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

Tsogtgerel Gantumur<br>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 are 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 multievel 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 conforming and nonconforming triangulations.


## 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
3 Lagrange finite elements ..... 15
3.1 Preliminaries ..... 15
3.2 Quasi-interpolation operators ..... 16
3.3 Multilevel approximation spaces ..... 22
3.4 Adaptive approximation ..... 27
3.5 Discontinuous piecewise polynomials ..... 31
4 Second order elliptic problems ..... 33
4.1 Introduction ..... 33
4.2 A characterization of adaptive approximation classes ..... 35

## 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 $\Omega \in \mathbb{R}^{2}$, 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 $\mathscr{A}^{s}$ 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 $c N^{-s}$, where $c=c\left(u, P_{0}, s\right) \geq 0$ 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 $\mathscr{A}^{s}$ if

$$
\begin{equation*}
\min _{\{P \in \mathscr{P}: \# P \leq N\}} \inf _{v \in S_{P}}\|u-v\|_{H^{1}} \leq c N^{-s}, \tag{1}
\end{equation*}
$$

for all $N \geq \# P_{0}$ and for some constant $c$, where $\mathscr{P}$ 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 $u \in \mathscr{A}^{s}$ but $u \notin \mathscr{A}^{\sigma}$ for any $\sigma>s$. Then as far as the exponent $\sigma$ in $c N^{-\sigma}$ is concerned, it is obvious that the best asymptotic bound on the error we can hope for is $c N^{-s}$, 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 $u \in \mathscr{A}^{s}$ 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 $\mathscr{A}^{s}$ 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 $\mathscr{A}^{s}$ 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 $B_{p, p}^{\alpha} \subset \mathscr{A}^{s} \subset B_{p, p}^{\sigma}$ for $\frac{2}{p}=\sigma<1+\frac{1}{p}$ and $\sigma<\alpha<\max \left\{2,1+\frac{1}{p}\right\}$ with $s=\frac{\alpha-1}{2}$. Here $B_{p, q}^{\alpha}$ 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 $B_{p, p}^{\alpha} \subset \mathscr{A}^{s}$ holds in the larger range $\sigma<\alpha<m+\max \left\{1, \frac{1}{p}\right\}$, 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 $\mathscr{A}^{s} \subset B_{p, p}^{\sigma}$ 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

$$
\begin{equation*}
\inf _{v \in S_{P_{k}}}\|u-v\|_{L_{p}} \leq c h_{k}^{\sigma} \tag{2}
\end{equation*}
$$

where $\left\{P_{k}\right\} \subset \mathscr{P}$ 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 $\mathscr{A}^{s} \subset A_{p, p}^{\sigma}$ is recovered for all $\sigma \leq \frac{2}{p}$.

In this paper, we prove the direct embedding $A_{p, p}^{\alpha} \subset \mathscr{A}^{s}$, so that the existing situation $B_{p, p}^{\alpha} \subset \mathscr{A}^{s} \subset A_{p, p}^{\sigma}$ is improved to $A_{p, p}^{\alpha} \subset \mathscr{A}^{s} \subset A_{p, p}^{\sigma,}$. It is a genuine improvement, since $A_{p, p}^{\alpha}(\Omega) \supsetneq B_{p, p}^{\alpha}(\Omega)$ for $\alpha \geq 1+\frac{1}{p}$. Moreover, as one stays entirely within an approximation theory framework, one can argue that the link between $\mathscr{A}^{s}$ and $A_{p, p}^{\alpha}$ is more natural than the link between $\mathscr{A}^{s}$ and $B_{p, p}^{\alpha}$. Once the link between $\mathscr{A}^{s}$ and $A_{p, p}^{\alpha}$ has been established,

[^0]one can then invoke the well known relationships between $A_{p, p}^{\alpha}$ and $B_{p, p}^{\alpha}$. 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 procedure and conforming triangulations, we deal with possibly nonconforming triangulations, and therefore are able to handle the red refinement procedure, as well as newest vertex bisections without the conformity requirement.

(a) If the space $B_{p, p}^{\alpha}$ is located strictly above the solid line and below the dashed line, then $B_{p, p}^{\alpha} \subset \mathscr{A}^{s}$ with $s=\frac{\alpha-1}{2}$. The inverse embeddings $\mathscr{A}^{s+\varepsilon} \subset B_{p, p}^{\alpha}$ hold on the solid line and below the (slanted) dotted line.

(b) If the space $B_{p, p}^{\alpha}$ is located above or on the solid line, and if $u \in \mathscr{A}^{s}$ and $\Delta u \in B_{p, p}^{\alpha}$ with $s=\frac{\alpha+1}{2}$, then $u \in \mathscr{A}_{*}^{s}$. It is as if the approximation of $\Delta u$ is taking place in $H^{-1}$, with the proviso that the shaded area is excluded from all considerations.

Fig. 1: Illustration of various embeddings. The point $\left(\frac{1}{p}, \alpha\right)$ represents the space $B_{p, p}^{\alpha}$.
The approximation classes $\mathscr{A}^{s}$ 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 $B_{p, p}^{\alpha}$, 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

$$
\begin{equation*}
\rho(u, v, P)=\left(\|u-v\|_{H^{1}}^{2}+\sum_{\tau \in P}(\operatorname{diam} \tau)^{2}\left\|f-\Pi_{\tau} f\right\|_{L_{2}(\tau)}^{2}\right)^{\frac{1}{2}} \tag{3}
\end{equation*}
$$

where $f=\Delta u$, and $\Pi_{\tau}: L_{2}(\tau) \rightarrow \mathbb{P}_{d}$ 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

$$
\begin{equation*}
\Delta u=f \quad \text { in } \Omega \quad \text { and }\left.\quad u\right|_{\Omega}=0 \tag{4}
\end{equation*}
$$

satisfies

$$
\begin{equation*}
\min _{\{P \in \mathscr{P}: \# P \leq N\}} \inf _{v \in S_{P}} \rho(u, v, P) \leq c N^{-s}, \tag{5}
\end{equation*}
$$

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 $\mathscr{A}_{*}^{s}$ defined by the condition (5) are more attuned to certain practical adaptive finite element methods than the standard approximation classes $\mathscr{A}^{s}$ defined by (1), see Section 4. Obviously, we have $\mathscr{A}_{*}^{s} \subset \mathscr{A}^{s}$ but we cannot expect the inclusion $\mathscr{A}^{s} \subset \mathscr{A}_{*}^{s}$ to hold in general. In Cascon et al. (2008), an effective characterization of $\mathscr{A}_{*}^{s}$ was announced as an important pending issue.

In the present paper, we establish a characterization of $\mathscr{A}_{*}^{s}$ 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 $u \in \mathscr{A}^{s}$ 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 $u \in \mathscr{A}_{*}^{s}$, 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 $B_{p, p}^{\alpha} \not \subset L_{2}$, hence the quantity (3) would be infinite in general for $f \in B_{p, p}^{\alpha}$.

At this point, the reader might be wondering if we can deduce $f \in B_{p, p}^{\alpha}$ from $u \in \mathscr{A}^{s}$. If so, everything would follow from the single assumption $u \in \mathscr{A}^{s}$, which would then render the theory more in line with the traditional results. However, $u$ can be in a space slightly smaller than $H^{1}$, and in this case we cannot guarantee $f \in L^{2}$, and hence the quantity (3) would not be defined. On the other hand, there are standard examples where $f$ is smooth but $u$ is barely in $H^{1+\varepsilon}$ for a small $\varepsilon$. Furthermore, since we are solving a PDE, and $f$ is "given", there would generally be much more information available about $f$ than about $u$, and so it is not an urgent matter to deduce the regularity of $f$ from that of $u$.

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 $\mathscr{A}_{*}^{s}$. 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 3.8, Theorem 3.9, and Theorem 3.12. Finally, in Section 4, we investigate approximation classes associated to certain adaptive finite element methods for variable coefficient second order boundary value problems. We emphasize that the actual results in Section 4 are in terms of the approximation spaces that are studied in Section 3, and in order to relate them to Besov spaces, one has to appeal to Section 3.

## 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 $\bar{\Omega}=\bigcup_{\tau \in P} \bar{\tau}$. We assume that a set $\mathscr{P}$ 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 $\{P \in \mathscr{P}: \# P \leq k\}$ is finite. In practice, $\mathscr{P}$ 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 class of important examples arises when we want to allow partitions with hanging nodes. In this case, an admissibility criterion on a partition has been discussed in Bonito and Nochetto (2010). 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. Hence note that the use of a two dimensional language does not mean that the results we discuss are valid only in two dimensions.

We will assume the existence of a refinement procedure satisfying certain requirements. Given a partition $P \in \mathscr{P}$ and a set $R \subset P$ of its elements, the refinement procedure produces $P^{\prime} \in \mathscr{P}$, such that $P \backslash P^{\prime} \supset R$, i.e., the elements in $R$ are refined at least once. Let us denote it by $P^{\prime}=\operatorname{refine}(P, R)$. 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 $\tau \in P^{\prime}$ and $\sigma \in R$ with $\tau \cap \sigma \neq \varnothing$. Note that we have $\lambda=2$ for red refinements, and $\lambda=\sqrt[n]{2}$ for the newest vertex bisection. Moreover, we assume the following on the efficiency of the refinement procedure: If $\left\{P_{k}\right\} \subset \mathscr{P}$ and $\left\{R_{k}\right\}$ are sequences such that $P_{k+1}=\operatorname{refine}\left(P_{k}, R_{k}\right)$ and $R_{k} \subset P_{k}$ for $k=0,1, \ldots$, then

$$
\begin{equation*}
\# P_{k}-\# P_{0} \lesssim \sum_{m=0}^{k-1} \# R_{m}, \quad k=1,2, \ldots \tag{6}
\end{equation*}
$$

Here and in what follows, we shall often dispense with giving explicit names to constants, and use the Vinogradov-style notation $X \lesssim Y$, which means that $X \leq C \cdot Y$ with some constant $C$ that is allowed to depend only on $\mathscr{P}$ and (the geometry of) the domain $\Omega$. Assumption (6) is justified for newest vertex bisection algorithm in Binev et al. (2004); Stevenson (2008), and the red refinement rule is treated in Bonito and Nochetto (2010).

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 $P \in \mathscr{P}$, 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 $P \in \mathscr{P}$. 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 $P \in \mathscr{P}$, we let

$$
\begin{equation*}
E\left(u, S_{P}\right)_{\rho}=\inf _{v \in S_{P}} \rho(u, v, P), \tag{7}
\end{equation*}
$$

which is the error of a best approximation of $u$ from $S_{P}$. Furthermore, we introduce

$$
\begin{equation*}
E_{k}(u)_{\rho}=\inf _{\left\{P \in \mathscr{P}: \# P \leq 2^{k} N\right\}} E\left(u, S_{P}\right)_{\rho}, \tag{8}
\end{equation*}
$$

for $u \in X_{0}$ and $k \in \mathbb{N}$, with the constant $N$ chosen sufficiently large in order to ensure that the set $\{P \in \mathscr{P}: \# P \leq 2 N\}$ 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 $2^{k} N$ triangles. Finally, we define the main object of our study, the (adaptive) approximation class

$$
\begin{equation*}
\mathscr{A}_{q}^{s}(\rho)=\mathscr{A}_{q}^{s}\left(\rho, \mathscr{P},\left\{S_{P}\right\}\right)=\left\{u \in X_{0}:|u|_{\mathscr{A} \mathscr{q}_{q}^{s}(\rho)}<\infty\right\}, \tag{9}
\end{equation*}
$$

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

$$
\begin{equation*}
|u|_{\mathscr{A}_{q}^{s}(\rho)}=\left\|\left(2^{k s} E_{k}(u)_{\rho}\right)_{k \in \mathbb{N}}\right\|_{\ell q}, \quad u \in X_{0} . \tag{10}
\end{equation*}
$$

In the following, we will use the abbreviation $\mathscr{A}^{s}(\rho)=\mathscr{A}_{\infty}^{s}(\rho)$. Note that $u \in \mathscr{A}_{q}^{s}(\rho)$ implies $E_{k}(u)_{\rho} \leq c 2^{-k s}$ 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 $\mathscr{A}_{q}^{s}(\rho) \subset \mathscr{A}_{r}^{\alpha}(\rho)$ 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 $\mathscr{P}$. However, in a typical situation, it is indeed a vector space equipped with the quasi-norm $\|\cdot\|_{\mathscr{A}_{q}^{s}(\rho)}=\|\cdot\|_{X_{0}}+|\cdot|_{\mathscr{A}}^{s}(\rho)$.
Remark 2.1. Suppose that $\rho$ satisfies

- $\rho(\alpha u, \alpha v, P)=|\alpha| \rho(u, v, P)$ for $\alpha \in \mathbb{R}$, and
- $\rho\left(u+u^{\prime}, v+v^{\prime}, P\right) \lesssim \rho(u, v, P)+\rho\left(u^{\prime}, v^{\prime}, P\right)$.

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 $\mathscr{A}_{q}^{s}\left(X_{0}\right) \equiv \mathscr{A}_{q}^{s}(\rho)$ and $\mathscr{A}^{s}\left(X_{0}\right) \equiv \mathscr{A}^{s}(\rho)$. These standard spaces are within the scope of the general theory of approximation spaces, cf., Pietsch (1981); DeVore and Lorentz (1993). However, to treat the more general spaces $\mathscr{A}^{s}(\rho)$, the standard theory needs to be reworked, which is the aim of this section.

We want to characterize $\mathscr{A}^{s}(\rho)$ 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 $X=B_{q, q}^{\alpha}$, with $\frac{\alpha}{n}>\frac{1}{q}-\frac{1}{p}$. We assume that the space $X$ has the following local structure: There exist a constant $0<q<\infty$, and a function $|\cdot|_{X(G)}: X \rightarrow \mathbb{R}^{+}$associated to each open set $G \subset \Omega$, such that

$$
\begin{equation*}
\sum_{k}|u|_{X\left(\tau_{k}\right)}^{q} \lesssim\|u\|_{X}^{q} \quad(u \in X) \tag{11}
\end{equation*}
$$

for any finite sequence $\left\{\tau_{k}\right\} \subset P$ of non-overlapping elements taken from any $P \in \mathscr{P}$. Finally, for any $\tau \in P$ with $P \in \mathscr{P}$, 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

$$
\begin{equation*}
\sum_{k}|u|_{X\left(P_{k}\left(\tau_{k}\right)\right)}^{q} \lesssim\|u\|_{X}^{q} \quad(u \in X) \tag{12}
\end{equation*}
$$

for any finite sequences $\left\{P_{k}\right\} \subset \mathscr{P}$ and $\left\{\tau_{k}\right\}$, with $\tau_{k} \in P_{k}$ and $\left\{\tau_{k}\right\}$ non-overlapping. A trivial example of such a structure is $X=L_{q}(\Omega)$ with $|\cdot|_{X(G)}=\|\cdot\|_{L_{q}(G)}$. Here the subadditivity (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 $X \subset \mathscr{A}^{s}(\rho)$ can be proved by exhibiting a direct estimate. A direct application of this criterion is mainly useful for deriving embeddings of the form $X \subset \mathscr{A}^{s}\left(X_{0}\right)$. 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 $\delta>0$. Assume (6) on the complexity of completion, and assume (12) on the local structure of $X$. Then for any $k \in \mathbb{N}$ sufficiently large there exists a partition $P \in \mathscr{P}$ with $\# P \leq k$ satisfying

$$
\begin{equation*}
\left(\sum_{\tau \in P}|\tau|^{p \delta}|u|_{X(\hat{\tau})}^{p}\right)^{\frac{1}{p}} \lesssim k^{-s}\|u\|_{X}, \tag{13}
\end{equation*}
$$

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

$$
\begin{equation*}
E\left(u, S_{P}\right)_{\rho} \lesssim\left(\sum_{\tau \in P}|\tau|^{p \delta}|u|_{X(\hat{\tau})}^{p}\right)^{\frac{1}{p}}, \tag{14}
\end{equation*}
$$

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

$$
\begin{equation*}
e(\tau, P)=|\tau|^{p \delta}|u|_{X(\hat{\tau})}^{p}, \tag{15}
\end{equation*}
$$

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

$$
\begin{equation*}
\sum_{\tau \in \mathscr{P}} e(\tau, P) \leq c^{\prime}(\# P) \varepsilon \tag{16}
\end{equation*}
$$

and

$$
\begin{equation*}
\# P-\# P_{0} \leq c \varepsilon^{-1 /(1+p s)}\|u\|_{X}^{p /(1+p s)}, \tag{17}
\end{equation*}
$$

where $c^{\prime}$ 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

$$
\begin{equation*}
\varepsilon=(c / k)^{1+p s}\|u\|_{X}^{p}, \tag{18}
\end{equation*}
$$

we would be able to guarantee a partition $P \in \mathscr{P}$ satisfying $\# P \leq \# P_{0}+k$ and

$$
\begin{equation*}
\sum_{\tau \in \mathscr{P}} e(\tau, P) \lesssim k^{-s p}\|u\|_{X}^{p} \tag{19}
\end{equation*}
$$

This would imply the lemma, as $k^{-s}$ can be replaced by $\left(\# P_{0}+k\right)^{-s}$ for, e.g., $k \geq \# P_{0}$.
Let $\varepsilon>0$ and let $P_{0} \in \mathscr{P}$. We then recursively define $R_{k}=\left\{\tau \in P_{k}: e\left(\tau, P_{k}\right)>\varepsilon\right\}$ and $P_{k+1}=\operatorname{refine}\left(P_{k}, R_{k}\right)$ for $k=0,1, \ldots$. For all sufficiently large $k$ we will have $R_{k}=\varnothing$ since $|u|_{X(\hat{\tau})} \lesssim\|u\|_{X}$ 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 $R_{k}=\varnothing$. Then recalling (15), and taking into account that $e\left(\tau, P_{k}\right) \leq \varepsilon$ 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 $\Lambda_{j}=\left\{\tau \in R: \mu^{j+1} \leq|\tau|<\mu^{j}\right\}$ 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 $\left(\mu, \mu^{-1}\right)$. This gives the trivial bound

$$
\begin{equation*}
m_{j} \leq \mu^{-j-1}|\Omega| . \tag{20}
\end{equation*}
$$

On the other hand, we have $e\left(\tau, P_{k}\right)>\varepsilon$ for $\tau \in \Lambda_{j}$ with some $k$, which gives

$$
\begin{equation*}
\varepsilon<|\tau|^{p \delta}|u|_{X(\hat{\tau})}^{p}<\mu^{j p \delta}|u|_{X(\hat{\tau})}^{p}, \tag{21}
\end{equation*}
$$

where $\hat{\tau}$ is defined with respect to $P_{k}$, and $k$ may depend on $\tau$. Summing over $\tau \in \Lambda_{j}$, we get

$$
\begin{equation*}
m_{j} \varepsilon^{q / p} \leq \mu^{j q \delta} \sum_{\tau \in \Lambda_{j}}|u|_{X(\hat{\tau})}^{q} \lesssim \mu^{j q \delta}\|u\|_{X}^{q}, \tag{22}
\end{equation*}
$$

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

$$
\begin{equation*}
\# R \leq \sum_{j=-\infty}^{\infty} m_{j} \lesssim \sum_{j=-\infty}^{\infty} \min \left\{\mu^{-j}, \varepsilon^{-q / p} \mu^{j q \delta}\|u\|_{X}^{q}\right\} \lesssim \varepsilon^{-q /(p+p q \delta)}\|u\|_{X}^{q /(1+q \delta)} \tag{23}
\end{equation*}
$$

which, in view of $(6)$ and $q /(1+q \delta)=p /(1+p s)$, establishes the bound (17).
We only sketch the case $p=\infty$ since the proof is essentially the same. We use

$$
\begin{equation*}
e(\tau, P)=|\tau|^{\delta}|u|_{X(\hat{\tau})}, \tag{24}
\end{equation*}
$$

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

$$
\begin{equation*}
\max _{\tau \in P} e(\tau, P) \leq \varepsilon . \tag{25}
\end{equation*}
$$

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

$$
\begin{equation*}
\# P-\# P_{0} \leq c \varepsilon^{-q /(1+q \delta)}\|u\|_{X}^{q /(1+q \delta)}=c \varepsilon^{-1 / s}\|u\|_{X}^{1 / s}, \tag{26}
\end{equation*}
$$

where $s=\delta+\frac{1}{q}$. The proof is complete.
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 $\mathscr{A}^{s}\left(X_{0}\right)$, 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 $\mathscr{P}$ to be the set of conforming triangulations of $\Omega$ obtained from a fixed conforming triangulation $P_{0}$ by means of newest vertex bisections. For $P \in \mathscr{P}$, 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 $P \in \mathscr{P}$ and $\tau \in P$, let $\hat{\tau}=P(\tau)$ be the interior of $\bigcup\{\bar{\sigma}: \sigma \in P, \bar{\sigma} \cap \bar{\tau} \neq \varnothing\}$. 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 $0<\alpha<m+\max \left\{1, \frac{1}{q}\right\}$ and $\delta>0$, cf. Binev et al. (2002); Gaspoz and Morin (2013). Hence the preceding lemma immediately implies that $B_{q, q}^{\alpha}(\Omega) \hookrightarrow \mathscr{A}^{s}\left(L_{p}(\Omega)\right)$ 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

$$
\begin{equation*}
K\left(u, t ; X_{0}, X\right)=\inf _{v \in X}\left(\|u-v\|_{X_{0}}+t\|v\|_{X}\right), \tag{27}
\end{equation*}
$$

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

$$
\begin{equation*}
|u|_{\left(X_{0}, X\right)_{\theta, \gamma}}=\left\|\left[\lambda^{\theta m} K\left(u, \lambda^{-m} ; X_{0}, X\right)\right]_{m \geq 0}\right\|_{\ell_{\gamma}} \tag{28}
\end{equation*}
$$

is finite. These are quasi-Banach spaces with the quasi-norms $\|\cdot\|_{X_{0}}+|\cdot|_{\left(X_{0}, X\right)_{\theta, \gamma}}$. 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 $s=\delta+\frac{1}{q}-\frac{1}{p}>0$. Assume

$$
\begin{equation*}
E\left(u, S_{P}\right)_{X_{0}} \lesssim\left(\sum_{\tau \in P}|\tau|^{p \delta}|u|_{X(\hat{\tau})}^{p}\right)^{\frac{1}{p}} \tag{29}
\end{equation*}
$$

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

Proof. For $u \in\left(X_{0}, X\right)_{\alpha / s, \gamma}$ and $v \in X$, we have

$$
\begin{equation*}
E_{k}(u)_{X_{0}} \lesssim\|u-v\|_{X_{0}}+E_{k}(v)_{X_{0}} \tag{30}
\end{equation*}
$$

Theorem 2.2 yields $E_{k}(v)_{X_{0}} \lesssim 2^{-k s}\|v\|_{X}$, leading to

$$
\begin{equation*}
E_{k}(u)_{X_{0}} \lesssim\|u-v\|_{X_{0}}+2^{-k s}\|v\|_{X} . \tag{31}
\end{equation*}
$$

After minimizing over $v \in X$, the right hand side gives the $K$-functional $K\left(u, 2^{-k s} ; X_{0}, X\right)$, and (28) implies that $u \in \mathscr{A}_{\gamma}^{\alpha}\left(X_{0}\right)$.

### 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 $T u=f$. Hence it should not come as a surprise that we assume the existence of a continuous operator $T: X_{0} \rightarrow Y_{0}$, with $Y_{0}$ a quasi-Banach space. An example to keep in mind is the Laplace operator sending $H_{0}^{1}$ 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 $\frac{\sigma}{n}>\frac{1}{r}-\frac{1}{2}$.

It is obvious that $\mathscr{A}^{s}(\rho) \subset \mathscr{A}^{s}\left(X_{0}\right)$, 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 affirming embeddings such as $\mathscr{A}^{s}\left(X_{0}\right) \cap T^{-1}(Y) \subset \mathscr{A}^{s}(\rho)$.

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

$$
\begin{equation*}
S_{P}+S_{Q} \subset S_{P \oplus Q}, \quad \text { and } \quad \#(P \oplus Q) \lesssim \# P+\# Q, \tag{32}
\end{equation*}
$$

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

$$
\begin{equation*}
\rho(u, v, P \oplus Q) \lesssim \rho(u, v, P) \tag{33}
\end{equation*}
$$

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 nonconforming partitions satisfying a certain admissibility criterion, cf. Bonito and Nochetto (2010).
Theorem 2.5. Let $0<p \leq \infty, \delta>0$, and let $s=\delta+\frac{1}{r}-\frac{1}{p}>0$. Assume (6) on the complexity of completion, as well as (12) on the local structure of $Y$, with r replacing $q$ there. There are no additional assumptions at this point on $\rho$, except (33) and the obvious conditions we have imposed in §2.1. Suppose that $u \in \mathscr{A}^{s}\left(X_{0}\right) \cap T^{-1}(Y)$ satisfies

$$
\begin{equation*}
E\left(u, S_{P}\right)_{\rho} \lesssim E\left(u, S_{P}\right)_{X_{0}}+\left(\sum_{\tau \in P}|\tau|^{p \delta}|T u|_{Y(\hat{\tau})}^{p}\right)^{\frac{1}{p}} \tag{34}
\end{equation*}
$$

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

$$
\begin{equation*}
\left(\sum_{\tau \in P \oplus Q}|\tau|^{p \delta}|T u|_{Y(\hat{\tau})}^{p}\right)^{\frac{1}{p}} \lesssim\left(\sum_{\tau \in P}|\tau|^{p \delta}|T u|_{Y(\hat{\tau})}^{p}\right)^{\frac{1}{p}} \tag{35}
\end{equation*}
$$

for any $P, Q \in \mathscr{P}$. Then we have $u \in \mathscr{A}^{s}(\rho)$ with $|u|_{\mathscr{A}^{s}(\rho)} \lesssim|u|_{\mathscr{A}^{s}\left(X_{0}\right)}+\|T u\|_{Y}$.
Proof. Let $k \in \mathbb{N}$ be an arbitrary number. Then by definition of $\mathscr{A}^{s}\left(X_{0}\right)$, there exists a partition $P^{\prime} \in \mathscr{P}$ such that

$$
\begin{equation*}
E\left(u, S_{P^{\prime}}\right)_{X_{0}} \leq 2^{-k s}|u|_{\mathscr{A}^{s}\left(X_{0}\right)}, \quad \text { and } \quad \# P^{\prime} \leq 2^{k} N \tag{36}
\end{equation*}
$$

Similarly, by applying Theorem 2.2 with $Y$ in place of $X$, and with $T u$ in place $u$, we can generate a partition $P^{\prime \prime} \in \mathscr{P}$ such that

$$
\begin{equation*}
\left(\sum_{\tau \in P^{\prime \prime}}|\tau|^{p \delta}|T u|_{Y(\hat{\tau})}^{p}\right)^{\frac{1}{p}} \lesssim 2^{-k s}\|T u\|_{Y}, \quad \text { and } \quad \# P^{\prime \prime} \leq 2^{k} N \tag{37}
\end{equation*}
$$

Then for $P=P^{\prime} \oplus P^{\prime \prime}$ we have $\# P \lesssim 2^{k}$ by (32). Moreover, (35) together with the obvious monotonicity

$$
\begin{equation*}
E\left(u, S_{P}\right)_{X_{0}} \leq E\left(u, S_{P^{\prime}}\right)_{X_{0}} \tag{38}
\end{equation*}
$$

guarantee that the right hand side of (34) is bounded by a multiple of $2^{-k s}\left(|u|_{\mathscr{A}^{s}\left(X_{0}\right)}+\right.$ $\|T u\|_{Y}$, which completes the proof.

Remark 2.6. Suppose that we replace the condition (34) in the statement of the preceding theorem by the new condition

$$
\begin{equation*}
E\left(u, S_{P}\right)_{\rho} \lesssim E\left(u, S_{P}\right)_{X_{0}}+E\left(u, S_{P}\right)_{\rho_{1}}+\left(\sum_{\tau \in P}|\tau|^{p \delta}|T u|_{Y(\hat{\tau})}^{p}\right)^{\frac{1}{p}} \tag{39}
\end{equation*}
$$

where $\rho_{1}$ is some distance function. Then by the same argument, we would be able to conclude that $u \in \mathscr{A}^{s}(\rho)$ with $|u|_{\mathscr{A}^{s}(\rho)} \lesssim|u|_{\mathscr{A}^{s}\left(X_{0}\right)}+|u|_{\mathscr{A}^{s}\left(\rho_{1}\right)}+\|T u\|_{Y}$. We will use similar strtaightforward extensions of the preceding theorem later in the paper, for instance, in 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 $\mathscr{P}$ as in Example 2.3, and for $P \in \mathscr{P}$, let $S_{P}$ be the Lagrange finite element space of continuous piecewise polynomials of degree not exceeding $m$, with the homogeneous Dirichlet boundary condition. Moreover, we set $T=\Delta$ the Laplace operator, sending $X_{0}=H_{0}^{1}(\Omega)$ onto $Y_{0}=H^{-1}(\Omega)$. Then in this
context, it is proved in Cascon et al. (2008) that certain adaptive finite element methods converge optimally with respect to the approximation classes $\mathscr{A}^{s}(\rho)$, with the distance function $\rho$ given by

$$
\begin{equation*}
\rho(u, v, P)=\left(\|u-v\|_{H^{1}}^{2}+\sum_{\tau \in P}|\tau|^{2 / n}\left\|f-\Pi_{\tau} f\right\|_{L_{2}(\tau)}^{2}\right)^{\frac{1}{2}} \tag{40}
\end{equation*}
$$

where $f=\Delta u$, and $\Pi_{\tau}: L_{2}(\tau) \rightarrow \mathbb{P}_{d}$ is the $L_{2}(\tau)$-orthogonal projection onto $\mathbb{P}_{d}$, 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 $\alpha<d+\max \left\{1, \frac{1}{r}\right\}$. Then we claim that for each $u \in H_{0}^{1}(\Omega)$ with $\Delta u \in B_{r, r}^{\alpha}(\Omega)$, there exists $u_{P} \in S_{P}$ such that

$$
\begin{equation*}
\rho\left(u, u_{P}, P\right) \lesssim \inf _{v \in S_{P}}\|u-v\|_{H^{1}}+\left(\sum_{\tau \in P}|\tau|^{2(\delta+1 / n)}|\Delta u|_{B_{r, r}^{\alpha}(\tau)}^{2}\right)^{\frac{1}{2}} \tag{41}
\end{equation*}
$$

for all $P \in \mathscr{P}$. In light of the preceding theorem, this would imply that each function $u \in \mathscr{A}^{s}\left(H_{0}^{1}(\Omega)\right)$ with $\Delta u \in B_{r, r}^{\alpha}(\Omega)$, satisfies $u \in \mathscr{A}^{s}(\rho)$, cf. Figure $1(\mathrm{~b})$. Note that since we can choose $d$ at will, the restriction $\alpha<d+\max \left\{1, \frac{1}{r}\right\}$ 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

$$
\begin{equation*}
\left\|u-u_{P}\right\|_{H^{1}} \lesssim \inf _{v \in S_{P}}\|u-v\|_{H^{1}} \tag{42}
\end{equation*}
$$

for all $P \in \mathscr{P}$. The oscillation term in (40) can be estimated as

$$
\begin{equation*}
\left\|f-\Pi_{\tau} f\right\|_{L_{2}(\tau)} \leq\|f-g\|_{L_{2}(\tau)} \lesssim|\tau|^{\delta}\|f-g\|_{L_{r}(\tau)}+|\tau|^{\delta}|f|_{B_{r, r}^{\alpha}(\tau)} \tag{43}
\end{equation*}
$$

for any $g \in \mathbb{P}_{d}$, where we have used continuity of the embedding $B_{r, r}^{\alpha}(\tau) \subset L_{2}(\tau)$ and the fact that $|g|_{B_{r, r}^{\alpha}(\tau)}=0$ 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

$$
\begin{equation*}
\|f-g\|_{L_{r}(\tau)} \lesssim \omega_{d+1}(f, \tau)_{r} \lesssim|f|_{B_{r, r}^{\alpha}(\tau)} \tag{44}
\end{equation*}
$$

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 $B_{r, r}^{\alpha}(\tau) \subset L_{2}(\tau)$ 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.

## 3 Lagrange finite elements

### 3.1 Preliminaries

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

$$
\begin{equation*}
\omega_{r}(u, t, \Omega)_{p}=\sup _{|h| \leq t}\left\|\Delta_{h}^{r} u\right\|_{L_{p}\left(\Omega_{r h}\right)} \tag{45}
\end{equation*}
$$

where $\Omega_{r h}=\{x \in \Omega:[x, x+r h] \subset \Omega\}$, and $\Delta_{h}^{r}$ is the $r$-th order forward difference operator defined recursively by $\left[\Delta_{h}^{1} u\right](x)=u(x+h)-u(x)$ and $\Delta_{h}^{k} u=\Delta_{h}^{1}\left(\Delta_{h}^{k-1}\right) u$, i.e.,

$$
\begin{equation*}
\Delta_{h}^{r} u(x)=\sum_{k=0}^{r}(-1)^{r+k}\binom{r}{k} u(x+k h) \tag{46}
\end{equation*}
$$

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

$$
\begin{equation*}
|u|_{B_{p, q ; r}^{\alpha}(\Omega)}=\left\|t \mapsto t^{-\alpha-1 / q} \omega_{r}(u, t, \Omega)_{p}\right\|_{L_{q}((0, \infty))} \tag{47}
\end{equation*}
$$

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

$$
\begin{equation*}
|u|_{B_{p, q ; r}^{\alpha}(\Omega)} \approx\left\|\left(\lambda^{j \alpha} \omega_{r}\left(u, \lambda^{-j}, \Omega\right)_{p}\right)_{j \geq 0}\right\|_{\ell_{q}} \tag{48}
\end{equation*}
$$

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 $\alpha>r+\max \left\{0, \frac{1}{p}-1\right\}$ 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 $r>\alpha-\max \left\{0, \frac{1}{p}-1\right\}$, 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 $\alpha=r+\max \left\{0, \frac{1}{p}-1\right\}$, then $B_{p, q ; r}^{\alpha}=\mathbb{P}_{r-1}$. The case $q=\infty$ gives nontrivial spaces: For instance, we have $B_{p, \infty ; r}^{r}(\Omega)=W^{r, p}(\Omega)$ 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.

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

Lemma 3.1. Let $\left(a_{j}\right)_{j \in \mathbb{Z}}$ and $\left(b_{k}\right)_{k \in \mathbb{Z}}$ be two sequences satisfying either

$$
\begin{equation*}
\left|a_{j}\right| \leq C\left(\sum_{k=j}^{\infty}\left|b_{k}\right|^{\mu}\right)^{1 / \mu}, \quad j \in \mathbb{Z} \tag{49}
\end{equation*}
$$

for some $\mu>0$ and $C>0$, or

$$
\begin{equation*}
\left|a_{j}\right| \leq C 2^{-\theta j}\left(\sum_{k=-\infty}^{j}\left|2^{\theta k} b_{k}\right|^{\mu}\right)^{1 / \mu}, \quad j \in \mathbb{Z} \tag{50}
\end{equation*}
$$

for some positive $\theta, \mu$, and $C$. Then we have

$$
\begin{equation*}
\left\|\left(2^{\alpha j} a_{j}\right)_{j}\right\|_{\ell_{q}} \lesssim C\left\|\left(2^{\alpha k} b_{k}\right)_{k}\right\|_{\ell_{q}} \tag{51}
\end{equation*}
$$

for all $0<q \leq \infty$ and $0<\alpha<\theta$, with the convention that $\theta=\infty$ if (49) holds, and with the implicit constant depending only on $q$ and $\alpha$.

### 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. This ensures that all partitions are shape regular. The set $\mathscr{P}$ can be chosen to be, in case of the newest vertex bisection, the set of all conforming triangulations arising from $P_{0}$. More generally, we want to deal with possibly nonconforming partitions, and we require that $\mathscr{P}$ satisfy the finite support condition (53) stated below. We remark that the results in this paper are insensitive to the exact definition of $\mathscr{P}$, so long as the family $\mathscr{P}$ satisfies (53). We note that for nonconforming partitions, 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

$$
\begin{equation*}
S_{P}=S_{P}^{m}=\left\{u \in C(\Omega):\left.u\right|_{\tau} \in \mathbb{P}_{m} \forall \tau \in P\right\}, \quad P \in \mathscr{P} \tag{52}
\end{equation*}
$$

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 $\tilde{Q}_{P}: L_{p_{0}}(\Omega) \rightarrow S_{P}$ for $p_{0}>0$ small. Their construction works almost verbatim here but we need to be a bit careful since we want to include partitions with hanging nodes into the analysis. Let $\tau_{0}=\left\{x \in \mathbb{R}^{n}: x_{1}+\ldots+x_{n}<m\right\} \cap(0, m)^{n}$ 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 $N_{\tau}=F\left(\bar{\tau}_{0} \cap \mathbb{Z}^{n}\right)$, where $F: \tau_{0} \rightarrow \tau$ is any invertible affine mapping. The nodal set $N_{P}$ of a possibly nonconforming partition $P \in \mathscr{P}$ 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 $\tau$ satisfying $\bar{\tau} \ni z$, see Figure 2. Furthermore, we define the nodal basis $\left\{\phi_{z}: z \in N_{P}\right\} \subset S_{P}$ of $S_{P}$ by $\phi_{z}\left(z^{\prime}\right)=\delta_{z, z^{\prime}}$ for $z, z^{\prime} \in N_{P}$.

The aforementioned finite support condition is as follows. We require that there is a constant $C>0$ independent of $P \in \mathscr{P}$ and $z \in N_{P}$, such that

$$
\begin{equation*}
\#\left\{\tau \in P: \tau \subset \operatorname{supp} \phi_{z}\right\} \leq C \tag{53}
\end{equation*}
$$

for all $P \in \mathscr{P}$ and $z \in N_{P}$. It is obvious that conforming triangulations satisfy this requirement. For partitions with hanging nodes, we refer to Bonito and Nochetto (2010).

Given $\tau \in P$ and $\sigma \in P$, let us write $\tau \sim \sigma$ if (and only if) there is $z \in N_{P}$ such that $\tau \cup \sigma \subset \operatorname{supp} \psi_{z}$. Then, taking into account the refinement rule and the definition of the nodal basis, one can show that the finite support condition is equivalent to the strong local finiteness condition:

$$
\begin{equation*}
\sup _{P \in \mathscr{P}} \sup _{\tau \in P} \#\{\sigma \in P: \sigma \sim \tau\}<\infty \tag{54}
\end{equation*}
$$

which in turn is equivalent to the strong gradedness condition:

$$
\begin{equation*}
\sup _{P \in \mathscr{P}}\left\{\frac{\operatorname{diam} \sigma}{\operatorname{diam} \tau}: \tau, \sigma \in P, \sigma \sim \tau\right\}<\infty . \tag{55}
\end{equation*}
$$

By defining the support extension $\hat{\tau}=P(\tau)$ of $\tau \in P$ as the interior of

$$
\begin{equation*}
\bigcup\{\sigma \in P: \sigma \sim \tau\} \tag{56}
\end{equation*}
$$

we can also write the strong gradedness condition as

$$
\begin{equation*}
\sup _{P \in \mathscr{P}} \sup _{\tau \in P} \frac{\operatorname{diam} \hat{\tau}}{\operatorname{diam} \tau}<\infty \tag{57}
\end{equation*}
$$

In what follows, the implicit constant in any of the aforementioned conditions will be referred to as an admissibility constant.

(a) Quadratic elements $(m=2)$.

(b) Cubic elements $(m=3)$.

Fig. 2: Examples of nodal sets.
Next, we introduce a basis dual to $\left\{\phi_{z}\right\}$. For each $\tau \in P$, we let

$$
\begin{equation*}
N_{P, \tau}=\left\{z \in N_{P}: \tau \subset \operatorname{supp} \phi_{z}\right\} \tag{58}
\end{equation*}
$$

and define $\eta_{\tau, z} \in \mathbb{P}_{m}, z \in N_{P, \tau}$, by the condition

$$
\begin{equation*}
\int_{\tau} \eta_{\tau, z} \xi_{\tau, z^{\prime}}=\delta_{z, z^{\prime}}, \quad z, z^{\prime} \in N_{P, \tau} \tag{59}
\end{equation*}
$$

Note that $\# N_{P, \tau}=\# N_{\tau}=\operatorname{dim} \mathbb{P}_{m}$, so that the set $\left\{\eta_{\tau, z}: z \in N_{P, \tau}\right\}$ is uniquely determined. Then for $z \in N_{P}$, we let

$$
\begin{equation*}
\tilde{\phi}_{z}=\frac{1}{n_{z}} \sum_{\left\{\tau \in P: \tau \subset \operatorname{supp} \phi_{z}\right\}} \chi_{\tau} \eta_{\tau, z} \tag{60}
\end{equation*}
$$

where $n_{z}=\#\left\{\tau \underset{\sim}{\in} P: \tau \subset \operatorname{supp} \phi_{z}\right\}$, and $\chi_{\tau}$ is the characteristic function of $\tau$. By construction, $\operatorname{supp} \tilde{\phi}_{z}=\operatorname{supp} \phi_{z}$ for $z \in N_{P}$, and we have the biorthogonality

$$
\begin{equation*}
\left\langle\tilde{\phi}_{z}, \phi_{z^{\prime}}\right\rangle=\int_{\Omega} \tilde{\phi}_{z} \phi_{z^{\prime}}=\delta_{z, z^{\prime}}, \quad z, z^{\prime} \in N_{P} \tag{61}
\end{equation*}
$$

Now we define the quasi-interpolation operator $Q_{P}: L_{1}(\Omega) \rightarrow S_{P}$ by

$$
\begin{equation*}
Q_{P} u=Q_{P}^{(\Omega)} u=\sum_{z \in N_{P}}\left\langle u, \tilde{\phi}_{z}\right\rangle \phi_{z} \tag{62}
\end{equation*}
$$

It is clear that $Q_{P}$ is linear and that $Q_{P} v=v$ for $v \in S_{P}$.
Lemma 3.2. For $1 \leq p \leq \infty$, we have

$$
\begin{equation*}
\left\|u-Q_{P} u\right\|_{L_{p}(\Omega)} \lesssim \inf _{v \in S_{P}}\|u-v\|_{L_{p}(\Omega)}, \quad u \in L_{p}(\Omega) \tag{63}
\end{equation*}
$$

with the implicit constant depending only on the shape regularity and admissibility constants of $\mathscr{P}$. Furthermore, for $0<p \leq \infty$ and $\tau \in P$, we have

$$
\begin{equation*}
\left\|Q_{P} v\right\|_{L_{p}(\tau)} \lesssim\|v\|_{L_{p}(\hat{\tau})}, \quad v \in \bar{S}_{P}^{m} \tag{64}
\end{equation*}
$$

where

$$
\begin{equation*}
\bar{S}_{P}^{m}=\left\{w \in L_{\infty}(\Omega):\left.w\right|_{\tau} \in \mathbb{P}_{m} \forall \tau \in P\right\} \tag{65}
\end{equation*}
$$

and $\hat{\tau}=P(\tau)$ is the support extension of $\tau$ as defined in (56).
Proof. For $1 \leq p \leq \infty$ and $u \in L_{p}(\Omega)$, we have

$$
\begin{equation*}
\left\|Q_{P} u\right\|_{L_{p}(\tau)} \leq \sum_{z \in N_{P, \tau}}\left|\left\langle u, \tilde{\phi}_{z}\right\rangle\right|\left\|\phi_{z}\right\|_{L_{p}} \leq\|u\|_{L_{p}(\hat{\tau})} \sum_{z \in N_{P, \tau}}\left\|\tilde{\phi}_{z}\right\|_{L_{q}}\left\|\phi_{z}\right\|_{L_{p}} \tag{66}
\end{equation*}
$$

where $\frac{1}{p}+\frac{1}{q}=1$. It is clear that $\left\|\phi_{z}\right\|_{L_{p}} \leq\left|\operatorname{supp} \phi_{z}\right|^{1 / p}$ and by a scaling argument one can deduce that $\left\|\tilde{\phi}_{z}\right\|_{L_{q}} \lesssim\left|\operatorname{supp} \tilde{\phi}_{z}\right|^{1 / q-1}$, with the implicit constant depending only on the shape regularity and admissibility constants of $\mathscr{P}$. Consequently, for $1 \leq p<\infty$, we infer

$$
\begin{equation*}
\left\|Q_{P} u\right\|_{L_{p}(\Omega)}^{p}=\sum_{\tau \in P}\left\|Q_{P} u\right\|_{L_{p}(\tau)}^{p} \lesssim \sum_{\tau \in P}\|u\|_{L_{p}(\hat{\tau})}^{p} \lesssim\|u\|_{L_{p}(\Omega)}^{p} \tag{67}
\end{equation*}
$$

by the strong local finiteness of the mesh. The case $p=\infty$ can be handled similarly, and we have $\left\|Q_{P} u\right\|_{L_{p}(\Omega)} \lesssim\|u\|_{L_{p}(\Omega)}$ for $1 \leq p \leq \infty$. Then a standard argument yields (63).

For $1 \leq p \leq \infty$, (66) implies (64). 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

$$
\begin{equation*}
\left\|u-\Pi_{p_{0}, \tau} u\right\|_{L_{p_{0}}(\tau)} \leq C_{m, p_{0}} \inf _{v \in \mathbb{P}_{m}}\|u-v\|_{L_{p_{0}}(\tau)}, \quad u \in L_{p_{0}}(\tau) \tag{68}
\end{equation*}
$$

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

$$
\begin{equation*}
\left\|\Pi_{p_{0}, \tau} u\right\|_{L_{p_{0}}(\tau)} \lesssim\|u\|_{L_{p_{0}}(\tau)}, \quad u \in L_{p_{0}}(\tau) \tag{69}
\end{equation*}
$$

i.e., the operator $\Pi_{p_{0}, \tau}: L_{p_{0}}(\tau) \rightarrow L_{p_{0}}(\tau)$ is bounded.
(iii) For any $u \in L_{p_{0}}(\tau)$ and $v \in \mathbb{P}_{m}$, we have

$$
\begin{equation*}
\Pi_{p_{0}, \tau}(u+v)=\Pi_{p_{0}, \tau} u+v \tag{70}
\end{equation*}
$$

In particular, $\Pi_{p_{0}, \tau} v=v$ for $v \in \mathbb{P}_{m}$.
Finally, we let

$$
\begin{equation*}
\Pi_{P} u=\sum_{\tau \in P} \chi_{\tau} \Pi_{p_{0}, \tau} u \tag{71}
\end{equation*}
$$

and define the operator $\tilde{Q}_{P}: L_{p_{0}}(\Omega) \rightarrow S_{P}$ by

$$
\begin{equation*}
\tilde{Q}_{P} u=Q_{P} \Pi_{P} u=\sum_{z \in N_{P}}\left\langle\Pi_{P} u, \tilde{\phi}_{z}\right\rangle \phi_{z} \tag{72}
\end{equation*}
$$

It is easy to see that $\tilde{Q}_{P} v=v$ for $v \in S_{P}$, and that $\left.\left(\tilde{Q}_{P} u\right)\right|_{\tau}$ depends only on $\left.u\right|_{\hat{\tau}}$, where $\hat{\tau}=P(\tau)$ is the support extension of $\tau$, as defined in (56). Furthermore, as a consequence of the linearity property (70), we have

$$
\begin{equation*}
\left.\left(\tilde{Q}_{P}(u+v)\right)\right|_{\tau}=\left.\left(\tilde{Q}_{P} u\right)\right|_{\tau}+\left.v\right|_{\tau}, \quad u, v \in L_{p_{0}}(\tau),\left.\quad v\right|_{\hat{\tau}} \in S_{P} \tag{73}
\end{equation*}
$$

for $\tau \in P$.
Lemma 3.3. Let $p_{0} \leq p \leq \infty$ and $P \in \mathscr{P}$. Then for $\tau \in P$ we have

$$
\begin{equation*}
\left\|\tilde{Q}_{P} u\right\|_{L_{p}(\tau)} \lesssim\|u\|_{L_{p}(\hat{\tau})}, \quad u \in L_{p}(\Omega) \tag{74}
\end{equation*}
$$

As a consequence, we have

$$
\begin{equation*}
\left\|u-\tilde{Q}_{P} u\right\|_{L_{p}(\tau)} \lesssim \inf _{v \in S_{P}}\|u-v\|_{L_{p}(\hat{\tau})}, \quad u \in L_{p}(\Omega) \tag{75}
\end{equation*}
$$

and

$$
\begin{equation*}
\left\|u-\tilde{Q}_{P} u\right\|_{L_{p}(\Omega)} \lesssim \inf _{v \in S_{P}}\|u-v\|_{L_{p}(\Omega)}, \quad u \in L_{p}(\Omega) \tag{76}
\end{equation*}
$$

Proof. An application of (64) gives

$$
\begin{equation*}
\left\|\tilde{Q}_{P} u\right\|_{L_{p}(\tau)}=\left\|Q_{P} \Pi_{P} u\right\|_{L_{p}(\tau)} \lesssim\left\|\Pi_{P} u\right\|_{L_{p}(\hat{\tau})} \tag{77}
\end{equation*}
$$

On the other hand, for $\sigma \in P$, we have

$$
\begin{equation*}
\left\|\Pi_{p_{0}, \sigma} u\right\|_{L_{p}(\sigma)} \lesssim|\sigma|^{\frac{1}{p}-\frac{1}{p_{0}}}\left\|\Pi_{p_{0}, \sigma} u\right\|_{L_{p_{0}}(\sigma)} \lesssim|\sigma|^{\frac{1}{p}-\frac{1}{p_{0}}}\|u\|_{L_{p_{0}}(\sigma)} \leq\|u\|_{L_{p}(\sigma)} \tag{78}
\end{equation*}
$$

where we have used scaling properties of polynomials in the first step, the boundedness (69) 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

$$
\begin{equation*}
\left\|\Pi_{P} u\right\|_{L_{p}(\hat{\tau})}=\left(\sum_{\{\sigma \in P: \sigma \subset \hat{\tau}\}}\left\|\Pi_{p_{0}, \sigma} u\right\|_{L_{p}(\sigma)}^{p}\right)^{\frac{1}{p}} \lesssim\left(\sum_{\{\sigma \in P: \sigma \subset \hat{\tau}\}}\|u\|_{L_{p}(\sigma)}^{p}\right)^{\frac{1}{p}}=\|u\|_{L_{p}(\hat{\tau})} \tag{79}
\end{equation*}
$$

establishing (74). Then (75) follows from the linearity property (73).
The estimate (76) is proved by first deriving the stability

$$
\begin{equation*}
\left\|\tilde{Q}_{P} u\right\|_{L_{p}(\Omega)}=\left(\sum_{\tau \in P}\left\|\tilde{Q}_{P} u\right\|_{L_{p}(\tau)}^{p}\right)^{\frac{1}{p}} \lesssim\left(\sum_{\tau \in P}\|u\|_{L_{p}(\hat{\tau})}^{p}\right)^{\frac{1}{p}} \lesssim\|u\|_{L_{p}(\Omega)} \tag{80}
\end{equation*}
$$

and then invoking the linearity property (73).
An important tool in approximation theory is the Whitney estimate

$$
\begin{equation*}
\inf _{v \in \mathbb{P}_{m}}\|u-v\|_{L_{p}(G)} \lesssim \omega_{m+1}(u, \operatorname{diam} G, G)_{p}, \quad u \in L_{p}(G) \tag{81}
\end{equation*}
$$

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 $P \in \mathscr{P}$, with the implicit constant additionally depending on the shape regularity constant of $\mathscr{P}$, see Gaspoz and Morin (2013).

Lemma 3.4. Let $p_{0} \leq p \leq \infty$ and let $P \in \mathscr{P}$ be conforming. Then we have

$$
\begin{equation*}
\left\|u-\tilde{Q}_{P} u\right\|_{L_{p}(\Omega)} \lesssim\left(\frac{\max _{\tau \in P} \operatorname{diam} \tau}{\min _{\tau \in P} \operatorname{diam} \tau}\right)^{\frac{n}{p}} \omega_{m+1}\left(u, \max _{\tau \in P} \operatorname{diam} \tau, \Omega\right)_{p}, \quad u \in L_{p}(\Omega) \tag{82}
\end{equation*}
$$

Proof. We start with the special case $p=\infty$. Note that since $P$ is conforming, the support extension $\hat{\tau}$ of $\tau$ coincides with the star around $\tau$. It is immediate from (75) and the Whitney estimate (81) that

$$
\begin{align*}
\left\|u-\tilde{Q}_{P} u\right\|_{L_{\infty}(\Omega)} & \lesssim \max _{\tau \in P}\left\|u-\tilde{Q}_{P} u\right\|_{L_{\infty}(\tau)} \lesssim \max _{\tau \in P} \inf _{v \in S_{P}}\|u-v\|_{L_{\infty}(\hat{\tau})} \\
& \lesssim \max _{\tau \in P} \omega_{m+1}(u, \operatorname{diam} \hat{\tau}, \hat{\tau})_{\infty} \lesssim \max _{\tau \in P} \omega_{m+1}\left(u, \mu^{-1} \operatorname{diam} \hat{\tau}, \hat{\tau}\right)_{\infty} \tag{83}
\end{align*}
$$

with $\mu>0$ sufficiently large, where in the last line we have used the property

$$
\begin{equation*}
\omega_{r}(u, \mu t, G)_{p} \leq(\mu+1)^{r} \omega_{r}(u, t, G)_{p} \tag{84}
\end{equation*}
$$

cf. (DeVore and Lorentz, 1993, §2.7). With $t=\mu^{-1} \max _{\tau \in P} \operatorname{diam} \hat{\tau}$, we proceed as

$$
\begin{align*}
\left\|u-\tilde{Q}_{P} u\right\|_{L_{\infty}(\Omega)} & \lesssim \max _{\tau \in P} \omega_{m+1}(u, t, \hat{\tau})_{\infty}=\max _{\tau \in P} \sup _{|h| \leq t}\left\|\Delta_{h}^{m+1} u\right\|_{L_{\infty}\left(\hat{\tau}_{r h}\right)} \\
& =\sup _{|h| \leq t} \max _{\tau \in P}\left\|\Delta_{h}^{m+1} u\right\|_{L_{\infty}\left(\hat{\tau}_{r h}\right)} \leq \sup _{|h| \leq t}\left\|\Delta_{h}^{m+1} u\right\|_{L_{\infty}\left(\Omega_{r h}\right)}, \tag{85}
\end{align*}
$$

which establishes (82) for $p=\infty$.
To handle the case $0<p<\infty$ we introduce the averaged $L_{p}$-modulus of smoothness

$$
\begin{equation*}
w_{r}(u, t, G)_{p}=\left(\frac{1}{t^{n}} \int_{[0, t]^{n}}\left\|\Delta_{h}^{r} u\right\|_{L_{p}\left(G_{r h}\right)}^{p} \mathrm{~d} h\right)^{1 / p} \tag{86}
\end{equation*}
$$

for any domain $G \subset \mathbb{R}^{n}$. When $G$ is Lipschitz, the averaged modulus is equivalent to the original one:

$$
\begin{equation*}
w_{r}(u, t, G)_{p} \sim \omega_{r}(u, t, G)_{p}, \quad \text { for } \quad t \lesssim 1 \tag{87}
\end{equation*}
$$

This equivalence is also true when $G=\tau$ or $G=\hat{\tau}$ for $\tau \in P$ with $P \in \mathscr{P}$, in the range $t \lesssim \operatorname{diam} G$, 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 $\mathscr{P}$, and the geometry of the underlying domain $\Omega$.

Let us get back to the proof of (82) for $0<p<\infty$. As in the case $p=\infty$, we have

$$
\begin{align*}
\left\|u-\tilde{Q}_{P} u\right\|_{L_{p}(\Omega)}^{p} & \lesssim \sum_{\tau \in P}\left\|u-\tilde{Q}_{P} u\right\|_{L_{p}(\tau)}^{p} \lesssim \sum_{\tau \in P} \inf _{v \in S_{P}}\|u-v\|_{L_{p}(\hat{\tau})}^{p} \\
& \lesssim \sum_{\tau \in P} \omega_{m+1}(u, \operatorname{diam} \hat{\tau}, \hat{\tau})_{p}^{p} \lesssim \sum_{\tau \in P} \omega_{m+1}\left(u, \mu^{-1} \operatorname{diam} \hat{\tau}, \hat{\tau}\right)_{p}^{p} \tag{88}
\end{align*}
$$

with $\mu>0$ sufficiently large. Now we employ (87), to get

$$
\begin{align*}
\left\|u-\tilde{Q}_{P} u\right\|_{L_{p}(\Omega)}^{p} & \lesssim \sum_{\tau \in P} w_{m+1}\left(u, \mu^{-1} \operatorname{diam} \hat{\tau}, \hat{\tau}\right)_{p}^{p} \\
& =\sum_{\tau \in P} \frac{1}{t(\tau)^{n}} \int_{[0, t(\tau)]^{n}} \int_{\hat{\tau}_{r h}}\left|\Delta_{h}^{r} u(x)\right|^{p} \mathrm{~d} x \mathrm{~d} h \tag{89}
\end{align*}
$$

where $t(\tau)=\mu^{-1} \operatorname{diam} \hat{\tau}$ and $r=m+1$. With $t_{0}=\mu^{-1} \min _{\tau \in P} \operatorname{diam} \hat{\tau}$ and $t_{1}=\mu^{-1} \max _{\tau \in P} \operatorname{diam} \hat{\tau}$,
we can switch the sum with the outer integration as follows.

$$
\begin{align*}
\left\|u-\tilde{Q}_{P} u\right\|_{L_{p}(\Omega)}^{p} & \lesssim \frac{1}{t_{0}^{n}} \int_{\left[0, t_{1}\right]^{n}} \sum_{\tau \in P} \int_{\hat{\tau}_{r h}}\left|\Delta_{h}^{r} u(x)\right|^{p} \mathrm{~d} x \mathrm{~d} h \\
& \lesssim \frac{1}{t_{0}^{n}} \int_{\left[0, t_{1}\right]^{n}} \int_{\Omega_{r h}}\left|\Delta_{h}^{r} u(x)\right|^{p} \mathrm{~d} x \mathrm{~d} h  \tag{90}\\
& =\frac{t_{1}^{n}}{t_{0}^{n}} w_{r}\left(u, t_{1}, \Omega\right)_{p}^{p}
\end{align*}
$$

The proof is completed upon using the equivalence (87) 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

$$
\begin{equation*}
E\left(u, S_{j}\right)_{L_{p}(G)}=\inf _{v \in S_{j}}\|u-v\|_{L_{p}(G)}, \quad u \in L_{p}(G) \tag{91}
\end{equation*}
$$

Note that the infimum is achieved since $S_{j}$ is a finite dimensional space. We define the multilevel approximation spaces

$$
\begin{equation*}
A_{p, q}^{\alpha}\left(\left\{S_{j}\right\}, G\right)=\left\{u \in L_{p}(G):|u|_{A_{p, q}^{\alpha}(G)}:=\left\|\left(\lambda^{j \alpha} E\left(u, S_{j}\right)_{L_{p}(G)}\right)_{j \geq 0}\right\|_{\ell_{q}}<\infty\right\} \tag{92}
\end{equation*}
$$

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

$$
\begin{equation*}
A_{p, q}^{\alpha}(G)=A_{p, q ; m}^{\alpha}(G)=A_{p, q}^{\alpha}\left(\left\{S_{j}\right\}, G\right) \tag{93}
\end{equation*}
$$

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

It is no coincidence that the aforementioned embedding relations are identical to those among Besov spaces. When reading the following theorem, keep in mind that $B_{p, q ; m+1}^{\alpha}(\Omega)$ is the classical Besov space $B_{p, q}^{\alpha}(\Omega)$ for $\alpha<m+\max \left\{1, \frac{1}{p}\right\}$.

Theorem 3.5. We have $B_{p, q ; m+1}^{\alpha}(\Omega) \hookrightarrow A_{p, q ; m}^{\alpha}(\Omega)$ for $0<p, q \leq \infty$, and $\alpha>0$. In the other direction, we have $A_{p, q ; m}^{\alpha,}(\Omega) \hookrightarrow B_{p, q ; m+1}^{\alpha}(\Omega)$ for $0<p, q \leq \infty$, and $0<\alpha<1+\frac{1}{p}$.

Proof. We follow the standard approach. The inclusion $B_{p, q ; m+1}^{\alpha}(\Omega) \hookrightarrow A_{p, q ; m}^{\alpha}(\Omega)$ is a direct consequence of (82) with $p_{0} \leq p$ and the norm equivalence (48).

For the second part, we start with the estimate

$$
\begin{equation*}
\omega_{m+1}\left(\phi_{z}, t\right)_{p} \lesssim \lambda^{-j n / p} \min \left\{1,\left(\lambda^{j} t\right)^{1+1 / p}\right\}, \tag{94}
\end{equation*}
$$

which holds for all nodal basis functions $\phi_{z}$ of $S_{j}$ and for all $j \geq 0$. This is Proposition 4.7 in Gaspoz and Morin (2013), which also holds for $p=\infty$. Hence for $0<p<\infty$ and for all $u_{j}=\sum_{z} b_{z} \phi_{z} \in S_{P_{j}}$, we infer

$$
\begin{align*}
\omega_{m+1}\left(u_{j}, t\right)_{p}^{p} & \lesssim \sum_{z}\left|b_{z}\right|^{p} \omega_{m+1}\left(\phi_{z}, t\right)_{p}^{p} \lesssim \sum_{z}\left|b_{z}\right|^{p} \lambda^{-j n} \min \left\{1,\left(\lambda^{j} t\right)^{p+1}\right\}  \tag{95}\\
& \lesssim \min \left\{1,\left(\lambda^{j} t\right)^{p+1}\right\}\left\|u_{j}\right\|_{L_{p}(\Omega)}^{p},
\end{align*}
$$

where we have used the finite overlap property of the nodal basis functions, the $L_{p}$ stability of finite elements and the estimate $\left\|\phi_{z}\right\|_{L_{p}} \bar{\sim} \lambda^{-j n / p}$. The same ingredients are used to perform the corresponding computation for $p=\infty$, as

$$
\begin{align*}
\omega_{m+1}\left(u_{j}, t\right)_{\infty} & \lesssim \max _{z}\left|b_{z}\right| \omega_{m+1}\left(\phi_{z}, t\right)_{\infty} \lesssim \max _{z}\left|b_{z}\right| \min \left\{1, \lambda^{j} t\right\}  \tag{96}\\
& \lesssim \min \left\{1, \lambda^{j} t\right\}\left\|u_{j}\right\|_{L_{\infty}(\Omega)} .
\end{align*}
$$

Now we write $u=\sum_{j \geq 0}\left(u_{j}-u_{j-1}\right)$ 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 (82). With $p^{*}=\min \{1, p\}$, we have

$$
\begin{align*}
\omega_{m+1}\left(u, \lambda^{-k}\right)_{p}^{p^{*}} & \lesssim \sum_{j \geq 0} \omega_{m+1}\left(u_{j}-u_{j-1}, \lambda^{-k}\right)_{p}^{p^{*}} \\
& \lesssim \sum_{j=0}^{k} \lambda^{(j-k)(1+1 / p) p^{*}}\left\|u_{j}-u_{j-1}\right\|_{L_{p}(\Omega)}^{p^{*}}+\sum_{j=k+1}^{\infty}\left\|u_{j}-u_{j-1}\right\|_{L_{p}(\Omega)}^{p^{*}}, \tag{97}
\end{align*}
$$

and an application of the discrete Hardy inequality (Lemma 3.1) gives

$$
\begin{equation*}
|u|_{B_{p, q ; m+1}^{\alpha}} \lesssim\left\|\left(\lambda^{j \alpha}\left\|u_{j}-u_{j-1}\right\|_{L_{p}(\Omega)}\right)_{j}\right\|_{\ell_{q}}, \tag{98}
\end{equation*}
$$

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 $u_{j}-u_{j-1}=\left(u-u_{j-1}\right)-\left(u-u_{j}\right)$.

Notice the gap between the two inclusions: While $B_{p, q ; m+1}^{\alpha}(\Omega) \hookrightarrow A_{p, q ; m}^{\alpha}(\Omega)$ 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 $A_{p, q ; m}^{\alpha}(\Omega)$, but there are functions in $S_{j}$ that are not in $B_{p, q ; m+1}^{\alpha}(\Omega)$, because the


Fig. 3: The inverse embedding $A_{p, q ; m}^{\alpha} \hookrightarrow B_{p, q ; m+1}^{\alpha}$ holds below the dashed line. The direct embedding $B_{p, q ; m+1}^{\alpha} \hookrightarrow A_{p, q ; m}^{\alpha}$ holds without restriction, but the spaces $B_{p, q ; m+1}^{\alpha}$ are nontrivial (and coincide with the classical Besov spaces $B_{p, q}^{\alpha}$ ) only below the solid line.
estimate (94) is saturated for small $t$. This leads to the expectation that for large $\alpha$, the difference $A_{p, q ; m}^{\alpha}(\Omega) \backslash B_{p, q ; m+1}^{\alpha}(\Omega)$ 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 $B_{p, q ; m+1}^{\alpha}(\Omega)$ coincides with the intersection of all $A_{p, q ; m}^{\alpha}(\Omega)$ 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 $A_{p, q}^{\alpha}(\Omega)$ coincide with the spaces $\hat{B}_{p, q}^{\alpha}(\Omega)$ considered in Gaspoz and Morin (2013), cf. Definition 7.1 and Corollary 4.14 therein.

Theorem 3.6. Let $p_{0} \leq p \leq \infty, 0<q \leq \infty$ and $\alpha>0$. Then we have

$$
\begin{align*}
|u|_{A_{p, q}^{\alpha}(\Omega)} & \sim\left\|\left(\lambda^{j \alpha}\left\|u-\tilde{Q}_{j} u\right\|_{L_{p}(\Omega)}\right)_{j \geq 0}\right\|_{\ell_{q}}  \tag{99}\\
& \sim\left\|\left(\lambda^{j \alpha}\left\|\tilde{Q}_{j+1} u-\tilde{Q}_{j} u\right\|_{L_{p}(\Omega)}\right)_{j \geq 0}\right\|_{\ell_{q}},
\end{align*}
$$

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 (76). The generalized triangle inequality

$$
\begin{equation*}
\left\|\tilde{Q}_{j+1} u-\tilde{Q}_{j} u\right\|_{L_{p}(\Omega)} \lesssim\left\|u-\tilde{Q}_{j} u\right\|_{L_{p}(\Omega)}+\left\|u-\tilde{Q}_{j+1} u\right\|_{L_{p}(\Omega)}, \tag{100}
\end{equation*}
$$

implies one of the directions of the second equivalence, while the other direction follows
from applying the discrete Hardy inequality (Lemma 3.1) to

$$
\begin{equation*}
\left\|u-\tilde{Q}_{j} u\right\|_{L_{p}(\Omega)} \leq\left(\sum_{k=j}^{\infty}\left\|\tilde{Q}_{k+1} u-\tilde{Q}_{k} u\right\|_{L_{p}(\Omega)}^{p^{*}}\right)^{\frac{1}{p^{*}}} \tag{101}
\end{equation*}
$$

where $p^{*}=\min \{1, p\}$.
The following technical result will be used later.
Theorem 3.7. Let $0<\alpha_{1}<\alpha_{2}<\infty$ and $0<p, q, q_{1}, q_{2} \leq \infty$. Then we have

$$
\begin{equation*}
\left[A_{p, q_{1}}^{\alpha_{1}}(G), A_{p, q_{2}}^{\alpha_{2}}(G)\right]_{\theta, q}=A_{p, q}^{\alpha}(G) \tag{102}
\end{equation*}
$$

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 (102) 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

$$
\begin{equation*}
|v|_{A_{p, q_{2}}^{\alpha_{2}}(G)}^{q_{2}}=\sum_{j=0}^{m-1} \lambda^{\alpha_{2} q_{2} j} E\left(v, S_{j}, G\right)_{p}^{q_{2}} \leq\|v\|_{L_{p}(G)}^{q_{2}} \sum_{j=0}^{m-1} \lambda^{\alpha_{2} q_{2} j} \leq \frac{\lambda^{\alpha_{2} q_{2} m}}{\lambda^{\alpha_{2} q_{2}}-1}\|v\|_{L_{p}(G)}^{q_{2}} \tag{103}
\end{equation*}
$$

It is also true for $q_{2}=\infty$ :

$$
\begin{equation*}
|v|_{A_{p, \infty}^{\alpha_{2}}(G)}=\max _{0 \leq j<m} \lambda^{\alpha_{2} j} E\left(v, S_{j}, G\right)_{p} \leq \lambda^{\alpha_{2} m}\|v\|_{L_{p}(G)} \tag{104}
\end{equation*}
$$

Another fact we will need is the following. We have the generalized triangle inequality

$$
\begin{equation*}
|u+v|_{A_{p, q_{2}}^{\alpha_{2}}(G)} \leq c|u|_{A_{p, q_{2}}^{\alpha_{2}}(G)}+c|v|_{A_{p, q_{2}}^{\alpha_{2}}(G)} \tag{105}
\end{equation*}
$$

with $c \geq 1$ depening only on $p$ and $q_{2}$. Then the Aoki-Rolewicz theorem (Bergh and Löfström, 1976, page 59) implies that

$$
\begin{equation*}
\left|v_{1}+\ldots+v_{k}\right|_{A_{p, q_{2}}(G)}^{\mu} \leq 2\left|v_{1}\right|_{A_{p, q_{2}}(G)}^{\mu}+\ldots+2\left|v_{k}\right|_{A_{p, q_{2}}(G)}^{\mu} \tag{106}
\end{equation*}
$$

for any $v_{1}, \ldots, v_{k} \in A_{p, q_{2}}^{\alpha_{2}}(G)$, with $\mu$ given by $(2 c)^{\mu}=2$.
With the abbreviation $K(u, t)=K\left(u, t ; A_{p, q_{1}}^{\alpha_{1}}(G), A_{p, q_{2}}^{\alpha_{2}}(G)\right)$, for $u \in A_{p, q_{1}}^{\alpha_{1}}(G)$, we have

$$
\begin{equation*}
K\left(u, \lambda^{-\left(\alpha_{2}-\alpha_{1}\right) m}\right) \leq\left|u-u_{m}\right|_{A_{p, q_{1}}^{\alpha_{1}}(G)}+\lambda^{-\left(\alpha_{2}-\alpha_{1}\right) m}\left|u_{m}\right|_{A_{p, q_{2}}^{\alpha_{2}}(G)} \tag{107}
\end{equation*}
$$

where $u_{m} \in S_{m}$ is an approximation satisfying $\left\|u-u_{m}\right\|_{L_{p}(G)}=E\left(u, S_{m}, G\right)_{p}$. We estimate the first term in the right hand side as

$$
\begin{align*}
\left|u-u_{m}\right|_{A_{p, q_{1}}(G)}^{q_{1}} & =\sum_{j=0}^{m} \lambda^{\alpha_{1} q_{1} j}\left\|u-u_{m}\right\|_{L_{p}(G)}^{q_{1}}+\sum_{j=m+1}^{\infty} \lambda^{\alpha_{1} q_{1} j}\left\|u-u_{j}\right\|_{L_{p}(G)}^{q_{1}} \\
& \lesssim \sum_{j=m}^{\infty} \lambda^{\alpha_{1} q_{1} j}\left\|u-u_{j}\right\|_{L_{p}(G)}^{q_{1}}, \tag{108}
\end{align*}
$$

with the implicit constant depending only on $\lambda^{\alpha_{1} q_{1}}$, and the second term as

$$
\begin{align*}
\left|u_{m}\right|_{A_{p, q_{2}}(G)}^{\alpha_{2}} & \leq 2 \sum_{j=1}^{m}\left|u_{j}-u_{j-1}\right|_{A_{p, q_{2}}(G)}^{\mu} \lesssim \sum_{j=1}^{m} \lambda^{\alpha_{2} \mu j}\left\|u_{j}-u_{j-1}\right\|_{L_{p}(G)}^{\mu}  \tag{109}\\
& \lesssim \sum_{j=0}^{m} \lambda^{\alpha_{2} \mu j}\left\|u-u_{j}\right\|_{L_{p}(G)}^{\mu},
\end{align*}
$$

where we have used the $\mu$-triangle inequality (106) in the first step, the inverse estimate (103) 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

$$
\begin{align*}
K\left(u, \lambda^{-\left(\alpha_{2}-\alpha_{1}\right) m}\right) \lesssim & \left(\sum_{j=m}^{\infty} \lambda^{\alpha_{1} q_{1} j}\left\|u-u_{j}\right\|_{L_{p}(G)}^{q_{1}}\right)^{\frac{1}{q_{1}}} \\
& +\lambda^{-\left(\alpha_{2}-\alpha_{1}\right) m}\left(\sum_{j=0}^{m} \lambda^{\alpha_{2} \mu j}\left\|u-u_{j}\right\|_{L_{p}(G)}^{\mu}\right)^{\frac{1}{\mu}}, \tag{110}
\end{align*}
$$

and then the discrete Hardy inequalities (Lemma 3.1) give

$$
\begin{equation*}
\left\|\left[\lambda^{\gamma m} K\left(u, \lambda^{-\left(\alpha_{2}-\alpha_{1}\right) m}\right)\right]_{m \geq 0}\right\|_{\ell_{q}} \lesssim\left\|\left[\lambda^{\left(\alpha_{1}+\gamma\right) m}\left\|u-u_{j}\right\|_{L_{p}(G)}\right]_{m \geq 0}\right\|_{\ell_{q}}, \tag{1111}
\end{equation*}
$$

for $0<\gamma<\alpha_{2}-\alpha_{1}$. The left hand side of this inequality is the (quasi) norm for $\left[A_{p, q_{1}}^{\alpha_{1}}(G), A_{p, q_{2}}^{\alpha_{2}}(G)\right]_{\gamma /\left(\alpha_{2}-\alpha_{1}\right), q}$, while the right hand side is the (quasi) norm for $A_{p, q}^{\alpha_{1}+\gamma}(G)$.

For the other direction, we start with

$$
\begin{equation*}
\left\|u-u_{j}\right\|_{L_{p}(G)} \leq\left\|u-w_{j}-v_{j}\right\|_{L_{p}(G)} \lesssim\left\|u-v-w_{j}\right\|_{L_{p}(G)}+\left\|v-v_{j}\right\|_{L_{p}(G)}, \tag{112}
\end{equation*}
$$

where $u \in A_{p, q_{1}}^{\alpha_{1}}(G)$ and $u_{j} \in S_{j}$ are as before, and $v \in A_{p, q_{2}}^{\alpha_{2}}(G), v_{j}, w_{j} \in S_{j}$ are arbitrary. Note that the implicit constant depends only on $p$. Optimizing over $v_{j}$ and $w_{j}$ gives

$$
\begin{equation*}
\min _{w_{j} \in S_{j}}\left\|u-v-w_{j}\right\|_{L_{p}(G)} \leq \lambda^{-\alpha_{1} j}|u-v|_{A_{p, q_{1}}^{\alpha_{1}}(G)}, \tag{113}
\end{equation*}
$$

and

$$
\begin{equation*}
\min _{v_{j} \in S_{j}}\left\|v-v_{j}\right\|_{L_{p}(G)} \leq \lambda^{-\alpha_{2} j}|v|_{A_{p, q_{2}}^{\alpha_{2}}(G)} \tag{114}
\end{equation*}
$$

and substituting these back, we get

$$
\begin{align*}
\left\|u-u_{j}\right\|_{L_{p}(G)} & \lesssim \inf _{v \in A_{p, q_{2}}^{\alpha_{2}(G)}}\left(\lambda^{-\alpha_{1} j}|u-v|_{A_{p, q_{1}}^{\alpha_{1}}(G)}+\lambda^{-\alpha_{2} j}|v|_{A_{p, q_{2}}(G)}^{\alpha_{2}(G)}\right)  \tag{115}\\
& =\lambda^{-\alpha_{1} j} K\left(u, \lambda^{-\left(\alpha_{2}-\alpha_{1}\right) j}\right) .
\end{align*}
$$

The proof is completed upon recalling the definition of $|\cdot|_{A_{p, q}^{\alpha}(G)}$.

### 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. (52). In Binev et al. (2002); Gaspoz and Morin (2013), among other things, it is proved that $B_{q, q}^{\alpha}(\Omega) \hookrightarrow \mathscr{A}_{\infty}^{s}\left(L_{p}(\Omega)\right)$ with $s=\frac{\alpha}{n}$, as long as $\frac{\alpha}{n}+\frac{1}{p}-\frac{1}{q}>0$ and $0<\alpha<m+\max \left\{1, \frac{1}{q}\right\}$. In the other direction, the same references give $\mathscr{A}_{q}^{s}\left(L_{p}(\Omega)\right) \hookrightarrow A_{q, q}^{\alpha}(\Omega)$ 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 $A_{q, q}^{\alpha}(\Omega) \hookrightarrow \mathscr{A}_{\infty}^{s}\left(L_{p}(\Omega)\right)$. This is a genuine improvement, since $A_{q, q}^{\alpha}(\Omega) \supsetneq B_{q, q}^{\alpha}(\Omega)$ 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 procedure and conforming triangulations, we deal with possibly nonconforming triangulations, and therefore are able to handle the red refinement procedure, as well as newest vertex bisections without the conformity requirement.

Theorem 3.8. 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 $\tilde{Q}_{P}$ depends on $p_{0}$ ), we have

$$
\begin{equation*}
\left\|u-\tilde{Q}_{P} u\right\|_{L_{p}(\Omega)} \lesssim\left(\sum_{\tau \in P}|\tau|^{p \delta}|u|_{A_{q, q}^{\alpha}(\hat{\tau})}^{p}\right)^{\frac{1}{p}}, \quad u \in A_{q, q}^{\alpha}(\Omega), \quad P \in \mathscr{P} \tag{116}
\end{equation*}
$$

where $\delta=\frac{\alpha}{n}+\frac{1}{p}-\frac{1}{q}$. In particular, we have $A_{q, q}^{\alpha}(\Omega) \hookrightarrow \mathscr{A}^{s}\left(L_{p}(\Omega)\right)$ with $s=\frac{\alpha}{n}$.
Proof. We have the sub-additivity property

$$
\begin{equation*}
\sum_{k}\|u\|_{A_{q, q}^{\alpha}\left(P_{k}\left(\tau_{k}\right)\right)}^{q} \lesssim\|u\|_{A_{q, q}^{\alpha}(\Omega)}^{q} \tag{117}
\end{equation*}
$$

for $0<q<\infty$ and for any finite sequences $\left\{P_{k}\right\} \subset \mathscr{P}$ and $\left\{\tau_{k}\right\}$, with $\tau_{k} \in P_{k}$ and $\left\{\tau_{k}\right\}$ non-overlapping. Recall that $\hat{\tau}=P(\tau)$ is the support extension of $\tau$, as defined in (56). Therefore the estimate (116) would imply the second statement by Theorem 2.2.

We shall prove (116). Every element $\tau \in P$ of any partition $P \in \mathscr{P}$ 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 $j(\tau)=\min \{[\sigma]: \sigma \in P, \sigma \subset \hat{\tau}\}$. Note that $|\tau| \sim \lambda^{-n[\tau]} \sim \lambda^{-n j(\tau)}$ and $\left.\left.S_{j(\tau)}\right|_{\hat{\tau}} \subset S_{P}\right|_{\hat{\tau}}$. By invoking (75), we infer

$$
\begin{align*}
\left\|u-\tilde{Q}_{P} u\right\|_{L_{p}(\Omega)}^{p} & =\sum_{\tau \in P}\left\|u-\tilde{Q}_{P} u\right\|_{L_{p}(\tau)}^{p} \lesssim \sum_{\tau \in P} \inf _{v \in S_{P}}\|u-v\|_{L_{p}(\hat{\tau})}^{p}  \tag{118}\\
& \leq \sum_{\tau \in P}\left\|u-u_{j(\tau)}\right\|_{L_{p}(\hat{\tau})}^{p},
\end{align*}
$$

where $u_{j} \in S_{j}(j \geq 0)$ is an approximation (that may depend on $\tau$ ) satisfying

$$
\begin{equation*}
\left\|u-u_{j}\right\|_{L_{q}(\hat{\tau})} \leq c E\left(u, S_{j}\right)_{L_{q}(\hat{\tau})} \tag{119}
\end{equation*}
$$

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 $p^{*}=\min \{1, p\}$, we have

$$
\begin{align*}
\left\|u-u_{j(\tau)}\right\|_{L_{p}(\hat{\tau})}^{p^{*}} & \leq \sum_{j=j(\tau)}^{\infty}\left\|u_{j+1}-u_{j}\right\|_{L_{p}(\hat{\tau})}^{p^{*}} \lesssim \sum_{j=j(\tau)}^{\infty} \lambda^{\left(\frac{1}{q}-\frac{1}{p}\right) j n p^{*}}\left\|u_{j+1}-u_{j}\right\|_{L_{q}(\hat{\tau})}^{p^{*}} \\
& \lesssim \sum_{j=j(\tau)}^{\infty} \lambda^{\left(\frac{1}{q}-\frac{1}{p}\right) j n p^{*}}\left\|u-u_{j}\right\|_{L_{q}(\hat{\tau})}^{p^{*}} \tag{120}
\end{align*}
$$

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

$$
\begin{align*}
\left\|u-u_{j(\tau)}\right\|_{L_{p}(\hat{\tau})}^{p^{*}} & \lesssim \sum_{j=j(\tau)}^{\infty} \lambda^{-j \delta n p^{*}} \lambda^{j \alpha p^{*}}\left\|u-u_{j}\right\|_{L_{q}(\hat{\tau})}^{p^{*}} \\
& \leq \lambda^{-j(\tau) \delta n p^{*}} \sum_{j=j(\tau)}^{\infty} \lambda^{j \alpha p^{*}}\left\|u-u_{j}\right\|_{L_{q}(\hat{\tau})}^{p^{*}}  \tag{121}\\
& \lesssim|\tau|^{\delta p^{*}}|u|_{A_{q, p^{*}}^{\alpha}(\hat{\tau})}^{p^{*}},
\end{align*}
$$

by (119). This establishes the theorem for $q \leq 1$, in which case we have $A_{q, q}^{\alpha}(\hat{\tau}) \hookrightarrow A_{q, p^{*}}^{\alpha}(\hat{\tau})$.
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 $u_{j}=Q_{P_{j}}^{(\hat{\tau})} u$, where $Q_{P_{j}}^{(\hat{\tau})}:\left.L_{1}(\hat{\tau}) \rightarrow S_{j}\right|_{\hat{\tau}}$ is the quasi-interpolation operator defined in (62), with $\hat{\tau}$ playing the role of $\Omega$. Then Lemma 3.2 guarantees the
property (119) with $c$ depending only on global geometric properties of $\mathscr{P}$. In particular, $c$ is bounded independently of $\tau$. Thus (121) gives

$$
\begin{equation*}
\left\|u-u_{j(\tau)}\right\|_{L^{p}(\hat{\tau})} \lesssim|\tau|^{\delta_{i}}|u|_{A_{q, 1}^{\alpha_{i}}(\hat{\tau})}, \tag{122}
\end{equation*}
$$

for $i=1,2$. Since the operators $Q_{P_{j}}^{(\hat{\tau})}$ are linear, so is the map $u \mapsto u-u_{j(\tau)}$, and hence interpolation and Theorem 3.7 yield

$$
\begin{equation*}
\left\|u-u_{j(\tau)}\right\|_{L^{p}(\hat{\tau})} \lesssim|\tau|^{\left(\delta_{1}+\delta_{2}\right) / 2}|u|_{\left[A_{q, 1}^{\alpha_{1}(\hat{\tau}), A_{q, 1}^{\left.\alpha_{2}(\hat{\tau})\right]_{1 / 2, p}}}\right.} \lesssim|\tau|^{\delta}|u|_{A_{q, p}^{\alpha}(\hat{\tau})} \lesssim|\tau|^{\delta}|u|_{A_{q, q}^{\alpha}(\hat{\tau})} \tag{123}
\end{equation*}
$$

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

(a) If the space $A_{q, q}^{\alpha}$ is located above the solid line, we have $A_{q, q}^{\alpha} \subset \mathscr{A}^{s}\left(L_{p}\right)$ with $s=\frac{\alpha}{n}$.

(b) If the space $A_{q, q}^{\alpha}$ is located above the solid line, we have $A_{q, q}^{\alpha} \subset \mathscr{A}^{s}\left(A_{p, p}^{\sigma}\right)$ with $s=\frac{\alpha-\sigma}{n}$.

Fig. 4: Illustration of Theorem 3.8 and Theorem 3.9.
Now we look at adaptive approximation in the space $A_{p, p}^{\sigma}(\Omega)$. Recall from Gaspoz and Morin (2013) that $\mathscr{A}_{q}^{s}\left(A_{p, p}^{\sigma}(\Omega)\right) \hookrightarrow A_{q, q}^{\alpha}(\Omega)$ for $s=\frac{\alpha-\sigma}{n}=\frac{1}{q}-\frac{1}{p}>0$ and $0<p, q<\infty$.
Theorem 3.9. 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

$$
\begin{equation*}
\left\|u-\tilde{Q}_{P} u\right\|_{A_{p, p}^{\sigma}(\Omega)} \lesssim\left(\sum_{\tau \in P}|\tau|^{p \delta}|u|_{A_{q, q}^{\alpha}(\hat{\tau})}^{p}\right)^{\frac{1}{p}}, \quad u \in A_{q, q}^{\alpha}(\Omega), \quad P \in \mathscr{P}, \tag{124}
\end{equation*}
$$

with $\delta=\frac{\alpha-\sigma}{n}+\frac{1}{p}-\frac{1}{q}$. In particular, we have $A_{q, q}^{\alpha}(\Omega) \hookrightarrow \mathscr{A}^{s}\left(A_{p, p}^{\sigma}(\Omega)\right)$ with $s=\frac{\alpha-\sigma}{n}$.

Proof. With $v=u-\tilde{Q}_{P} u$ and $\tilde{Q}_{j}=\tilde{Q}_{P_{j}}$, we have

$$
\begin{equation*}
\|v\|_{A_{p, p}^{\sigma}(\Omega)} \leq\left(\sum_{j \geq 0} \lambda^{j \sigma p}\left\|v-\tilde{Q}_{j} v\right\|_{L_{p}(\Omega)}^{p}\right)^{\frac{1}{p}}=\left(\sum_{\tau \in P} \sum_{j \geq 0} \lambda^{j \sigma p}\left\|v-\tilde{Q}_{j} v\right\|_{L_{p}(\tau)}^{p}\right)^{\frac{1}{p}}, \tag{125}
\end{equation*}
$$

with the usual modification for $p=\infty$. Let $j(\tau)=\max \{[\sigma]: \sigma \in P, \sigma \subset \tau\}$ for $\tau \in P$, as in the preceding proof, with $[\sigma]$ denoting the generation number (or the level) of $\sigma$. Then for $\tau \in P$ and $j \geq j(\tau)$ we have $\left.\left.S_{P}\right|_{\hat{\tau}} \subset S_{j}\right|_{\hat{\tau}}$, and hence

$$
\begin{equation*}
\tilde{Q}_{j}\left(u-\tilde{Q}_{P} u\right)=\tilde{Q}_{j} u-\tilde{Q}_{P} u \quad \text { on } \tau, \tag{126}
\end{equation*}
$$

by the linearity property (73). 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

$$
\begin{align*}
\left\|u-\tilde{Q}_{j} u\right\|_{L_{p}(\tau)}^{p^{*}} & \leq \sum_{k=j}^{\infty}\left\|\tilde{Q}_{k+1} u-\tilde{Q}_{k} u\right\|_{L_{p}(\tau)}^{p^{*}} \lesssim \sum_{k=j}^{\infty} \lambda^{\left(\frac{1}{q}-\frac{1}{p}\right) k n p^{*}}\left\|\tilde{Q}_{k+1} u-\tilde{Q}_{k} u\right\|_{L_{q}(\tau)}^{p^{*}} \\
& \lesssim \sum_{k=j}^{\infty} \lambda^{\left(\frac{1}{q}-\frac{1}{p}\right) k n p^{*}}\left\|u-\tilde{Q}_{k} u\right\|_{L_{q}(\tau)}^{p^{*}} . \tag{127}
\end{align*}
$$

Then the discrete Hardy inequality yields

$$
\begin{align*}
\sum_{j \geq j(\tau)} \lambda^{j \sigma q}\left\|u-\tilde{Q}_{j} u\right\|_{L_{p}(\tau)}^{q} & \lesssim \sum_{k \geq j(\tau)} \lambda^{k \sigma q} \lambda^{\left(\frac{1}{q}-\frac{1}{p}\right) k n q}\left\|u-\tilde{Q}_{k} u\right\|_{L_{q}(\tau)}^{q} \\
& \leq \lambda^{-\delta n q j(\tau)} \sum_{k=j(\tau)}^{\infty} \lambda^{k \alpha q}\left\|u-\tilde{Q}_{k} u\right\|_{L_{q}(\tau)}^{q}  \tag{128}\\
& \lesssim|\tau|^{\delta q}|u|_{A_{q, q}^{\alpha}(\tau)}^{q},
\end{align*}
$$

where we have taken into account the relation $\frac{\sigma}{n}+\frac{1}{q}-\frac{1}{p}=\frac{\alpha}{n}-\delta$. 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 (125) into two sums according to $j<j(\tau)$ or $j \geq j(\tau)$. We rewrite the other sum (or maximum) as

$$
\begin{align*}
\left(\sum_{\tau \in P} \sum_{\{j<j(\tau)\}} \lambda^{j \sigma p}\left\|v-\tilde{Q}_{j} v\right\|_{L_{p}(\tau)}^{p}\right)^{\frac{1}{p}} & =\left(\sum_{j \geq 0} \sum_{\{\tau \in P: j(\tau)>j\}} \lambda^{j \sigma p}\left\|v-\tilde{Q}_{j} v\right\|_{L_{p}(\tau)}^{p}\right)^{\frac{1}{p}}  \tag{129}\\
& =\left(\sum_{j \geq 0} \lambda^{j \sigma p}\left\|v-\tilde{Q}_{j} v\right\|_{L_{p}\left(\Omega_{j}\right)}^{p}\right)^{\frac{1}{p}}
\end{align*}
$$

where $\Omega_{j}=\bigcup\{\tau \in P: j(\tau)>j\}$. Note that $\Omega_{j} \supset \Omega_{j}^{0}$ with $\Omega_{j}^{0}=\bigcup\{\tau \in P:[\tau]>j\}$, 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 $\Omega_{j}^{0}=\bigcup\left\{\tau \in R_{j}^{0}\right\}$ up to a zero measure set. The triangles $\tau \in P$ with $\tau \not \subset \Omega_{j}^{0}$ are at the level $j$ or less, and hence there is $R_{j} \subset P_{j}$ such that $\Omega_{j}=\bigcup\left\{\tau \in R_{j}\right\}$ up to a zero measure set. Now, by the stability property (74), we get

$$
\begin{equation*}
\left\|v-\tilde{Q}_{j} v\right\|_{L_{p}\left(\Omega_{j}\right)} \lesssim\|v\|_{L_{p}\left(\Omega_{j}\right)}+\left\|\tilde{Q}_{j} v\right\|_{L_{p}\left(\Omega_{j}\right)} \lesssim\|v\|_{L_{p}\left(\hat{\Omega}_{j}\right)} \tag{130}
\end{equation*}
$$

where $\hat{\Omega}_{j}=\bigcup\left\{\tau \in P_{j}: \bar{\tau} \cap \Omega_{j} \neq \varnothing\right\}$. Obviously, $\hat{\Omega}_{j}$ is a subset of $\hat{\Omega}_{j}^{\prime}=\bigcup\left\{\tau \in P: \bar{\tau} \cap \Omega_{j} \neq\right.$ $\varnothing\}$, that can also be described as $\hat{\Omega}_{j}^{\prime}=\bigcup\left\{\tau \in P: j^{2}(\tau)>j\right\}$, with $j^{2}(\tau)=\max \{j(\sigma)$ : $\sigma \in P, \bar{\sigma} \cap \bar{\tau} \neq \varnothing\}$ for $\tau \in P$. All this yields

$$
\begin{align*}
\left(\sum_{\tau \in P} \sum_{\{j<j(\tau)\}} \lambda^{j \sigma p}\left\|v-\tilde{Q}_{j} v\right\|_{L_{p}(\tau)}^{p}\right)^{\frac{1}{p}} & \lesssim\left(\sum_{\tau \in P} \sum_{\left\{j<j^{2}(\tau)\right\}} \lambda^{j \sigma p}\left\|u-\tilde{Q}_{P} u\right\|_{L_{p}(\tau)}^{p}\right)^{\frac{1}{p}}  \tag{131}\\
& \lesssim\left(\sum_{\tau \in P}|\tau|^{\sigma p / n}\left\|u-\tilde{Q}_{P} u\right\|_{L_{p}(\tau)}^{p}\right)^{\frac{1}{p}}
\end{align*}
$$

where we have taken into account the geometric growth of $\lambda^{j \sigma p}$ in $j$, and the fact that $\lambda^{j^{2}(\tau)} \sim|\tau|^{1 / n}$. Then once we recall from the proof of the preceding theorem that

$$
\begin{equation*}
\left\|u-\tilde{Q}_{P} u\right\|_{L_{p}(\tau)} \lesssim|\tau|^{\delta^{\prime}}|u|_{A_{q, q}^{\alpha}(\hat{\tau})} \tag{132}
\end{equation*}
$$

with $\delta^{\prime}=\frac{\alpha}{n}+\frac{1}{p}-\frac{1}{q}=\delta-\frac{\sigma}{n}$, 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 $\mathscr{P},\left\{P_{j}\right\}$, etc., will mean the same things as before. For $P \in \mathscr{P}$, let

$$
\begin{equation*}
\bar{S}_{P}=\bar{S}_{P}^{d}=\left\{v \in L_{\infty}(\Omega):\left.v\right|_{\tau} \in \mathbb{P}_{d} \forall \tau \in P\right\}, \tag{133}
\end{equation*}
$$

where $d$ is a nonnegative integer, and let $\bar{S}_{j}=\bar{S}_{P_{j}}$ for all $j$. Then with $G \subset \Omega$ a domain consisting of elements from some $P_{j}$, we define the multilevel approximation spaces $A_{p, q}^{\alpha}\left(\left\{\bar{S}_{j}\right\}, G\right)$ by (92), with the sequence $\left\{\bar{S}_{j}\right\}$ replacing $\left\{S_{j}\right\}$. We will also use the shorthand notations

$$
\begin{equation*}
\bar{A}_{p, q}^{\alpha}(G)=\bar{A}_{p, q ; d}^{\alpha}(G)=A_{p, q}^{\alpha}\left(\left\{\bar{S}_{j}\right\}, G\right) . \tag{134}
\end{equation*}
$$

The analogue of Theorem 3.5 is the following.

Theorem 3.10. We have $B_{p, q ; d+1}^{\alpha}(\Omega) \hookrightarrow \bar{A}_{p, q ; d}^{\alpha}(\Omega)$ for $0<p, q \leq \infty$, and $\alpha>0$. In the other direction, we have $\bar{A}_{p, q ; d}^{\alpha}(\Omega) \hookrightarrow B_{p, q ; d+1}^{\alpha}(\Omega)$ 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 3.7.

Theorem 3.11. Let $0<\alpha_{1}<\alpha_{2}<\infty$ and $0<p, q, q_{1}, q_{2} \leq \infty$. Then we have

$$
\begin{equation*}
\left[\bar{A}_{p, q_{1}}^{\alpha_{1}}(G), \bar{A}_{p, q_{2}}^{\alpha_{2}}(G)\right]_{\theta, q}=\bar{A}_{p, q}^{\alpha}(G) \tag{135}
\end{equation*}
$$

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 $\mathscr{P}$. Given $0<p \leq \infty, \theta \in \mathbb{R}$, and $P \in \mathscr{P}$, we define the norm

$$
\begin{equation*}
\|u\|_{L_{p}^{\theta}(\Omega)}=\left(\sum_{\tau \in P}|\tau|^{\frac{\theta p}{n}}\|u\|_{L_{p}(\tau)}^{p}\right)^{\frac{1}{p}}, \quad \text { for } \quad u \in L_{p}(\Omega) \tag{136}
\end{equation*}
$$

with the obvious modification for $p=\infty$, and denote by $L_{p}^{\theta}(\Omega)$ the space $L_{p}(\Omega)$ equipped with this norm. Then we define the approximation class $\overline{\mathscr{A}}_{q ; d}^{s}\left(L_{p}^{\theta}(\Omega)\right)$ exactly as $\mathscr{A}_{q}^{s}\left(L_{p}(\Omega)\right)$, by replacing $S_{P}^{m}$ with $\bar{S}_{P}^{d}$, and by using the distance function

$$
\begin{equation*}
\rho(u, v, P)=\|u-v\|_{L_{p}^{\theta}(\Omega)} . \tag{137}
\end{equation*}
$$

More precisely, recalling the definition (9), let

$$
\begin{equation*}
\overline{\mathscr{A}}_{q ; d}^{s}\left(L_{p}^{\theta}(\Omega)\right)=\mathscr{A}_{q}^{s}\left(\rho, \mathscr{P},\left\{\bar{S}_{P}^{d}\right\}\right), \tag{138}
\end{equation*}
$$

with $\rho$ given by (137). It is for later reference that we have introduced the mesh dependent weight in the distance function. We write $\overline{\mathscr{A}}_{q ; d}^{s}\left(L_{p}(\Omega)\right)=\overline{\mathscr{A}}_{q ; d}^{s}\left(L_{p}^{0}(\Omega)\right)$.

We have the following direct embedding result.
Theorem 3.12. 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 $\bar{A}_{q, q ; d}^{\alpha}(\Omega) \hookrightarrow \overline{\mathscr{A}}_{\infty ; d}^{s}\left(L_{p}^{\theta}(\Omega)\right)$ with $s=\frac{\alpha+\theta}{n}$.
Proof. Let $u \in L_{p}(\Omega)$ and let $P \in \mathscr{P}$. Then with $\Pi_{P}$ the projection operator defined in (71) with $m:=d$, we have

$$
\begin{equation*}
\left\|u-\Pi_{P} u\right\|_{L_{p}^{\theta}(\Omega)}=\left(\sum_{\tau \in P}|\tau|^{\frac{\theta p}{n}}\left\|u-\Pi_{P} u\right\|_{L_{p}(\tau)}^{p}\right)^{\frac{1}{p}} \lesssim\left(\sum_{\tau \in P}|\tau|^{\frac{\theta_{p}}{n}} \inf _{v \in \bar{S}_{P}}\|u-v\|_{L_{p}(\tau)}^{p}\right)^{\frac{1}{p}} \tag{139}
\end{equation*}
$$

Now proceeding exactly as in the proof of Theorem 3.8, we get

$$
\begin{equation*}
\left\|u-\Pi_{P} u\right\|_{L_{p}^{\theta}(\Omega)} \lesssim\left(\sum_{\tau \in P}|\tau|^{\frac{\theta p}{n}}|\tau|^{\delta p}|u|_{\bar{A}_{q, q ; d}^{\alpha}(\tau)}\right)^{\frac{1}{p}} \tag{140}
\end{equation*}
$$

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


Fig. 5: Illustration of Theorem 3.12 and Theorem 3.13. If the space $\bar{A}_{q, q}^{\alpha}$ is located above or on the solid line, then $\bar{A}_{q, q}^{\alpha} \subset \overline{\mathscr{A}}^{s}\left(L_{p}^{\theta}\right)$ with $s=\frac{\alpha+\theta}{n}$. It is as if the approximation is taking place in a space such as $B_{p, p}^{-\theta}$, but instead of $\frac{\alpha+\theta}{n}>\frac{1}{q}-\frac{1}{p}$ (dashed line) we have the condition $\frac{\alpha}{n} \geq \frac{1}{q}-\frac{1}{p}$ (solid line). On the other hand, the inverse embedding takes the form $\overline{\mathscr{A}}_{q}^{s}\left(L_{p}^{\theta}\right) \cap L_{p} \subset \bar{A}_{q, q}^{\alpha}$, which holds on the part of the dashed line with $\alpha>0$.

We close this section by stating an inverse embedding theorem (A proof can be found in arXiv version 1 of the current paper).

Theorem 3.13. 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 $\overline{\mathscr{A}}_{q ; d}^{s}\left(L_{p}^{\theta}(\Omega)\right) \cap L_{p}(\Omega) \subset \bar{A}_{q, q ; d}^{\alpha}(\Omega)$.

## 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 $\mathscr{P}$ 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 $\mathscr{P}$ satisfies the admissibility criterion (53).

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 $P \in \mathscr{P}$, 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 $H_{\Gamma}^{1}\left(\mathbb{R}^{n}\right)$ as the closure of $\mathscr{D}\left(\mathbb{R}^{n} \backslash \bar{\Gamma}\right)$ in $H^{1}\left(\mathbb{R}^{n}\right)$, and $H_{\Gamma}^{1}=H_{\Gamma}^{1}(\Omega)$ as the restriction of functions from $H_{\Gamma}^{1}\left(\mathbb{R}^{n}\right)$ to $\Omega$. Note that $S_{P}=H_{\Gamma}^{1} \cap S_{P}^{m}$.

The operator $T$ is given as

$$
\begin{equation*}
T u=-a_{i j} \partial_{i} \partial_{j} u+b_{k} \partial_{k} u+c u, \tag{141}
\end{equation*}
$$

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

$$
\begin{equation*}
\langle T u, v\rangle=\langle f, v\rangle, \quad \text { for all } \quad v \in H_{\Gamma}^{1} . \tag{142}
\end{equation*}
$$

Here $\langle\cdot, \cdot\rangle$ is the duality pairing between $\left(H_{\Gamma}^{1}\right)^{\prime}$ and $H_{\Gamma}^{1}$, and $f \in L_{2}(\Omega) \hookrightarrow\left(H_{\Gamma}^{1}\right)^{\prime}$ 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 $T: H_{\Gamma}^{1} \rightarrow\left(H_{\Gamma}^{1}\right)^{\prime}$ the operator defined in (142).

Given an $n$-simplex $\tau$, let us denote by $E_{\tau}$ the union of the ( $n-1$ )-dimensional open faces of $\tau$. In other words, $E_{\tau}$ is the boundary of $\tau$ with all but the ( $n-1$ )-dimensional faces removed. Then for $P \in \mathscr{P}$, let

$$
\begin{equation*}
E_{P}=\left\{E_{\tau} \cap E_{\sigma} \cap(\bar{\Omega} \backslash \Gamma): \tau, \sigma \in P\right\}, \tag{143}
\end{equation*}
$$

be the set of faces not intersecting the Dirichlet piece $\Gamma$. Note that if $P$ is nonconforming, then only the faces of the "smaller" simplices go into $E_{P}$. Given $P \in \mathscr{P}, u \in T^{-1}\left(L_{2}(\Omega)\right)$, and $v \in S_{P}$, define the element residual $r_{\tau}=\left.(T u-T v)\right|_{\tau}$ 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 $a_{i j} \partial_{j} v$ across the edge $e$. Finally, we define the a posteriori error estimator

$$
\begin{equation*}
(\eta(u, v, P))^{2}=\sum_{\tau \in P} h_{\tau}^{2}\left\|r_{\tau}\right\|_{L_{2}(\tau)}^{2}+\sum_{e \in E_{P}} h_{e}\left\|r_{e}\right\|_{L_{2}(e)}^{2} . \tag{144}
\end{equation*}
$$

A typical adaptive finite element method that uses (144) as its error indicator converges optimally with respect to the approximation classes $\mathscr{A}^{s}(\eta)$, in the sense that if the solution $u$ of the problem (142) satisfies $u \in \mathscr{A}^{s}(\eta)$ for some $s>0$, then the adaptive method reduces the quantity $\eta\left(u, u_{P}, P\right)$ 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 (144) is equivalent to the total error

$$
\begin{equation*}
\left(\rho_{d}(u, v, P)\right)^{2}=\|u-v\|_{H^{1}}^{2}+\left(\operatorname{osc}_{d}(u, v, P)\right)^{2}, \tag{145}
\end{equation*}
$$

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

$$
\begin{equation*}
\left(\operatorname{osc}_{d}(u, v, P)\right)^{2}=\sum_{\tau \in P} h_{\tau}^{2}\left\|\left(1-\Pi_{\tau}\right) r_{\tau}\right\|_{L_{2}(\tau)}^{2}+\sum_{e \in E_{P}} h_{e}\left\|\left(1-\Pi_{e}\right) r_{e}\right\|_{L_{2}(e)}^{2}, \tag{146}
\end{equation*}
$$

with $\Pi_{\tau}: L_{2}(\tau) \rightarrow \mathbb{P}_{d}$ and $\Pi_{e}: L_{2}(e) \rightarrow \mathbb{P}_{d+1}$ 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 $\mathscr{A}^{s}\left(\rho_{d}\right)$ has also been proved, cf. Cascon et al. (2008). Ideally, one would like to have optimality with respect to the classes $\mathscr{A}^{s}\left(H_{\Gamma}^{1}\right)$ that correspond to the energy error. In particular, it is conceivable that for certain functions $u$, the energy error $\left\|u-u_{P}\right\|_{H^{1}}$ decays faster than the oscillation $\operatorname{osc}_{d}\left(u, u_{P}, P\right)$, so that the class $\mathscr{A}^{s}\left(H_{\Gamma}^{1}\right)$ is strictly larger than both $\mathscr{A}^{s}\left(\rho_{d}\right)$ and $\mathscr{A}^{s}(\eta)$. However, if the error estimator (144) 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 $\mathscr{A}^{s}\left(\rho_{d}\right)$ and $\mathscr{A}^{s}(\eta)$ 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 $u \in H_{\Gamma}^{1}$ to be in $\mathscr{A}^{s}\left(\rho_{d}\right)$. These conditions will be in terms of memberships of $u$ and $T u$ 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 $g \in \overline{\mathscr{A}}_{\infty ; d}^{s}\left(L_{\infty}^{\theta}(\Omega)\right)$, where the latter space is defined in (138), and again these spaces can be cast in terms of Besov spaces with the help of Theorem 3.12 and Theorem 3.10.

Theorem 4.1. Let $s>0$ and let $d \geq m-2$. Assume that $a_{i j} \in \overline{\mathscr{A}}_{\infty ; d+2-m}^{s}\left(L_{\infty}(\Omega)\right)$, $b_{i} \in \overline{\mathscr{A}}_{\infty ; d+1-m}^{s}\left(L_{\infty}^{1}(\Omega)\right)$, and $c \in \overline{\mathscr{A}}_{\infty ; d-m}^{s}\left(L_{\infty}^{2}(\Omega)\right)$. Then we have

$$
\begin{equation*}
\mathscr{A}^{s}\left(\rho_{d}\right)=\mathscr{A}^{s}\left(H_{\Gamma}^{1}\right) \cap T^{-1}\left(\mathscr{\mathscr { A }}_{\infty ; d}^{s}\left(L_{2}^{1}(\Omega)\right)\right) . \tag{147}
\end{equation*}
$$

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 3.12 that $\bar{A}_{q, q ; d}^{\sigma}(\Omega) \subset \overline{\mathscr{A}}_{\infty ; d}^{s}\left(L_{2}^{1}(\Omega)\right)$ for $\sigma=s n-1$ and $0 \leq \frac{1}{q}-\frac{1}{2} \leq \frac{\sigma}{n}$, and from Theorem 3.10 that $B_{q, q}^{\sigma}(\Omega) \subset \bar{A}_{q, q ; d}^{\sigma}(\Omega)$ for $\sigma<d+\max \left\{1, \frac{1}{q}\right\}$.

Second, while the approximation classes $\mathscr{A}^{s}\left(H_{\Gamma}^{1}\right)$ are associated to the finite element spaces $S_{P}=H_{\Gamma}^{1} \cap S_{P}^{m}$, the approximation classes we considered in the preceding section are associated to the spaces $S_{P}^{m}$ with no boundary conditions. In view of applying Theorem 3.9 and Theorem 3.5, 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

$$
\begin{equation*}
\mathscr{A}^{s}\left(H_{\Gamma}^{1}\right) \equiv \mathscr{A}^{s}\left(H^{1}, \mathscr{P},\left\{H_{\Gamma}^{1} \cap S_{P}^{m}\right\}\right)=H_{\Gamma}^{1} \cap \mathscr{A}^{s}\left(A_{2,2}^{1}(\Omega), \mathscr{P},\left\{S_{P}^{m}\right\}\right) \tag{148}
\end{equation*}
$$

In particular, we have $H_{\Gamma}^{1} \cap A_{p, p}^{\alpha}(\Omega) \subset \mathscr{A}^{s}\left(H_{\Gamma}^{1}\right)$ for $\alpha=s n+1$ and $\frac{1}{p}<s+\frac{1}{2}$.
Proof. Let $u \in H_{\Gamma}^{1}$, and let $u_{P} \in H_{\Gamma}^{1} \cap S_{P}^{m}$ be the Scott-Zhang interpolator of $u$ adapted to the Dirichlet boundary condition on $\Gamma$, cf. Scott and Zhang (1990). We have

$$
\begin{equation*}
\inf _{v \in H_{\Gamma}^{1} \cap S_{P}^{m}}\|u-v\|_{H^{1}(\Omega)} \leq\left\|u-u_{P}\right\|_{H^{1}(\Omega)} \lesssim \inf _{v \in S_{P}^{m}}\|u-v\|_{H^{1}(\Omega)} \tag{149}
\end{equation*}
$$

by standard properties of the Scott-Zhang interpolator. Since $H^{1}=B_{2,2}^{1}=A_{2,2}^{1}$ by Theorem 3.5, this implies (148). Then the second assertion of the theorem follows from a direct application of Theorem 3.9.

The inclusion $\mathscr{A}^{s}\left(H_{\Gamma}^{1}\right) \cap T^{-1}\left(\overline{\mathscr{A}}_{\infty ; d}^{s}\left(L_{2}^{1}(\Omega)\right)\right) \subset \mathscr{A}^{s}\left(\rho_{d}\right)$ of Theorem 4.1 is a consequence of the following lemma.

Lemma 4.3. For $u \in H_{\Gamma}^{1}$ and $P \in \mathscr{P}$, there exists $v \in S_{P}$ such that

$$
\begin{align*}
& \rho_{d}(u, v, P)^{2} \lesssim E\left(u, S_{P}\right)_{H^{1}(\Omega)}^{2}+E\left(T u, \bar{S}_{P}^{d}\right)_{L_{2}^{1}(\Omega)}^{2} \\
& \quad+\left(E\left(a_{i j}, \bar{S}_{P}^{d+2-m}\right)_{L_{\infty}(\Omega)}^{2}+E\left(b_{i}, \bar{S}_{P}^{d+1-m}\right)_{L_{\infty}^{1}(\Omega)}^{2}+E\left(c, \bar{S}_{P}^{d-m}\right)_{L_{\infty}^{2}(\Omega)}^{2}\right)|u|_{H^{1}(\Omega)}^{2} \tag{150}
\end{align*}
$$

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

$$
\begin{equation*}
\|u-v\|_{H^{1}} \lesssim \inf _{w \in S_{P}}\|u-w\|_{H^{1}} \tag{151}
\end{equation*}
$$

for all $P \in \mathscr{P}$. 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 $\left.a_{i j}\right|_{\tau} \in \mathbb{P}_{d+2-m},\left.b_{i}\right|_{\tau} \in \mathbb{P}_{d+1-m}$, and $\left.c\right|_{\tau} \in \mathbb{P}_{d-m}$ 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 $f=T u$, we have

$$
\begin{equation*}
\left\|\left(1-\Pi_{\tau}\right)(f-T v)\right\|_{L_{2}(\tau)} \leq\left\|\left(1-\Pi_{\tau}\right) f\right\|_{L_{2}(\tau)}+\left\|\left(1-\Pi_{\tau}\right) T v\right\|_{L_{2}(\tau)} \tag{152}
\end{equation*}
$$

and the last term is zero because $T v \in \mathbb{P}_{d}$. The remaining term gives rise to

$$
\begin{equation*}
\sum_{\tau \in P} h_{\tau}^{2}\left\|\left(1-\Pi_{\tau}\right) f\right\|_{L_{2}(\tau)}^{2}=E\left(f, \bar{S}_{P}^{d}\right)_{L_{2}^{1}(\Omega)}^{2} \tag{153}
\end{equation*}
$$

which yields the desired result.

In the general case, the edge residuals and the terms $\left.T v\right|_{\tau}$ can be nonpolynomial. Let us treat $\left.T v\right|_{\tau}=-a_{i j} \partial_{i} \partial_{j} v+b_{k} \partial_{k} v+c v$ term by term. We have

$$
\begin{equation*}
\left(1-\Pi_{\tau}\right) a_{i j} \partial_{i} \partial_{j} v=\left(1-\Pi_{\tau}\right)\left(a_{i j}-\bar{a}_{i j}\right) \partial_{i} \partial_{j} v, \quad \bar{a}_{i j} \in \mathbb{P}_{d+2-m}, \tag{154}
\end{equation*}
$$

which implies

$$
\begin{equation*}
\left\|\left(1-\Pi_{\tau}\right) a_{i j} \partial_{i} \partial_{j} v\right\|_{L_{2}(\tau)}=\left\|\left(a_{i j}-\bar{a}_{i j}\right) \partial_{i} \partial_{j} v\right\|_{L_{2}(\tau)} \leq\left\|a_{i j}-\bar{a}_{i j}\right\|_{L_{\infty}(\tau)}\left\|\partial_{i} \partial_{j} v\right\|_{L_{2}(\tau)} \tag{155}
\end{equation*}
$$

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

$$
\begin{align*}
\sum_{\tau \in P} h_{\tau}^{2}\left\|\left(1-\Pi_{\tau}\right) a_{i j} \partial_{i} \partial_{j} v\right\|_{L_{2}(\tau)}^{2} & \leq \sum_{\tau \in P} h_{\tau}^{2}\left\|a_{i j}-\bar{a}_{i j}\right\|_{L_{\infty}(\tau)}^{2}\left\|\partial_{i} \partial_{j} v\right\|_{L_{2}(\tau)}^{2} \\
& \leq E\left(a_{i j}, \bar{S}_{P}^{d+2-m}\right)_{L_{\infty}(\Omega)}^{2} \sum_{\tau \in P} h_{\tau}^{2}\left\|\partial_{i} \partial_{j} v\right\|_{L_{2}(\tau)}^{2}  \tag{156}\\
& \lesssim E\left(a_{i j}, \bar{S}_{P}^{d+2-m}\right)_{L_{\infty}(\Omega)}^{2}\|\nabla v\|_{L_{2}(\Omega)}^{2},
\end{align*}
$$

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 (150).

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

$$
\begin{equation*}
\left(1-\Pi_{\tau}\right) c v=\left(1-\Pi_{\tau}\right)(c-\bar{c}) v=\left(1-\Pi_{\tau}\right)(c-\bar{c})(v-\bar{v}), \tag{157}
\end{equation*}
$$

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

$$
\begin{align*}
\sum_{\tau \in P} h_{\tau}^{2}\left\|\left(1-\Pi_{\tau}\right) c v\right\|_{L_{2}(\tau)}^{2} & \leq \sum_{\tau \in P} h_{\tau}^{2}\|c-\bar{c}\|_{L_{\infty}(\tau)}^{2}\|v-\bar{v}\|_{L_{2}(\tau)}^{2} \\
& \lesssim \sum_{\tau \in P} h_{\tau}^{4}\|c-\bar{c}\|_{L_{\infty}(\tau)}^{2}\|\nabla v\|_{L_{2}(\tau)}^{2}  \tag{158}\\
& \leq E\left(c, \bar{S}_{P}^{d-m}\right)_{L_{\infty}(\Omega)}^{2}\|\nabla v\|_{L_{2}(\Omega)}^{2},
\end{align*}
$$

where we have used the Poincaré inequality in the second line. Estimation of the term involving $b_{i} \partial_{i} v$ 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

$$
\begin{align*}
\left\|\left(1-\Pi_{e}\right) a_{i j} \partial_{j} v\right\|_{L_{2}(e)} & =\left\|\left(1-\Pi_{e}\right)\left(a_{i j}-\bar{a}_{i j}\right) \partial_{j} v\right\|_{L_{2}(e)} \\
& \leq\left\|\left(a_{i j}-\bar{a}_{i j}\right) \partial_{j} v\right\|_{L_{2}(e)} \\
& \leq\left\|a_{i j}-\bar{a}_{i j}\right\|_{L_{\infty}(e)}\left\|\partial_{j} v\right\|_{L_{2}(e)}  \tag{159}\\
& \lesssim h_{e}^{-\frac{1}{2}}\left\|a_{i j}-\bar{a}_{i j}\right\|_{L_{\infty}(\tau)}\|\nabla v\|_{L_{2}(\tau)},
\end{align*}
$$

for any $\bar{a}_{i j} \in \mathbb{P}_{d+2-m}$, which shows that the contribution of the edge oscillations to the final estimate (150) is identical to that of (156).

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 (156), (157), and (159). This means that the regularity conditions on the coefficients $a_{i j}, b_{k}$, and $c$ in Theorem 4.1 can be relaxed slightly, if the conclusion of the theorem is to be changed to $\mathscr{A}^{s}\left(H_{\Gamma}^{1}\right) \cap H^{t}(\Omega) \cap T^{-1}\left(\overline{\mathscr{A}}_{\infty ; d}^{s}\left(L_{2}^{1}(\Omega)\right)\right) \subset \mathscr{A}^{s}\left(\rho_{d}\right)$ with $1<t<\frac{3}{2}$.

Lemma 4.5. For any $u \in H_{\Gamma}^{1}, P \in \mathscr{P}$ and $v \in S_{P}$, we have

$$
\begin{align*}
& E\left(T u, \bar{S}_{P}^{d}\right)_{L_{2}^{1}(\Omega)}^{2}+\|u-v\|_{H^{1}(\Omega)}^{2} \lesssim \rho_{d}(u, v, P)^{2} \\
& \quad+\left(E\left(a_{i j}, \bar{S}_{P}^{d+2-m}\right)_{L_{\infty}(\Omega)}^{2}+E\left(b_{i}, \bar{S}_{P}^{d+1-m}\right)_{L_{\infty}^{1}(\Omega)}^{2}+E\left(c, \bar{S}_{P}^{d-m}\right)_{L_{\infty}^{2}(\Omega)}^{2}\right)|v|_{H^{1}(\Omega)}^{2} \tag{160}
\end{align*}
$$

In particular, under the hypotheses of Theorem 4.1, we have the inclusion $\mathscr{A}^{s}\left(\rho_{d}\right) \subset$ $\mathscr{A}^{s}\left(H_{\Gamma}^{1}\right) \cap T^{-1}\left(\overline{\mathscr{A}}_{\infty ; d}^{s}\left(L_{2}^{1}(\Omega)\right)\right)$.

Proof. All the ingredients for establishing the estimate (160) is already given in the proof of the preceding lemma. Namely, we start with the bound

$$
\begin{align*}
\left(\operatorname{osc}_{d}(u, v, P)\right)^{2} \lesssim & \sum_{\tau \in P} h_{\tau}^{2}\left\|\left(1-\Pi_{\tau}\right) T u\right\|_{L_{2}(\tau)}^{2}+\sum_{\tau \in P} h_{\tau}^{2}\left\|\left(1-\Pi_{\tau}\right) T v\right\|_{L_{2}(\tau)}^{2} \\
& +\sum_{e \in E_{P}} h_{e}\left\|\left(1-\Pi_{e}\right) r_{e}\right\|_{L_{2}(e)}^{2} \tag{161}
\end{align*}
$$

and use the estimates $(156),(157)$, and (159), etc., on the last two terms to get (160).
As for the second assertion, let $\left\{P_{k}\right\} \subset \mathscr{P}$ and $\left\{v_{k}\right\}$ be two sequences with $v_{k} \in S_{P_{k}}$ such that $\# P_{k} \lesssim 2^{k}$ and $\rho_{d}\left(u, v_{k}, P_{k}\right) \lesssim 2^{-k s}$. Then since $\left\|u-v_{k}\right\|_{H^{1}} \leq \rho_{d}\left(u, v_{k}, P_{k}\right)$, we have $\left\|v_{k}\right\|_{H^{1}} \lesssim\|u\|_{H^{1}}$. Hence, by employing overlay of partitions, without loss of generality, we can suppose that the right hand side of (160) with $P=P_{k}$ and $v=v_{k}$ is bounded by a constant multiple of $2^{-k s}$. Looking at the left hand side then reveals that $T u \in \mathscr{A}_{\infty ; d}^{s}\left(L_{2}^{1}(\Omega)\right)$ and $u \in \mathscr{A}^{s}\left(H_{\Gamma}^{1}\right)$.

## Acknowledgements

This work was supported by NSERC Discovery Grants Program and FQRNT New University Researchers Start Up Program.

## References

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 $W_{p}^{\alpha}$. Mat. Sb. (N.S.), 73(3):331-355, 1967.

Andrea Bonito and Ricardo H. Nochetto. Quasi-optimal convergence rate of an adaptive discontinuous Galerkin method. SIAM J. Numer. Anal., 48(2):734-771, 2010. ISSN 0036-1429. doi: 10.1137/08072838X. URL http://dx.doi.org/10.1137/08072838X.
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.

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, 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.

Albrecht Pietsch. Approximation spaces. J. Approx. Theory, 32(2):115-134, 1981. ISSN 0021-9045. doi: 10.1016/0021-9045(81)90109-X. URL http://dx.doi.org/10.1016/ 0021-9045 (81) 90109-X.
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.


[^0]:    ${ }^{1}$ 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).

