Convergence rates of adaptive methods, Besov spaces, and multilevel approximation

Tsogtgerel Gantumur
McGill University

Abstract

This paper concerns characterizations of approximation classes associated to adaptive finite element methods with isotropic $h$-refinements. It is known from the seminal work of Binev, Dahmen, DeVore and Petrushev that such classes hold related to Besov spaces. The range of parameters for which the inverse embedding results hold is rather limited, and recently, Gaspoz and Morin have shown, among other things, that this limitation disappears if we replace Besov spaces by suitable approximation spaces associated to finite element approximation from uniformly refined triangulations. We call the latter spaces multilevel approximation spaces, and argue that these spaces are placed naturally halfway between adaptive approximation classes and Besov spaces, in the sense that it is more natural to relate multilevel approximation spaces with either Besov spaces or adaptive approximation classes, than to go directly from adaptive approximation classes to Besov spaces. In particular, we prove embeddings of multilevel approximation spaces into adaptive approximation classes, complementing the inverse embedding theorems of Gaspoz and Morin.

Furthermore, in the present paper, we initiate a theoretical study of adaptive approximation classes that are defined using a modified notion of error, the so-called total error, which is the energy error plus an oscillation term. Such approximation classes have recently been shown to arise naturally in the analysis of adaptive algorithms. We first develop a sufficiently general approximation theory framework to handle such modifications, and then apply the abstract theory to second order elliptic problems discretized by Lagrange finite elements, resulting in characterizations of modified approximation classes in terms of memberships of the problem solution and data into certain approximation spaces, which are in turn related to Besov spaces. Finally, it should be noted that throughout the paper we paid equal attention to both the newest vertex bisection and the red refinement procedures.

Contents

1 Introduction .......................................................... 2
2 General theorems ..................................................... 6
2.1 The setup ......................................................... 6
2.2 Direct embeddings for standard approximation classes .... 9
2.3 Direct embeddings for general approximation classes ...... 12
1 Introduction

Among the most important achievements in theoretical numerical analysis during the last
decade was the development of mathematical techniques for analyzing the performance of
adaptive finite element methods. A crucial notion in this theory is that of approximation
classes, which we discuss here in a simple but very paradigmatic setting. Given a polygonal
domain \( \Omega \subseteq \mathbb{R}^2 \), a conforming triangulation \( P_0 \) of \( \Omega \), and a number \( s > 0 \), we say that a
function \( u \) on \( \Omega \) belongs to the approximation class \( \mathcal{A}_s \) if for each \( N \), there is a conforming
triangulation \( P \) of \( \Omega \) with at most \( N \) triangles, such that \( P \) is obtained by a sequence of
newest vertex bisections from \( P_0 \), and that \( u \) can be approximated by a continuous piecewise
affine function subordinate to \( P \) with the error bounded by \( cN^{-s} \), where \( c = c(u,P_0,s) \geq 0 \)
is a constant independent of \( N \). In a typical setting, the error is measured in the \( H^1 \)-norm,
which is the natural energy norm for second order elliptic problems. To reiterate and to
remove any ambiguities, we say that \( u \in H^1(\Omega) \) belongs to \( \mathcal{A}_s \) if
\[
\min_{\{P \in \mathcal{P} : \# P \leq N\}} \inf_{v \in S_P} \| u - v \|_{H^1} \leq cN^{-s},
\]
for all \( N \geq \# P_0 \) and for some constant \( c \), where \( \mathcal{P} \) is the set of conforming triangulations
of \( \Omega \) that are obtained by a sequence of newest vertex bisections from \( P_0 \), and \( S_P \) is the
space of continuous piecewise affine functions subordinate to the triangulation \( P \).

Approximation classes can be used to reveal a theoretical barrier on any procedure that
is designed to approximate \( u \) by means of piecewise polynomials and a fixed refinement rule
such as the newest vertex bisection. Suppose that we start with the initial triangulation
\( P_0 \), and generate a sequence of conforming triangulations by using newest vertex bisections.
Suppose also that we are trying to capture the function \( u \) by using continuous piecewise
linear functions subordinate to the generated triangulations. Finally, assume that \( u \in \mathcal{A}_s \)
but \( u \notin \mathcal{A}_\sigma \) for any \( \sigma > s \). Then as far as the exponent \( \sigma \) in \( cN^{-\sigma} \) is concerned, it is
obvious that the best asymptotic bound on the error we can hope for is \( cN^{-s} \), where \( N \) is
the number of triangles. Now supposing that \( u \) is given as the solution of a boundary value
problem, a natural question is if this convergence rate can be achieved by any practical algorithm, and it was answered in the seminal works of Binev, Dahmen, and DeVore (2004) and Stevenson (2007): These papers established that the convergence rates of certain adaptive finite element methods are optimal, in the sense that if \( u \in \mathcal{A}^s \) for some \( s > 0 \), then the method converges with the rate not slower than \( s \). One must mention the earlier developments Dörfler (1996); Morin, Nochetto, and Siebert (2000); Cohen, Dahmen, and DeVore (2001); Gantumur, Harbrecht, and Stevenson (2007), which paved the way for the final achievement.

Having established that the smallest approximation class \( \mathcal{A}^s \) in which the solution \( u \) belongs to essentially determines how fast adaptive finite element methods converge, the next issue is to determine how large these classes are and if the solution to a typical boundary value problem would belong to an \( \mathcal{A}^s \) with large \( s \). In particular, one wants to compare the performance of adaptive methods with that of non-adaptive ones. A first step towards addressing this issue is to characterize the approximation classes in terms of classical smoothness spaces, and the main work in this direction so far appeared is Binev, Dahmen, DeVore, and Petrushev (2002), which, upon tailoring to our situation and a slight simplification, tells that \( B_{p,p}^\alpha \subset \mathcal{A}^s \subset B_{p,p}^\sigma \) for \( 2 \leq \sigma < 1 + \frac{1}{p} \) and \( \sigma < \alpha < \max\{2,1 + \frac{1}{p}\} \) with \( \sigma = \frac{\alpha - 1}{2} \). Here \( B_{p,q}^\sigma \) are the standard Besov spaces defined on \( \Omega \). This result has recently been generalized to higher order Lagrange finite elements by Gaspoz and Morin (2013). In particular, they show that the direct embedding \( B_{p,p}^\alpha \subset \mathcal{A}^s \) holds in the larger range \( \sigma < \alpha < m + \max\{1, \frac{1}{p}\} \), where \( m \) is the polynomial degree of the finite element space, see Figure 1(a). However, the restriction \( \sigma < 1 + \frac{1}{p} \) on the inverse embedding \( \mathcal{A}^s \subset B_{p,p}^\sigma \) cannot be removed, since for instance, any finite element function whose derivative is discontinuous cannot be in \( B_{p,p}^\sigma \) if \( \sigma \geq 1 + \frac{1}{p} \) and \( p < \infty \). To get around this problem, Gaspoz and Morin proposed to replace the Besov space \( B_{p,p}^\sigma \) by the approximation space \( A_{p,p}^\sigma \) associated to uniform refinements\(^1\). We call the spaces \( A_{p,p}^\sigma \) multilevel approximation spaces, and their definition will be given in Subsection 3.3. For the purposes of this introduction, and roughly speaking, the space \( A_{p,p}^\sigma \) is the collection of functions \( u \in L_p \) for which

\[
\inf_{v \in S_{P_k}} \|u - v\|_{L_p} \leq ch_k^\sigma, \tag{2}
\]

where \( \{P_k\} \subset \mathcal{P} \) is a sequence of triangulations such that \( P_{k+1} \) is the uniform refinement of \( P_k \), and \( h_k \) is the diameter of a typical triangle in \( P_k \). Note for instance that finite element functions are in every \( A_{p,p}^\sigma \). With the multilevel approximation spaces at hand, the inverse embedding \( \mathcal{A}^s \subset A_{p,p}^\sigma \) is recovered for all \( \sigma = \frac{2}{p} \).

In this paper, we prove the direct embedding \( A_{p,p}^\alpha \subset \mathcal{A}^s \), so that the existing situation \( B_{p,p}^\alpha \subset \mathcal{A}^s \subset A_{p,p}^\alpha \) is improved to \( A_{p,p}^\alpha \subset A_{p,p}^s \subset A_{p,p}^\alpha \). It is a genuine improvement, since \( A_{p,p}^\alpha(\Omega) \supseteq B_{p,p}^\alpha(\Omega) \) for \( \alpha \geq 1 + \frac{1}{p} \). Moreover, as one stays entirely within an approximation

\(^1\) This space is denoted by \( \hat{B}_{p,p}^\sigma \) in Gaspoz and Morin (2013). In the present paper we are adopting the notation of Oswald (1994).
theory framework, one can argue that the link between $A^s$ and $A^\alpha_{p,p}$ is more natural than the link between $A^s$ and $B^\alpha_{p,p}$. Once the link between $A^s$ and $A^\alpha_{p,p}$ has been established, one can then invoke the well known relationships between $A^\alpha_{p,p}$ and $B^\alpha_{p,p}$. It seems that this two step process offers more insight into the underlying phenomenon. We also remark that while the existing results are only for the newest vertex bisection, we take into account the red refinement procedure as well.

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{embedding_diagram.png}
\caption{Illustration of various embeddings. The point \((\frac{1}{p}, \alpha)\) represents the space \(B^\alpha_{p,p}\).}
\end{figure}

Fig. 1: Illustration of various embeddings. The point \((\frac{1}{p}, \alpha)\) represents the space \(B^\alpha_{p,p}\).

The approximation classes $\mathcal{A}^s$ defined by (1) are associated to measuring the error of an approximation in the $H^1$-norm. Of course, this can be generalized to other function space norms, such as $L_p$ and $B^\alpha_{p,p}$, which we will consider in Section 3. However, we will not stop there, and consider more general approximation classes corresponding to ways of measuring the error between a general function $u$ and a discrete function $v \in S_P$ by a quantity $\rho(u, v, P)$ that may depend on the triangulation $P$ and is required to make sense merely for discrete functions $v \in S_P$. An example of such an error measure is

\[
\rho(u, v, P) = \left( \|u - v\|^2_{H^1} + \sum_{\tau \in P} (\text{diam } \tau)^2 \|f - \Pi_T f\|^2_{L^2(\tau)} \right)^{\frac{1}{2}},
\]

where $f = \Delta u$, and $\Pi_T : L_2(\tau) \rightarrow \mathbb{P}_d$ is the $L_2(\tau)$-orthogonal projection onto $\mathbb{P}_d$, the space of polynomials of degree not exceeding $d$. It has been shown in Cascon, Kreuzer, Nochetto,
and Siebert (2008) that if the solution \( u \) of the boundary value problem

\[
\Delta u = f \quad \text{in } \Omega \quad \text{and} \quad u|_{\partial \Omega} = 0,
\]

satisfies

\[
\min_{\{P \in \mathcal{A} : \#P \leq N\}} \inf_{v \in S_P} \rho(u, v, P) \leq c N^{-s},
\]

for all \( N \geq \#P_0 \) and for some constants \( c \) and \( s > 0 \), then a typical adaptive finite element method for solving (4) converges with the rate not slower than \( s \). Moreover, there are good reasons to consider that the approximation classes \( \mathcal{A}^s \) defined by the condition (5) are more attuned to certain practical adaptive finite element methods than the standard approximation classes \( \mathcal{A}^s \) defined by (1), see Section 4. Obviously, we have \( \mathcal{A}^s \subset \mathcal{A}^s \) but we cannot expect the inclusion \( \mathcal{A}^s \subset \mathcal{A}^s \) to hold in general. In Cascon et al. (2008), an effective characterization of \( \mathcal{A}^s \) was announced as an important pending issue.

In the present paper, we establish a characterization of \( \mathcal{A}^s \) in terms of memberships of \( u \) and \( f = \Delta u \) into suitable approximation spaces, which in turn are related to Besov spaces. For instance, we show that if \( u \in \mathcal{A}^s \) and \( f \in B_{p,p}^{\alpha} \) with \( \frac{\alpha}{2} \geq \frac{1}{p} - \frac{1}{2} \) and \( s = \frac{\alpha+1}{2} \), then \( u \in \mathcal{A}^s \), see Figure 1(b). Note that the approximation rate \( s = \frac{\alpha+1}{2} \) is as if we were approximating \( f \) in the \( H^{-1} \)-norm, which is illustrated by the arrow downwards. However, the parameters must satisfy \( \frac{\alpha}{2} \geq \frac{1}{p} - \frac{1}{2} \) (above or on the solid line), which is more restrictive compared to \( \frac{\alpha+1}{2} > \frac{1}{p} - \frac{1}{2} \) (above the dashed line), the latter being the condition we would expect if the approximation was indeed taking place in \( H^{-1} \). This situation cannot be improved in the sense that if \( \frac{\alpha}{2} < \frac{1}{p} - \frac{1}{2} \) then \( B_{p,p}^{\alpha} \subset L_2 \), hence the quantity (3) would be infinite in general for \( f \in B_{p,p}^{\alpha} \). In Section 4, we treat a more general class of variable coefficient second order equations. As a rule, the actual results therein are in terms of the approximation spaces that are studied in Section 3, which means that Besov spaces are rarely mentioned, and one has to appeal to Section 3 in order to deduce statements such as the one we have just described.

The results of Section 4 and some of the results of Section 3 are proved by invoking abstract theorems that are established in Section 2. These theorems extend some of the standard results from approximation theory to deal with generalized approximation classes such as \( \mathcal{A}^s \). We decided to consider a fairly general setting in the hope that the theorems will be used for establishing characterizations of other approximation classes. For example, adaptive boundary element methods and adaptive approximation in finite element exterior calculus seem to be amenable to our abstract framework, although checking the details poses some technical challenges.

This paper is organized as follows. In Section 2, we introduce an abstract framework that is more general than usually considered in approximation theory of finite element methods, and collect some theorems that can be used to prove embedding theorems between adaptive approximation classes and other function spaces. In Section 3, we recall some standard results on multilevel approximation spaces and their relationships with Besov
spaces, and then prove direct embedding theorems between multilevel approximation spaces and adaptive approximation classes. The main results of this section are Theorem 3.7, Theorem 3.8, Theorem 3.11, and Theorem 3.12. In Section 4, we investigate approximation classes associated to certain adaptive finite element methods for second order boundary value problems. Finally, in the appendix, we establish some technical estimates on the complexity of completions of triangulations obtained by the red refinement rule.

2 General theorems

2.1 The setup

Let $M$ be an $n$-dimensional topological manifold, equipped with a compatible measure, in the sense that all Borel sets are measurable. What we have in mind here is $M = \mathbb{R}^n$ with the Lebesgue measure on it, or a piecewise smooth surface $M \subset \mathbb{R}^N$ with its canonical Hausdorff measure. With $\Omega \subset M$ a bounded domain, we consider a class of partitions (triangulations) of $\Omega$, and finite element type functions defined over those partitions. Ultimately, we are interested in characterizing those functions on $\Omega$ that can be well approximated by such finite element type functions. In order to make these concepts precise, we will use in this section a fairly abstract setting, which we believe to be a good compromise between generality and readability.

By a partition of $\Omega$ we understand a collection $P$ of finitely many disjoint open subsets of $\Omega$, satisfying $\overline{\Omega} = \bigcup_{\tau \in P} \tau$. We assume that a set $\mathcal{P}$ of partitions of $\Omega$ is given, which we call the set of admissible partitions. For simplicity, we will assume that for any $k \in \mathbb{N}$ the set $\{P \in \mathcal{P} : \#P \leq k\}$ is finite. In practice, $\mathcal{P}$ would be, for instance, the set of all conforming triangulations obtained from a fixed initial triangulation $P_0$ by repeated applications of the newest vertex bisection procedure. Another important example arises from the red refinement rule. In this case, the admissibility criterion on a partition would be either that the number of hanging nodes per edge is bounded by a prescribed finite number, or that the diameter ratio between neighbouring elements stays bounded by a prescribed constant. Here and in the following, we often write triangles and edges et cetera to mean $n$-simplices and $(n-1)$-dimensional faces et cetera, which seems to improve readability.

We will assume the existence of a refinement procedure satisfying certain requirements. Given a partition $P \in \mathcal{P}$ and a set $R \subset P$ of its elements, the refinement procedure produces $P' \in \mathcal{P}$, such that $P \setminus P' \supseteq R$, i.e., the elements in $R$ are refined at least once. Let us denote it by $P' = \text{refine}(P,R)$. In practice, this is implemented by a usual naive refinement possibly producing a non-admissible partition, followed by a so-called completion procedure. We assume the existence of a constant $\lambda > 1$ such that $|\tau| \leq \lambda^{-n}|\sigma|$ for all $\tau \in P'$ and $\sigma \in R$ with $\tau \cap \sigma \neq \emptyset$. Note that we have $\lambda = 2$ for red refinements, and $\lambda = \sqrt{2}$ for the newest vertex bisection. Moreover, we assume the following on the efficiency of the refinement procedure: If $\{P_k\} \subset \mathcal{P}$ and $\{R_k\}$ are sequences such that
$P_{k+1} = \text{refine}(P_k, R_k)$ and $R_k \subset P_k$ for $k = 0, 1, \ldots$, then

$$\#P_k - \#P_0 \lesssim \sum_{m=0}^{k-1} \#R_m, \quad k = 1, 2, \ldots. \quad (6)$$

This assumption is justified for newest vertex bisection algorithm in Binev et al. (2004); Stevenson (2008), and demonstrated for a 1D refinement rule in Aurada, Feischl, Führer, Karkulik, and Praetorius (2012). The general red refinement procedure is treated in the appendix of this paper.

Next, we shall introduce an abstraction of finite element spaces. To this end, we assume that there is a quasi-Banach space $X_0$, and for each $P \in \mathcal{P}$, there is a nontrivial, finite dimensional subspace $S_P \subset X_0$. The space $X_0$ models the function space over $\Omega$ in which the approximation takes place, such as $X_0 = H^t(\Omega)$ and $X_0 = L_p(\Omega)$. The spaces $S_P$ are, as the reader might have guessed, models of finite element spaces, from which we approximate general functions in $X_0$. Obviously, a natural notion of error between an element $u \in X_0$ and its approximation $v \in S_P$ is the quasi-norm $\|u - v\|_{X_0}$. However, we need a bit more flexibility in how to measure such errors, and so we suppose that there is a function $\rho(u, v, P) \in [0, \infty]$ defined for $u \in X_0$, $v \in S_P$, and $P \in \mathcal{P}$. Note that this error measure, which we call a distance function, can depend on the partition $P$, and it is only required to make sense for functions $v \in S_P$. We allow the value $\rho = \infty$ to leave open the possibility that for some $u \in X_0$ we have $\rho(u, \cdot, \cdot) = \infty$. The most important distance function is still $\rho(u, v, P) = \|u - v\|_{X_0}$, but other examples will appear later in the paper, see e.g., Example 2.7, Subsection 3.5 and Section 4.

Given $u \in X_0$ and $P \in \mathcal{P}$, we let

$$E(u, S_P)_{\rho} = \inf_{v \in S_P} \rho(u, v, P), \quad (7)$$

which is the error of a best approximation of $u$ from $S_P$. Furthermore, we introduce

$$E_k(u)_{\rho} = \inf_{\{P \in \mathcal{P} : \#P \leq 2^k N\}} E(u, S_P)_{\rho}, \quad (8)$$

for $u \in X_0$ and $k \in \mathbb{N}$, with the constant $N$ chosen sufficiently large in order to ensure that the set $\{P \in \mathcal{P} : \#P \leq 2N\}$ is nonempty. In a certain sense, $E_k(u)_{\rho}$ is the best approximation error when one tries to approximate $u$ within the budget of $2^kN$ triangles. Finally, we define the main object of our study, the (adaptive) approximation class

$$\mathcal{A}_s^q(\rho) = \mathcal{A}_s^q(\rho, \mathcal{P}, \{S_P\}) = \{u \in X_0 : |u|_{\mathcal{A}_s^q(\rho)} < \infty\}, \quad (9)$$

where $s > 0$ and $0 < q \leq \infty$ are parameters, and

$$|u|_{\mathcal{A}_s^q(\rho)} = \|(2^{ks} E_k(u)_{\rho})_{k \in \mathbb{N}}\|_{\ell_q}, \quad u \in X_0. \quad (10)$$
In the following, we will use the abbreviation \( \mathscr{A}_q^s(\rho) = \mathscr{A}^s_{q,q}(\rho) \). Note that \( u \in \mathscr{A}_q^s(\rho) \) implies \( E_k(u) \leq c2^{-ks} \) for all \( k \) and for some constant \( c \), and these two conditions are equivalent if \( q = \infty \). We have \( \mathscr{A}_q^s(\rho) \subset \mathscr{A}_r^s(\rho) \) for \( q \leq r \), and \( \mathscr{A}_q^s(\rho) \subset \mathscr{A}^s_{q,q}(\rho) \) for \( s > \alpha \) and for any \( 0 < q, r \leq \infty \). The set \( \mathscr{A}_q^s(\rho) \) is not a linear space without further assumptions on \( \rho \) and \( \mathcal{P} \). However, in a typical situation, it is indeed a vector space equipped with the quasi-norm \( \| \cdot \|_{\mathscr{A}_q^s(\rho)} = \| \cdot \|_X + \| \cdot \|_{\mathscr{A}_q^s(\rho)} \).

Remark 2.1. Suppose that \( \rho \) satisfies

- \( \rho(\alpha u, \alpha v, P) = |\alpha| \rho(u, v, P) \) for \( \alpha \in \mathbb{R} \), and
- \( \rho(u + u', v + v', P) \lesssim \rho(u, v, P) + \rho(u', v', P) \).

Then \( | \cdot |_{\mathscr{A}_q^s(\rho)} \) is a quasi-seminorm, in the sense that it is positive homogeneous and satisfies the generalized triangle inequality. Moreover, \( \mathscr{A}_q^s(\rho) \) is a quasi-normed vector space. If only the second condition holds, then \( \mathscr{A}_q^s(\rho) \) would be a quasi-normed abelian group, in the sense of Bergh and Löfström (1976). Even though we will not use this fact, it is worth noting that \( \rho \) has the aforementioned properties for all applications we have in mind.

We call the approximation classes associated to \( \rho(u, v, \cdot) = \| u - v \|_{X_0} \) standard approximation classes, and write \( \mathscr{A}_q^s(X_0) \equiv \mathscr{A}_q^s(\rho) \) and \( \mathscr{A}(X_0) \equiv \mathscr{A}(\rho) \).

We want to characterize \( \mathscr{A}_q^s(\rho) \) in terms of an auxiliary quasi-Banach space \( X \hookrightarrow X_0 \). The main examples to keep in mind are \( X_0 = L^p \) and \( X = B^\alpha_{q,q} \), with \( \frac{1}{n} > \frac{1}{q} - \frac{1}{p} \). We assume that the space \( X \) has the following local structure: There exist a constant \( 0 < q < \infty \), and a function \( | \cdot |_{X(G)} : X \to \mathbb{R}^+ \) associated to each open set \( G \subset \Omega \), such that

\[
\sum_k |u|_{X(\tau_k)}^q \lesssim \|u\|_{X}^q \quad (u \in X),
\]

for any finite sequence \( \{\tau_k\} \subset P \) of non-overlapping elements taken from any \( P \in \mathcal{P} \). Finally, for any \( \tau \in P \) with \( P \in \mathcal{P} \), we let \( \hat{\tau} \subset \Omega \) be a domain containing \( \tau \), which will, in a typical situation, be the union of elements of \( P \) surrounding \( \tau \). We express the dependence of \( \hat{\tau} \) on \( P \) as \( \hat{\tau} = P(\tau) \). Then as an extension of the above sub-additivity property, we assume that

\[
\sum_k |u|_{X(P_k(\tau_k))}^q \lesssim \|u\|_{X}^q \quad (u \in X),
\]

for any finite sequences \( \{P_k\} \subset \mathcal{P} \) and \( \{\tau_k\} \), with \( \tau_k \in P_k \) and \( \{\tau_k\} \) non-overlapping. A trivial example of such a structure is \( X = L_q(\Omega) \) with \( | \cdot |_{X(G)} = \| \cdot \|_{L_q(G)} \). Here the sub-additivity (12) can be guaranteed if the underlying triangulations satisfy a certain local finiteness property.
2.2 Direct embeddings for standard approximation classes

The following theorem shows that the inclusion $X \subset A^s(\rho)$ can be proved by exhibiting a direct estimate. A direct application of this criterion is mainly useful for deriving embeddings of the form $X \subset A^s(X_0)$. In the next subsection, it will be generalized to a criterion that is valid in a more complex situations.

**Theorem 2.2.** Let $0 < p \leq \infty$ and let $\delta > 0$. Then for any $k \in \mathbb{N}$ sufficiently large there exists a partition $P \in \mathcal{P}$ with $\#P \leq k$ satisfying

$$
\left( \sum_{\tau \in P} |\tau|^{p\delta} |u|_{X(\hat{\tau})}^p \right)^{\frac{1}{p}} \lesssim k^{-s} \|u\|_X,
$$

with $s = \delta + \frac{1}{q} - \frac{1}{p} > 0$, where $\hat{\tau} = P(\tau)$ is as in (12), and the case $p = \infty$ must be interpreted in the usual way (with a maximum replacing the discrete $p$-norm). In particular, if $u \in X$ satisfies

$$
E(u, S_P)_\rho \lesssim \left( \sum_{\tau \in P} |\tau|^{p\delta} |u|_{X(\hat{\tau})}^p \right)^{\frac{1}{p}},
$$

for all $P \in \mathcal{P}$, then we have $u \in A^s(\rho)$ with $|u|_{A^s(\rho)} \lesssim \|u\|_X$.

**Proof.** What follows is a slight abstraction of the proof of Proposition 5.2 in Binev et al. (2002); we include it here for completeness. We first deal with the case $0 < p < \infty$. Let

$$
e(\tau, P) = |\tau|^{p\delta} |u|_{X(\hat{\tau})}^p,
$$

for $\tau \in P$ and $P \in \mathcal{P}$. Then for any given $\varepsilon > 0$, and any $P_0 \in \mathcal{P}$, below we will specify a procedure to generate a partition $P \in \mathcal{P}$ satisfying

$$
\sum_{\tau \in \mathcal{P}} e(\tau, P) \leq c'(\#P)\varepsilon,
$$

and

$$
\#P - \#P_0 \leq c\varepsilon^{-1/(1+ps)} \|u\|_{X}^{p/(1+ps)},
$$

where $c'$ depends only on the implicit constant of (14), and $c$ depends only on $|\Omega|$, $\lambda$, and the implicit constants of (6), and (12). Then, for any given $k > 0$, by choosing

$$
\varepsilon = (c/k)^{1+ps} \|u\|_{X}^{p},
$$

we would be able to guarantee a partition $P \in \mathcal{P}$ satisfying $\#P \leq \#P_0 + k$ and

$$
\sum_{\tau \in \mathcal{P}} e(\tau, P) \lesssim k^{-sp} \|u\|_{X}^{p}.
$$
This would imply the lemma, as $k^{-s}$ can be replaced by $(\#P_0 + k)^{-s}$ for, e.g., $k \geq \#P_0$.

Let $\varepsilon > 0$ and let $P_0 \in \mathcal{P}$. We then recursively define $R_k = \{\tau \in P_k : e(\tau, P_k) > \varepsilon\}$ and $P_{k+1} = \text{refine}(P_k, R_k)$ for $k = 0, 1, \ldots$ For all sufficiently large $k$ we will have $R_k = \emptyset$ since $|u|_{X(\tilde{\tau})} \lesssim \|u\|_X$ by (12), and $|\tau|$ is reduced by a constant factor $\mu = \lambda^{-n} < 1$ at each refinement. Let $P = P_k$, where $k$ marks the first occurrence of $R_k = \emptyset$. Then combining (14) with (15), and taking into account that $e(\tau, P_k) \leq \varepsilon$ for $\tau \in P_k$, we obtain (16).

In order to get a bound on $\#P$, we estimate the cardinality of $R = R_0 \cup R_1 \cup \ldots \cup R_{k-1}$, and use (6). Let $\Lambda_j = \{\tau \in R : \mu^{j+1} \leq |\tau| < \mu^j\}$ for $j \in \mathbb{Z}$, and let $m_j = \#\Lambda_j$. Note that the elements of $\Lambda_j$ (for any fixed $j$) are disjoint, since if any two elements intersect, then they must come from different $R_k$’s as each $R_k$ consists of disjoint elements, and hence by assumption on the refinement procedure, the ratio between the measures of the two elements must lie outside $(\mu, \mu^{-1})$. This gives the trivial bound

$$m_j \leq \mu^{-j-1}|\Omega|.$$  \hfill (20)

On the other hand, we have $e(\tau, P_k) > \varepsilon$ for $\tau \in \Lambda_j$ with some $k$, which gives

$$\varepsilon < |\tau|^p|\tilde{u}|_X^p < \mu^{jp\delta}|\tilde{u}|_X^p,$$  \hfill (21)

where $\tilde{\tau}$ is defined with respect to $P_k$, and $k$ may depend on $\tau$. Summing over $\tau \in \Lambda_j$, we get

$$m_j\varepsilon^{q/p} \leq \mu^{qj\delta} \sum_{\tau \in \Lambda_j} |u|^q_{X(\tilde{\tau})} \lesssim \mu^{qj\delta} \|u\|_X^q,$$  \hfill (22)

where we have used (12). Finally, summing for $j$, we obtain

$$\#R \leq \sum_{j=-\infty}^{\infty} m_j \lesssim \sum_{j=-\infty}^{\infty} \min \left\{ \mu^{-j}, \varepsilon^{-q/p}\mu^{qj\delta}\|u\|_X^q \right\} \lesssim \varepsilon^{-q/(1+q\delta)}\|u\|_X^{q/(1+q\delta)},$$  \hfill (23)

which, in view of (6) and $q/(1+q\delta) = p/(1+ps)$, establishes the bound (17).

We only sketch the case $p = \infty$ since the proof is essentially the same. We use

$$e(\tau, P) = |\tau|^p|\tilde{u}|_X^p,$$  \hfill (24)

instead of (15), and run the same algorithm. This guarantees that the resulting partition $P$ satisfies

$$\max_{\tau \in P} e(\tau, P) \leq \varepsilon.$$  \hfill (25)

We bound the cardinality of $P$ in the same way, which formally amounts to putting $p = 1$ into (21) and proceeding. The final result is

$$\#P - \#P_0 \leq c\varepsilon^{-q/(1+q\delta)}\|u\|_X^{q/(1+q\delta)} = c\varepsilon^{-1/s}\|u\|_X^{1/s},$$  \hfill (26)

where $s = \delta + \frac{1}{q}$. The proof is complete. \hfill □
Example 2.3. The main argument of the preceding proof can be traced back to Birman and Solomyak (1967). Recently, in Binev et al. (2002), this argument was applied to obtain an embedding of a Besov space into $\mathcal{A}^a(X_0)$, i.e., the case where the distance function $\rho$ is given by $\rho(u, v, \cdot) = \|u - v\|_{X_0}$. We want to include here one such application. Let $\Omega \subset \mathbb{R}^n$ be a bounded polyhedral domain with Lipschitz boundary, and take $\mathcal{P}$ to be the set of conforming triangulations of $\Omega$ obtained from a fixed conforming triangulation $P_0$ by means of newest vertex bisections. For $P \in \mathcal{P}$, let $S_P$ be the Lagrange finite element space of continuous piecewise polynomials of degree not exceeding $m$. Thus, for instance, the piecewise linear finite elements would correspond to $m = 1$. Moreover, for $P \in \mathcal{P}$ and $\tau \in P$, let $\hat{\tau} = P(\tau)$ be the interior of $\bigcup \{ \sigma : \sigma \in P, \sigma \cap \tau \neq \emptyset \}$. Finally, let us put $X_0 = L_p(\Omega)$, $X = B_{q,q}^\alpha(\Omega)$, and $\rho(u, v, \cdot) = \|u - v\|_{L_p(\Omega)}$. Then in this setting, the estimate (14) holds with the parameters $p$ and $\delta = \frac{\alpha}{n} + \frac{1}{p} - \frac{1}{q}$, as long as $0 < \alpha < m + \max\{1, \frac{1}{q}\}$ and $\delta > 0$, cf. Binev et al. (2002); Gaspoz and Morin (2013). Hence the preceding lemma immediately implies that $B_{q,q}^\alpha(\Omega) \hookrightarrow \mathcal{A}^a(L_p(\Omega))$ with $s = \frac{\alpha}{n}$.

In the rest of this subsection, we want to record some results involving interpolation spaces. For $u \in X_0$ and $t > 0$, the $K$-functional is

$$K(u, t; X_0, X) = \inf_{v \in X} (\|u - v\|_{X_0} + t\|v\|_X),$$

and for $0 < \theta < 1$ and $0 < \gamma \leq \infty$, we define the (real) interpolation space $(X_0, X)_{\theta, \gamma}$ as the space of functions $u \in X_0$ for which the quantity

$$\|u\|(X_0, X)_{\theta, \gamma} = \\|\|[\lambda^m K(u, \lambda^{-n}; X_0, X)]_{m \geq 0}\|_{L_\gamma},$$

is finite. These are quasi-Banach spaces with the quasi-norms $\| \cdot \|_{X_0} + \| \cdot \|(X_0, X)_{\theta, \gamma}$. The parameter $\lambda > 1$ can be chosen at one’s convenience, because the resulting quasi-norms are all pairwise equivalent.

Corollary 2.4. Let $0 < p \leq \infty$, $\delta > 0$ and let $s = \delta + \frac{1}{q} - \frac{1}{p} > 0$. Let

$$E(u, S_P)_{X_0} \lesssim \left( \sum_{\tau \in P} |\tau|^{p\delta} \|u\|_{X(\hat{\tau})}^p \right)^{\frac{1}{p}},$$

for all $u \in X$ and $P \in \mathcal{P}$. Then we have $(X_0, X)_{\alpha/s, \gamma} \subset \mathcal{A}^a(X_0)$ for $0 < \alpha < s$ and $0 < \gamma \leq \infty$.

Proof. For $u, v \in (X_0, X)_{\alpha/s, \gamma}$, we have

$$E_k(u)_{X_0} \lesssim \|u - v\|_{X_0} + E_k(v)_{X_0},$$

By Theorem 2.2, for $v \in X$ we have $E_k(v)_{X_0} \lesssim 2^{-ks}\|v\|_X$, leading to

$$E_k(u)_{X_0} \lesssim \|u - v\|_{X_0} + 2^{-ks}\|v\|_X.$$
2.3 Direct embeddings for general approximation classes

As mentioned in the introduction, our study is motivated by algorithms for approximating the solution of the operator equation $Tu = f$. Hence it should not come as a surprise that we assume the existence of a continuous operator $T : X_0 \to Y_0$, with $Y_0$ a quasi-Banach space. An example to keep in mind is the Laplace operator sending $H^1_0$ onto $H^{-1}$. In this subsection we do not assume linearity, although all examples of $T$ we have in this paper are linear. We also need an auxiliary quasi-Banach space $Y \hookrightarrow Y_0$, satisfying the properties analogous to that of $X$, in particular, (12) with some $0 < r < \infty$ replacing $q$ there. If $Y_0 = H^{-1}$, a typical example of $Y$ would be $B_{r,r}^{\sigma - 1}$ with $\sigma_n > 1 - \frac{1}{2}r$.

It is obvious that $\mathscr{A}^s(\rho) \subset \mathscr{A}^s(X_0)$, provided we have $\|u - v\|_{X_0} \lesssim \rho(u, v, \cdot)$. The latter condition is satisfied for all practical applications we have in mind. We will shortly present a theorem providing a criterion for answering questions such as $\mathscr{A}^s(X_0) \cap T^{-1}(Y) \subset \mathscr{A}^s(\rho)$.

Before stating the theorem, we need to introduce a bit more structure on the set $P$. The structure we need is that of overlay of partitions: We assume that there is an operation $\oplus : P \times P \to P$ satisfying

$$S_P + S_Q \subset S_{P \oplus Q}, \quad \text{and} \quad \#(P \oplus Q) \lesssim \#P + \#Q,$$

(32)

for $P, Q \in P$. In addition, we will assume that

$$\rho(u, v, P \oplus Q) \lesssim \rho(u, v, P).$$

(33)

In the conforming world, $P \oplus Q$ can be taken to be the smallest and common conforming refinement of $P$ and $Q$, for which (32) is demonstrated in Stevenson (2007), see also Cascon et al. (2008). The same argument works for the red refinement rule, cf. Aurada et al. (2012).

**Theorem 2.5.** Let $0 < p \leq \infty$, $\delta > 0$, and let $s = \delta + \frac{1}{r} - \frac{1}{p} > 0$. Assume that $u \in \mathscr{A}^s(X_0) \cap T^{-1}(Y)$ satisfies

$$E(u, S_P)_{X_0} \lesssim E(u, S_P)_{X_0} + \left( \sum_{\tau \in P} |\tau|^{p\delta} |Tu|^p_{Y(\hat{\tau})} \right)^\frac{1}{p},$$

(34)

for all $P \in P$ (in particular $E(u, \cdot)_{\rho}$ is always finite). Suppose also that

$$\left( \sum_{\tau \in P} |\tau|^{p\delta} |Tu|^p_{Y(\hat{\tau})} \right)^\frac{1}{p} \lesssim \left( \sum_{\tau \in P} |\tau|^{p\delta} |Tu|^p_{Y(\hat{\tau})} \right)^\frac{1}{p},$$

(35)

for any $P, Q \in P$. Then we have $u \in \mathscr{A}^s(\rho)$ with $|u|_{\mathscr{A}^s(\rho)} \lesssim |u|_{\mathscr{A}^s(X_0)} + \|Tu\|_Y$.

**Proof.** Let $k \in \mathbb{N}$ be an arbitrary number. Then by definition of $\mathscr{A}^s(X_0)$, there exists a partition $P' \in P$ such that

$$E(u, S_{P'})_{X_0} \leq 2^{-ks} |u|_{\mathscr{A}^s(X_0)}, \quad \text{and} \quad \#P' \leq 2^k N.$$

(36)
Similarly, by applying Theorem 2.2 with the roles of $X$ and $Y$, and those of $u$ and $Tu$ reversed, respectively, we can generate a partition $P'' \in \mathcal{P}$ such that
\[
\left( \sum_{\tau \in P''} |\tau|^{b_d}|Tu|_{Y(\tau)}^p \right)^{\frac{1}{p}} \lesssim 2^{-k\delta}||Tu||_Y, \quad \text{and} \quad \#P'' \leq 2^k N. \tag{37}
\]

Then for $P = P' \oplus P''$ we have $\#P \lesssim 2^k$ by (32). Moreover, (35) together with the obvious monotonicity
\[
E(u, S_P)_{X_0} \leq E(u, S_{P'})_{X_0}, \tag{38}
\]
guarantee that the right hand side of (34) is bounded by a multiple of $2^{-ks}(|u|_{\mathcal{A}^s(X_0)} + ||Tu||_Y)$, which completes the proof.

Remark 2.6. By the same argument, one can establish more general results, say, with an additional term involving $E(u, S_P)_{\rho_1}$ in the right hand side of (34), where $\rho_1$ is some distance function. We do not state such generalizations, but will use them on occasions later in the paper, cf. the proof of Theorem 4.1.

Example 2.7. We would like to illustrate the usefulness of Theorem 2.5 by sketching a simple application. For full details, we refer to Section 4, as the current example is a special case of the results derived there. We take $\Omega$ and $\mathcal{P}$ as in Example 2.3, and for $P \in \mathcal{P}$, let $S_P$ be the $L^2(\tau)$-orthogonal projection onto $P_d$, with $d \geq m - 2$ fixed. The sum involving $f - \Pi_\tau f$ is known as the oscillation term.

Let $0 < r, \alpha < \infty$ satisfy $\delta = \frac{\alpha}{n} - \frac{1}{r} + \frac{1}{2} \geq 0$ and $\alpha < d + \max\{1, \frac{1}{r}\}$. Then we claim that for each $u \in H^1_0(\Omega)$ with $\Delta u \in B^{\alpha}_{r,r}(\Omega)$, there exists $u_P \in S_P$ such that
\[
\rho(u, u_P, P) \lesssim \inf_{v \in S_P} \|u - v\|_{H^1} + \left( \sum_{\tau \in P} |\tau|^{2(\delta+1/n)}|\Delta u|_{B^{\alpha}_{r,r}(\tau)}^{2} \right)^{\frac{1}{2}}, \tag{40}
\]
for all $P \in \mathcal{P}$. In light of the preceding theorem, this would imply that each function $u \in \mathcal{A}^s(H^1_0(\Omega))$ with $\Delta u \in B^{\alpha}_{r,r}(\Omega)$, satisfies $u \in \mathcal{A}^s(\rho)$, cf. Figure 1(b). Note that since we can choose $d$ at will, the restriction $\alpha < d + \max\{1, \frac{1}{r}\}$ is immaterial.
To prove the claim, we take \( u_P \) to be the Scott-Zhang interpolator of \( u \), preserving the Dirichlet boundary condition. Then we have

\[
\|u - u_P\|_{H^1} \lesssim \inf_{v \in S_P} \|u - v\|_{H^1},
\]

for all \( P \in \mathcal{P} \). The oscillation term in (39) can be estimated as

\[
\|f - \Pi_T u\|_{L^2(\tau)} \leq \|f - g\|_{L^2(\tau)} \lesssim |\tau|^\delta \|f - g\|_{L^\infty(\tau)} + |\tau|^\delta \|f\|_{B^\alpha_{\tau}(\tau)},
\]

for any \( g \in \mathbb{P}_d \), where we have used continuity of the embedding \( B^\alpha_{\tau}(\tau) \subset L^2(\tau) \) and the fact that \( |g|_{B^\alpha_{\tau}(\tau)} = 0 \) when the Besov seminorm is defined using \( \omega_{d+1} \). Furthermore, if \( g \) is a best approximation of \( f \) in the \( L^r(\tau) \) sense, the Whitney estimate gives

\[
\|f - g\|_{L^r(\tau)} \lesssim \omega_{d+1}(f, \tau)_{r} \lesssim |f|_{B^\alpha_{\tau}(\tau)},
\]

which yields the desired result. In closing the example, we note that for this argument to work, the constants in the Whitney estimates and in the embeddings \( B^\alpha_{\tau,X}(\tau) \subset L^2(\tau) \) must be uniformly bounded independently of \( \tau \). While such investigations on Whitney estimates can be found in Dekel and Leviatan (2004); Gaspoz and Morin (2013), it seems difficult to locate similar studies on Besov space embeddings. To remove any doubt, the arguments in the following sections are arranged so that we do not use Besov space embeddings. Instead, we use embeddings between approximation spaces, and give a self contained proof that the embedding constants are suitably controlled.

**Remark 2.8.** With \( \mathcal{A}^s(Y_0) \) denoting the approximation class analogous to \( \mathcal{A}^s(X_0) \), where we replace \( X_0 \) by \( Y_0 \) and \( S_P \) by \( A(S_P) \), it is clear that \( \mathcal{A}^s(\rho) \subset T^{-1}(\mathcal{A}^s(Y_0)) \), provided we have \( \|Tu - Tv\|_{Y_0} \lesssim \rho(u, v, \cdot) \). This implies that if \( \|u - v\|_{X_0} + \|Tu - Tv\|_{Y_0} \lesssim \rho(u, v, \cdot) \), which is typically true in practice, then \( \mathcal{A}^s(\rho) \subset \mathcal{A}^s(X_0) \cap T^{-1}(\mathcal{A}^s(Y_0)) \). For the other direction, one can easily prove that if

\[
E(u, P) \lesssim \inf_{v \in S_P} \|u - v\|_{X_0} + \inf_{v \in S_P} \|Tu - Tv\|_{Y_0},
\]

for \( P \in \mathcal{P} \) and for \( u \in \mathcal{A}^s(X_0) \cap T^{-1}(\mathcal{A}^s(Y_0)) \), then \( \mathcal{A}^s(X_0) \cap T^{-1}(\mathcal{A}^s(Y_0)) \subset \mathcal{A}^s(\rho) \). Note that if \( T : X_0 \to Y_0 \) is an invertible bounded linear operator, then \( u \in \mathcal{A}^s(X_0) \) would automatically imply \( u \in T^{-1}(\mathcal{A}^s(Y_0)) \), so that \( \mathcal{A}^s(X_0) \subset \mathcal{A}^s(\rho) \) provided (44) holds for all \( u \in \mathcal{A}^s(X_0) \). However, (44) is not true in general, since, for example thinking of (39), \( \rho(u, \cdot, P) \) can be infinite even for some \( u \in S_{P'} \) with a sufficiently fine mesh \( P' \).

### 2.4 Inverse embeddings

In the preceding subsection, we have derived general criteria to get the inclusions \( X \cap T^{-1}(Y) \subset \mathcal{A}^s(\rho) \) and \( \mathcal{A}^s(X_0) \cap T^{-1}(Y) \subset \mathcal{A}^s(\rho) \). The inclusion \( \mathcal{A}^s(\rho) \subset \mathcal{A}^s(X_0) \) being
trivial, now we want to investigate to what extent the condition $u \in X \cap T^{-1}(Y)$ is necessary in order for $u \in \mathcal{A}^s(\rho)$.

From now on, the operator $T$ will be linear. We consider some auxiliary quasi-Banach spaces $X_1$ and $Y_1$, and assume that $X_0 \hookrightarrow X_1$ and $Y_0 \hookrightarrow Y_1$. The latter assumption is not really necessary, but we consider it here for convenience.

**Theorem 2.9.** Let $u \in \mathcal{A}^s_q(\rho)$ for some $s > 0$ and $0 < q \leq \infty$, and suppose that there is $P_0 \in \mathcal{P}$ with $\rho(u, 0, P_0) < \infty$. Moreover, assume that for each $P \in \mathcal{P}$ there exists $v \in S_P$ satisfying

$$\rho(u, v, P) \lesssim E(u, S_P)_\rho, \quad (45)$$

and

$$\|u - v\|_{X_1} + \|T(u - v)\|_{Y_1} \lesssim (\#P)^{\delta} \rho(u, v, P), \quad (46)$$

with some constant $\delta < s$. Finally, let $\sigma > s$, and assume that

$$\|v - w\|_{X_1} + \|T(v - w)\|_{Y_1} \lesssim (\#P)^{\sigma} (\rho(u, v, P) + \rho(u, w, P)), \quad (47)$$

for $P \in \mathcal{P}$ and $v, w \in S_P$. Then we have $u \in (Z_1, Z)_{(s-\delta)/(s-\delta), q}$ with

$$|u|_{(Z_1, Z)_{(s-\delta)/(s-\delta), q}} \lesssim \rho(u, 0, P_0) + |u|_{\mathcal{A}^s_q(\rho)}, \quad (48)$$

where $Z_1 = X_1 \cap T^{-1}(Y_1)$ and $Z = X \cap T^{-1}(Y)$. Note that we have the trivial inclusion $(Z_1, Z)_{\theta, q} \hookrightarrow (X_1, X)_{\theta, q} \cap T^{-1}((Y_1, Y)_{\theta, q})$ for all $0 < \theta < 1$.

**Proof.** The following is an adaptation of the standard argument, e.g., from DeVore and Lorentz (1993). For $k \in \mathbb{N}$, let $P_k \in \mathcal{P}$ be such that

$$E(u, P_k)_\rho = E_k(u)_\rho \quad \text{and} \quad \#P_k \leq 2^k N, \quad (49)$$

and let $v_k \in S_{P_k}$ satisfy (45) and (46) with respect to $P_k$. Then for any $m \in \mathbb{N}$, we have

$$K(u, 2^{-m(\sigma-\delta)}, X_1 \cap T^{-1}(Y_1), X \cap T^{-1}(Y))$$

$$\leq \|u - v_m\|_{X_1} + \|T(u - T v_m)\|_{Y_1} + 2^{-m(\sigma-\delta)} (\|v_m\|_{X_1} + \|Tv_m\|_{Y_1})$$

$$\lesssim 2^{m\delta} \rho(u, v_m, P_m) + 2^{-m(\sigma-\delta)} (\|v_m\|_{X_1} + \|Tv_m\|_{Y_1}), \quad (50)$$

where we have used (27) and (46). Since $v_k - v_{k-1} \in S_{P_k} \oplus P_{k-1}$, by (47), (32), and (33) we obtain

$$\|v_k - v_{k-1}\|_{X_1} \lesssim 2^{k\sigma} (\rho(u, v_k, P_k \oplus P_{k-1}) + \rho(u, v_{k-1}, P_k \oplus P_{k-1}))$$

$$\lesssim 2^{k\sigma} (\rho(u, v_k, P_k) + \rho(u, v_{k-1}, P_{k-1}))$$

$$\lesssim 2^{k\sigma} (E_k(u)_\rho + E_{k-1}(u)_\rho), \quad (51)$$
with the understanding that \( v_0 = 0 \) and \( E_0(u)_\rho = \rho(u,0,P_0) \). By the Aoki-Rolewicz theorem (Bergh and L"ofstr"om, 1976, page 59) it is no loss of generality to assume that the quasi-norm \( \| \cdot \|_X \) satisfies the \( \mu \)-triangle inequality
\[
\|w + v\|_X^\mu \leq \|w\|_X^\mu + \|v\|_X^\mu, \quad w, v \in X,
\] (52)
for some \( \mu > 0 \). Applying this to the sum \( v_m = \sum_{k=1}^{m} (v_k - v_{k-1}) \), we get
\[
\|v_m\|_X^\mu \leq \sum_{k=1}^{m} \|v_k - v_{k-1}\|_X^\mu \lesssim \sum_{k=0}^{m} [2^{k\sigma} E_k(u)_\rho]^\mu. \tag{53}
\]
We can do the same computation for the term \( \|Tv_m\|_Y \) in (50), to derive the bound
\[
\|Tv_m\|_Y^\nu \lesssim \sum_{k=0}^{m} [2^{k\sigma} E_k(u)_\rho]^\nu, \tag{54}
\]
with some \( \nu > 0 \). In fact, we can suppose \( \mu = \nu \) by taking the minimum of the two. Therefore we conclude
\[
2^{-m\delta} K(u, 2^{-m(\sigma-\delta)}, X_1 \cap T^{-1}(Y_1), X \cap T^{-1}(Y)) \lesssim 2^{-m\sigma} \left( \sum_{k=0}^{m} [2^{k\sigma} E_k(u)_\rho]^\mu \right)^{1/\mu}. \tag{55}
\]
The proof is complete upon using the formula (28), and the discrete Hardy inequality, recalled below in Lemma 2.12.

A simple example of a distance function that lies outside the standard theory but can be handled by the preceding theorem is \( \rho(u,v,P) = \| u - v \|_P \), where \( \| \cdot \|_P \) is some quasi-norm on \( X_0 \) possibly depending on \( P \in \mathcal{P} \).

**Corollary 2.10.** In the setting of the previous paragraph, suppose that for each \( u \in X_0 \) and \( P \in \mathcal{P} \) there is \( v \in S_P \) satisfying
\[
\|u - v\|_P \lesssim E(u, u, S_P)_\rho \quad \text{and} \quad \|u - v\|_{X_1} \lesssim (\#P)^\delta \|u - v\|_P, \tag{56}
\]
with some constant \( \delta \geq 0 \). Moreover, let \( \sigma > \delta \), and let
\[
\|v\|_X \lesssim (\#P)^\sigma \|v\|_P, \quad \text{for} \quad P \in \mathcal{P} \quad \text{and} \quad v \in S_P. \tag{57}
\]
Then we have \( \mathcal{A}_q^s(\rho) \cap X_0 \hookrightarrow (X_1, X)(s-\delta)/(\sigma-\delta)_q \) for each \( \delta < s < \sigma \) and \( 0 < q \leq \infty \).

**Proof.** We apply the theorem with \( T = 0 \). The estimate (57) easily implies (47), as
\[
\|v - w\|_X \lesssim (\#P)^\sigma \|v - w\|_P \lesssim (\#P)^\sigma (\|v - u\|_P + \|u - w\|_P), \tag{58}
\]
for \( v, w \in S_P \). The existence of \( P_0 \in \mathcal{P} \) with \( \rho(u,0,P_0) < \infty \) is trivial, since \( \|u\|_P \) is finite for any \( P \in \mathcal{P} \) and any \( u \in X_0 \). Hence all the hypotheses of the theorem are satisfied for each \( u \in \mathcal{A}_q^s(\rho) \cap X_0 \).

\[\square\]
Example 2.11. Continuing Example 2.3, recall from there that \( B^\alpha_{q,q}(\Omega) \hookrightarrow \mathcal{A}^s_{\infty}(L_p(\Omega)) \) with \( s = \frac{\alpha}{n} \), as long as \( \frac{\alpha}{n} + \frac{1}{p} - \frac{1}{q} > 0 \) and \( 0 < \alpha < m + \max\{1, \frac{1}{q}\} \). In light of the preceding corollary with \( X_0 = X_1 \), inclusions in the other direction depend on the inverse estimate (57). Indeed, this estimate has been proved in Binev et al. (2002); Gaspoz and Morin (2013) with \( \|\cdot\|_p = \|\cdot\|_{L_p} \) and \( X = A^{ns}_{q,q}(\Omega) \) for all \( s > 0 \) and \( 0 < q < \infty \) satisfying \( \frac{1}{q} = s + \frac{1}{p} \). The space \( A^{\alpha}_{q,q}(\Omega) \) coincides with the Besov space \( B^\alpha_{q,q}(\Omega) \) when \( \alpha < 1 + \frac{1}{q} \).

We refer to Section 3 for more details on these spaces. In any case, from the preceding corollary we infer \( A^s_{q,q}(L_p(\Omega)) \hookrightarrow B^\alpha_{q,q}(\Omega) \) for \( s = \frac{\alpha}{n} = \frac{1}{q} - \frac{1}{p} \), \( \alpha < 1 + \frac{1}{q} \), and \( 0 < p, q < \infty \).

The following result, often called the discrete Hardy inequality, was used in the proof of Theorem 2.9, and will be used many times later. We include the statement here for convenience. A proof can be found in (DeVore and Lorentz, 1993, page 27).

Lemma 2.12. Let \((a_j)_{j \in \mathbb{Z}}\) and \((b_k)_{k \in \mathbb{Z}}\) be two sequences satisfying either
\[
|a_j| \leq C \left( \sum_{k=j}^{\infty} |b_k|^\mu \right)^{1/\mu}, \quad j \in \mathbb{Z},
\]
for some \( \mu > 0 \) and \( C > 0 \), or
\[
|a_j| \leq C 2^{-\theta j} \left( \sum_{k=-\infty}^{j} 2^{\theta k} |b_k|^\mu \right)^{1/\mu}, \quad j \in \mathbb{Z},
\]
for some positive \( \theta \), \( \mu \), and \( C \). Then we have
\[
\|(2^{\alpha j}a_j)\|_{\ell_q} \lesssim C \|(2^{\alpha k}b_k)\|_{\ell_q},
\]
for all \( 0 < q \leq \infty \) and \( 0 < \alpha < \theta \), with the convention that \( \theta = \infty \) if (59) holds, and with the implicit constant depending only on \( q \) and \( \alpha \).

3 Lagrange finite elements

3.1 Besov spaces on bounded domains

Let \( \Omega \subset \mathbb{R}^n \) be a bounded domain. Then for \( 0 < p \leq \infty \), we define the \( r \)-th order \( L_p \)-modulus of smoothness
\[
\omega_r(u, t, \Omega)_p = \sup_{|h| \leq t} \|\Delta^r_h u\|_{L_p(\Omega_r h)}
\]
where \( \Omega_r h = \{x \in \Omega : [x + rh] \subset \Omega\} \), and \( \Delta^r_h \) is the \( r \)-th order forward difference operator defined recursively by \( [\Delta^1_h u](x) = u(x + h) - u(x) \) and \( \Delta^k_h u = \Delta^1_h (\Delta^{k-1}_h u) \), i.e.,
\[
\Delta^r_h u(x) = \sum_{k=0}^{r} (-1)^{r+k} \binom{r}{k} u(x + kh).
\]
Furthermore, for $0 < p, q \leq \infty$, $\alpha \geq 0$, and $r \in \mathbb{N}$, the Besov space $B_{p,q;r}^\alpha(\Omega)$ consists of those $u \in L_p(\Omega)$ for which

$$|u|_{B_{p,q;r}^\alpha(\Omega)} = \| t \mapsto t^{-\alpha - 1/q} \omega_r(u, t, \Omega)_p \|_{L_q((0,\infty))}, \quad (64)$$

is finite. Since $\Omega$ is bounded, being in a Besov space is a statement about the size of $\omega_r(u, t, \Omega)_p$ only for small $t$. From this it is easy to derive the useful equivalence

$$|u|_{B_{p,q;r}^\alpha(\Omega)} \approx \| (\lambda^{\alpha_j} \omega_r(u, \lambda^{-j}, \Omega)_p)_{j \geq 0} \|_{\ell_q}, \quad (65)$$

for any constant $\lambda > 1$. The mapping $\| \cdot \|_{B_{p,q;r}^\alpha(\Omega)} = \| \cdot \|_{L_p(\Omega)} + \| \cdot \|_{B_{p,q;r}^\alpha(\Omega)}$ defines a norm when $p, q \geq 1$ and a quasi-norm in general. If $\alpha > r + \max\{0, \frac{1}{p} - 1\}$ then the space $B_{p,q;r}^\alpha$ is trivial in the sense that $B_{p,q;r}^\alpha = \mathbb{P}_{r-1}$. On the other hand, so long as $r > \alpha - \max\{0, \frac{1}{p} - 1\}$, different choices of $r$ will result in quasi-norms that are equivalent to each other, and in this case we have the classical Besov spaces $B_{p,q}^\alpha(\Omega) = B_{p,q;r}^\alpha(\Omega)$. In the borderline case, the situation depends on the index $q$. If $0 < q < \infty$ and $\alpha = r + \max\{0, \frac{1}{p} - 1\}$, then $B_{p,q;r}^\alpha = \mathbb{P}_{r-1}$. The case $q = \infty$ gives nontrivial spaces: For instance, we have $B_{p,\infty;r}^\alpha(\Omega) = W^{r,p}(\Omega)$ for $p > 1$. A proof can be found in (DeVore and Lorentz, 1993, page 53) for the one dimensional case, and the same proof works in multi-dimensions.

Now we describe various embedding relationships among the Besov and Sobolev spaces, even though we do not use them in this paper. Since $\Omega$ is bounded, it is clear that $B_{p,q}^\alpha(\Omega) \hookrightarrow B_{p,q}^{\alpha'}(\Omega)$ for any $\alpha \geq 0$, $0 < q \leq \infty$ and $\infty \geq p > p' > 0$. From the equivalence (65), we have the lexicographical ordering $B_{p,q}^\alpha(\Omega) \hookrightarrow B_{p,q}^{\alpha'}(\Omega)$ for $\alpha > \alpha'$ with any $0 < q, q' \leq \infty$, and $B_{p,q}^\alpha(\Omega) \hookrightarrow B_{p,q}^{\alpha'}(\Omega)$ for $0 < q < q' \leq \infty$. Nontrivial embeddings are

$$B_{p,q}^\alpha(\Omega) \hookrightarrow B_{p,q}^{\alpha'}(\Omega), \quad \text{for} \quad \frac{\alpha - \alpha'}{n} = \frac{1}{p} - \frac{1}{p'} > 0, \quad (66)$$

and

$$B_{q,q}^\alpha(\Omega) \hookrightarrow L^p(\Omega), \quad \text{for} \quad \frac{\alpha}{n} = \frac{1}{q} - \frac{1}{p} > 0. \quad (67)$$

Finally, we recall the fact that $B_{2,2}^\alpha(\Omega) = H^\alpha(\Omega)$ for all $\alpha > 0$.

### 3.2 Quasi-interpolation operators

Let $\Omega \subset \mathbb{R}^n$ be a bounded polyhedral domain with Lipschitz boundary, and fix a conforming partition $P_0$ of $\Omega$. We also fix a refinement rule, which is either the newest vertex bisection or the red refinement. The set $\mathcal{P}$ will be, in case of the newest vertex bisection, the set of all conforming triangulations arising from $P_0$, and in case of the red refinement, a family arising from $P_0$ which has the gradedness property

$$\sup \left\{ \frac{\text{diam } \sigma}{\text{diam } \tau} : \sigma, \tau \in P, \sigma \cap \tau \neq \emptyset, P \in \mathcal{P} \right\} < \infty. \quad (68)$$
In the latter case, we remark that the results in this paper are insensitive to the exact definition of \( \mathcal{P} \), so long as the family \( \mathcal{P} \) is graded. For example, we can define \( \mathcal{P} \) by the requirement that the number of hanging nodes per edge is bounded by 3, or by the requirement that the diameter ratio between neighbouring elements stays bounded by 10. As a consequence, we will have to deal with nonconforming partitions, but the degrees of freedom will be so arranged that they give rise to \( H^1 \)-conforming finite element spaces. In this regard, the terminology “nonconforming partition” may be a bit confusing.

We define the Lagrange finite element spaces

\[
S_P = S^m_P = \{ u \in C(\Omega) : u|_\tau \in \mathbb{P}_m \forall \tau \in P \}, \quad P \in \mathcal{P},
\]

where \( \mathbb{P}_m \) is the space of polynomials of degree not exceeding \( m \). Thus, for instance, the piecewise linear finite elements would correspond to \( m = 1 \).

Following Gaspoz and Morin (2013), we will now construct a quasi-interpolation operator \( \tilde{Q}_P : L^p_0(\Omega) \to S_P \) for \( p_0 > 0 \) small. Their construction works verbatim here but we need to be a bit careful since we want to include partitions with hanging nodes into the analysis. Let \( \tau_0 = \{ x \in \mathbb{R}^n : x_1 + \ldots + x_n < m \} \cap (0,m)^n \) be the standard simplex. Then an \( n \)-simplex \( \tau \subset \mathbb{R}^n \) is the image of \( \tau_0 \) under an invertible affine mapping. To each \( n \)-simplex \( \tau \), we associate its nodal set \( N_\tau = F(\bar{\tau}_0 \cap \mathbb{Z}^n) \), where \( F : \tau_0 \to \tau \) is any invertible affine mapping. For each \( \tau \in P \), we define the set \( \{ \xi_{\tau,z} : z \in N_{\tau} \} \subset \mathbb{P}_m \) of shape functions by \( \xi_{\tau,z}(z') = \delta_{z,z'} \) for \( z, z' \in N_{\tau} \), where \( \delta_{z,z'} \) is the Kronecker delta. The nodal set \( N_P \) of a possibly nonconforming partition \( P \in \mathcal{P} \) is defined by the requirement that \( z \in \bigcup_{\tau \in P} N_{\tau} \) is in \( N_P \) if and only if \( z \in N_{\tau} \) for all \( \tau \) with \( \bar{\tau} \ni z \), see Figure 2. Furthermore, we define the nodal basis \( \{ \phi_z : z \in N_P \} \subset S_P \) of \( S_P \) by \( \phi_z(z') = \delta_{z,z'} \) for \( z, z' \in N_P \).

![Fig. 2: Examples of nodal sets.](image)

(a) Quadratic elements \((m = 2)\). (b) Cubic elements \((m = 3)\).

Next, we introduce a basis dual to \( \{ \phi_z \} \). For each \( \tau \in P \) and \( z \in N_{\tau} \), we let \( \eta_{\tau,z} \in \mathbb{P}_m \) be such that

\[
\int_{\tau} \eta_{\tau,z} \xi_{\tau,z'} = \delta_{z,z'}, \quad z' \in N_P \cap \bar{\tau},
\]
and for \( z \in N_P \), define
\[
\tilde{\phi}_z = \frac{1}{n_z} \sum_{\{\tau \in P : \tau \ni z\}} \chi_{\tau} \eta_{\tau,z},
\tag{71}
\]
where \( n_z = \#\{\tau \in P : \tau \ni z\} \), and \( \chi_{\tau} \) is the characteristic function of \( \tau \). By construction, \( \text{supp} \tilde{\phi}_z \subset \text{supp} \phi_z \) for \( z \in N_P \) and we have the biorthogonality
\[
\langle \tilde{\phi}_z, \phi_{z'} \rangle = \int_{\Omega} \tilde{\phi}_z \phi_{z'} = \delta_{z,z'}, \quad z, z' \in N_P.
\tag{72}
\]
Now we define the quasi-interpolation operator \( Q_P : L_1(\Omega) \to S_P \) by
\[
Q_P u = Q^{(\Omega)}_P u = \sum_{z \in N_P} \langle u, \tilde{\phi}_z \rangle \phi_z,
\tag{73}
\]
It is clear that \( Q_P \) is linear and that \( Q_P v = v \) for \( v \in S_P \).

**Lemma 3.1.** For \( 1 \leq p \leq \infty \), we have
\[
\|u - Q_P u\|_{L_p(\Omega)} \lesssim \inf_{v \in S_P} \|u - v\|_{L_p(\Omega)}, \quad u \in L_p(\Omega),
\tag{74}
\]
with the implicit constant depending only on the shape regularity and the gradedness constants of \( \mathcal{P} \). Furthermore, for \( 0 < p \leq \infty \) and \( \tau \in P \), we have
\[
\|Q_P v\|_{L_p(\tau)} \lesssim \|v\|_{L_p(\hat{\tau})}, \quad v \in \tilde{S}^m_P,
\tag{75}
\]
where
\[
\tilde{S}^m_P = \{w \in L_\infty(\Omega) : w|_{\tau} \in \mathbb{P}_m \forall \tau \in P\},
\tag{76}
\]
and \( \hat{\tau} = P(\tau) \) is the interior of \( \bigcup\{\eta : \sigma \in P, \sigma \cap \tau \neq \emptyset\} \).

**Proof.** For \( 1 \leq p \leq \infty \) and \( u \in L_p(\Omega) \), we have
\[
\|Q_P u\|_{L_p(\tau)} \leq \sum_{z \in N_P \cap \tau} \|\langle u, \tilde{\phi}_z \rangle\|_{L_p(\tau)} \|\phi_z\|_{L_p(\tau)} \leq \|u\|_{L_p(\hat{\tau})} \sum_{z \in N_P \cap \tau} \|\tilde{\phi}_z\|_{L_q(\tau)} \|\phi_z\|_{L_p},
\tag{77}
\]
where \( \frac{1}{p} + \frac{1}{q} = 1 \). It is clear that \( \|\phi_z\|_{L_p} \leq |\text{supp} \phi_z|^{1/p} \) and by a scaling argument one can deduce that \( \|\tilde{\phi}_z\|_{L_q} \lesssim |\text{supp} \tilde{\phi}_z|^{1/q-1} \), with the implicit constant depending only on the shape regularity and the gradedness constants of \( \mathcal{P} \). Using this result, for \( 1 \leq p < \infty \), we infer
\[
\|Q_P u\|_{L_p(\Omega)}^p = \sum_{\tau \in P} \|Q_P u\|_{L_p(\tau)}^p \lesssim \sum_{\tau \in P} \|u\|_{L_p(\hat{\tau})}^p \lesssim \|u\|_{L_p(\Omega)}^p,
\tag{78}
\]
by the local finiteness of the mesh. The case \( p = \infty \) can be handled similarly, and we have \( \|Q_P u\|_{L_p(\Omega)} \lesssim \|u\|_{L_p(\Omega)} \) for \( 1 \leq p \leq \infty \). Then a standard argument yields (74).

For \( 1 \leq p \leq \infty \), (77) implies (75). The proof for \( 0 < p < 1 \) follows exactly the same lines as those in the proof of (Gaspoz and Morin, 2013, Lemma 3.2). □
In the following, we fix $p_0 > 0$, and for $\tau \subset \mathbb{R}^n$ a domain, let $\Pi_{p_0, \tau} : L_{p_0}(\tau) \rightarrow \mathbb{P}_m$ be the local polynomial approximation operator given in Definition 3.7 of Gaspoz and Morin (2013). We recall the following important properties of this operator, cf. (Gaspoz and Morin, 2013, Theorem 3.8).

(i) There is a constant $C_{m,p_0}$ depending only on $m$ and $p_0$, such that

$$\|u - \Pi_{p_0, \tau} u\|_{L_{p_0}(\tau)} \leq C_{m,p_0} \inf_{v \in \mathbb{P}_m} \|u - v\|_{L_{p_0}(\tau)}, \quad u \in L_{p_0}(\tau).$$  \hspace{1cm} (79)

In other words, $\Pi_{p_0, \tau} u$ is a near-best approximation of $u$ from $\mathbb{P}_m$ in $L_{p_0}(\tau)$.

(ii) We have

$$\|\Pi_{p_0, \tau} u\|_{L_{p_0}(\tau)} \lesssim \|u\|_{L_{p_0}(\tau)}, \quad u \in L_{p_0}(\tau),$$  \hspace{1cm} (80)

i.e., the operator $\Pi_{p_0, \tau} : L_{p_0}(\tau) \rightarrow L_{p_0}(\tau)$ sends bounded sets to bounded sets.

(iii) For any $u \in L_{p_0}(\tau)$ and $v \in \mathbb{P}_m$, we have

$$\Pi_{p_0, \tau}(u + v) = \Pi_{p_0, \tau} u + v.$$  \hspace{1cm} (81)

In particular, $\Pi_{p_0, \tau} v = v$ for $v \in \mathbb{P}_m$.

Finally, we let

$$\Pi_P u = \sum_{\tau \in P} \chi_{\tau} \Pi_{p_0, \tau} u,$$  \hspace{1cm} (82)

and define the operator $\tilde{Q}_P : L_{p_0}(\Omega) \rightarrow S_P$ by

$$\tilde{Q}_P u = Q_P \Pi_P u = \sum_{z \in N_P} (\Pi_P u, \tilde{\phi}_z) \phi_z.$$  \hspace{1cm} (83)

It is easy to see that $\tilde{Q}_P v = v$ for $v \in S_P$, and that $(\tilde{Q}_P u)|_{\tilde{\tau}}$ depends only on $u|_{\tilde{\tau}}$, where $\tilde{\tau} = P(\tau)$ is the interior of $\bigcup\{\sigma : \sigma \in P, \sigma \cap \tau \neq \emptyset\}$. Furthermore, as a consequence of the linearity (81), we have

$$(\tilde{Q}_P(u + v))|_{\tilde{\tau}} = (\tilde{Q}_P u)|_{\tilde{\tau}} + v|_{\tilde{\tau}}, \quad u, v \in L_{p_0}(\tau), \quad v|_{\tilde{\tau}} \in S_P,$$  \hspace{1cm} (84)

for $\tau \in P$.

**Lemma 3.2.** Let $p_0 \leq p \leq \infty$ and $P \in \mathcal{P}$. Then for $\tau \in P$ we have

$$\|\tilde{Q}_P u\|_{L_p(\tau)} \lesssim \|u\|_{L_p(\tilde{\tau})}, \quad u \in L_p(\Omega).$$  \hspace{1cm} (85)

As a consequence, we have

$$\|u - \tilde{Q}_P u\|_{L_p(\tau)} \lesssim \inf_{v \in S_P} \|u - v\|_{L_p(\tilde{\tau})}, \quad u \in L_p(\Omega),$$  \hspace{1cm} (86)

and

$$\|u - \tilde{Q}_P u\|_{L_p(\Omega)} \lesssim \inf_{v \in S_P} \|u - v\|_{L_p(\Omega)}, \quad u \in L_p(\Omega).$$  \hspace{1cm} (87)
Proof. An application of (75) gives
\[ \|\tilde{Q}_P u\|_{L_p(\tau)} = \|Q_P \Pi_P u\|_{L_p(\tau)} \lesssim \|\Pi_P u\|_{L_p(\hat{\tau})}. \] (88)
On the other hand, for \( \sigma \in P \), we have
\[ \|\Pi_{p_0,\sigma} u\|_{L_p(\sigma)} \lesssim |\sigma|^{\frac{1}{p} - \frac{1}{p_0}} \|\Pi_{p_0,\sigma} u\|_{L_{p_0}(\sigma)} \lesssim |\sigma|^{\frac{1}{p} - \frac{1}{p_0}} \|u\|_{L_{p_0}(\sigma)} \leq \|u\|_{L_p(\sigma)}, \] (89)
where we have used scaling properties of polynomials in the first step, the boundedness (80) of \( \Pi_{p_0,\sigma} \) in the second step, and the Hölder inequality in the final step. Using this, with the usual modifications for \( p = \infty \), we infer
\[ \|\Pi_P u\|_{L_p(\hat{\tau})} = \left( \sum_{\{\sigma \in P : \sigma \subset \hat{\tau}\}} \|\Pi_{p_0,\sigma} u\|_{L_p(\sigma)}^p \right)^{\frac{1}{p}} \lesssim \left( \sum_{\{\sigma \in P : \sigma \subset \hat{\tau}\}} \|u\|_{L_p(\sigma)}^p \right)^{\frac{1}{p}} = \|u\|_{L_p(\hat{\tau})}, \] (90)
establishing (85). Then (86) follows from the linearity property (84).

The estimate (87) is proved by first deriving the stability
\[ \|\tilde{Q}_P u\|_{L_p(\Omega)} = \left( \sum_{\tau \in P} \|\tilde{Q}_P u\|_{L_p(\tau)}^p \right)^{\frac{1}{p}} \lesssim \left( \sum_{\tau \in P} \|u\|_{L_p(\hat{\tau})}^p \right)^{\frac{1}{p}} \lesssim \|u\|_{L_p(\Omega)}, \] (91)
and then invoking the linearity property (84). \( \square \)

An important tool in approximation theory is the Whitney estimate
\[ \inf_{v \in \mathbb{P}_m} \|u - v\|_{L_p(G)} \lesssim \omega_{m+1}(u, \text{diam } G, G)_p, \quad u \in L_p(G), \] (92)
that holds for any convex domain \( G \subset \mathbb{R}^n \), with the implicit constant depending only on \( n, m \), and \( 0 < p \leq \infty \), see Dekel and Leviatan (2004). The same estimate is also true when \( G \) is the star around \( \tau \in P \) for some partition \( P \in \mathcal{P} \), with the implicit constant additionally depending on the shape regularity constant of \( \mathcal{P} \), see Gaspoz and Morin (2013).

**Lemma 3.3.** Let \( p_0 \leq p \leq \infty \) and \( P \in \mathcal{P} \). Then we have
\[ \|u - \tilde{Q}_P u\|_{L_p(\Omega)} \lesssim \left( \frac{\max_{\tau \in P} \text{diam } \tau}{\min_{\tau \in P} \text{diam } \tau} \right)^{\frac{p}{2}} \omega_{m+1}(u, \text{max } \text{diam } \tau, \Omega)_p, \quad u \in L_p(\Omega). \] (93)

**Proof.** We start with the special case \( p = \infty \). It is immediate from (86) and the Whitney estimate (92) that
\[ \|u - \tilde{Q}_P u\|_{L_\infty(\hat{\tau})} \lesssim \max_{\tau \in P} \|u - \tilde{Q}_P u\|_{L_\infty(\tau)} \lesssim \max_{\tau \in P} \inf_{v \in \mathbb{P}_p} \|u - v\|_{L_\infty(\hat{\tau})} \lesssim \max_{\tau \in P} \omega_{m+1}(u, \text{diam } \hat{\tau}, \hat{\tau})_\infty \lesssim \max_{\tau \in P} \omega_{m+1}(u, \mu^{-1} \text{diam } \hat{\tau}, \hat{\tau})_\infty, \] (94)
with $\mu > 0$ sufficiently large, where in the last line we have used the property
\[
\omega_r(u, \mu t, G)_p \leq (\mu + 1)^r \omega_r(u, t, G)_p, \tag{95}
\]
cf. (DeVore and Lorentz, 1993, §2.7). With $t = \mu^{-1} \max_{\tau \in P} \text{diam } \tau$, we proceed as
\[
\|u - \tilde{Q} P u\|_{L_\infty(\Omega)} \lesssim \max_{\tau \in P} \omega_{m+1}(u, t, \hat{\tau})_\infty = \max_{\tau \in P} \sup_{|h| \leq t} \|\Delta_{h}^{m+1} u\|_{L_\infty(\hat{\tau}_h)}
\]
\[
= \sup_{|h| \leq t} \max_{\tau \in P} \|\Delta_{h}^{m+1} u\|_{L_\infty(\hat{\tau}_h)} \leq \sup_{|h| \leq t} \|\Delta_{h}^{m+1} u\|_{L_\infty(\Omega_h)}, \tag{96}
\]
which establishes (93) for $p = \infty$.

To handle the case $0 < p < \infty$ we introduce the averaged $L_p$-modulus of smoothness
\[
w_r(u, t, G)_p = \left(\frac{1}{t^n} \int_{[0, t]^n} \|\Delta_{h}^{r} u\|_{L_p(G_h)}^p \, dh\right)^{1/p}, \tag{97}
\]
for any domain $G \subset \mathbb{R}^n$. When $G$ is Lipschitz, the averaged modulus is equivalent to the original one:
\[
w_r(u, t, G)_p \sim \omega_r(u, t, G)_p, \quad \text{for } t \lesssim 1. \tag{98}
\]
This equivalence is also true when $G = \tau$ or $G = \hat{\tau}$ for $\tau \in P$ with $P \in \mathcal{P}$, in the range $t \lesssim \text{diam } G$, cf. Corollary 4.3 of Gaspoz and Morin (2013). In the latter case, the implicit constants depend only on $p$, $r$, the shape regularity constant of $P$, and the geometry of the underlying domain $\Omega$.

Let us get back to the proof of (93) for $0 < p < \infty$. As in the case $p = \infty$, we have
\[
\|u - \tilde{Q} P u\|_{L_p(\Omega)}^p \leq \sum_{\tau \in P} \|u - \tilde{Q} P u\|_{L_p(\tau)}^p \leq \sum_{\tau \in P} \inf_{v \in S_\tau} \|u - v\|_{L_p(\hat{\tau})}^p
\]
\[
\lesssim \sum_{\tau \in P} \omega_{m+1}(u, \text{diam } \hat{\tau}, \hat{\tau})_p \lesssim \sum_{\tau \in P} \omega_{m+1}(u, \mu^{-1} \text{diam } \hat{\tau}, \hat{\tau})_p, \tag{99}
\]
with $\mu > 0$ sufficiently large. Now we employ (98), to get
\[
\|u - \tilde{Q} P u\|_{L_p(\Omega)}^p \lesssim \sum_{\tau \in P} w_{m+1}(u, \mu^{-1} \text{diam } \hat{\tau}, \hat{\tau})_p^p
\]
\[
= \sum_{\tau \in P} \frac{1}{t(\tau)^n} \int_{[0, t(\tau)]^n} \int_{\hat{\tau}_h} |\Delta_{h}^{r} u(x)|^p \, dx \, dh, \tag{100}
\]
where $t(\tau) = \mu^{-1} \text{diam } \hat{\tau}$ and $r = m+1$. With $t_0 = \mu^{-1} \min_{\tau \in P} \text{diam } \hat{\tau}$ and $t_1 = \mu^{-1} \max_{\tau \in P} \text{diam } \hat{\tau}$,
we can switch the sum with the outer integration as follows.

\[
\|u - \tilde{Q}_P u\|_{L^p(\Omega)}^p \lesssim \frac{1}{t^p_0} \int_{[0,t_1]^n} \sum_{\tau \in P} \int_{\tau_h} |\Delta_h^e u(x)|^p dx \, dh \\
\lesssim \frac{1}{t^p_0} \int_{[0,t_1]^n} \int_{\Omega_h} |\Delta_h^e u(x)|^p dx \, dh \\
= \frac{t^p_0}{t^p_0} w_r(u, t_1, \Omega)^p.
\]

The proof is completed upon using the equivalence (98) for \( G = \Omega \).

### 3.3 Multilevel approximation spaces

In this subsection, we study approximation from uniformly refined Lagrange finite element spaces. We keep the setting of the preceding subsection intact, and define the partitions \( P_j \) for \( j = 1, 2, \ldots \) recursively as \( P_{j+1} \) is the uniform refinement of \( P_j \). Let \( G \subset \Omega \) be a domain consisting of elements from some \( P_j \). More precisely, let \( G \) be the interior of \( \bigcup_{\tau \in Q} \bar{\tau} \) for some \( Q \subset P_j \) and \( j \). Then with \( S_j = S_{P_j} \), and \( 0 < p \leq \infty \), we let

\[
E(u, S_j)_{L^p(G)} = \inf_{v \in S_j} \|u - v\|_{L^p(G)}, \quad u \in L^p(G).
\]

Note that the infimum is achieved since \( S_j \) is a finite dimensional space. We define the multilevel approximation spaces

\[
A^\alpha_{p,q}(\{S_j\}, G) = \left\{ u \in L^p(G) : |u|_{A^\alpha_{p,q}(G)} := \left\| \left( \lambda^{\alpha} E(u, S_j)_{L^p(G)} \right)_{j \geq 0} \right\|_{\ell^2_q} < \infty \right\},
\]

for \( 0 < p, q \leq \infty \), and \( \alpha > 0 \), where \( \lambda = 2 \) for red refinements and \( \lambda = \sqrt{2} \) for newest vertex bisections. We will also use the shorthand notations

\[
A^\alpha_{p,q}(G) = A^\alpha_{p,q,m}(G) = A^\alpha_{p,q}(\{S_j\}, G).
\]

These spaces are quasi-Banach spaces with the quasi-norms \( \| \cdot \|_{L^p(G)} + | \cdot |_{A^\alpha_{p,q}(G)} \). Since \( \Omega \) is bounded, it is clear that \( A^\alpha_{p,q}(\Omega) \hookrightarrow A^\alpha_{p',q}(\Omega) \) for any \( \alpha \geq 0 \), \( 0 < q \leq \infty \) and \( \infty \geq p > p' > 0 \). We also have the lexicographical ordering: \( A^\alpha_{p,q}(\Omega) \hookrightarrow A^{\alpha'}_{p,q}(\Omega) \) for \( \alpha > \alpha' \) with any \( 0 < q, q' \leq \infty \), and \( A^\alpha_{p,q}(\Omega) \hookrightarrow A^{\alpha}_{p,q'}(\Omega) \) for \( 0 < q < q' \leq \infty \).

It is no coincidence that the aforementioned embedding relations are identical to those among Besov spaces. To help digesting the following theorem, recall that \( B^\alpha_{p,q;m+1}(\Omega) \) is the classical Besov space \( B^\alpha_{p,q}(\Omega) \) for \( \alpha < m + \max\{1, \frac{1}{p}\} \).

**Theorem 3.4.** We have \( B^\alpha_{p,q;m+1}(\Omega) \hookrightarrow A^\alpha_{p,q;m}(\Omega) \) for \( 0 < p, q \leq \infty \), and \( \alpha > 0 \). In the other direction, we have \( A^\alpha_{p,q;m}(\Omega) \hookrightarrow B^\alpha_{p,q;m+1}(\Omega) \) for \( 0 < p, q \leq \infty \), and \( 0 < \alpha < 1 + \frac{1}{p} \).
Proof. We follow the standard approach. The inclusion \( B^\alpha_{p,q,m+1}(\Omega) \rightarrow A^\alpha_{p,q,m}(\Omega) \) is a direct consequence of (93) with \( p_0 \leq p \) and the norm equivalence (65).

For the second part, we start with the estimate

\[
\omega_{m+1}(\phi, t)_p \lesssim \lambda^{-jn/p} \min\{1, (\lambda^j t)^{1+1/p}\},
\]

(105)

which holds for all nodal basis functions \( \phi \) of \( S_j \) and for all \( j \geq 0 \). This is Proposition 4.7 in Gaspoz and Morin (2013), which also holds for \( \omega \) of finite elements and the estimate \( \|\phi\|_{L_p} \approx \lambda^{-jn/p} \). The same ingredients are used to perform the corresponding computation for \( p = \infty \), as

\[
\omega_{m+1}(u_j, t)_\infty \lesssim \max z |b_z| \omega_{m+1}(\phi, t)_\infty \lesssim \max z |b_z| \min\{1, \lambda^j t\}
\]

(107)

where we have used the finite overlap property of the nodal basis functions, the \( L_p \) stability of finite elements and the estimate \( \|\phi\|_{L_p} \approx \lambda^{-jn/p} \). The same ingredients are used to perform the corresponding computation for \( p = \infty \), as

\[
\omega_{m+1}(u_j, t)_\infty \lesssim \max z |b_z| \omega_{m+1}(\phi, t)_\infty \lesssim \max z |b_z| \min\{1, \lambda^j t\}
\]

(107)

Now we write \( u = \sum_{j \geq 0} (u_j - u_{j-1}) \) with \( u_j \in S_j \) a best approximation to \( u \) from \( S_j \) for \( j \geq 0 \) and \( u_{-1} = 0 \). Note that the series converges in \( L_p \) by (93). We have

\[
\omega_{m+1}(u, \lambda^{-k})_p \lesssim \sum_{j \geq 0} \omega_{m+1}(u_j - u_{j-1}, \lambda^{-k})_p
\]

(108)

\[
\lesssim \sum_{j=0}^k \lambda^{(j-k)(1+1/p)} \|u_j - u_{j-1}\|_{L_p(\Omega)} + \sum_{j=k+1}^\infty \|u_j - u_{j-1}\|_{L_p(\Omega)},
\]

and an application of the discrete Hardy inequality (Lemma 2.12) gives

\[
|u|_{B^\alpha_{p,q,m+1}} \lesssim \left\| \left( \lambda^j \|u_j - u_{j-1}\|_{L_p(\Omega)} \right)_j \right\|_{\ell_q},
\]

(109)

for \( 0 < p, q \leq \infty \), and \( 0 < \alpha < 1 + \frac{1}{p} \). Finally, to go from \( u_j - u_{j-1} \) to \( u - u_j \) in the right hand side, we can apply the triangle inequality to \( u_j - u_{j-1} = (u - u_{j-1}) - (u - u_j) \).

Notice the gap between the two inclusions: While \( B^\alpha_{p,q,m+1}(\Omega) \rightarrow A^\alpha_{p,q,m}(\Omega) \) holds for all \( \alpha > 0 \), the reverse inclusion is proved only for \( 0 < \alpha < 1 + \frac{1}{p} \). In fact, if \( \alpha \geq 1 + \frac{1}{p} \) and \( p < \infty \), the forward inclusion is strict: Any function from \( S_j \) would be an element of all \( A^\alpha_{p,q,m}(\Omega) \), but there are functions in \( S_j \) that are not in \( B^\alpha_{p,q,m+1}(\Omega) \), because the
estimate (105) is saturated for small $t$. This leads to the expectation that for large $\alpha$, the difference $A^\alpha_{p,q,m}(\Omega) \setminus B^\alpha_{p,q,m+1}(\Omega)$ should be “skewed” considerably depending on the initial mesh $P_0$. We will not pursue this issue here, but we conjecture that the Besov space $B^\alpha_{p,q,m+1}(\Omega)$ coincides with the intersection of all $A^\alpha_{p,q,m}(\Omega)$ as one considers all possible initial triangulations $P_0$.

We quote the following standard result, in order to assure the reader of the fact that the multilevel approximation spaces $A^\alpha_{p,q}(\Omega)$ coincide with the spaces $\hat{B}^\alpha_{p,q}(\Omega)$ considered in Gaspoz and Morin (2013), cf. Definition 7.1 and Corollary 4.14 therein.

**Theorem 3.5.** Let $p_0 \leq p \leq \infty$, $0 < q \leq \infty$ and $\alpha > 0$. Then we have

$$
|u|_{A^\alpha_{p,q}(\Omega)} \sim \left\| \left( \lambda^{j\alpha} \| u - \tilde{Q}_j u \|_{L_p(\Omega)} \right)_{j \geq 0} \right\|_{\ell_q}
$$

$$
\sim \left\| \left( \lambda^{j\alpha} \| \tilde{Q}_{j+1} u - \tilde{Q}_j u \|_{L_p(\Omega)} \right)_{j \geq 0} \right\|_{\ell_q},
$$

(110)

for $u \in L_p(\Omega)$, where we have used the abbreviation $\tilde{Q}_j = \tilde{Q}_{P_j}$ for all $j$.

**Proof.** The first equivalence is immediate from (87). The generalized triangle inequality

$$
\| \tilde{Q}_{j+1} u - \tilde{Q}_j u \|_{L_p(\Omega)} \lesssim \| u - \tilde{Q}_j u \|_{L_p(\Omega)} + \| u - \tilde{Q}_{j+1} u \|_{L_p(\Omega)},
$$

(111)

implies one of the directions of the second equivalence, while the other direction follows...
from applying the discrete Hardy inequality (Lemma 2.12) to
\[
\|u - Q_j u\|_{L_p(\Omega)} \leq \left( \sum_{k=j}^{\infty} \|Q_{k+1} u - Q_k u\|_{L_p(\Omega)}^{p^*} \right)^{1/p},
\]
where \(p^* = \min\{1, p\} \).

The following technical result will be used later.

**Theorem 3.6.** Let \(0 < \alpha_1 < \alpha_2 < \infty\) and \(0 < p, q_1, q_2 \leq \infty\). Then we have
\[
[A_{p,q_1}^{\alpha_1}(G), A_{p,q_2}^{\alpha_2}(G)]_{\theta,q} = A_{p,q}^{\alpha}(G),
\]
for \(\alpha = (1 - \theta)\alpha_1 + \theta\alpha_2\) and \(0 < \theta < 1\), with the equivalence constants of quasi-norms depending only on the parameters \(\alpha, \alpha_1, \alpha_2, p, q, q_1\) and \(q_2\).

**Proof.** The equivalence (113) is standard, but we want to keep track of the equivalence constants. So we sketch a proof here. First, for \(v \in S_m\), we observe the inverse inequality
\[
|v|_{A_{p,q_2}^{\alpha_2}(G)}^{q_2} = \sum_{j=0}^{m-1} \lambda^{\alpha_2 q_2 j} E(v, S_j, G)^{q_2} \leq \|v\|_{L_p(G)}^{q_2} \sum_{j=0}^{m-1} \lambda^{\alpha_2 q_2 j} \leq \frac{\lambda^{\alpha_2 q_2 m}}{\lambda^{\alpha_2 q_2} - 1} \|v\|_{L_p(G)}^{q_2}.
\]
It is also true for \(q_2 = \infty\):
\[
|v|_{A_{p,\infty}^{\alpha_2}(G)} = \max_{0 \leq j < m} \lambda^{\alpha_2 j} E(v, S_j, G) \leq \lambda^{\alpha_2 m} \|v\|_{L_p(G)}.
\]
Another fact we will need is the following. We have the generalized triangle inequality
\[
|u + v|_{A_{p,q_2}^{\alpha_2}(G)} \leq c|u|_{A_{p,q_2}^{\alpha_2}(G)} + c|v|_{A_{p,q_2}^{\alpha_2}(G)},
\]
with \(c \geq 1\) depending only on \(p\) and \(q_2\). Then the Aoki-Rolewicz theorem (Bergh and L"ofstr"om, 1976, page 59) implies that
\[
|v_1 + \ldots + v_k|_{A_{p,q_2}^{\alpha_2}(G)}^{\mu} \leq 2|v_1|_{A_{p,q_2}^{\alpha_2}(G)}^{\mu} + \ldots + 2|v_k|_{A_{p,q_2}^{\alpha_2}(G)}^{\mu},
\]
for any \(v_1, \ldots, v_k \in A_{p,q_2}^{\alpha_2}(G)\), with \(\mu\) given by \((2c)^{\mu} = 2\).

With the abbreviation \(K(u, t) = K(u, t; A_{p,q_1}^{\alpha_1}(G), A_{p,q_2}^{\alpha_2}(G))\), for \(u \in A_{p,q_1}^{\alpha_1}(G)\), we have
\[
K(u, \lambda^{-(\alpha_2 - \alpha_1)m}) \leq |u - u_m|_{A_{p,q_1}^{\alpha_1}(G)} + \lambda^{-(\alpha_2 - \alpha_1)m} |u_m|_{A_{p,q_2}^{\alpha_2}(G)},
\]
where \( u_m \in S_m \) is an approximation satisfying \( \|u-u_m\|_{L_p(G)} = E(u, S_m, G)_p \). We estimate the first term in the right hand side as

\[
|u - u_m|_{A_{p,q_1}^q(G)} = \sum_{j=0}^{m} \lambda^{\alpha_1 q_1 j} \|u - u_m\|_{L_p(G)}^{q_1} + \sum_{j=m+1}^{\infty} \lambda^{\alpha_1 q_1 j} \|u - u_j\|_{L_p(G)}^{q_1} \lesssim \sum_{j=m}^{\infty} \lambda^{\alpha_1 q_1 j} \|u - u_j\|_{L_p(G)}^{q_1},
\]

(119)

with the implicit constant depending only on \( \lambda^{\alpha_1 q_1} \), and the second term as

\[
|u_m|^\mu_{A_{p,q_2}^q(G)} \lesssim 2 \sum_{j=1}^{m} |u_j - u_{j-1}|_{A_{p,q_2}^q(G)}^{\mu} \lesssim \sum_{j=1}^{m} \lambda^{\alpha_2 \mu j} \|u_j - u_{j-1}\|_{L_p(G)}^{\mu} \lesssim \sum_{j=0}^{m} \lambda^{\alpha_2 \mu j} \|u - u_j\|_{L_p(G)}^{\mu},
\]

(120)

where we have used the \( \mu \)-triangle inequality (117) in the first step, the inverse estimate (114) in the second step, and the (generalized) triangle inequality for the \( L_p \)-quasi-norm in the third step. Note that the implicit constants depend only on \( \lambda^{\alpha_2 q_2}, \lambda^{\alpha_2 \mu}, \) and \( p \). Putting everything together, we have

\[
K(u, \lambda^{-(\alpha_2 - \alpha_1)m}) \lesssim \left( \sum_{j=m}^{\infty} \lambda^{\alpha_1 q_1 j} \|u - u_j\|_{L_p(G)}^{q_1} \right)^{\frac{1}{q_1}} + \lambda^{-(\alpha_2 - \alpha_1)m} \left( \sum_{j=0}^{m} \lambda^{\alpha_2 \mu j} \|u - u_j\|_{L_p(G)}^{\mu} \right)^{\frac{1}{\mu}},
\]

(121)

and then the discrete Hardy inequalities (Lemma 2.12) give

\[
\left\| \|\lambda^m K(u, \lambda^{-(\alpha_2 - \alpha_1)m})\|_{L_p(G)} \right\|_{L_q(G)} \lesssim \left\| \|\lambda^{(\alpha_1 + \gamma)m} u - u^j\|_{L_p(G)} \|_{L_q(G)}^{m \geq 0} \right\|_{L_q(G)},
\]

(122)

for \( 0 < \gamma < \alpha_2 - \alpha_1 \). The left hand side of this inequality is the (quasi) norm for \([A_{p,q_1}^q(G), A_{p,q_2}^q(G)]_{\gamma/(\alpha_2 - \alpha_1), q}\), while the right hand side is the (quasi) norm for \( A_{p,q}^q(G) \).

For the other direction, we start with

\[
\|u - u_j\|_{L_p(G)} \lesssim \|u - v - w_j\|_{L_p(G)} \lesssim \|u - v\|_{L_p(G)} + \|v - w_j\|_{L_p(G)},
\]

(123)

where \( u \in A_{p,q_1}^q(G) \) and \( u_j \in S_j \) are as before, and \( v \in A_{p,q_2}^q(G) \), \( v_j, w_j \in S_j \) are arbitrary. Note that the implicit constant depends only on \( p \). Optimizing over \( v_j \) and \( w_j \) gives

\[
\min_{w_j \in S_j} \|u - v - w_j\|_{L_p(G)} \leq \lambda^{\alpha_1 j} \|u - v\|_{A_{p,q_1}^q(G)},
\]

(124)
and
\[ \min_{v_j \in S_j} \|v - v_j\|_{L_p(G)} \leq \lambda^{-\alpha_2 j} |v|_{A_{p,q_2}^2(G)}, \quad (125) \]
and substituting these back, we get
\[ \|u - u_j\|_{L_p(G)} \lesssim \inf_{v \in A_{p,q_2}^2(G)} \left( \lambda^{-\alpha_1 j} |u - v|_{A_{p,q_1}^1(G)} + \lambda^{-\alpha_2 j} |v|_{A_{p,q_2}^2(G)} \right) \]
\[ = \lambda^{-\alpha_1 j} K(u, \lambda^{-(\alpha_2 - \alpha_1) j}). \quad (126) \]
The proof is completed upon recalling the definition of $| \cdot |_{A_{p,q}^s(G)}$.

### 3.4 Adaptive approximation

In this subsection, we consider the approximation problem from adaptively generated Lagrange finite element spaces. We study various approximation classes associated to the finite element spaces $S_P$, cf. (69). In Binev et al. (2002); Gaspoz and Morin (2013), among other things, it is proved that $B_{q,q}^s(\Omega) \hookrightarrow A_{s}^\infty(L_p(\Omega))$ with $s = \frac{\alpha}{n}$, as long as $\frac{\alpha}{n} + \frac{1}{p} - \frac{1}{q} > 0$ and $0 < \alpha < m + \max \{ 1, \frac{1}{q} \}$. In the other direction, the same references give $A_{p,q}^s(L_p(\Omega)) \hookrightarrow B_{q,q}^s(\Omega)$ for $s = \frac{\alpha}{n} = \frac{1}{q} - \frac{1}{p} > 0$ and $0 < p, q < \infty$. Below we complement these results by establishing direct embeddings of the form $A_{p,q}^s(\Omega) \hookrightarrow A_{s}^\infty(L_p(\Omega))$. This is a genuine improvement, since $A_{p,q}^s(\Omega) \supseteq B_{q,q}^s(\Omega)$ for $\alpha \geq 1 + \frac{1}{q}$. Moreover, it seems natural to relate adaptive approximation to multilevel approximation first, and then bring in the relationships between multilevel approximation and Besov spaces. We also remark that while the existing results are only for the newest vertex bisection, we take into account the red refinement procedure as well.

**Theorem 3.7.** Let $0 < q \leq p \leq \infty$ and $\alpha > 0$ satisfy $\frac{\alpha}{n} + \frac{1}{p} - \frac{1}{q} > 0$ and $q < \infty$. Then for any $0 < p_0 < q$ (Recall that $Q_P$ depends on $p_0$), we have
\[ \|u - \tilde{Q}_P u\|_{L_p(\Omega)} \lesssim \left( \sum_{\tau \in P} |\tau|^{p \delta} |u|_{A_{p,q}^{\delta}(\hat{\tau})}^p \right)^{\frac{1}{p}}, \quad u \in A_{p,q}^\alpha(\Omega), \quad P \in \mathcal{P}, \quad (127) \]
where $\delta = \frac{\alpha}{n} + \frac{1}{p} - \frac{1}{q}$. In particular, we have $A_{q,q}^\alpha(\Omega) \hookrightarrow A_{s}^\infty(L_p(\Omega))$ with $s = \frac{\alpha}{n}$.

**Proof.** We have the sub-additivity property
\[ \sum_k \|u|_{A_{q,q}^{\alpha}(P_k(\tau_k))}^q \lesssim \|u|_{A_{q,q}^{\alpha}(\Omega)}^q, \quad (128) \]
for $0 < q < \infty$ and for any finite sequences $\{P_k\} \subset \mathcal{P}$ and $\{\tau_k\}$, with $\tau_k \in P_k$ and $\{\tau_k\}$ non-overlapping. Recall that $\hat{\tau} = P(\tau)$ is the interior of $\bigcup \{ \sigma : \sigma \in P, \sigma \cap \tau \neq \emptyset \}$. Therefore the estimate (127) would imply the second statement by Theorem 2.2.
3 Lagrange finite elements

We shall prove (127). Every element \( \tau \in P \) of any partition \( P \in \mathcal{P} \) is an element of a unique \( P_j \), with the number \( j \) counting how many refinements one needs in order to arrive at \( \tau \). We call \( j \) the generation or the level of \( \tau \), and write \( j = [\tau] \). We will also need \( j(\tau) = \min\{ [\sigma] : \sigma \in P, \sigma \cap \tau \neq \emptyset \} \). Note that \( |\tau| \sim \lambda^{-n[\tau]} \sim \lambda^{-n j(\tau)} \) and \( S_{j(\tau)} \mid \hat{\tau} \subset S_P \mid \hat{\tau} \).

By invoking (86), we infer

\[
\| u - \hat{Q}_P u \|_{L_p(\Omega)}^p = \sum_{\tau \in P} \| u - \hat{Q}_P u \|_{L_p(\tau)}^p \lesssim \sum_{\tau \in P} \inf_{v \in \mathcal{S}_P} \| u - v \|_{L_p(\tau)}^p \leq \sum_{\tau \in P} \| u - u_j(\tau) \|_{L_p(\tau)}^p,
\]

(129)

where \( u_j \in S_j \ (j \geq 0) \) is an approximation (that may depend on \( \tau \)) satisfying

\[
\| u - u_j \|_{L_q(\hat{\tau})} \leq c E(u, S_j)_{L_q(\hat{\tau})},
\]

(130)

with some constant \( c \geq 1 \). The same is true for \( p = \infty \) with obvious modifications. For an individual term in the right hand side, with \( p^* = \min\{1, \infty \} \), we have

\[
\| u - u_j(\tau) \|_{L_p(\hat{\tau})}^{p^*} \leq \sum_{j=j(\tau)}^{\infty} \| u_{j+1} - u_j \|_{L_p(\hat{\tau})}^{p^*} \lesssim \sum_{j=j(\tau)}^{\infty} \lambda^{(\frac{q}{p} - \frac{1}{p})jnp^*} \| u_{j+1} - u_j \|_{L_q(\hat{\tau})}^{p^*}
\]

\[
\lesssim \sum_{j=j(\tau)}^{\infty} \lambda^{(\frac{q}{p} - \frac{1}{p})jnp^*} \| u - u_j \|_{L_q(\hat{\tau})}^{p^*},
\]

(131)

where we have estimated \( u - u_j(\tau) \) as a telescoping sum in the first line, and used the estimate \( \lambda^{jn/p}\| v \|_{L_p(\hat{\tau})} \sim \lambda^{jn/q}\| v \|_{L_q(\hat{\tau})} \) for \( v \in S_{j+1} \) in the second line. We continue by noting the relation \( \frac{1}{q} - \frac{1}{p} = \frac{\alpha}{n} - \delta \), which yields

\[
\| u - u_j(\tau) \|_{L_p(\hat{\tau})}^{p^*} \lesssim \sum_{j=j(\tau)}^{\infty} \lambda^{-j\delta np^*} \lambda^{j\alpha p^*} \| u - u_j \|_{L_q(\hat{\tau})}^{p^*}
\]

\[
\leq \lambda^{-j(\tau)\delta np^*} \sum_{j=j(\tau)}^{\infty} \lambda^{j\alpha p^*} \| u - u_j \|_{L_q(\hat{\tau})}^{p^*}
\]

\[
\lesssim |\tau|^{\delta p^*} |u|_{A^0_{q,p^*}(\hat{\tau})}^{p^*},
\]

(132)

by (130). This establishes the theorem for \( q \leq 1 \), in which case we have \( A^0_{q,q^*}(\hat{\tau}) \subset A^0_{q,p^*}(\hat{\tau}) \).

If \( q > 1 \), choose \( 0 < \alpha_1 < \alpha < \alpha_2 \) satisfying \( \alpha = \frac{\alpha_1 + \alpha_2}{2} \) and \( \delta_i = \frac{\alpha_i}{n} + \frac{1}{p} - \frac{1}{q} > 0 \) for \( i = 1, 2 \). Moreover, we put \( u_j = Q_{P_j}^{(\tau)} u \), where \( Q_{P_j}^{(\tau)} : L_1(\hat{\tau}) \rightarrow S_j \) is the quasi-interpolation operator defined in (73), with \( \hat{\tau} \) playing the role of \( \Omega \). Then Lemma 3.1 guarantees the
property (130) with $c$ depending only on global geometric properties of $\mathcal{P}$. In particular, $c$ is bounded independently of $\tau$. Thus (132) gives

$$\|u - u_j(\tau)\|_{L^p(\hat{\tau})} \lesssim |\tau|^{\delta_i} |u|_{A_{q,i,1}^\alpha(\hat{\tau})},$$

(133)

for $i = 1, 2$. Since the operators $Q_{\hat{\tau}}^{(\tau)} P_j$ are linear, so is the map $u \mapsto u - u_j(\tau)$, and hence interpolation and Theorem 3.6 yield

$$\|u - u_j(\tau)\|_{L^p(\hat{\tau})} \lesssim |\tau|^{\delta_i} |u|_{A_{q,i,1}^\alpha(\hat{\tau})} / 2 \lesssim |\tau|^{\delta_i} |u|_{A_{q,i,1}^\alpha(\hat{\tau})},$$

(134)

with the implicit constants depending only on global geometric properties of $\mathcal{P}$ and on the indices of the spaces involved. This completes the proof.

Fig. 4: Illustration of Theorem 3.7 and Theorem 3.8.

Now we look at adaptive approximation in the space $A_{p,p}^\sigma(\Omega)$. Recall from Gaspoz and Morin (2013) that $\mathbb{G}^s_q(A_{p,p}^\sigma(\Omega)) \hookrightarrow A_{q,q}^\alpha(\Omega)$ for $s = \frac{\alpha - \sigma}{n} = \frac{1}{q} - \frac{1}{p} > 0$ and $0 < p, q < \infty$.

**Theorem 3.8.** Let $0 < q \leq p \leq \infty$, and $\alpha, \sigma > 0$ satisfy $\frac{\alpha - \sigma}{n} + \frac{1}{p} - \frac{1}{q} > 0$ and $q < \infty$. Then for any $0 < p_0 < q$, we have

$$\|u - \tilde{Q}_P u\|_{A_{p,p}^\sigma(\Omega)} \lesssim \left( \sum_{\tau \in P} |\tau|^{\delta_i} |u|_{A_{q,i,1}^\alpha(\hat{\tau})}^p \right)^{\frac{1}{p}} , \quad u \in A_{q,q}^\alpha(\Omega), \quad P \in \mathcal{P},$$

(135)

with $\delta = \frac{\alpha - \sigma}{n} + \frac{1}{p} - \frac{1}{q}$. In particular, we have $A_{q,q}^\alpha(\Omega) \hookrightarrow \mathbb{G}^s_q(A_{p,p}^\sigma(\Omega))$ with $s = \frac{\alpha - \sigma}{n}$. 
Proof. With \( v = u - \tilde{Q}_P u \) and \( \tilde{Q}_j = \tilde{Q}_{P_j} \), we have

\[
\|v\|_{A^p_{Q,\tau}(\Omega)} \leq \left( \sum_{j \geq 0} \lambda^{j\sigma p} \|v - \tilde{Q}_j v\|_{L^p(\Omega)}^{p} \right)^{\frac{1}{p}} = \left( \sum_{\tau \in P} \sum_{j \geq 0} \lambda^{j\sigma p} \|v - \tilde{Q}_j v\|_{L^p(\tau)}^{p} \right)^{\frac{1}{p}}, \tag{136}
\]

with the usual modification for \( p = \infty \). Let \( j(\tau) = \max\{[\sigma] : \sigma \in P, \sigma \cap \tau \neq \emptyset\} \) for \( \tau \in P \), with \([\sigma]\) denoting the generation number (or the level) of \( \sigma \). Then for \( \tau \in P \) and \( j \geq j(\tau) \) we have \( S_{\tau} \mid M \subseteq S_j \mid \tau \), and hence

\[
\tilde{Q}_j (u - \tilde{Q}_P u) = \tilde{Q}_j u - \tilde{Q}_j u \quad \text{on} \quad \tau, \tag{137}
\]

by the linearity property (84). This implies that \( v - \tilde{Q}_j v = u - \tilde{Q}_j u \) on \( \tau \), for all \( j \geq j(\tau) \).

Now, proceeding exactly as in the preceding proof, with \( p^* = \min\{1, p\} \), we infer

\[
\|u - \tilde{Q}_j u\|_{L^p(\tau)}^{p^*} \leq \sum_{k=j}^{\infty} \lambda^{\frac{(1-\frac{1}{p})}{\frac{1}{p}} k^np^*} \|\tilde{Q}_k u - \tilde{Q}_k u\|_{L^p(\tau)}^{p^*} \leq \sum_{k=j}^{\infty} \lambda^{\frac{(1-\frac{1}{p})}{\frac{1}{p}} k^np^*} \|u - \tilde{Q}_k u\|_{L^p(\tau)}^{p^*} \tag{138}
\]

Then the discrete Hardy inequality yields

\[
\sum_{j \geq j(\tau)} \lambda^{j\sigma q} \|u - \tilde{Q}_j u\|_{L^p(\tau)}^q \leq \sum_{k \geq j(\tau)} \lambda^{k\sigma q} \lambda^{\frac{(1-\frac{1}{p})}{\frac{1}{p}} k^np^*} \|u - \tilde{Q}_k u\|_{L^q(\tau)}^q \leq \lambda^{-\delta n q j(\tau)} \sum_{k \geq j(\tau)} \lambda^{k\sigma q} \|u - \tilde{Q}_k u\|_{L^q(\tau)}^q \tag{139}
\]

where we have taken into account the relation \( \frac{\sigma}{n} + \frac{1}{q} - \frac{1}{p} = \frac{\sigma}{n} - \delta \). Notice that the discrete Hardy inequality made the use of interpolation unnecessary, to compare the present arguments with the proof of the preceding theorem. This takes care of one of the sums (or maximums) when we split the sum in the right hand side of (136) into two sums according to \( j < j(\tau) \) or \( j \geq j(\tau) \). We rewrite the other sum (or maximum) as

\[
\left( \sum_{\tau \in P} \sum_{j \geq j(\tau)} \lambda^{j\sigma p} \|v - \tilde{Q}_j v\|_{L^p(\tau)}^p \right)^{\frac{1}{p}} = \left( \sum_{j \geq 0} \sum_{\{\tau \in P : j(\tau) < j\}} \lambda^{j\sigma p} \|v - \tilde{Q}_j v\|_{L^p(\tau)}^p \right)^{\frac{1}{p}} \tag{140}
\]

\[
= \left( \sum_{j \geq 0} \lambda^{j\sigma p} \|v - \tilde{Q}_j v\|_{L^p(\Omega_j)}^p \right)^{\frac{1}{p}},
\]
where \( \Omega_j = \bigcup \{ \tau \in P : j(\tau) > j \} \). Note that \( \Omega_j \supset \Omega_j^0 \) with \( \Omega_j^0 = \{ \tau \in P : [\tau] > j \} \), and that \( \Omega_j^0 \) consists of triangles from \( P_j \), in the sense that there is \( R_j^0 \subset P_j \) such that \( \Omega_j^0 = \bigcup \{ \tau \in R_j^0 \} \) up to a zero measure set. The triangles \( \tau \in P \) with \( \tau \cap \Omega_j^0 \) are at the level \( j \) or less, and \( \Omega_j \setminus \Omega_j^0 \) consists of precisely those triangles that have a neighbour in \( \Omega_j^0 \). Hence there is \( R_j \subset P_j \) such that \( \Omega_j = \bigcup \{ \tau \in R_j \} \) up to a zero measure set. Now, by the stability property (85), we get

\[
\|v - \hat{Q}_j v\|_{L^p(\Omega_j)} \lesssim \|v\|_{L^p(\Omega_j)} + \|\hat{Q}_j v\|_{L^p(\Omega_j)} \lesssim \|v\|_{L^p(\hat{\Omega}_j)},
\]

where \( \hat{\Omega}_j = \bigcup \{ \tilde{\tau} : \tau \in R_j \} \) with \( \tilde{\tau} = P_j(\tau) \). Obviously, \( \hat{\Omega}_j \) is a subset of \( \hat{\Omega}_j^0 = \bigcup \{ \tau \in P : \tilde{\tau} \cap \Omega_j \neq \emptyset \} \), that can also be described as \( \hat{\Omega}_j^0 = \bigcup \{ \tau \in P : j^2(\tau) > j \} \), with \( j^2(\tau) = \max \{ j(\sigma) : \sigma \in P, \sigma \cap \tilde{\tau} \neq \emptyset \} \) for \( \tau \in P \). All this yields

\[
\left( \sum_{\tau \in P \setminus \{ j \leq j(\tau) \}} \sum_{i \in \mathbb{N}} \lambda_i^{j(\tau)} \|v - \hat{Q}_j v\|_{L^p(\tau)}^p \right)^{1/p} \lesssim \left( \sum_{\tau \in P \setminus \{ j \leq j(\tau) \}} \sum_{i \in \mathbb{N}} \lambda_i^{j(\tau)} \|u - \hat{Q}_j u\|_{L^p(\tau)}^p \right)^{1/p} \lesssim \left( \sum_{\tau \in P} \|\tau\|^{\sigma_p/\alpha} \|u - \hat{Q}_j u\|_{L^p(\tau)}^p \right)^{1/p},
\]

where we have taken into account the geometric growth of \( \lambda_i^{j(\tau)} \) in \( \tilde{\tau} \), and the fact that \( \lambda_i^{j(\tau)} \sim |\tau|^{1/n} \). Then once we recall from the proof of the preceding theorem that

\[
\|u - \hat{Q}_j u\|_{L^p(\tau)} \lesssim |\tau|^{-1/\alpha} \|u\|_{A^\alpha_{p,q}(\tau)},
\]

with \( \delta' = \frac{\alpha}{n} + \frac{1}{p} - \frac{1}{q} = \delta - \frac{\alpha}{n} \), the proof is complete.

### 3.5 Discontinuous piecewise polynomials

All that has been said on multilevel and adaptive approximation for continuous Lagrange finite elements have analogues in the world of discontinuous polynomials subordinate to triangulations. The theory is in fact much simpler due to the absence of the continuity requirement across elements. Thus we will state here the relevant results and only sketch or omit the proofs.

The notations \( \mathcal{P}, \{ P_j \} \), etc., will mean the same things as before. For \( P \in \mathcal{P} \), let

\[
\hat{S}_P = \hat{S}_P^d = \{ v \in L^\infty(\Omega) : v|_\tau \in P_{\delta} \forall \tau \in P \},
\]

where \( d \) is a nonnegative integer, and let \( \hat{S}_j = \hat{S}_{P_j} \) for all \( j \). Then with \( G \subset \Omega \) a domain consisting of elements from some \( P_j \), we define the multilevel approximation spaces \( \hat{A}^\alpha_{p,q}(\{ \hat{S}_j \}, G) \) by (103), with the sequence \( \{ \hat{S}_j \} \) replacing \( \{ S_j \} \). We will also use the shorthand notations

\[
\hat{A}^\alpha_{p,q}(G) = \hat{A}^\alpha_{p,q,d}(G) = A^\alpha_{p,q}(\{ \hat{S}_j \}, G).
\]

The analogue of Theorem 3.4 is the following.
Theorem 3.9. We have $B_{p,q;d+1}^\alpha(\Omega) \hookrightarrow \tilde{A}_{p,q,d}^\alpha(\Omega)$ for $0 < p, q \leq \infty$, and $\alpha > 0$. In the other direction, we have $\tilde{A}_{p,q;d}^\alpha(\Omega) \hookrightarrow B_{p,q;d+1}^\alpha(\Omega)$ for $0 < p, q \leq \infty$, and $0 < \alpha < \frac{1}{p}$.

Note that due to the lack of continuity the inverse inclusion holds in a very small range of indices. We also have the analogue of Theorem 3.6.

Theorem 3.10. Let $0 < \alpha_1 < \alpha_2 < \infty$ and $0 < p, q, q_1, q_2 \leq \infty$. Then we have

$$\left[ \tilde{A}_{p,q_1}^{\alpha_1}(G), \tilde{A}_{p,q_2}^{\alpha_2}(G) \right]_{\theta,q} = \tilde{A}_{p,q}^\alpha(G),$$

for $\alpha = (1 - \theta)\alpha_1 + \theta\alpha_2$ and $0 < \theta < 1$, with the equivalence constants of quasi-norms depending only on the parameters $\alpha, \alpha_1, \alpha_2, p, q, q_1$ and $q_2$.

Finally, we want to record some results on adaptive approximation by discontinuous polynomials subordinate to the partitions in $\mathcal{S}$. Let us define the approximation class $\mathcal{A}_{q;d}(L_p^0(\Omega))$ exactly as $\mathcal{A}_q^s(L_p^0(\Omega))$, by replacing $S_p^0$ with $S_p^d$, and by using the distance function

$$\rho(u, v, P) = \|u - v\|_{L_p^0(\Omega)} = \left( \sum_{\tau \in P} |\tau| \frac{\rho_p}{n} \|u - v\|_{L_p^0(\tau)} \right)^\frac{1}{p},$$

with the obvious modification for $p = \infty$. It is for later reference that we have introduced the mesh dependent weight in the distance function. We write $\mathcal{A}_{q;d}(L_p^0(\Omega)) = \mathcal{A}_{q;d}^s(L_p^0(\Omega))$.

We have the following direct embedding result.

Theorem 3.11. Let $0 < q \leq p \leq \infty$, $\alpha > 0$ and $\theta \geq 0$ satisfy $\frac{\alpha}{n} + \frac{1}{p} - \frac{1}{q} > 0$ and $q < \infty$, with $\frac{\alpha}{n} + \frac{1}{p} - \frac{1}{q} = 0$ allowed if $\theta > 0$. Then we have $\tilde{A}_{q;d}(\Omega) \hookrightarrow \mathcal{A}_{q;d}^s(L_p^0(\Omega))$ with $s = \frac{\alpha + \theta}{n}$.

Proof. Let $u \in L_p^0(\Omega)$ and let $P \in \mathcal{S}$. Then with $\Pi_P$ the projection operator defined in (82) with $m := d$, we have

$$\|u - \Pi_P u\|_{L_p^0(\Omega)} = \left( \sum_{\tau \in P} |\tau| \frac{\rho_p}{n} \|u - \Pi_P u\|_{L_p^0(\tau)} \right)^\frac{1}{p} \lesssim \left( \sum_{\tau \in P} |\tau| \frac{\rho_p}{n} \sup_{v \in S_P} \|u - v\|_{L_p^0(\tau)} \right)^\frac{1}{p}. \tag{148}$$

Now proceeding exactly as in the proof of Theorem 3.7, we get

$$\|u - \Pi_P u\|_{L_p^0(\Omega)} \lesssim \left( \sum_{\tau \in P} |\tau| \frac{\rho_p}{n} |\tau|^{\delta_p} |u|_{\tilde{A}_{q;d}^\alpha(\tau)} \right)^\frac{1}{p}, \tag{149}$$

with $\delta = \frac{\alpha}{n} + \frac{1}{p} - \frac{1}{q}$. Then an application of Theorem 2.2 finishes the proof. □

We close this section by proving an inverse embedding theorem.

Theorem 3.12. Let $0 < q \leq p < \infty$, $\alpha, \theta > 0$, and let $s = \frac{\alpha + \theta}{n} = \frac{1}{q} - \frac{1}{p}$. Then we have $\mathcal{A}_{q;d}^s(L_p^0(\Omega)) \cap L_p(\Omega) \subset \tilde{A}_{q;d}^\alpha(\Omega)$. 

We shall apply Corollary 2.10 with $X_0 = L_p(\Omega)$, $X_1 = \tilde{A}^{-\theta}_{p,p,d}(\Omega)$, $X = \tilde{A}^{2\alpha}_{\gamma,\gamma,d}(\Omega)$ and $\|\cdot\|_P = \|\cdot\|_{L_p(\Omega)}$, where $\frac{1}{r} = \frac{1}{p} + \frac{\theta}{n}$ and $\frac{1}{\gamma} = \frac{2\alpha + \theta}{n} + \frac{1}{p}$. We will only need to apply the quasi-seminorm of $\tilde{A}^{-\theta}_{p,p,d}(\Omega)$ to functions from $L_p(\Omega)$, and the space $\tilde{A}^{-\theta}_{p,p,d}(\Omega)$ itself will remain undefined. Pick $p_0 < \gamma$, and let $\Pi_P$ be the projection operator defined in (82) with $m := d$. Then for $u \in \mathcal{A}^s_q(L_p(\Omega)) \cap L_p(\Omega)$, we have

$$|u - \Pi_P u|_{\tilde{A}^{-\theta}_{p,p,d}(\Omega)}^p = \sum_{j \geq 0} \lambda^{-j\theta p} \|w - \Pi_P w\|^p_{L_p(\Omega)}$$

$$= \sum_{\tau \in P} \sum_{j < [\tau]} \lambda^{-j\theta p} \|w - \Pi_P w\|^p_{L_p(\tau)} + \sum_{\tau \in P} \sum_{j \geq [\tau]} \lambda^{-j\theta p} \|w - \Pi_P w\|^p_{L_p(\tau)}, \quad (150)$$

where $w = u - \Pi_P u$. For the first term in the right hand side, applying the same argument as that starts around (140) in the proof of Theorem 3.8, we get

$$\sum_{\tau \in P} \sum_{j < [\tau]} \lambda^{-j\theta p} \|w - \Pi_P w\|^p_{L_p(\tau)} \lesssim \sum_{\tau \in P} \sum_{j \geq [\tau]} \lambda^{-j\theta p} \|w - \Pi_P w\|^p_{L_p(\tau)}, \quad (151)$$

For the second term, noting that $(\Pi_P \Pi_P u)_{\tau} = (\Pi_P u)_{\tau}$ for $j \geq [\tau]$, and using the linearity
property (81), we have
\[
\sum_{j \geq [\tau]} \lambda^{-j\theta_p} \|w - \Pi_j w\|_{L_p(\tau)}^p = \sum_{j \geq [\tau]} \lambda^{-j\theta_p} \|u - \Pi_j u\|_{L_p(\tau)}^p \\
\lesssim \|u - \Pi[\tau] u\|_{L_p(\tau)}^p \sum_{j \geq [\tau]} \lambda^{-j\theta_p} \tag{152}
\]
which confirms (56) with \(\delta = 0\).

To verify (57), let \(v \in S_P\). Then it is clear that \((\Pi_j v)|_{\tau} = v|_{\tau}\) for \(j \geq [\tau]\) and \(\tau \in P\). Using this, we have
\[
|v|^\gamma_{A_{\gamma,\gamma,d}^2}(\Omega) \leq \sum_{j \geq 0} \lambda^{2\alpha j\gamma} \|v - \Pi_j v\|_{L_{\gamma}(\Omega)}^\gamma = \sum_{j \in P} \sum_{\tau < [\tau]} \lambda^{2\alpha j\gamma} \|v - \Pi_j v\|_{L_{\gamma}(\tau)}^\gamma \\
\lesssim \sum_{\tau \in P} |\tau|^{-\frac{2\alpha + \theta}{p} - 1} \|v\|_{L_{\gamma}(\tau)}^\gamma \leq \sum_{\tau \in P} |\tau|^{-\frac{2\alpha + \theta}{p} - 1} \|v\|_{L_{\gamma}(\tau)}^\gamma = \sum_{\tau \in P} |\tau|^{-\frac{\sigma}{p}} \|v\|_{L_{\gamma}(\tau)}^\gamma \tag{153}
\]
\[
\leq \left( \sum_{\tau \in P} \left( \sum_{\tau \in P} |\tau|^{-\frac{\sigma}{p}} \|v\|_{L_{\gamma}(\tau)}^p \right)^{\frac{\gamma}{p}} \right) = (\# P)^{\gamma} \left( \sum_{\tau \in P} |\tau|^{-\frac{\sigma}{p}} \|v\|_{L_{\gamma}(\tau)}^p \right)^{\frac{\gamma}{p}},
\]
where \(\sigma = \frac{2\alpha + \theta}{n}\). Here we have used the stability of \(\Pi_j\) in the third step, the Hölder inequality in the fourth step, and the discrete Hölder inequality in the penultimate step.

Now an application of Corollary 2.10 (or rather, the proof of Theorem 2.9) yields
\[
\left\| \left[ 2^{mK} (u, 2^{-m\sigma}, A_{\gamma,\gamma,d}^2(\Omega), A_{\gamma,\gamma,d}^{2\alpha}(\Omega)) \right]_{m \geq 0} \right\|_{\ell_q} \lesssim \|u\|_{L_{\gamma,d}^{\sigma}(\Omega)}, \tag{154}
\]
where
\[
K'(u, t, A_{\gamma,\gamma,d}^{2\alpha}(\Omega), A_{\gamma,\gamma,d}^{2\alpha}(\Omega)) = \inf_{v \in A_{\gamma,\gamma,d}^{2\alpha}(\Omega)} \left( |u - v|_{A_{\gamma,\gamma,d}^{2\alpha}(\Omega)} + t|v|_{A_{\gamma,\gamma,d}^{2\alpha}(\Omega)} \right). \tag{155}
\]
On the other hand, we have
\[
[\ell_{\gamma,\gamma}^{-\theta}(X), \ell_{\gamma,\gamma}^{2\alpha}(Y)]_{\ell_q} = \ell_q^{\alpha}([X, Y]_{\ell_q}), \tag{156}
\]
with equivalent quasi-norms, if \(\alpha = -\theta(1 - \varepsilon) + 2\alpha \varepsilon\) and \(\frac{1}{q} = \frac{1-\varepsilon}{p} + \frac{\varepsilon}{q}\), where \(\ell_q^\alpha(Z)\) is the space of \(Z\)-valued sequences \(\{a_k\}\) satisfying
\[
\|(2^{\alpha k})a_k\|_{Z, k}\ell_q < \infty, \tag{157}
\]
4 Second order elliptic problems

4.1 Introduction

In this section, we will apply the abstract theory of Section 2 to second order elliptic boundary value problems. As far as the domain $\Omega$ and the family of triangulations $\mathcal{P}$ are concerned, we will keep the setting of the previous section intact. In particular, we fix a refinement rule, which is either the newest vertex bisection or the red refinement, and assume that the family $\mathcal{P}$ has the gradedness property (68).

Let $\Gamma \subset \partial \Omega$ be an open piece (or the whole) of the boundary, consisting of faces of the initial triangulation $P_0$. For $P \in \mathcal{P}$, the space $S_P$ will be the Lagrange finite element space of continuous piecewise polynomials of degree not exceeding $m$, with the homogeneous
Dirichlet condition on $\Gamma$. We also define $H^1_\Gamma(\mathbb{R}^n)$ as the closure of $\mathcal{D}(\mathbb{R}^n \setminus \Gamma)$ in $H^1(\mathbb{R}^n)$, and $H^1_\Gamma = H^1_\Gamma(\Omega)$ as the restriction of functions from $H^1_\Gamma(\mathbb{R}^n)$ to $\Omega$. Note that $S_P = H^1_\Gamma \cap S^m_P$.

The operator $T$ is given as

$$ Tu = -a_{ij} \partial_i \partial_j u + b_k \partial_k u + cu, \quad (163) $$

where the repeated indices are summed over. The coefficients $a_{ij}$ are Lipschitz continuous, and $b_k, c \in L_\infty(\Omega)$. The problem we consider is to find $u \in H^1_\Gamma$ satisfying

$$ (Tu, v) = (f, v), \quad \text{for all } v \in H^1_\Gamma. \quad (164) $$

Here $(\cdot, \cdot)$ is the duality pairing between $(H^1_\Gamma)'$ and $H^1_\Gamma$, and $f \in L_2(\Omega) \hookrightarrow (H^1_\Gamma)'$ is given. This is of course the variational formulation of the mixed Dirichlet-Neumann problem with the homogenous Dirichlet data on $\Gamma$. We will also denote by $T : H^1_\Gamma \rightarrow (H^1_\Gamma)'$ the operator defined in (164).

For $P \in \mathcal{P}$, let

$$ E_P = \{ \partial \tau \cap \partial \sigma \cap (\Omega \setminus \Gamma) : \tau, \sigma \in P \}, \quad (165) $$

be the set of edges not intersecting the Dirichlet piece $\Gamma$. Note that if $P$ is nonconforming, then only the edges of the “smaller” triangles go into $E_P$. Given $P \in \mathcal{P}$, $u \in T^{-1}(L_2(\Omega))$, and $v \in S_P$, define the element residual $r_\tau = (Tu - Tv)\tau$ for $\tau \in P$, and the edge residual $r_e \in L_2(e)$ for $e \in E_P$ as the jump of the normal component of the vector field $a_{ij} \partial_j v$ across the edge $e$. Finally, we define the a posteriori error estimator

$$ (\eta(u, v, P))^2 = \sum_{\tau \in P} h^2_\tau \| r_\tau \|_{L_2(\tau)}^2 + \sum_{e \in E_P} h_e \| r_e \|_{L_2(e)}^2. \quad (166) $$

A typical adaptive finite element method that uses (166) as its error indicator converges optimally with respect to the approximation classes $\mathcal{A}^s(\eta)$, in the sense that if the solution $u$ of the problem (164) satisfies $u \in \mathcal{A}^s(\eta)$ for some $s > 0$, then the adaptive method reduces the quantity $\eta(u, u_P, P)$ with the rate $s$, where $u_P$ is the Galerkin approximation of $u$ from $S_P$, cf. Feischl, Führer, and Praetorius (2014). Moreover, it is well known that the estimator (166) is equivalent to the total error

$$ (\rho_d(u, v, P))^2 = \| u - v \|_{H^1}^2 + (\text{osc}_d(u, v, P))^2, \quad (167) $$

when $v = u_P$ and for any fixed $d \geq m - 2$, where the oscillation is defined as

$$ (\text{osc}_d(u, v, P))^2 = \sum_{\tau \in P} h^2_\tau \| (1 - \Pi_\tau) r_\tau \|_{L_2(\tau)}^2 + \sum_{e \in E_P} h_e \| (1 - \Pi_e) r_e \|_{L_2(e)}^2, \quad (168) $$

with $\Pi_\tau : L_2(\tau) \rightarrow P_d$ and $\Pi_e : L_2(e) \rightarrow P_{d+1}$ being $L_2$-orthogonal projections onto polynomial spaces, see e.g., Nochetto, Siebert, and Veeser (2009). Optimality of adaptive finite element methods with respect to the approximation classes $\mathcal{A}^s(\rho_d)$ has also been
proved, cf. Cascon et al. (2008). Ideally, one would like to have optimality with respect to the classes $\mathcal{A}^s(H^1_\Gamma)$ that correspond to the energy error. In particular, it is conceivable that for certain functions $u$, the energy error $\|u - u_P\|_{H^1}$ decays faster than the oscillation $\text{osc}_d(u, u_P, P)$, so that the class $\mathcal{A}^s(H^1_\Gamma)$ is strictly smaller than both $\mathcal{A}^s(p_d)$ and $\mathcal{A}^s(\eta)$. However, if the error estimator (166) is the only source of information used by the algorithm in its stopping criterion (or in the marking of triangles for refinement), then it is clear that one has to reduce the oscillation anyway. It appears therefore that the approximation classes $\mathcal{A}^s(p_d)$ and $\mathcal{A}^s(\eta)$ are completely natural from the perspective of adaptive finite element methods.

4.2 A characterization of adaptive approximation classes

In this subsection, we give necessary and sufficient conditions for $u \in H^1_\Gamma$ to be in $\mathcal{A}^s(p_d)$. These conditions will be in terms of memberships of $u$ and $Tu$ into suitable approximation classes, which, in light of the preceding section, are related to Besov spaces. The coefficients of $T$ are required to satisfy conditions of the form $g \in \mathcal{A}^s_{\infty; d}(L^\theta_\infty(\Omega))$, and again they can be cast in terms of Besov spaces with the help of Theorem 3.11 and Theorem 3.9.

**Theorem 4.1.** Let $s > 0$ and let $d \geq m - 2$. Assume that $a_{ij} \in \mathcal{A}^s_{\infty; d+1-m}(L^1_\infty(\Omega))$, $b_i \in \mathcal{A}^s_{\infty; d+1-m}(L^1_\infty(\Omega))$, and $c \in \mathcal{A}^s_{\infty; d-m}(L^2_\infty(\Omega))$. Then we have

$$\mathcal{A}^s(p_d) = \mathcal{A}^s(H^1_\Gamma) \cap T^{-1}(\mathcal{A}^s_{\infty; d}(L^2_\infty(\Omega))).$$

(169)

The proof of this theorem will be given below in Lemma 4.3 and Lemma 4.5. Before proving those lemmata, let us make a few points on the conditions of the theorem.

First, recall from Theorem 3.11 that $A^\sigma_{q,q;d}(\Omega) \subset \mathcal{A}^\sigma_{\infty;d}(L^1_\infty(\Omega))$ for $\sigma = sn - 1$ and $0 \leq \frac{1}{q} - \frac{1}{2} \leq \frac{\sigma}{m}$, and from Theorem 3.9 that $B^\sigma_{q,q;d}(\Omega) \subset \mathcal{A}^\sigma_{\infty;d}(L^2_\infty(\Omega))$ for $\sigma < d + \max\{1, \frac{1}{q}\}$.

Second, while the approximation classes $\mathcal{A}^s(H^1_\Gamma)$ are associated to the finite element spaces $S_P = H^1_\Gamma \cap S^m_P$, the approximation classes we considered in the preceding section are associated to the spaces $S^P$ with no boundary conditions. In view of applying Theorem 3.8 and Theorem 3.4, we need the latter type of approximation classes. The following lemma provides a link between the two types.

**Lemma 4.2.** For $s > 0$ we have

$$\mathcal{A}^s(H^1_\Gamma) \equiv \mathcal{A}^s(H^1, \mathcal{P}, \{H^1_\Gamma \cap S^m_P\}) = H^1_\Gamma \cap \mathcal{A}^s(A^0_{\frac{1}{2}; 2}(\Omega), \mathcal{P}, \{S^P\}).$$

(170)

In particular, we have $H^1_\Gamma \cap A^\alpha(p; \Omega) \subset \mathcal{A}^s(H^1_\Gamma)$ for $\alpha = sn + 1$ and $\frac{1}{p} < s + \frac{1}{2}$.

**Proof.** Let $u \in H^1_\Gamma$, and let $u_P \in H^1_\Gamma \cap S^m_P$ be the Scott-Zhang interpolator of $u$ adapted to the Dirichlet boundary condition on $\Gamma$, cf. Scott and Zhang (1990). We have

$$\inf_{v \in H^1_\Gamma \cap S^m_P} \|u - v\|_{H^1(\Omega)} \leq \|u - u_P\|_{H^1(\Omega)} \leq \inf_{v \in S^m_P} \|u - v\|_{H^1(\Omega)}.$$
by standard properties of the Scott-Zhang interpolator. Since \( H^1 = B^1_{2,2} = A^1_{2,2} \) by Theorem 3.4, this implies (170). Then the second assertion of the theorem follows from a direct application of Theorem 3.8.

The inclusion \( \mathcal{A}^s(H^1_1) \cap T^{-1}(\mathcal{A}^s_{\infty,d}(L^1_2(\Omega))) \subset \mathcal{A}^s(\rho_d) \) of Theorem 4.1 is a consequence of the following lemma.

**Lemma 4.3.** For \( u \in H^1_1 \) and \( P \in \mathcal{P} \), there exists \( v \in S_P \) such that

\[
\rho_d(u, v, P)^2 \lesssim E(u, S_P)^2_{H^1(\Omega)} + E(Tu, S_P^d)_{L^2_1(\Omega)}^2 + \left( E(a_{ij}, S^d_{P, d+2-m}) + E(b_i, S^d_{P, d+1-m})_{L^\infty_2(\Omega)} + E(c, S^d_{P, d-m})_{L^\infty_2(\Omega)} \right) |u|_{H^1(\Omega)}^2.
\]

(172)

**Proof.** We take \( v \) to be the Scott-Zhang interpolator of \( u \) adapted to the Dirichlet boundary condition on \( \Gamma \), cf. Scott and Zhang (1990). We have

\[
\|u - v\|_{H^1} \lesssim \inf_{w \in S_P} \|u - w\|_{H^1},
\]

(173)

for all \( P \in \mathcal{P} \). It remains to bound the oscillation term.

First, let us consider the special case where the coefficients of \( T \) are piecewise polynomials subordinate to \( P \). More specifically, assume that \( a_{ij} \in \mathbb{P}_{d+2-m}, b_i \in \mathbb{P}_{d+1-m}, \) and \( c \in \mathbb{P}_{d-m} \) for each \( \tau \in P \). In this case, the oscillations associated to edges vanish, because the edge residuals \( r_e \) are polynomials of degree not exceeding \( d+1 \). For the element residuals, with the shorthand \( f = Tu \), we have

\[
\|(1 - \Pi_\tau)(f - Tu)\|_{L^2_2(\tau)} \leq \|(1 - \Pi_\tau)f\|_{L^2_2(\tau)} + \|(1 - \Pi_\tau)Tv\|_{L^2_2(\tau)},
\]

(174)

and the last term is zero because \( Tv \in \mathbb{P}_d \). The remaining term gives rise to

\[
\sum_{\tau \in P} h^2_\tau \|(1 - \Pi_\tau)f\|_{L^2_2(\tau)}^2 = E(f, S^d_P)_{L^2_1(\Omega)},
\]

(175)

which yields the desired result.

In the general case, the edge residuals and the terms \( Tv \) can be nonpolynomial. Let us treat \( Tv_{|\tau} = -a_{ij}\partial_i \partial_j v + b_k \partial_k v + cv \) term by term. We have

\[
(1 - \Pi_\tau)a_{ij}\partial_i \partial_j v = (1 - \Pi_\tau)(a_{ij} - \bar{a}_{ij})\partial_i \partial_j v, \quad \bar{a}_{ij} \in \mathbb{P}_{d+2-m},
\]

(176)

which implies

\[
\|(1 - \Pi_\tau)a_{ij}\partial_i \partial_j v\|_{L^2_2(\tau)} = \|(a_{ij} - \bar{a}_{ij})\partial_i \partial_j v\|_{L^2_2(\tau)} \leq \|a_{ij} - \bar{a}_{ij}\|_{L^\infty(\tau)}\|\partial_i \partial_j v\|_{L^2_2(\tau)},
\]

(177)
for any $a_{ij} \in \mathbb{P}_{d+2-m}$. Now we think of $\bar{a}_{ij}$ as a function in $\tilde{S}^{d+2-m}_P$ that approximates $a_{ij}$ in each element $\tau \in P$ with the best $L_\infty(\tau)$-error. As a result, we get

$$\sum_{\tau \in P} h_\tau^2 \|(1 - \Pi_\tau) a_{ij} \partial_i \partial_j v\|_{L_2(\tau)}^2 \leq \sum_{\tau \in P} h_\tau^2 \|a_{ij} - \bar{a}_{ij}\|_{L_\infty(\tau)}^2 \|\partial_i \partial_j v\|_{L_2(\tau)}^2$$

$$\leq E(a_{ij}, \tilde{S}^{d+2-m}_P(\Omega)) \sum_{\tau \in P} h_\tau^2 \|\partial_i \partial_j v\|_{L_2(\tau)}^2$$

$$\lesssim E(a_{ij}, \tilde{S}^{d+2-m}_P(\Omega)) \|\nabla v\|_{L_2(\Omega)}^2 \tag{178}$$

where we have used an inverse inequality in the last step. In light of the $H^1$-stability of the Scott-Zhang projector, this is one of the terms in the right hand side of (172).

Similarly, let $\bar{c}$ be a function in $\tilde{S}_P^{d-m}$ that approximates $c$ in each element $\tau \in P$ with the best $L_\infty(\tau)$-error. Then we have

$$(1 - \Pi_\tau)cv = (1 - \Pi_\tau)(c - \bar{c})v = (1 - \Pi_\tau)(c - \bar{c})(v - \bar{v}), \tag{179}$$

in each $\tau \in P$, where $\bar{v}$ is the average of $v$ over $\tau$. This yields

$$\sum_{\tau \in P} h_\tau^2 \|(1 - \Pi_\tau) cv\|_{L_2(\tau)}^2 \leq \sum_{\tau \in P} h_\tau^2 \|c - \bar{c}\|_{L_\infty(\tau)}^2 \|v - \bar{v}\|_{L_2(\tau)}^2$$

$$\lesssim \sum_{\tau \in P} h_\tau^2 \|c - \bar{c}\|_{L_\infty(\tau)}^2 \|\nabla v\|_{L_2(\tau)}^2 \tag{180}$$

$$\leq E(c, \tilde{S}_P^{d-m}(\Omega)) \|\nabla v\|_{L_2(\Omega)}^2,$$

where we have used the Poincaré inequality in the second line. Estimation of the term involving $b_k \partial_i v$ is more straightforward, which we omit.

As for the edge oscillations, let $\tau \in P$, and let $e$ be an edge of $\tau$. Then we have

$$\|(1 - \Pi_\tau)e_{ij} \partial_j v\|_{L_2(e)} = \|(1 - \Pi_\tau)(a_{ij} - \bar{a}_{ij}) \partial_j v\|_{L_2(e)}$$

$$\leq \|(a_{ij} - \bar{a}_{ij}) \partial_j v\|_{L_2(e)}$$

$$\leq \|a_{ij} - \bar{a}_{ij}\|_{L_\infty(e)} \|\partial_j v\|_{L_2(e)} \tag{181}$$

$$\lesssim h_e^{-\frac{1}{2}} \|a_{ij} - \bar{a}_{ij}\|_{L_\infty(\tau)} \|\nabla v\|_{L_2(\tau)},$$

for any $\bar{a}_{ij} \in \mathbb{P}_{d+2-m}$, which shows that the contribution of the edge oscillations to the final estimate (172) is identical to that of (178).

**Remark 4.4.** By using the fact that the Scott-Zhang projector is bounded in $H^t(\Omega)$ for $t < \frac{3}{2}$, we could have introduced extra powers of $h_\tau$ or $h_e$ into the estimates (178), (179), and (181). This means that the regularity conditions on the coefficients $a_{ij}$, $b_k$, and $c$ in Theorem 4.1 can be relaxed slightly, if the conclusion of the theorem is to be changed to $\mathcal{A}^s(H^1) \cap H^t(\Omega) \cap T^{-1}(\mathcal{A}^{\bar{s}}_{\infty,q}(L_2^1(\Omega))) \subset \mathcal{A}^s(\rho_d)$ with $1 < t < \frac{3}{2}$.
Lemma 4.5. For any $u \in H^1_\Omega$, $P \in \mathcal{P}$ and $v \in S_P$, we have
\[
E(Tu, S^d_P)_{L^2(\Omega)}^2 + \|u - v\|_{H^1(\Omega)}^2 \lesssim \rho_d(u, v, P)^2 \\
+ \left( E(a_{ij}, S^d_P)_{L^\infty(\Omega)}^2 + E(b_1, S^d_P_{L^1(\Omega)}^2 + E(e, S^d_P - m)_{L^\infty(\Omega)}^2 + E(e, S^d_P - m)_{L^1(\Omega)}^2 \right) |v|_{H^1(\Omega)}^2.
\]
In particular, under the hypotheses of Theorem 4.1, we have the inclusion $\mathcal{A}^s(\rho_d) \subset \mathcal{A}^s(H^1_\Omega \cap T^{-1}(\mathcal{A}^s_{\infty;d}(L^1_2(\Omega))))$.

Proof. All the ingredients for establishing the estimate (182) is already given in the proof of the preceding lemma. Namely, we start with the bound
\[
(\text{osc}_d(u, v, P))^2 \lesssim \sum_{\tau \in P} h^2_\tau (1 - \Pi_\tau) T u \|_{L^2(\tau)}^2 + \sum_{\tau \in P} h^2_\tau (1 - \Pi_\tau) T v \|_{L^2(\tau)}^2 \\
+ \sum_{e \in E_P} h_e (1 - \Pi_e) r_e \|_{L^2(e)}^2
\]
and use the estimates (178), (179), and (181), etc., on the last two terms to get (182).

As for the second assertion, let $\{P_k\} \subset \mathcal{P}$ and $\{v_k\}$ be two sequences with $v_k \in S_{P_k}$ such that $\#P_k \lesssim 2^k$ and $\rho_d(u, v_k, P_k) \lesssim 2^{-k}\rho$. Then since $\|u - v_k\|_{H^1} \lesssim \rho_d(u, v_k, P_k)$, we have $\|v_k\|_{H^1} \lesssim \|u\|_{H^1}$. Hence, by employing overlay of partitions, without loss of generality, we can suppose that the right hand side of (182) with $P = P_k$ and $v = v_k$ is bounded by a constant multiple of $2^{-k}\rho$. Looking at the left hand side then reveals that $Tu \in \mathcal{A}^s_{\infty;d}(L^1_2(\Omega))$ and $u \in \mathcal{A}^s(H^1_\Omega)$. \qed

A Tree completion and the red refinement

In this appendix, we will justify the assumption (6) for the red refinement rule. There is no doubt that this result is known to the experts but it does not seem to have appeared in writing. The proof relies on arguments from Binev et al. (2004) and Stevenson (2008), and below we make an attempt at presenting those arguments in a reusable manner as an abstract theorem for trees. We remark that this form has also appeared in the author’s PhD thesis Gantumur (2006).

Let $\nabla$ be a countable set, and let a parent-child relation be defined on $\nabla$. We assume that every element $\lambda \in \nabla$ has a uniformly bounded number of children, and has at most one parent. We say that $\lambda \in \nabla$ is a descendant of $\mu \in \nabla$ and write $\lambda \succ \mu$ if $\lambda$ is a child of a descendant of $\sigma$ or $\mu$ is a child of $\mu$. The level or generation of an element $\lambda \in \nabla$, denoted by $|\lambda| \in \mathbb{N}_0$, is the number of its ascendants. Obviously, $\lambda \succ \mu$ implies $|\lambda| > |\mu|$. We call the set $\nabla_0 := \{\lambda \in \nabla : |\lambda| = 0\}$ the root, and assume that $1 \leq \#\nabla_0 < \infty$.

Example A.1. Let $\Omega \subset \mathbb{R}^n$ be a polyhedral domain, and let $\nabla_0$ be a conforming partition of $\Omega$ into finitely many $n$-simplices. We form the set $\nabla$ by collecting all $n$-simplices created
by a (possibly trivial) finite sequence of red (or dyadic) refinements of an initial simplex \( \lambda \in \nabla_0 \). The parent-child relation on \( \nabla \) is defined by saying that \( \lambda \in \nabla \) is a child of \( \mu \in \nabla \) if \( \lambda \) is created by one elementary red refinement of \( \mu \). In the notations of the main body of the current article, once the red refinement rule has been fixed, we have \( \nabla_0 = P_0 \) and \( \nabla = \{ \tau \in P_k, k = 0, 1, \ldots \} \). Recall that we denoted the level by \( [\tau] \), because the notation \( |\tau| \) was reserved for the volume of \( \tau \).

A subset \( \Lambda \subseteq \nabla \) is said to be a tree if with every member \( \lambda \in \Lambda \) all its ascendants are included in \( \Lambda \). For a tree \( \Lambda \), those \( \lambda \in \Lambda \) whose children are not contained in \( \Lambda \) are called leaves of \( \Lambda \). Similarly, those \( \lambda \notin \Lambda \) whose parent belongs to \( \Lambda \) is called outer leaves of \( \Lambda \) and the set of all outer leaves of \( \Lambda \) is denoted by \( \partial \Lambda \).

**Example A.2.** In the context of Example A.1, every partition of \( \Omega \) that is obtained by a finite sequence of red refinements corresponds to a tree. The leaves of this tree are exactly the triangles of the partition, and the outer leaves are the triangles that are “one refinement away” from appearing as part of the partition.

We assume the existence of functions \( d : \nabla \to \mathbb{R} \) and \( d : \nabla \times \nabla \to \mathbb{R} \) satisfying the following conditions:

(i) For any \( \lambda \in \nabla \), with some absolute constants \( C_d, \chi \geq 0 \), it holds that
\[
0 < d(\lambda) \leq C_d 2^{-\chi |\lambda|};
\]

(ii) For any \( \lambda, \mu \in \nabla \), we have \( d(\lambda, \mu) = d(\mu, \lambda) \geq 0 \), and \( d(\lambda, \mu) = 0 \) if \( \lambda \succeq \mu \);

(iii) For any \( \lambda, \mu, \nu \in \nabla \), there holds a triangle inequality:
\[
d(\lambda, \nu) \leq d(\lambda, \mu) + d(\mu, \nu);
\]

(iv) Let \( L \in \mathbb{N}_0 \) and \( C > 0 \) be arbitrary but fixed constants. Then for any fixed \( \mu \in \nabla \), \( \ell \in \mathbb{N}_0 \) with \( \ell \leq |\mu| + L \), there exists a uniformly bounded number of \( \lambda \in \nabla \) with \( d(\lambda, \mu) \leq C 2^{-\chi \ell} \).

**Example A.3.** Continuing Example A.1, it is straightforward to verify that the functions \( d(\cdot) = \text{diam}(\cdot) \) and \( d(\cdot, \cdot) = \text{dist}(\cdot, \cdot) \) satisfy the aforementioned conditions with \( \chi = 1 \).

Let \( T \) denote the set of all finite trees, and let \( \hat{T} \subset T \) be a subset such that \( \nabla_0 \in \hat{T} \). We call \( \hat{T} \) the set of admissible trees. It is a simple operation to append an outer leave \( \mu \in \partial \Lambda \) to an existing tree \( \Lambda \in \hat{T} \). However, if one wants to stay in the class of admissible trees, then in general one needs to append additional nodes to \( \Lambda \). We model this operation by a map \( R \) that sends the pair of a tree \( \Lambda \in \hat{T} \) and any of its outer leaves \( \mu \in \partial \Lambda \) to a set
\( \mu \in R(\Lambda, \mu) \subset \nabla \) such that \( R(\Lambda, \mu) \cap \Lambda = \emptyset \), and \( R(\Lambda, \mu) \cup \Lambda \) is an admissible tree. We assume that for any \( \lambda \in R(\Lambda, \mu) \) it holds that
\[
d(\lambda, \mu) \leq C_R 2^{-\chi(|\lambda|)}, \quad \text{and} \quad |\lambda| \leq |\mu| + L_R, \tag{184}
\]
where \( C_R \) and \( L_R \in \mathbb{N}_0 \) are constants.

**Example A.4.** In the context of Example A.3, we fix a positive integer \( t \in \mathbb{N} \), and define an admissible partition by the condition that the level difference between any two neighbouring simplices cannot be larger than \( t \). Here \( \lambda \) and \( \mu \) are understood to be neighbours if the intersection of the closures of \( \lambda \) and \( \mu \) is an \( n - 1 \) dimensional simplex. We let \( \tilde{T} \) be the set of trees corresponding to admissible partitions. To define the map \( \tilde{t} \), simplices cannot be larger than admissible partition by the condition that the level difference between any two neighbouring \( \lambda \) given. We call a sequence \( \{\lambda_k\} \subset P \) a path if \( \lambda_k \) and \( \lambda_{k+1} \) are neighbours for each \( k \), and the length of a path is defined as the number of simplices in the path. Singletones are considered as paths of length 1. Then with \( \lambda_1 \in P \) being the parent of \( \mu \), we define \( M_k \subset P \) by the condition that \( \lambda \in P \) is in \( M_k \) iff the length of a shortest path between \( \lambda \) and \( \lambda_1 \) is \( k \). Now we associate a number \( f(\lambda) \) to each \( \lambda \in P \), as follows. We set \( f(\lambda_1) = |\lambda_1| + 1 \), and let us assume that \( f \) has been defined on \( M_1 \cup \ldots \cup M_k \). Let \( \lambda \in M_{k+1} \). We set \( f(\lambda) = |\lambda| \) if \( f(\lambda') - |\lambda| \leq t \) for each \( \lambda' \in M_k \cap N(\lambda) \), where \( N(\lambda) \subset P \) is the set of neighbours of \( \lambda \). Otherwise, we set \( f(\lambda) = |\lambda| + 1 \). Finally, we obtain a new partition \( P' \) by refining each \( \lambda \in P \) for which \( f(\lambda) > |\lambda| \). We claim that \( P' \) is admissible. Indeed, by construction, admissibility is maintained across \( M_k \) and \( M_{k+1} \) for all \( k \). So assume that \( \lambda \in M_k \) is refined, i.e., \( \lambda \in P \setminus P' \), and that \( \lambda' \in M_k \cap N(\lambda) \) is not refined, with \( |\lambda'| + t = |\lambda| \), meaning that admissibility is violated once \( \lambda \) is refined. Then there exists a path of length \( k \) connecting \( \lambda \) with \( \lambda_1 \), whose every element is refined, implying that \( |\lambda| + (k - 1)t = |\lambda_1| \). This means that if \( |\lambda'| < |\lambda_1| - (k - 1)t = |\lambda_1| \), then \( \lambda' \) would have to be refined by the influence of its neighbours from \( M_{k-1} \), yielding a contradiction. We also remark that \( P' \) is the smallest admissible refinement of \( P \) such that \( \mu \in P' \), because every refinement we had performed was necessary for maintaining admissibility. After this not-so-short preparation, we get back to our original goal of defining the map \( R \), and set \( R(\Lambda, \mu) = P' \setminus P \). For this map, the second property in (184) is clearly satisfied with \( L_R = t \). As for the first property, if \( \lambda \in R(\Lambda, \mu) \), then by the aforementioned reasoning, there exists a path connecting \( \lambda \) with \( \lambda_1 \), whose every element is refined. This means that the diameter of the elements in this path decreases geometrically, and since the distance between \( \lambda \) and \( \mu \) is smaller than the sum of those diameters, we get the desired estimate.

The following theorem is an easy extension of (Stevenson, 2008, Theorem 6.1) and (Binev et al., 2004, Theorem 2.4), and we include the proof for reader’s convenience. Specialized to the setting of Example A.4, this theorem justifies the assumption (6).

**Theorem A.5.** Let \( \{\Lambda_k\} \subset \tilde{T} \) and \( \{\mu_k\} \subset \nabla \) be sequences such that \( \Lambda_0 = \nabla_0 \) and
\[
\Lambda_{k+1} = R(\Lambda_k, \mu_k), \quad \text{and} \quad \mu_k \in \partial \Lambda_k, \quad \text{for} \quad k = 0, 1, \ldots \tag{185}
\]
Then we have
\[ \#\Lambda_k - \#\Lambda_0 \lesssim k. \]  
(186)

**Proof.** The proof will closely follow the proof of Theorem 6.1 in Stevenson (2008). Let \( a : \mathbb{N}_0 \cup \{-1, \ldots, -L_R\} \to (0, \infty) \) and \( b : \mathbb{N} \to (1, \infty) \) be some sequences with
\[
\sum_p a(p) < \infty, \quad \sum_p b(p) 2^{-xp} < \infty, \quad \text{and} \quad \inf_{p \geq 1} (b(p) - 1)a(p) > 0.
\]  
(187)

For instance, \( a(p) = (p + L_R + 1)^{-2} \) and \( b(p) = 1 + 2^{\kappa p} \) with a constant \( \kappa \in (0, \chi) \) satisfy these conditions.

Let \( M = \{\mu_0, \ldots, \mu_k\} \), let \( A = C_R + (2\chi C_R + 2\chi C_d + C_d)\sum_p b(p) 2^{-xp} \), and define the function \( f : \Lambda \times M \to \mathbb{R} \) by
\[
f(\lambda, \mu) = \begin{cases} 
a(|\mu| - |\lambda|) & \text{if } d(\lambda, \mu) < A 2^{-\chi|\lambda|} \text{ and } |\mu| - |\lambda| \geq -L_R, \\ 0 & \text{otherwise}. \end{cases}
\]  
(188)

From condition (iv), for any \( \mu \in M \) we have
\[
\sum_{\lambda \in \Lambda} f(\lambda, \mu) = \sum_{\ell=0}^{|\mu|+L_R} \sum_{|\lambda|=\ell} f(\lambda, \mu) \lesssim \sum_{\ell=0}^{|\mu|+L_R} a(|\mu| - \ell) \leq \sum_p a(p) \lesssim 1, 
\]  
(189)

implying that \( \sum_{\mu \in M} \sum_{\lambda \in \Lambda} f(\lambda, \mu) \lesssim \#M \).

We claim that for any \( \lambda \in \Lambda \setminus \nabla_0 \),
\[
\sum_{\mu \in M} f(\lambda, \mu) \gtrsim 1, 
\]  
(190)

which would imply that
\[
\#(\Lambda \setminus \nabla_0) \lesssim \sum_{\lambda \in \Lambda \setminus \nabla_0} \sum_{\mu \in M} f(\lambda, \mu) \lesssim \sum_{\mu \in M} \sum_{\lambda \in \Lambda} f(\lambda, \mu) \lesssim \#M, 
\]  
(191)

as required. Now we will prove this claim.

The claim is true for \( \lambda \in M \) since \( f(\lambda, \lambda) = a(0) \gtrsim 1 \). Let \( \lambda_0 \in \Lambda \setminus (M \cup \nabla_0) \). For \( j \geq 0 \), assume that \( \lambda_j \) has been defined and let \( \lambda_j' \) be the parent of \( \lambda_j \) for \( j \geq 1 \), and \( \lambda_0' := \lambda_0 \).

Then we define \( \lambda_{j+1} \in M \) such that \( \lambda_j' \in R(\Lambda', \lambda_{j+1}) \) with some tree \( \Lambda' \). Let \( s \) be the smallest positive integer such that \( |\lambda_s| \in I := \{|\lambda_0| - L_R, \ldots, |\lambda_0|\} \). Note that such an \( s \) exists. Indeed, the sequence \( \{\lambda_j\} \) ends with some \( \lambda_J \in \partial \nabla_0 \) thus with \( |\lambda_J| = 1 \gtrsim |\lambda_0| \), and from the properties of \( R \) we have \( |\lambda_j'| \leq |\lambda_{j+1}| + L_R \) or \( |\lambda_{j+1}| \geq |\lambda_j| - L_R - 1 \) for \( j \geq 1 \) and
|λ_1| ≥ |λ_0| − L_R, meaning that if not |λ_1| ∈ I, we have |λ_1| > |λ_0|. Therefore the interval I can not be skipped by j → λ_j. Now for 1 ≤ j ≤ s, we have

\[ d(λ_0, λ_j) ≤ d(λ_0, λ_1) + d(λ_1, λ_j) ≤ \sum_{k=1}^{j} d(λ_k-1, λ_k) + \sum_{k=1}^{j-1} d(λ_k) \]

\[ ≤ \sum_{k=1}^{j} d(λ_k, λ_k) + \sum_{k=1}^{j-1} d(λ_k) \]

\[ ≤ C_R 2^{-|λ_0|} + C_R \sum_{k=1}^{j-1} 2^{-|λ_k|} + C_d \sum_{k=1}^{j-1} 2^{-|λ_k|} + 2^{-|λ_k|} \]

\[ = C_R 2^{-|λ_0|} + (2^C R + 2^C d + C_d) \sum_{k=1}^{j-1} 2^{-|λ_k|} \]

(192)

where m(p, j) denotes the number of k ∈ {1, ..., j − 1} with |λ_k| = |λ_0| + p. Note that m(p, 1) = 0 for any p.

In case m(p, s) ≤ b(p) for all p ≥ 1, then by the definition of the constant A we have d(λ_0, λ_s) < A 2^{-|λ_0|}. Since −L_R ≤ |λ_s| − |λ_0| ≤ 0, we have f(λ_0, λ_s) = a(|λ_s| − |λ_0|) ≥ 1, which proves the claim in this case.

Otherwise, there exist p with m(p, s) > b(p). For each of those p, there exists a smallest j = j(p) with m(p, j(p)) > b(p) because m(p, j) ≥ m(p, j − 1). With j* := min_{p≥1} j(p), let p* be such that j(p*) = j*. So we have m(p, j* − 1) ≤ b(p) for all p ≥ 1, and m(p*, j* − 1) ≥ m(p*, j*) − 1 > b(p*) − 1 ≥ 0. This implies that j* − 1 ≥ 1. As in the above case, we find that for all 1 ≤ k ≤ j* − 1, d(λ_0, λ_k) < A 2^{-|λ_0|} and f(λ_0, λ_k) = a(|λ_k| − |λ_0|).

Finally by using the definition of m(·, ·) we have

\[ \sum_{1 ≤ k ≤ j* − 1; |λ_k| = |λ_0| + p} f(λ_0, λ_k) = m(p*, j* − 1) a(p*) \]

\[ > [b(p*) − 1] a(p*) ≥ \inf_{p ≥ 0} [b(p) − 1] a(p) ≥ 1, \]

(193)

which proves the claim.

\[ \square \]

References

Markus Aurada, Michael Feischl, Thomas Führer, Michael Karkulik, and Dirk Praetorius. Efficiency and optimality of some weighted-residual error estimator for adaptive 2d boundary element methods. ASC Report 15/2012, 2012. URL http://publik.tuwien.
A Tree completion and the red refinement

Jöran Bergh and Jörgen Löfström. *Interpolation spaces. An introduction*. Springer-Verlag, Berlin, 1976. Grundlehren der Mathematischen Wissenschaften, No. 223.

Peter Binev, Wolfgang Dahmen, Ronald DeVore, and Pencho Petrushev. Approximation classes for adaptive methods. *Serdica Math. J.*, 28(4):391–416, 2002. Dedicated to the memory of Vassil Popov on the occasion of his 60th birthday.

Peter Binev, Wolfgang Dahmen, and Ron DeVore. Adaptive finite element methods with convergence rates. *Numer. Math.*, 97(2):219–268, 2004. URL http://dx.doi.org/10.1007/s00211-003-0492-7.

Mikhail Shlemovich Birman and Mikhail Zakharovich Solomyak. Piecewise polynomial approximations of functions of classes $W^p_α$. *Mat. Sb. (N.S.*), 73(3):331–355, 1967.

J. Manuel Cascon, Christian Kreuzer, Ricardo H. Nochetto, and Kunibert G. Siebert. Quasi-optimal convergence rate for an adaptive finite element method. *SIAM J. Numer. Anal.*, 46(5):2524–2550, 2008. URL http://dx.doi.org/10.1137/07069047X.

Albert Cohen, Wolfgang Dahmen, and Ronald DeVore. Adaptive wavelet methods for elliptic operator equations: convergence rates. *Math. Comp.*, 70(233):27–75, 2001. ISSN 0025-5718. doi: 10.1090/S0025-5718-00-01252-7. URL http://dx.doi.org/10.1090/S0025-5718-00-01252-7.

Shai Dekel and Dany Leviatan. Whitney estimates for convex domains with applications to multivariate piecewise polynomial approximation. *Found. Comput. Math.*, 4(4):345–368, 2004. ISSN 1615-3375. doi: 10.1007/s10208-004-0096-3. URL http://dx.doi.org/10.1007/s10208-004-0096-3.

Ronald A. DeVore and George G. Lorentz. *Constructive approximation*, volume 303 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1993. ISBN 3-540-50627-6.

Ronald A. DeVore and Vasil A. Popov. Interpolation of Besov spaces. *Trans. Amer. Math. Soc.*, 305(1):397–414, 1988. ISSN 0002-9947. doi: 10.2307/2001060. URL http://dx.doi.org/10.2307/2001060.

Willy Dörfler. A convergent adaptive algorithm for Poisson’s equation. *SIAM J. Numer. Anal.*, 33(3):1106–1124, 1996. URL http://dx.doi.org/10.1137/0733054.

Michael Feischl, Thomas Führer, and Dirk Praetorius. Adaptive FEM with optimal convergence rates for a certain class of nonsymmetric and possibly nonlinear problems. *SIAM J. Numer. Anal.*, 52(2):601–625, 2014. ISSN 0036-1429. doi: 10.1137/120897225. URL http://dx.doi.org/10.1137/120897225.
Tsogtgerel Gantumur. *Adaptive wavelet algorithms for solving operator equations.* PhD thesis, Utrecht University, The Netherlands, 2006. URL http://www.math.mcgill.ca/~gantumur/research.html.

Tsogtgerel Gantumur, Helmut Harbrecht, and Rob Stevenson. An optimal adaptive wavelet method without coarsening of the iterands. *Math. Comp.*, 76(258):615–629, 2007. ISSN 0025-5718. doi: 10.1090/S0025-5718-06-01917-X. URL http://dx.doi.org/10.1090/S0025-5718-06-01917-X.

Fernando D. Gaspoz and Pedro Morin. Approximation classes for adaptive higher order finite element approximation. *Math. Comp.*, 2013. To appear.

Pedro Morin, Ricardo H. Nochetto, and Kunibert G. Siebert. Data oscillation and convergence of adaptive FEM. *SIAM J. Numer. Anal.*, 38(2):466–488 (electronic), 2000. URL http://dx.doi.org/10.1137/S0036142999360044.

Ricardo H. Nochetto, Kunibert G. Siebert, and Andreas Veeser. Theory of adaptive finite element methods: an introduction. In *Multiscale, nonlinear and adaptive approximation*, pages 409–542. Springer, Berlin, 2009. URL http://dx.doi.org/10.1007/978-3-642-03413-8_12.

Peter Oswald. *Multilevel finite element approximation.* Teubner Skripten zur Numerik. [Teubner Scripts on Numerical Mathematics]. B. G. Teubner, Stuttgart, 1994. ISBN 3-519-02719-4. Theory and applications.

L. Ridgway Scott and Shangyou Zhang. Finite element interpolation of nonsmooth functions satisfying boundary conditions. *Math. Comp.*, 54(190):483–493, 1990. URL http://dx.doi.org/10.2307/2008497.

Rob Stevenson. Optimality of a standard adaptive finite element method. *Found. Comput. Math.*, 7(2):245–269, 2007. URL http://dx.doi.org/10.1007/s10208-005-0183-0.

Rob Stevenson. The completion of locally refined simplicial partitions created by bisection. *Math. Comp.*, 77(261):227–241, 2008. URL http://dx.doi.org/10.1090/S0025-5718-07-01959-X.