the adaptively generated solution will also use $O(n)$ subdivisions (and floating point computations) and have the same rate of convergence. The question arises whether the class of functions $\mathcal{A}^s$ with this approximation rate can be described by classical measures of smoothness. The purpose of the present paper is to describe such approximation classes $\mathcal{A}^s$ by Besov smoothness.

AMS subject classification: 42C40, 46B70, 26B35, 42B25.

Key Words: adaptive finite element methods, adaptive approximation, n-term approximation, degree of approximation, approximation classes, Besov spaces

This article is in memory of Vasil Popov. He was our colleague and collaborator - but more than that he was our friend.

*This work has been supported by the Office of Naval Research Contract Nr. N0014-91-J1343, the Army Research Office Contract Nr. DAAD 19-02-1-0028, the National Science Foundation grants DMS-0221642 and DMS-0200665, the Deutsche Forschungsgemeinschaft grant SFB 401, the IHP Network "Breaking Complexity" funded by the European Commission and the Alexander von Humboldt Foundation.

1 Introduction

Adaptive finite element methods (AFEM) are used to numerically approximate the solution to a PDE. These methods have been experimentally shown to outperform standard finite element methods for many problems. However, there is only now emerging a rigorous analysis which establishes the increased performance of these methods. For example, the papers of Dörfler [10] and Morin, Nochetto, and Siebert [13] prove the convergence of certain AFEMs which use piecewise linear approximation on partitions adaptively generated using newest vertex bisection (see §2 for a discussion of newest vertex bisection). Even here the results have only been established for Laplace's equation on polygonal domains in $\mathbb{R}^2$ although certain principles carry over to more general settings. Building on these results, an AFEM was introduced by Binev, Dahmen, and DeVore [2] and shown to have an optimal rate of convergence in the following sense. If, for some $s > 0$ and for each $n = 1, 2, \ldots$, the solution u can be approximated in the energy norm to order $O(n^{-s})$ by piecewise linear functions on a partition P obtained from n newest vertex bisections, then the adaptively generated solution will also use $O(n)$ subdivisions and have the same rate of convergence. The question arises whether the class of functions $\mathcal{A}^s$ with this approximation rate can be described by classical measures of smoothness. The purpose of the present paper is to prove theorems that help describe the approximation classes $\mathcal{A}^s$ by using Besov smoothness.

The analysis of AFEMs is somewhat complicated by the issue of so called 'hanging nodes'. If a partition of Ω into triangular cells has hanging nodes then these are removed by a completion process which consists of refining additional triangles. This allows one to retain standard data structures. To prove optimal or near optimal theorems concerning the approximation rates for a specific AFEM, it is necessary to control the number of new triangles added in the completion process. Such an analysis of completion was done in the case of the refinement rule called 'newest vertex bisection'(see §2 in [2]). We also know that it is possible to complete certain other subdivision rules (for example the rule which subdivides each triangle into four triangles by bisecting each edge). However, the completion in these other cases is of a different generic type from the original subdivision rule. That is, there is one rule for subdividing in the adaptive refinement and a different rule for subdividing in the completion. This adds a new layer of complexity in the analysis - at least in the notation. We want to avoid this and therefore, we shall in principle only discuss newest vertex bisection.

Although our analysis could be extended in a straightforward manner to higher order conforming Lagrange finite elements, it will be carried out in detail only in the case of piecewise linear trial spaces on polygonal domains in $\mathbb{R}^2$ to keep the exposition as simple as possible. Despite these restrictions, the results we obtain in this paper should be viewed as a guide on how to obtain similar results in more general settings whenever the necessary results about complexity of removing hanging nodes and existence of locally supported bases are in hand.

The results are similar in spirit to the developments in [14]. There are, however, two main distinctions. Whereas we discuss here approximations from spaces on admissible partitions which necessitates the above mentioned completion processes, in [14, 15] the rate of best N–term approximation for multilevel nodal finite element bases is related to Besov smoothness without such mesh constraints. Moreover, we shall consider here approximation errors that are not only measured in L_p norms but also in Besov norms.

This paper is organized as follows. In §2, we introduce the concept of an adaptive triangulation and related issues such as completion for the newest vertex bisection subdivision rule. In §3 we study the properties of piecewise linear functions on a partition P. In §4.2 we introduce quasi-interpolant operators which project onto these spaces. In §4, we introduce the Besov spaces and give some of their properties including embedding theorems. In §4.4, we introduce certain multiscale decompositions based on Courant elements. These decompositions are very similar to wavelet decompositions except that the collection of Courant elements is redundant. Using these multiscale decompositions, we will establish the equivalence of certain sequence norms with the Besov norms. These sequence norms are then used as a tool in proving our approximation results. The main results of this paper appear in the last two sections. In §5, we prove an embedding of Besov spaces into $\mathcal{A}^s$. For example, we show that whenever $1/\tau < s/2+1/p$, then any $f \in B^s_q(L_\tau(\Omega))$ is in $\mathcal{A}^{s/2}(L_p)$. That is, f can be approximated to accuracy $O(n^{-s/2})$ in the $L_p(\Omega)$-norm by piecewise linear functions on partitions generated by n newest vertex bisections. Similar results are also proved when the approximation takes place in a Besov space $B_0 := B^\alpha_p(L_p(\Omega))$ rather than $L_p(\Omega)$. In the final section we prove certain inverse results. Namely, we show that whenever a function is in $\mathcal{A}^{s/2}_\tau(B_0)$ then it is automatically in the Besov space $B^{s+\alpha}_\tau(L_\tau(\Omega))$ with $1/\tau = s/2 + 1/p$. In both of these theorems there are restrictions on s which come from the fact that we are using piecewise linear functions in the approximation.

2 Adaptive refinement and completion

In this section, we shall introduce the subdivision rule known as *newest vertex bisection.* A complete discussion of this rule can be found in [2]. Let Ω be a polygonal domain in $\mathbb{R}^2$. We shall use P to denote a partition of Ω into triangular cells Δ. This means that $\Omega = \bigcup_{\Delta\in P} \Delta$ and any two Δ, $\Delta' \in P$ satisfy $|\Delta \cap \Delta'| = 0$ where here and later in this paper $|G|$ denotes the Lebesgue measure of $G \subset \mathbb{R}^2$. We denote by $\mathcal{E}_P$ the set of edges of P and by $\dot{\mathcal{E}}_P$ the set of *interior edges.* Thus, $E \in \dot{\mathcal{E}}_P$ means that E is an edge of some $\Delta \in P$ and that the interior of E is in the interior of Ω. All other edges are called *boundary edges.* We also denote by $\mathcal{V}_P$ the set of all vertices v of P and by $\dot{\mathcal{V}}_P$ the set of *interior vertices.* Thus, $v \in \dot{\mathcal{V}}_P$ means that v is a vertex of one of the $\Delta \in P$ and v is in the interior of Ω. All other vertices are called *boundary vertices.*

A typical AFEM generates a sequence of partitions $P_0, P_1, \dots, P_n$ by using rules for subdividing triangles. Given the partition P_k, the algorithm marks a certain collection $\mathcal{M}_k$ of the triangular cells $\Delta \in P_k$ for subdivision. These marked cells are subdivided using the specified subdivision rule. This process may create *hanging nodes*. We say that $v \in \mathcal{V}_P$ is a hanging node for $\Delta \in P$ if v appears in the interior of one of the sides of Δ. Hanging nodes are an impediment to both the numerical computation in AFEMs and also their analysis. We shall say a partition is admissible if it has no hanging nodes. Hanging nodes are eliminated by subdividing an additional collection $\mathcal{M}'_k$ of cells from P_k. This part of the algorithm is called *completion*. The result after both of these sets of subdivisions have been made is the partition P_{k+1} which has no hanging nodes. The adaptive procedure is then repeated.

To execute the analysis we put forward in this paper, we need to know that the completion process does not seriously inflate the number of triangular cells that have been added. We would also like (mostly for notational convenience) that the completion process uses the same subdivision rules as the original process. These properties have been established in the case of a certain method of subdivision known as *newest vertex bisection* for triangular partitions on $\mathbb{R}^2$ in [2]. We shall restrict our presentation in this paper to this setting. The analysis we put forward could be implemented for other subdivision rules for triangular partitions in $\mathbb{R}^d$ provided the corresponding properties are established.

We next describe newest vertex bisection in the form we shall use it. Let Ω be a polygonal domain in $\mathbb{R}^2$ and let P_0 be any admissible partition of Ω into triangular cells Δ. To each edge E of a triangular cell $\Delta \in P_0$, we assign a label of 0 or 1 in such a way that for any $\Delta \in P_0$, exactly one of its edges $E(\Delta)$ has been labelled by 0 and the other two are labelled with 1. The vertex $v(\Delta)$ opposite the side $E(\Delta)$ is call the *newest vertex* for that cell. One can show that such a labelling exists for *any* initial triangulation P_0 (see [2]).

This gives the labelling of newest vertices for the cells Δ in the initial partition P_0. Each triangular cell that arises in the adaptive process will also have exactly one of its vertices designated as a newest vertex. If this cell is to be subdivided then the subdivision is a simple bisection of the newest vertex and the side $E(\Delta)$ opposite. Thus the cell produces two new cells and their newest vertex (assigned to each new triangular cell) is by definition the midpoint of $E(\Delta)$.

We now give a rule to label any edges that arise from the subdivision-completion process. There will be two main properties of this labelling. The first is that each triangular cell will have sides with labels $(i, i, i-1)$ for some positive integer i. The second is that the newest vertex for this cell will be the vertex opposite the side with lowest label. Certainly the edges in P_0 have such a labelling as we have just shown.

Suppose that we have such a labelling for the edges in P_k and let us describe how to label the edges in P_{k+1}. Suppose that a triangular cell $\Delta \in P_k$ has sides which have been labelled $(i, i, i-1)$ and the newest vertex for this cell is the one opposite the side labelled $i-1$. When this cell is subdivided (using newest vertex bisection) the side labelled $i-1$ is bisected and we label each of the two new sides $i+1$. We also label the *bisector* by $i+1$, i.e., the new edge connecting the newest vertex of Δ with the midpoint of the edge $E(\Delta)$ labelled by $i-1$. Thus each new triangle now has sides labelled $(i, i+1, i+1)$ with the newest vertex opposite the side with the lowest label. We note the important fact that if a cell has label $(i+1, i+1, i)$ then it is of generation i (i.e., it has been obtained from a cell in P_0 by i subdivisions). Therefore, specifying that the generation of the cell is i is the same as specifying that its label is $(i+1, i+1, i)$.

The partitions which arise when using newest vertex bisection satisfy a uniform minimal angle condition, i.e. the minimal angle in any triangle that belongs to a partition P generated by any sequence of newest vertex bisections is bounded from below by some positive constant $\beta > 0$ depending only on the initial partition P_0. This is established by showing that all triangles that arise in newest vertex bisection can be classified into a set of similarity classes depending only on the initial partition P_0 (see Mitchell [12]). Also note that if a partition P is created by a sequence of newest vertex bisections and if P has no hanging nodes, then any two neighboring cells have comparable size.

We can represent newest vertex bisection subdivision by an infinite binary tree T_* (which we call the *master tree*). The master tree T_* consists of all triangular cells which can be obtained by a sequence of subdivisions. The roots of the master tree are the triangular cells in P_0. When a cell Δ is subdivided, it produces two new cells which are called the children of Δ and Δ is their parent. It is very important to note that, no matter how a cell arises in a subdivision process, its associated newest vertex is unique and only depends on the initial assignment of newest vertices in P_0. This means that the children of Δ are uniquely determined and do not depend on how Δ arose in the subdivision process, i.e., it does not depend on the preceding sequence of subdivisions. The reason for this is that any subdivision only assigns newest vertices for the new triangular cells produced by the subdivision and does not alter any previous assignment. It follows that T_* is unique and does not depend at all on the order of subdivisions.

The *generation* of a triangular cell Δ is the number $g(\Delta)$ of ancestors it has in the master tree. Thus cells in P_0 have generation 0, their children have generation 1 and so on. The generation of a cell is also the number of subdivisions necessary to create this cell from its corresponding root cell in P_0.

A *subtree* $T \subset T_*$ is a collection of triangular cells $\Delta \in T_*$ with the following two properties: (i) whenever $\Delta \in T$ then its sibling is also in the tree; (ii) when $\Delta \subset \Delta'$ are both in the tree then each triangular cell $\bar{\Delta} \in T_*$ with $\Delta \subset \bar{\Delta} \subset \Delta'$ is also in T. The roots of T are all the cells $\Delta \in T$ whose parents are not in T. We say that T is *proper* if it has the same roots as T_*, i.e., it contains all $\Delta \in P_0$.

If $T \subset T_*$ is a finite subtree, we say $\Delta \in T$ is a *leaf* of T if T contains none of the children of Δ. The set of leaves of a tree T forms a partition of Ω into triangular cells which we will call $P(T)$.

For a proper subtree T, we define $N(T)$ to be the number of subdivisions made to produce T.

Any partition $P = P_n$ which is obtained by the application of an adaptive procedure based on newest vertex bisection (such as the algorithms we consider in this paper) can be associated to a proper subtree $T = T(P)$ of T_* consisting of all triangular cells that were created during the algorithm, i.e., all of the cells in $P_0, \dots, P_n$. The set of leaves of T form the final partition $P = P(T) = P_n$.

We shall say that $T = T(P)$ is admissible if P is admissible. We denote the class of all proper trees by $\mathcal{T}$ and all admissible trees by $\mathcal{T}^a$. We also let $\mathcal{T}_n$ be the set of all proper trees T with $N(T) = n$ and by $\mathcal{T}_n^a$ the corresponding class of admissible trees from $\mathcal{T}_n$. We denote by $\mathcal{P}$ the class of all partitions P that can be generated by newest vertex bisection and by $\mathcal{P}^a$ the set of all admissible partitions. Similarly, $\mathcal{P}_n$ and $\mathcal{P}_n^a$ are the subclasses of those partitions that are obtained from P_0 by using n subdivisions. There is a precise identification between $\mathcal{P}_n$ and $\mathcal{T}_n$. Any $P \in \mathcal{P}_n$ can be given by a tree, i.e., $P = P(T)$ for some $T \in \mathcal{T}_n$. Conversely any $T \in \mathcal{T}_n$ determines a $P = P(T)$ in $\mathcal{P}_n$. The same can be said about admissible partitions and trees.

As we have already mentioned, we shall need control on the number of additional subdivisions used to remove the hanging nodes. For this we shall use the following result of [2]. Suppose that $P_0, \dots, P_n$ is a sequence of partitions generated as described above. Then, there is a constant $c > 0$ depending only on P_0 such that

$$\#(P_n) \le \#(P_0) + c(\#(\mathcal{M}_0) + \dots + \#(\mathcal{M}_{n-1})). \tag{2.1}$$

In other words, the total number of cells in P_n is bounded in terms of the original subdivisions. This can be reformulated in another way. Suppose that P' is obtained from P_0 by performing N newest vertex bisections. Then P' can be completed to an admissible partition P which satisfies

$$\#(P) \le c\#(P') \tag{2.2}$$

with c an absolute constant. In the remainder of this paper we shall frequently refer to the following constants

$$\beta := \min\{\text{angle}\,(\Delta) : \Delta \in T_*\} > 0, \quad d := \max\{\text{diam}\,(\Delta) : \Delta \in T_*\}, \tag{2.3}$$

that depend only on the initial partition P_0.

3 Piecewise linear functions

In this section, we introduce spaces of piecewise linear functions on admissible partitions and analyze some of their properties. Given a partition $P \in \mathcal{P}^a$, we let $\mathcal{S}(P)$ denote the space of continuous piecewise linear functions on P. A basis for this space is given by the Courant elements $\phi_v = \phi_{v,P}$, defined for all vertices $v \in \mathcal{V}_P$. The function ϕ_v takes the value 1 at v and the value 0 at all other vertices.

Suppose now that B_0 is $L_p(\Omega)$ for some $0 < p \le \infty$ or a Besov space defined on Ω (for the definition of Besov spaces and some of their properties see §4). Given a function $f \in B_0$ we are interested in finding how well it can be approximated in the norm of B_0 by linear combinations of finite elements defined on an admissible partition P of complexity n. For each $n \geq 0$, we define

$$\sigma_n(f)_{B_0} := \inf_{P \in \mathcal{P}_n^a} \inf_{S \in \mathcal{S}(P)} \|f - S\|_{B_0} . \tag{3.1}$$

Thus, $\sigma_n(f)_{B_0}$ measures how well f can be approximated in the metric of B_0 by piecewise linear functions on partitions that are obtained by at most n newest vertex bisections.

For a fixed B_0 and any $s > 0$ let $\mathcal{A}^s = \mathcal{A}^s(B_0)$ denote the set of all $f \in B_0$ for which $\sigma_n(f)_{B_0}$ decays at least like n^{-s}. Then,

$$\|f\|_{\mathcal{A}^s} := \|f\|_{\mathcal{A}^s(B_0)} := \sup_{n \in \mathbb{N}} n^s \sigma_n(f)_{B_0} \tag{3.2}$$

defines the norm for $\mathcal{A}^s$. The central objective of this paper is to study the spaces $\mathcal{A}^s$, $s > 0$. More precisely, we shall show that $\mathcal{A}^s$ is closely related to some Besov space.

There are also more general approximation classes that are useful when we wish to make fine distinctions. Let $0 < q < \infty$, then we define the space $\mathcal{A}^s_q(B_0)$ as all $f \in B_0$ such that

$$|f|^q_{\mathcal{A}^s_q(B_0)} := \sum_{n=0}^{\infty} [2^{ns} \sigma_{2^n}(f)_{B_0}]^q \tag{3.3}$$

is finite. We obtain the norm for this space by adding $\|f\|_{B_0}$ to the above semi-norm. For $q = \infty$, we define $\mathcal{A}^s_\infty := \mathcal{A}^s$ and $|\cdot|_{\mathcal{A}^s_\infty} := |\cdot|_{\mathcal{A}^s}$. We shall use these more general approximation spaces when we discuss inverse approximation theorems in §6.

There are simple embeddings of the approximation spaces: if $0 < q_1 \leq q_2 \leq \infty$ then [1].

$$|f|_{\mathcal{A}^s_{q_2}} \lesssim |f|_{\mathcal{A}^s_{q_1}} . \tag{3.4}$$

These embeddings follow from the corresponding embeddings for ℓ_q spaces.

4 Smoothness spaces

Our goal in this paper is to describe the approximation classes $\mathcal{A}^s$, $s > 0$, by means of Besov spaces. There are several equivalent definitions of Besov spaces. We shall give several of these equivalent definitions in this section after collecting in the following two subsections some prerequisites.

[1] Throughout the paper we use the notation $A \lesssim B$ to mean that $A \leq CB$ with a constant C that does not depend on the variables of A and B. If apropriate, we shall indicate on which quantities the constants C depend. We also use $A \simeq B$ for $A \lesssim B \lesssim A$

4.1 Local polynomial approximation and moduli of smoothness

Let Π_r denote the set of all algebraic polynomials in two variables of total degree $< r$. For a function $f \in L_p(G)$, $G \subset \mathbb{R}^2$, and $0 < p \le \infty$, we denote by $E_r(f, G)_p$ the error of L_p-approximation of f from Π_r on G, i.e.,

$$E_r(f, G)_p := \inf_{\pi \in \Pi_r} \|f - \pi\|_{L_p(G)}. \tag{4.1}$$

An important characterization of this error functional is based on the r-th modulus of smoothness of f in $L_p(G)$ which is defined by

$$\omega_r(f, t)_p = \omega_r(f, t, G)_p := \sup_{|h| \le t} \|\Delta_h^r(f, \cdot)\|_{L_p(G)}, \quad t > 0, \tag{4.2}$$

where

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

Also, we denote by $\omega_r(f, G)_p$ the r-th local modulus of smoothness of f on G:

$$\omega_r(f, G)_p := \sup_{t > 0} \omega_r(f, t, G)_p = \sup_{h \in \mathbb{R}^2} \|\Delta_h^r(f, \cdot)\|_{L_p(G)}. \tag{4.3}$$

Whitney's theorem is an important tool in piecewise polynomial approximation. We shall give it in the form we need it. Suppose P is an admissible partition of Ω (with min angle$(\Delta) \ge \beta > 0$ for $\Delta \in P$, see (2.3)). For $\Delta \in P$, we denote by $\widehat{\Delta}$ the union of all triangles from P which have a common vertex with Δ. If $f \in L_p(G)$, $0 < p \le \infty$, where $G = \Delta$ or $G = \widehat{\Delta}$, and $r \ge 1$, then

$$E_r(f, G)_p \le c\, \omega_r(f, G)_p, \tag{4.4}$$

where $c = c(p, r)$ if $G = \Delta$ and $c = c(p, r, \beta)$ if $G = \widehat{\Delta}$. Note that this estimate holds for much more general regions G, but then the constant $c = c(G)$ may become hard to control. For this reason we shall restrict ourselves to using (4.4) only on simple regions G ($G = \Delta$ or $G = \widehat{\Delta}$).

Another important technical tool is the averaged modulus of smoothness which is defined by

$$\mathrm{w}_r(f, t)_p^p = \mathrm{w}_r(f, t, G)_p^p := \frac{1}{t^2} \int_{[0,t]^2} \int_{\Omega} |\Delta_h^r(f, x, G)|^p \, dx\, dh. \tag{4.5}$$

It is well known that $\mathrm{w}_r(f, t)_p$ is equivalent to $\omega_r(f, t)_p$:

$$c_1 \mathrm{w}_r(f, t)_p \le \omega_r(f, t)_p \le c_2 \mathrm{w}_r(f, t)_p, \quad t > 0, \tag{4.6}$$

where $c_1, c_2 > 0$ depend only on p and r (see, e.g., [6] or [16] for the proof of this in the univariate case; the same proof applies in the multivariate case as well).

We shall often use the equivalence of different norms of polynomials. For instance, if $\pi \in \Pi_r$ and $G = \widehat{\Delta}$, then

$$\|\pi\|_{L_p(G)} \simeq |G|^{1/p-1/q} \|\pi\|_{L_q(G)} \tag{4.7}$$

with constants of equivalence depending only on p, q, r, and β.

When dealing with Besov spaces, we operate in L_p spaces with $p \geq 1$ as well as $0 < p < 1$. For the latter case, we need the concept of *near best approximation* (see [7]): A polynomial $\pi \in \Pi_r$ is said to be a *near best L_p-approximation* to f from Π_r on $G \subset \mathbb{R}^2$ if

$$\|f - \pi\|_{L_p(G)} \leq A E_r(f, G)_p$$

with a constant $A \geq 1$. Exactly as in [7], one can prove the following result:

Lemma 4.1. *Suppose $0 < \rho \leq p \leq \infty$ and $\pi \in \Pi_r$ is a near best L_ρ-approximation to f on $G = \Delta$ or $G = \widehat{\Delta}$ with a constant A. Then π is a near best L_p-approximation to f on G with a constant cA, $c = c(r, p, \beta, \rho)$.*

4.2 Courant basis and quasi-interpolants

Recall that for a partition $P \in \mathcal{P}^a$ of Ω, $\mathcal{S}(\mathcal{P})$ denotes the space of all continuous piecewise linear functions on P and that a basis for this space is given by the Courant elements $\phi_v = \phi_{v,P}$, defined for all $v \in \mathcal{V}_P$. We stress that ϕ_v is always normalized in L_∞, i.e. the function ϕ_v takes value 1 at v and zero at all other vertices of P. We shall denote by $\theta = \theta_v$ the support of ϕ_v, that is the union of all triangles from P which share v as a vertex. It will be convenient for us (especially when dealing with multiscale sequences of Courant bases) to use the support θ ($\theta = \theta_v$) to index the corresponding Courant element $\phi =: \phi_\theta$. We shall denote by $\Theta = \Theta(P)$ the set of all (cells) supports of Courant elements generated by P.

Appropriate dual functionals to the Courant basis functions will serve as an important tool. Let $\langle f, g\rangle := \int_{\mathbb{R}^2} fg$ and denote by v_θ the "central" point, i.e. the interior vertex of θ while m_θ is the valence of v_θ. Defining $\tilde{\lambda}_{\Delta,\theta}$ as the linear polynomial which assumes values $\frac{9}{m_\theta|\Delta|}$ at v_θ and $-\frac{3}{m_\theta|\Delta|}$ at the other two vertices of Δ, let $\tilde{\phi}_\theta$ be defined by

$$\tilde{\phi}_\theta := \sum_{\Delta \in P,\, \Delta \subset \theta} \mathbb{1}_\Delta \cdot \tilde{\lambda}_{\Delta,\theta}.$$

Simple calculations show that

$$\langle \phi_\theta, \tilde{\phi}_{\theta'}\rangle = \delta_{\theta,\theta'}, \quad \theta, \theta' \in \Theta(P),$$

where δ is the Kronecker delta. One can use the dual basis to show that for any $\Delta \in P$,

$$\|S\|_{L_p(\Delta)} \simeq \Big(\sum_{\theta \in \Theta : \Delta \subset \theta} \|a_\theta \phi_\theta\|_p^p \Big)^{1/p}. \tag{4.8}$$

This in turn can be used to show that the Courant basis is a stable basis for $\mathcal{S}(P)$ in L_p, $0 < p \le \infty$. By this we mean that for any $S = \sum_{\theta\in\Theta(P)} a_\theta\phi_\theta$ one has

$$\|S\|_{L_p(\Omega)} \simeq (\sum_{\theta\in\Theta} \|a_\theta\phi_\theta\|_p^p)^{1/p} \tag{4.9}$$

with constants of equivalence depending only on p and β.

The well known *quasi-interpolant*:

$$Q_P^\diamond(f) := \sum_{\theta\in\Theta(P)} \langle f, \tilde{\phi}_\theta\rangle\phi_\theta, \tag{4.10}$$

is a linear projector mapping $L_p(\Omega)$ onto $\mathcal{S}(P)$ for $1 \le p \le \infty$. It is local which means that for $f \in L_p(\Omega)$ $(1 \le p \le \infty)$ and $\Delta \in P$, one has

$$\|Q_P^\diamond(f)\|_{L_p(\Delta)} \le c\|f\|_{L_p(\widehat{\Delta})}, \quad c = c(p,\beta). \tag{4.11}$$

Also, if $g = \sum_{\Delta\in P} 1\!\!1_\Delta \cdot \pi_\Delta$ with $\pi_\Delta \in \Pi_2$, then, for $0 < p \le \infty$ and $\Delta \in P$,

$$\|Q_P^\diamond(g)\|_{L_p(\Delta)} \le c\|g\|_{L_p(\widehat{\Delta})}, \quad c = c(p,\beta). \tag{4.12}$$

We now extend $Q_P^\diamond$ in a standard way to a projector from $L_p(\Omega)$ into $\mathcal{S}(P)$ for all $0 < p \le \infty$. Let $\pi_{p,\Delta} : L_p(\Delta) \to \Pi_r$ be a projector (linear if $1 \le p \le \infty$ and nonlinear if $0 < p < 1$) such that

$$\|f - \pi_{p,\Delta}\|_{L_p(\Delta)} \le AE_2(f,\Delta)_p, \quad f \in L_p(\Delta), \tag{4.13}$$

where $A \ge 1$ is a uniform constant and, therefore, $\pi_{p,\Delta}$ is a near best $L_p(\Delta)$-approximation to f from Π_r.

We define

$$\pi_{p,P}(f) := \sum_{\Delta\in P} 1\!\!1_\Delta \cdot \pi_{p,\Delta}(f) \tag{4.14}$$

and set

$$Q_{p,P}(f) := Q_P^\diamond(\pi_{p,P}(f)). \tag{4.15}$$

Evidently, $Q_{p,P} : L_p(\Omega) \to \mathcal{S}(P)$ is a projector (linear if $1 \le p \le \infty$ and nonlinear if $0 < p < 1$). We shall need the following well known property of the quasi-interpolant: If $f \in L_p(\Omega)$ and $0 < \rho \le p \le \infty$, then

$$\|f - Q_{\rho,P}\|_{L_p(\Delta)} \le cE_2(f,\widehat{\Delta})_p, \quad \Delta \in P. \tag{4.16}$$

Indeed, let $\pi_{\widehat{\Delta}} \in \Pi_2$ be such that $\|f - \pi_{\widehat{\Delta}}\|_{L_p(\widehat{\Delta})} = E_2(f,\widehat{\Delta})_p$. Using (4.12) and (4.15), we obtain

$$\begin{aligned}
\|f - Q_{\rho,P}\|_{L_p(\Delta)} &\lesssim \|f - \pi_{\widehat{\Delta}}\|_{L_p(\Delta)} + \|Q_P^\diamond(\pi_{\widehat{\Delta}} - \pi_{\rho,P}(f))\|_{L_p(\Delta)} \\
&\lesssim E_2(f,\widehat{\Delta})_p + \|\pi_{\widehat{\Delta}} - \pi_{\rho,P}(f)\|_{L_p(\widehat{\Delta})} \\
&\lesssim E_2(f,\widehat{\Delta})_p + \Big(\sum_{\Delta\subset\widehat{\Delta}} \|f - \pi_{\rho,\Delta}(f)\|_{L_p(\widehat{\Delta})}^p\Big)^{1/p} \\
&\lesssim E_2(f,\widehat{\Delta})_p,
\end{aligned}$$

where we also used that $\pi_{\rho,\Delta}(f)$ is a near best $L_p(\Delta)$-approximation to f from Π_2 (see Lemma 4.1).

4.3 Besov spaces via moduli of smoothness

We shall restrict ourselves to the setting that Ω is a bounded polygonal domain in $\mathbb{R}^2$ with admissible initial partition P_0 such that min angle $(\Delta) \geq \beta > 0$ for $\Delta \in T_*$. The constants which occur in our main estimates later on will depend on the parameters β and $|\Omega|$. We define $n_0 := \#P_0$. Note that Ω is not necessarily connected.

The Besov space $B^s_q(L_p(\Omega))$, $s > 0$, $0 < q, p \leq \infty$ is the collection of all functions $f \in L_p(\Omega)$ such that

$$|f|_{B^s_q(L_p(\Omega))} := \begin{cases} \left(\int_0^\infty [t^{-s}\omega_r(f,t)_p]^q \frac{dt}{t}\right)^{1/q}, & \text{if } 0 < q < \infty \\ \sup_{t>0} t^{-s}\omega_r(f,t)_p, & \text{if } q = \infty \end{cases} \tag{4.17}$$

is finite, where $r > s$ (usually $r := [s] + 1$).

The norm in $B^s_q(L_p(\Omega))$ is defined by

$$\|f\|_{B^s_q(L_p(\Omega))} := |f|_{B^s_q(L_p(\Omega))} + \|f\|_{L_p(\Omega)}. \tag{4.18}$$

Definition (4.17) is independent of r in the following sense. If r in (4.17) is replaced by $r' > s$, then the resulting space would be the same with equivalent norms. The condition $s < r$ should not be viewed as a restriction on s but rather as the need to accommodate the value of r. However, the situation changes when we allow to have $p < 1$ in that for fixed r the definition (4.17) is still meaningful for a larger range of s. In fact, now we may have $\omega_r(f,t)_p = O(t^{r-1+1/p})$ $(r - 1 + 1/p > r)$ for a nontrivial function f $(f \notin \Pi_r)$ so that the restriction $s < r$ in definition (4.17) of $B^s_q(L_p(\Omega))$ is no longer needed when $p < 1$ but should be replaced by $0 < s < r - 1 + 1/p$.

In this article, we shall only consider nonlinear approximation from piecewise linear functions. For this reason, we fix $r = 2$ and will consider the Besov spaces $B^s_p(L_p(\Omega))$ defined by (4.17), where $r = 2$, $0 < p < \infty$, and $0 < s < \max(2, 1 + 1/p)$. It is clear now that these are "classical" Besov spaces when $p \geq 1$ and "nonclassical" whenever $p < 1$ and $s \geq 2$.

It is well known that the need for measuring smoothness in L_p for $p < 1$ arises naturally in nonlinear approximation. Moreover, there is no essential difference whether the approximation itself takes place in L_p for $p \geq 1$ or $p < 1$. For this reason, we do not restrict our considerations in any respect to $p \geq 1$.

It is often convenient to use the following equivalence

$$|f|_{B^s_p(L_p(\Omega))} \simeq \left(\sum_{m\in\mathbb{Z}} 2^{msp}\omega_2(f, 2^{-m})_p^p\right)^{1/p} \tag{4.19}$$

which is immediate from (4.17) using the properties of $\omega_2(f,t)_p$.

Another equivalent semi-norm in $B^s_p(L_p(\Omega))$ can be deduced from (4.17) by using the equivalence from (4.6). We have

$$|f|_{B^s_p(L_p(\Omega))} \simeq |f|^A_{B^s_p(L_p(\Omega))} := \Big(\int_\Omega \int_0^\infty \int_{[0,t]^2} |\Delta^2_h(f,x,\Omega)|^p t^{-sp-3}\, dh\, dt\, dx\Big)^{1/p} \tag{4.20}$$

with constants of equivalence depending only on p. (Notice that $\Delta^2_h(f,x,\Omega) := 0$ if $[x, x+2h]$ is not entirely contained in Ω.)

4.4 Besov spaces via multiscale decompositions

It will be useful to work with several alternative characterizations of Besov spaces based on multiscale decompositions that are defined through a hierarchy of nested triangulations. To describe this, let P be an arbitrary admissible partition of Ω (that may or may not be equal to P_0, the initial partition of Ω). We inductively define a sequence $(P^{[m]})_{m=0}^\infty$ of uniform refinements of P. We set $P^{[0]} := P$. Suppose that $P^{[0]}$, $P^{[1]}$, $\ldots$, $P^{[m]}$ have already been defined. Then we construct $P^{[m+1]}$ by applying newest vertex bisection to each triangle in $P^{[m]}$ twice. Thus each $\Delta \in P^{[m]}$ is subdivided into four grandchildren. Hence each edge in $P^{[m]}$ is bisected so that $P^{[m+1]}$ is indeed admissible.

We denote by $\Theta_m := \Theta(P^{[m]})$ the set of all support cells of the Courant elements at level $P^{[m]}$ and set $\Theta := \bigcup_{m=0}^\infty \Theta_m$.

For a given $0 < p < \infty$, we select ρ such that $0 < \rho \le p$ and denote $Q_m := Q_{\rho, P^{[m]}}$, the quasi-interpolant from (4.15). (One can use the quasi-interpolant $Q^\diamond_{P^{[m]}}$ from (4.10) instead, if $p \ge 1$. The purpose of such a lower bound ρ is to employ the same projector for a whole range of p's just bounded from below by ρ.)

We set $q_m = q_{\rho,m} := Q_m - Q_{m-1}$, where $Q_{-1} := 0$. For a given function $f \in L_p(\Omega)$, we define $(b_\theta(f))_{\theta\in\Theta_m} = (b_{\rho,\theta}(f))_{\theta\in\Theta_m}$ from

$$q_m(f) =: \sum_{\theta\in\Theta_m} b_\theta(f)\phi_\theta. \tag{4.21}$$

By (4.16), it follows that $\|f - Q_m(f)\|_{L_p(\Omega)} \to 0$ as $m \to 0$, and hence

$$f = \sum_{m=0}^\infty q_m(f) = \sum_{m=0}^\infty \sum_{\theta\in\Theta_m} b_\theta(f)\phi_\theta \quad \text{in } L_p \quad (b_\theta(f) := b_{\rho,\theta}(f)). \tag{4.22}$$

Definition of norms in $B^s_p(L_p(\Omega))$ via multiscale Courant bases. Suppose $(P^{[m]})_{m=0}^\infty$ is a sequence of uniform refinements of P_0, the initial partition of Ω. We define, for $0 < p < \infty$,

$$\|f\|^Q_{B^s_p(L_p(\Omega))} := \Big(\sum_{\theta\in\Theta} |\theta|^{-\frac{sp}{2}} \|b_\theta(f)\phi_\theta\|_p^p\Big)^{1/p}, \tag{4.23}$$

where $\{b_\theta(f)\}$ are from (4.21)-(4.22).

Note that the coefficients $b_\theta(f)$ encode "difference information" since they represent the update needed when progressing to the next higher level of resolution. There is an alternative to the above decomposition which would remove the dependence on the quasi-interpolant operators. Namely, we can define another equivalent norm by

$$\|f\|^*_{B^s_p(L_p(\Omega))} := \inf_{f=\sum_{\theta\in\Theta} a_\theta\phi_\theta} \Big(\sum_{\theta\in\Theta} |\theta|^{-\frac{sp}{2}} \|a_\theta\phi_\theta\|_p^p\Big)^{1/p}, \tag{4.24}$$

where the infimum is taken over all representations of f: $f = \sum_{\theta\in\Theta} a_\theta\phi_\theta$ in L_p.

Theorem 4.2. *The Besov norms* $\|\cdot\|_{B^s_p(L_p(\Omega))}$ *from* (4.18), $\|\cdot\|^Q_{B^s_p(L_p(\Omega))}$ *from* (4.23), *and* $\|\cdot\|^*_{B^s_p(L_p(\Omega))}$ *from* (4.24), *are equivalent with constants of equivalence depending only on* p, s, β, *and* $|\Omega|$.

For the proof of this theorem, see [14, 15], and also the proof of Lemma 4.3 in Section 4.6 below.

4.5 Embedding results

There are various embedding theorems for Besov spaces. It is important to note that these embedding results involve constants which depend in an essential way on the shape of the domain which they refer to. To keep the constants under control we shall use such embedding theorems only on regions $G = \Delta$ or $G = \widehat{\Delta}$. In the present context the following variants are relevant.

Suppose $0 < p < \infty$, $0 < \alpha < 1 + 1/p$, $s > 0$, $1/\tau \le s/2 + 1/p$, $\alpha + s < \max(1 + 1/\tau, 2)$, and $\delta := s/2 + 1/p - 1/\tau$. Then

$$|f|_{B^\alpha_p(L_p(G))} \le c|G|^\delta |f|_{B^{\alpha+s}_\tau(L_\tau(G))}, \tag{4.25}$$

where c depends only on the above parameters and β.

The case $\alpha = 0$ is represented by the following estimate for local polynomial approximation. Let $0 < p < \infty$, $s > 0$, $1/\tau \le s/2 + 1/p$, and $\delta := s/2 + 1/p - 1/\tau$, then for $s \le \max(1 + 1/\tau, 2)$ one has

$$E_2(f, G)_p \le c|G|^\delta |f|_{B^s_\tau(L_\tau(G))} \tag{4.26}$$

with c depending only on the above parameters and β, see e.g. [9].

4.6 Two auxiliary results

For the proof of our direct estimates, we shall need the following localization of the error of good piecewise linear approximations with respect to Besov norms.

Lemma 4.3. *Let $f \in B^s_p(L_p(\Omega))$, where $0 < p < \infty$ and $0 < s < 1 + 1/p$. Suppose $0 < \rho \le p$ and $Q_P := Q_{\rho,P}$ is the quasi-interpolant from (4.15) (or $Q_P := Q^\diamond_P$ from (4.10), if $p \ge 1$). Then for any admissible partition P of Ω, we have*

$$|f - Q_P(f)|^p_{B^s_p(L_p(\Omega))} \le c \sum_{\Delta \in P} |f|^p_{B^s_p(L_p(\widehat{\Delta}))}, \tag{4.27}$$

where $c = c(p, \rho, s, \beta)$, and

$$\|f - Q_P(f)\|^p_{L_p(\Omega)} \le c \sum_{\Delta \in P} |f|^p_{B^s_p(L_p(\widehat{\Delta}))}, \tag{4.28}$$

where $c = c(p, \rho, s, \beta, d)$, see (2.3).

Proof. Suppose $(P^{[m]})_{m=0}^\infty$ is a sequence of uniform refinements of P as described in the beginning of §4.4. We also adhere to all other notation from §4.4. Thus $\Theta_m := \Theta(P^{[m]})$ will be the set of all support cells θ at level $P^{[m]}$ and $\Theta := \bigcup_{m=0}^\infty \Theta_m$. Further, we define $\Theta' := \bigcup_{m=1}^\infty \Theta_m$. We have

$$f - Q_P(f) = \sum_{m=1}^\infty q_m(f) = \sum_{m=1}^\infty \sum_{\theta \in \Theta_m} b_\theta(f)\phi_\theta \quad \text{in } L_p, \tag{4.29}$$

where $q_m := Q_m - Q_{m-1}$ and $(b_\theta(f))_{\theta \in \Theta_m}$ are defined as in (4.21)-(4.22).

To prove (4.27), we shall use the equivalence from (4.19). We define $\mathcal{X}_j := \{\theta \in \Theta : 2^{-j-1} < |\theta|^{1/2} \le 2^{-j}\}$ and $g_j := \sum_{\theta \in \mathcal{X}_j} b_\theta \phi_\theta$, where $b_\theta := b_\theta(f)$. Since there exists some fixed constant $c(\beta)$, depending only on the minimal angle, such that at most $c(\beta)$ cells $\theta \in \mathcal{X}_j$ may overlap, one has

$$\omega_2(g_j, t)^p_p \lesssim \|g_j\|^p_p \lesssim \sum_{\theta \in \mathcal{X}_j} \|b_\theta \phi_\theta\|^p_p, \quad t > 0, \tag{4.30}$$

see (4.9).

Using that $\|\phi_\theta\|_\infty = 1$, simple calculations show that

$$\omega_2(\phi_\theta, t)^p_p \simeq \begin{cases} |\theta|^{(1-p)/2} t^{1+p}, & \text{if } 0 < t < |\theta|^{1/2}, \\ |\theta|, & \text{if } t \ge |\theta|^{1/2}. \end{cases} \tag{4.31}$$

If $j < m$, then by (4.31) and using again that at most $c(\beta)$ cells $\theta \in \mathcal{X}_j$ may overlap, we obtain

$$\begin{aligned} \omega_2(g_j, 2^{-m})^p_p &\lesssim \sum_{\theta \in \mathcal{X}_j} \omega_2(b_\theta \phi_\theta, 2^{-m})^p_p \\ &\lesssim 2^{-m(1+p)} \cdot 2^{-j(1-p)} \sum_{\theta \in \mathcal{X}_j} |b_\theta|^p \\ &\simeq 2^{-(m-j)(1+p)} \sum_{\theta \in \mathcal{X}_j} \|b_\theta \phi_\theta\|^p_p, \end{aligned} \tag{4.32}$$

where we used that $\|\phi_\theta\|_p \simeq |\theta|^{1/p}$.

Now set $p^* := \min\{p, 1\}$. By (4.30) and (4.32), we deduce that

$$\omega_2(f - Q_P(f), 2^{-m})_p^{p^*} \le \sum_{j\in\mathbb{Z}} \omega_2(g_j, 2^{-m})_p^{p^*}$$

$$\lesssim \sum_{j=m+1}^{\infty} \Big(\sum_{\theta\in\mathcal{X}_j} \|b_\theta\phi_\theta\|_p^p\Big)^{p^*/p} + \sum_{j=-\infty}^{m} 2^{-(m-j)(1+p)p^*/p}\Big(\sum_{\theta\in\mathcal{X}_j} \|b_\theta\phi_\theta\|_p^p\Big)^{p^*/p}$$

$$\simeq \sum_{j=m+1}^{\infty} 2^{-jsp^*}\Big(\sum_{\theta\in\mathcal{X}_j} |\theta|^{-sp/2}\|b_\theta\phi_\theta\|_p^p\Big)^{p^*/p}$$

$$+ \sum_{j=-\infty}^{m} 2^{-(m-j)(1+p)p^*/p} 2^{-jsp^*}\Big(\sum_{\theta\in\mathcal{X}_j} |\theta|^{-sp/2}\|b_\theta\phi_\theta\|_p^p\Big)^{p^*/p}.$$

We put $A_\theta := |\theta|^{-sp/2}\|b_\theta\phi_\theta\|_p^p$ and substitute the above estimate in (4.19) to obtain

$$\begin{aligned}
|f - Q_P(f)|^p_{B^s_p(L_p(\Omega))} &\lesssim \sum_{m\in\mathbb{Z}} 2^{msp}\Big[\sum_{j=m+1}^{\infty} 2^{-jsp^*}\Big(\sum_{\theta\in\mathcal{X}_j} A_\theta\Big)^{p^*/p}\Big]^{p/p^*} \\
&+ \sum_{m\in\mathbb{Z}} 2^{msp}\Big[\sum_{j=-\infty}^{m} 2^{-(m-j)(1+p)p^*/p} 2^{-jsp^*}\Big(\sum_{\theta\in\mathcal{X}_j} A_\theta\Big)^{p^*/p}\Big]^{p/p^*} \\
&\lesssim \sum_{m\in\mathbb{Z}}\Big[\sum_{j=m+1}^{\infty} 2^{-(j-m)sp^*}\Big(\sum_{\theta\in\mathcal{X}_j} A_\theta\Big)^{p^*/p}\Big]^{p/p^*} \\
&+ \sum_{m\in\mathbb{Z}}\Big[\sum_{j=-\infty}^{m} 2^{-(m-j)[1+p-sp]p^*/p}\Big(\sum_{\theta\in\mathcal{X}_j} A_\theta\Big)^{p^*/p}\Big]^{p/p^*}.
\end{aligned}$$

We now use that $1+p-sp > 0$ because $s < 1+1/p$. Therefore, using the well known Hardy's inequalities (see, e.g., Lemma 3.4 in [6] and Lemma 3.10 in [16]) we obtain

$$|f - Q_P(f)|^p_{B^s_p(L_p(\Omega))} \lesssim \sum_{j\in\mathbb{Z}}\sum_{\theta\in\mathcal{X}_j} A_\theta \lesssim \sum_{\theta\in\Theta'} |\theta|^{-sp/2}\|b_\theta\phi_\theta\|_p^p. \tag{4.33}$$

In going further, suppose that $\theta \in \Theta_m$ ($m \ge 1$) and let $\Delta \in P^{[m]}$ be a triangle such that $\Delta \subset \theta$. We denote by Δ' the grandparent of Δ in $P^{[m-1]}$, i.e., the unique triangle $\Delta' \in P^{[m-1]}$ with the property $\Delta \subset \Delta'$. We now use (4.8), (4.16), and that $q_m := Q_m - Q_{m-1}$ to obtain

$$\begin{aligned}
\|b_\theta\phi_\theta\|_p \lesssim \|q_m(f)\|_{L_p(\Delta)} &\lesssim \|f - Q_m(f)\|_{L_p(\Delta)} + \|f - Q_{m-1}(f)\|_{L_p(\Delta')} \\
&\lesssim E_2(f, \widehat{\Delta})_p + E_2(f, \widehat{\Delta'})_p.
\end{aligned}$$

Inserting this estimate in (4.33) and taking into account that $|\theta| \simeq |\Delta| \simeq |\Delta'|$, we find

$$|f - Q_P(f)|^p_{B^s_p(L_p(\Omega))} \lesssim \sum_{\Delta\in\mathbb{P}} |\Delta|^{-sp/2} E_2(f, \widehat{\Delta})_p^p, \tag{4.34}$$

where $\mathbb{P} := \bigcup_{m=0}^{\infty} P^{[m]}$. Clearly, for each $\Delta \in \mathbb{P}$ there is at least one triangle $\Delta^\diamond \in P^{[0]} = P$ such that $\widehat{\Delta} \subset \widehat{\Delta^\diamond}$. Therefore, we can split the terms in the right sum of (4.34) and deduce

$$\begin{aligned}|f - Q_P(f)|^p_{B^s_p(L_p(\Omega))} &\lesssim \sum_{\Delta^\diamond \in P} \sum_{\Delta \in \mathbb{P}:\, \widehat{\Delta} \subset \widehat{\Delta^\diamond}} |\Delta|^{-sp/2} E_2(f, \widehat{\Delta})^p_p \\ &\lesssim \sum_{\Delta^\diamond \in P} \sum_{\Delta \in \mathbb{P}:\, \widehat{\Delta} \subset \widehat{\Delta^\diamond}} |\Delta|^{-sp/2} \omega_2(f, \widehat{\Delta})^p_p. \end{aligned} \tag{4.35}$$

Fix $\Delta^\diamond \in P$ and denote

$$\mathcal{Z}_j := \{\Delta \in \mathbb{P} : \widehat{\Delta} \subset \widehat{\Delta^\diamond},\ 2^{-j-1} < |\Delta| \le 2^{-j}\}.$$

We use (4.6) to obtain

$$\sum_{\Delta \in \mathcal{Z}_j} |\Delta|^{-sp/2} \omega_2(f, \widehat{\Delta})^p_p \lesssim 2^{sjp} \sum_{\Delta \in \mathcal{Z}_j} \mathrm{w}_2(f, 2^{-j}, \widehat{\Delta})^p_p \lesssim 2^{sjp} \omega_2(f, 2^{-j}, \widehat{\Delta^\diamond})^p_p,$$

where we used that only $\le c(\beta)$ triangles from $\mathcal{Z}_j$ may overlap at a time. This implies

$$\sum_{\Delta \in \mathbb{P}:\, \widehat{\Delta} \subset \widehat{\Delta^\diamond}} |\Delta|^{-sp/2} \omega_2(f, \widehat{\Delta})^p_p \lesssim \sum_{j \in \mathbb{Z}} 2^{sjp} \omega_2(f, 2^{-j}, \widehat{\Delta^\diamond})^p_p \lesssim |f|^p_{B^s_p(L_p(\widehat{\Delta^\diamond}))},$$

which combined with (4.35) and Lemma 4.4 completes the proof of (4.27).

The proof of estimate (4.28) is much easier and can be carried out along the following lines. In view of (4.16) and (4.26), we have

$$\|f - Q_P(f)\|^p_p \le c \sum_{\Delta \in P} E_2(f, \widehat{\Delta})^p_p \le c \sum_{\Delta \in P} |\widehat{\Delta}|^{ps/2} |f|^p_{B^s_p(L_p(\widehat{\Delta}))},$$

which confirms (4.28) with a constant c that depends now on the maximal diameter d from (2.3). □

Lemma 4.4. *Let $f \in B^s_p(L_p(\Omega))$, where $0 < p < \infty$. Then for any admissible partition P of Ω, we have*

$$\sum_{\Delta \in P} |f|^p_{B^s_p(L_p(\widehat{\Delta}))} \le c |f|^p_{B^s_p(L_p(\Omega))}, \tag{4.36}$$

where $c = c(p, s, \beta)$.

Proof: Estimate (4.36) follows immediately by using the Besov semi-norm $|\cdot|^A_{B^s_p(L_p(G))}$ from (4.20) taking into account that at most $c(\beta)$ rings $\widehat{\Delta}$ may overlap at a time when $\Delta \in P$. □

5 Direct theorems

In this section, we shall derive embeddings of Besov classes into the approximation classes $\mathcal{A}^s$ introduced earlier in §3. As we have throughout the paper, we fix a polygonal domain Ω and the space $B_0 = L_p(\Omega)$ or $B_0 = B_p^\alpha(L_p(\Omega))$ in which we are going to measure the approximation error. We also fix the subdivision method for generating adaptive partitions to be newest vertex bisection as discussed in §2.

We recall from §4 that a Besov space $B = B_\tau^{\alpha+s}(L_\tau(\Omega))$ is compactly embedded in B_0 if and only if

$$\delta := \frac{s}{2} + \frac{1}{p} - \frac{1}{\tau} > 0, \tag{5.1}$$

and δ is called the discrepancy for B relative to B_0.

Theorem 5.1. *Let $B_0 := B_p^\alpha(L_p(\Omega))$, $0 < p < \infty$, $0 \le \alpha < 1 + 1/p$ or $B_0 = L_p(\Omega)$ if $\alpha = 0$. If $f \in B := B_\tau^{s+\alpha}(L_\tau(\Omega))$ with $1/\tau < s/2 + 1/p$, $s > 0$, and $\alpha + s \le \max(1 + 1/\tau, 2)$, then*

$$\sigma_n(f)_{B_0} \le cn^{-s/2}|f|_B, \quad n \ge 1, \tag{5.2}$$

where $c = c(p, \alpha, s, \tau, \beta, d)|\Omega|^\delta$. Therefore, $f \in B$ implies that $f \in \mathcal{A}^{s/2}$.

Proof: We shall prove this theorem only for $\alpha > 0$, since the proof for $\alpha = 0$ is essentially the same yet simpler because the L_p norm localizes trivially.

The proof of (5.2) is based on the following observation.

Proposition 5.2. *For every $\varepsilon > 0$ there exists an admissible partition $P \in \mathcal{P}_n^a$, obtained from P_0 by n newest vertex bisections,*

such that

$$\|f - Q_P(f)\|_{B_0} \le c(n + n_0)^{1/p}\varepsilon \tag{5.3}$$

and

$$n \le c\Big(\varepsilon^{-1}|\Omega|^\delta|f|_B\Big)^{\tau/(1+\delta\tau)}, \tag{5.4}$$

where $Q_P := Q_{p,P}$ is the quasi-interpolant from (4.15) and $c = c(p, s, \tau, \delta, \beta, d)$.

To see that Theorem 5.1 follows from Proposition 5.2, for any fixed $n \ge n_0$, we define

$$\varepsilon := |\Omega|^\delta|f|_B n^{-(1+\delta\tau)/\tau},$$

provided $|f|_B > 0$ (the case $|f|_B = 0$ is trivial). Then by (5.3)-(5.4), we obtain

$$\sigma_n(f)_{B_0} \le \|f - Q_P(f)\|_{B_0} \le c(n + n_0)^{1/p}\varepsilon \le cn^{-s/2}|f|_B$$

where we used that $n \ge n_0$. Also, by (4.25) and (4.27), we have $\sigma_0(f)_{B_0} \lesssim \|f - Q_{P_0}(f)\|_{B_0} \lesssim |f|_B$ and these two estimates imply (5.2).

To prove Proposition 5.2 we shall use a local error indicator which is defined for any Δ in an admissible partition P by

$$e(\Delta, P) := |\Delta|^\delta |f|_{B(\widehat{\Delta})}, \tag{5.5}$$

where $|f|_{B(\widehat{\Delta})}$ is the B-norm of f on $\widehat{\Delta}$ and, as before, $\widehat{\Delta} = \widehat{\Delta}(P)$ is the ring of Δ in the partition P. Recall that by the remarks in §2, we still have that $|\widehat{\Delta}|/|\Delta|$ remains bounded by a constant depending only on the minimal angle β in P_0.

To construct the desired partition we fix any target accuracy $\varepsilon > 0$ and adaptively generate a tree $T = T_\varepsilon$ as follows. We start with the root nodes in T_0. We let $\mathcal{M}_0$ be the set of those nodes Δ for which $e(\Delta, P_0) > \varepsilon$. We let T_1' denote the tree obtained from T_0 by subdividing the nodes in $\mathcal{M}_0$ and leaving all other nodes in T_0 untouched and let T_1 denote the completion of T_1' (see §2). Now, we repeat this adaptive subdivision process on T_1. In other words we subdivide the nodes $\Delta \in \mathcal{M}_1$, where $\mathcal{M}_1$ is the set of all leaves Δ of T_1 for which $e(\Delta, P(T_1)) > \varepsilon$. We continue in this way and terminate the process when $\mathcal{M}_k = \emptyset$. The process terminates because $|f|_{B(\widehat{\Delta})} \le |f|_B$. We set $T' = T_k$, form the completion T of T', and set $P := P(T)$.

Let us first check that (5.3) holds. When the algorithm terminates, for each cell Δ in T satisfies $e(\Delta) = e(\Delta, T) \le \epsilon$. Thus we can invoke (4.25) and (5.5) to conclude that

$$|f|_{B_0(\widehat{\Delta})} \lesssim |\Delta|^\delta |f|_{B(\widehat{\Delta})} \lesssim \varepsilon. \tag{5.6}$$

Using this and Lemma 4.3, we obtain

$$\|f - Q_P(f)\|_{B_0}^p \lesssim \sum_{\Delta \in P} |f|_{B_0(\widehat{\Delta}))}^p \lesssim \sum_{\Delta \in P} |\Delta|^{p\delta} |f|_{B(\widehat{\Delta})}^p \lesssim (\#P)\varepsilon^p, \tag{5.7}$$

where the involved constants have the asserted dependence on the parameters p, s, β, d. This establishes (5.3).

We want to count next the number of cells that were marked during the subdivision process. Let $\mathcal{M} := \mathcal{M}_0 \cup \cdots \cup \mathcal{M}_k$ and for each $j \in \mathbb{Z}$, let Λ_j be the collection of all $\Delta \in \mathcal{M}$ which satisfy

$$2^{-j-1} \le |\Delta| < 2^{-j}. \tag{5.8}$$

We want to estimate the cardinality m_j of Λ_j. First we have the trivial estimate

$$m_j \le 2^{j+1}|\Omega|. \tag{5.9}$$

This follows from the fact that the cells in Λ_j are disjoint: if two cells of T_* intersect then one is contained in the other and the smaller cell has measure at most half of the larger.

To obtain our second estimate for m_j, we start with the fact that for each $\Delta \in \Lambda_j$, we have

$$\varepsilon < |\Delta|^\delta |f|_{B(\widehat{\Delta})} \le 2^{-j\delta} |f|_{B(\widehat{\Delta})} \tag{5.10}$$

because this cell was subdivided. Recall that $|\widehat{\Delta}| \leq c|\Delta|$. Hence, using again that the $\Delta \in \Lambda_j$ are pairwise disjoint, we have that a point $x \in \Omega$ appears in at most c of the cells $\widehat{\Delta}$ with $\Delta \in \Lambda_j$. Thus, returning to (5.10), we deduce

$$m_j \varepsilon^\tau \leq 2^{-j\delta\tau} \sum_{\Delta \in \Lambda_j} |f|^\tau_{B(\widehat{\Delta})} \leq c 2^{-j\delta\tau} |f|^\tau_B, \tag{5.11}$$

where the last inequality uses Lemma 4.4.

Let j_0 be the smallest integer such that $2^{j_0} > |\Omega|$. Then

$$\#(\mathcal{M}) \leq \sum_{j=-j_0}^{\infty} m_j \leq c \sum_{j=-j_0}^{\infty} \min(2^j|\Omega|, \varepsilon^{-\tau} 2^{-j\delta\tau} |f|^\tau_B) \leq c(\varepsilon^{-1} |\Omega|^\delta |f|_B)^{\tau/(1+\delta\tau)}. \tag{5.12}$$

Here, the last inequality is established by breaking the sum into two sums determined by where the minimum turns from $2^j|\Omega|$ to $\varepsilon^{-\tau} 2^{-j\delta\tau} |f|^\tau_B$.

From (2.1), we obtain that the number n of subdivisions needed to produce $P = P(T)$ satisfies

$$n \leq c(\varepsilon^{-1} |\Omega|^\delta |f|_B)^{\tau/(1+\delta\tau)} \tag{5.13}$$

which is (5.4). □

6 Inverse Theorems

In this section, we want to prove inverse estimates to the results of §5. We begin by proving the following Bernstein inequality for n-term adaptive approximation.

Theorem 6.1. (a) *Let $0 < p < \infty$, $0 < \alpha < 1 + 1/p$, $s > 0$, $1/\tau = s/2 + 1/p$, and $\alpha + s < 1 + 1/\tau$. If $S \in \mathcal{S}(P)$ for some partition $P \in \mathcal{P}^a$ with $\#(P) \leq n$, then*

$$\|S\|_{B^{s+\alpha}_\tau(L_\tau(\Omega))} \leq c n^{s/2} \|S\|_{B^\alpha_p(L_p(\Omega))}, \tag{6.1}$$

where c depends only on p, α, s, β, and $|\Omega|$.

(b) *Let $0 < p < \infty$, $s > 0$, $1/\tau = s/2 + 1/p$, and $s < 1 + 1/\tau$. If $S \in \mathcal{S}(P)$ for some partition $P \in \mathcal{P}^a$ with $\#(P) \leq n$, then*

$$|S|_{B^s_\tau(L_\tau(\Omega))} \leq c n^{s/2} \|S\|_{L_p(\Omega)}, \tag{6.2}$$

where c depends only on p, s, and β.

Moreover, the above two estimates hold under the same conditions if S is replaced by $S_1 - S_2$, where $S_1, S_2 \in \mathcal{S}(P)$ with $\#(P) \leq n$ as above.

Proof. We shall prove only part (a) of the theorem. The proof of part (b) is easier and can be found in [3]. We shall use the multiscale representation (4.22) starting from the *initial* partition P_0 for $f = S$, i.e.,

$$S = \sum_{m \geq 0} \sum_{\theta \in \Theta_m} b_\theta(S) \phi_\theta = \sum_{\theta \in \mathcal{M}} b_\theta(S) \phi_\theta, \tag{6.3}$$

where $b_\theta(S) = b_{\rho,\theta}(S)$ with $\rho := \tau < p$ and $\mathcal{M} \subset \Theta$ is the set of all nonzero coefficients. With this selection of ρ, we can use representation (6.3) to describe both $\|S\|_{B_p^\alpha(L_p(\Omega))}$ and $\|S\|_{B_\tau^{\alpha+s}(L_\tau(\Omega))}$ by using the norm from (4.23).

It is important to know how many coefficients b_θ in (6.3) are different from zero. To count them we shall consider the tree $T = T(P)$ that corresponds to the triangulation P on which S is defined. Let $N(P)$ be the number of subdivisions used to produce P starting from P_0. The definition (4.21) gives that $b_\theta \neq 0$ is possible for $\theta \in \Theta_k = \Theta(P_0^{[k]})$ only if $(Q_{k-1}(S))(v) \neq (Q_k(S))(v)$, where v is the central vertex for θ. In the case $v \in \mathcal{V}_{P_0^{[k-1]}}$, this can happen only if the element $\theta' \in \Theta_{k-1}$, centered at v, includes a triangle Δ that is subdivided in P, i.e. Δ corresponds to an internal node of the tree T. The case $v \notin \mathcal{V}_{P_0^{[k-1]}}$ means that v is a midpoint of an edge connecting two vertices $v', v'' \in \mathcal{V}_{P_0^{[k-1]}}$ and one of them has to be a vertex of a triangle from $P_0^{[k-1]}$ that corresponds to an internal node of the tree T. Adding to this count all the elements from Θ_0 whose number does not exceed $3\#(P_0)$, we receive

$$\#(\mathcal{M}) \le 3N(P) + 3m_0 N(P) + 3\#(P_0) \lesssim n\,, \tag{6.4}$$

where m_0 is the maximal valence of the vertices in the triangulations from $\mathcal{P}$.

We now use the norm from (4.23) (see Theorem 4.2) and Hölder's inequality to obtain

$$\begin{aligned}
\|S\|_{B_\tau^{\alpha+s}(L_\tau(\Omega))} &\simeq \Big(\sum_{\theta\in\mathcal{M}} |\theta|^{-(\alpha+s)\tau/2} \|b_\theta(S)\phi_\theta\|_\tau^\tau\Big)^{1/\tau} \\
&\simeq \Big(\sum_{\theta\in\mathcal{M}} (|\theta|^{-\alpha/2} \|b_\theta(S)\phi_\theta\|_p)^\tau\Big)^{1/\tau} \\
&\lesssim (\#\mathcal{M})^{(1-\tau/p)/\tau} \Big(\sum_{\theta\in\mathcal{M}} (|\theta|^{-\alpha/2} \|b_\theta(S)\phi_\theta\|_p)^p\Big)^{1/p} \\
&\lesssim n^{s/2} \|S\|_{B_p^\alpha(L_p(\Omega))},
\end{aligned}$$

where we have used that $\|\phi_\theta\|_\tau \simeq |\theta|^{1/\tau-1/p}\|\phi_\theta\|_p$ and (6.4). $\square$

It is well known how to derive inverse theorems once Bernstein inequalities have been established. In our case, the setting is the same as the inverse theorems for free knot spline approximation given in Chapter 12 of [6]. For this reason we shall be brief in our exposition.

In what follows, we assume that we have one of the following two settings:

(a) $B_0 := B_p^\alpha(L_p(\Omega))$, $0 < p < \infty$, $0 < \alpha < 1 + 1/p$, and $B := B_\tau^{s+\alpha}(L_\tau(\Omega))$, where $s > 0$, $1/\tau = s/2 + 1/p$, and $\alpha + s < 1 + 1/\tau$.

(b) $B_0 := L_p(\Omega)$ with $0 < p < \infty$ and $B := B_\tau^s(L_\tau(\Omega))$ with $s > 0$, $1/\tau = s/2 + 1/p$, and $s < 1 + 1/\tau$.

The K-functional for the pair (B_0, B) is defined by

$$K(f, t; B_0, B) := \inf_{g \in B} \|f - g\|_{B_0} + t\|g\|_B. \tag{6.5}$$

The following theorem relates $K(f, t; B_0, B)$ and $\sigma_n(f)_{B_0}$.

Theorem 6.2. *We have, for $f \in B_0$,*

$$K(f, 2^{-sn/2}; B_0, B) \leq c2^{-sn/2}\Big[\Big(\sum_{m=0}^{n}(2^{sm/2}\sigma_{2^m}(f)_{B_0})^{\tau^*}\Big)^{1/\tau^*} + \|f\|_{B_0}\Big], \tag{6.6}$$

where $\tau^* := \min\{\tau, 1\}$ and c depends on the corresponding parameters.

Proof. The proof of this theorem is standard: one writes f as a telescoping sum of best approximations to f which realize $\sigma_{2^k}(f)$ and then employ the Bernstein estimates from Theorem 6.1 on each of the terms. □

The following inverse theorem is our main result in this section.

Theorem 6.3. *Let $0 < p < \infty$, $0 < \alpha < 1 + 1/p$, $s > 0$, $1/\tau := s/2 + 1/p$, and $\alpha + s < 1 + 1/\tau$. If $f \in \mathcal{A}_\tau^{s/2}(B_0)$, then $f \in B_\tau^{s+\alpha}(L_\tau(\Omega))$.*

Proof. We choose the pair (s_1, τ_1) with $s_1 > s$, $\tau_1 < \tau$, $1/\tau_1 = s_1/2 + 1/p$, and such that $0 < s_1 + \alpha < 1 + 1/\tau$ with the additional requirement that $\tau_1 \geq 1$ if $\tau > 1$.

It is well known (see [1]) that $B := B_\tau^{s+\alpha}(L_\tau(\Omega))$ is equal (with equivalent norms) to the interpolation space $[B_0, B_1]_{\frac{s}{s_1},\tau}$ between $B_0 = B_\tau^\alpha(L_p(\Omega))$ and $B_1 := B_{\tau_1}^{s_1+\alpha}(L_{\tau_1(\Omega)})$. Its norm satisfies

$$\|f\|_{[B_0,B_1]_{s/s_1,\tau}} \simeq \left(\sum_{k=0}^{\infty}[2^{ks/2}K(f, 2^{-ks_1/2}; B_0, B_1)]^\tau\right)^{1/\tau}. \tag{6.7}$$

Next, we use Theorem 6.2 in the form

$$K(f, 2^{-ks_1/2}; B_0, B_1) \lesssim 2^{-ks_1/2}\left(\sum_{m=0}^{k}\left(2^{ms_1/2}\sigma_{2^m}(f)\right)^{\tau_1^*}\right)^{1/\tau_1^*} + 2^{-ks_1/2}\|f\|_{B_0}\ , \tag{6.8}$$

where as usual $\tau_1^* := \min\{\tau_1, 1\}$. Since $\|f\|_B \lesssim \|f\|_{(B_0,B_1)_{s/s_1,\tau}}$, we have

$$\|f\|_B^\tau \lesssim \sum_{k=0}^{\infty} 2^{k\tau(s-s_1)/2}\left(\sum_{m=0}^{k}\left(2^{ms_1/2}\sigma_{2^m}(f)\right)^{\tau_1^*}\right)^{\tau/\tau_1^*} + \sum_{k=0}^{\infty}\left(2^{k(s-s_1)/2}\|f\|_{B_0}\right)^\tau\ . \tag{6.9}$$

An application of Hardy's inequality gives that the first sum is dominated by constant times the semi-norm $|f|^\tau_{\mathcal{A}_\tau^{s/2}}$ given by (3.3). The second sum is bounded by $c\|f\|^\tau_{B_0}$ and therefore $\|f\|_B \lesssim \|f\|_{\mathcal{A}_\tau^{s/2}}$. This proves the embedding $\mathcal{A}_\tau^{s/2} \subset B$ and the theorem. □

References

[1] J. Bergh and J. Löfström, *Interpolation Spaces, An Introduction*, Springer, 1976.

[2] P. Binev, W. Dahmen, R. DeVore, *Adaptive finite element methods with convergence rates*, IGPM Report # 219, RWTH Aachen, June 2002.

[3] W. Dahmen, I. Graham, B. Faehrmann, W. Hackbusch, S. Sauter, *Inverse Inequalities on Non-Quasiuniform Meshes and Application to the Mortar Element Method*, IGPM Report # 201, RWTH Aachen, April 2001.

[4] O. Davydov, P. Petrushev, *Nonlinear approximation from differentiable piecewise polynomials*, IMI Preprint Series 2002:12, University of South Carolina, 2002.

[5] R. DeVore, B. Jawerth, and V. Popov, *Compression of wavelet decompositions*, Amer. J. Math., **114** (1992), 737–785.

[6] R. DeVore, G.G. Lorentz, *Constructive Approximation*, Springer, New York, 1993.

[7] R. DeVore, V. Popov, *Interpolation of Besov spaces*, Trans. Amer. Math. Soc., **305** (1998), 297–314.

[8] R. DeVore, and R. Sharpley, *Besov spaces on domains in* $\mathbb{R}^d$, TAMS, **335**(1993), 843–864.

[9] R. DeVore, R. Sharpley, *Maximal functions measuring smoothness*, Memoirs of the American Mathematical Society, 47 (No 293), 1984.

[10] W. Dörfler, *A convergent adaptive algorithm for Poisson's equation*, SIAM J. Numer. Anal., **33**(1996), pp. 1106-1124.

[11] B. Karaivanov, P. Petrushev, *Nonlinear piecewise polynomial approximation beyond Besov spaces*, IMI Preprint Series 2001:13, University of South Carolina, 2001.

[12] W.F. Mitchell, *A comparison of adaptive refinement techniques for elliptic problems*, ACM Transaction on Math. Software, **15**(1989), 326–347.

[13] P. Morin, R. Nochetto, K. Siebert, *Data Oscillation and Convergence of Adaptive FEM*, SIAM J. Numer. Anal., **38**(2000), pp. 466-488.

[14] P. Oswald, *On the degree of nonlinear spline approximation in Besov-Sobolev spaces*, J. Approx. Theory, **61**(1990), 131–157.

[15] P. Oswald, *Multilevel Finite Element Approximation*, Teubner, Stuttgart, 1994.

[16] P. Petrushev, V. Popov, *Rational approximation of real functions*, Cambridge University Press, 1987.

Peter Binev
Department of Mathematics
University of South Carolina
Columbia, SC 29208
U.S.A.
e–mail: `binev@math.sc.edu`

Wolfgang Dahmen
Institut für Geometrie und Praktische Mathematik
RWTH Aachen
Templergraben 55
52056 Aachen
Germany
e–mail: `dahmen@igpm.rwth-aachen.de`

Ronald DeVore
Department of Mathematics
University of South Carolina
Columbia, SC 29208
U.S.A.
e–mail: `devore@math.sc.edu`

Pencho Petrushev
Department of Mathematics
University of South Carolina
Columbia, SC 29208
U.S.A.
e–mail: `pencho@math.sc.edu`