Convergence Rates of AFEM with H^{-1} Data

Albert Cohen, Ronald DeVore, and Ricardo H. Nochetto [*]

April 20, 2011

Abstract

This paper studies Adaptive Finite Element Methods (AFEMs), based on piecewise linear elements and newest vertex bisection, for solving second order elliptic equations with piecewise constant coefficients on a polygonal domain $\Omega \subset \mathbb{R}^2$. The main contribution is to build algorithms that hold for a general right hand side $f \in H^{-1}(\Omega)$. Prior work assumes almost exclusively that $f \in L^2(\Omega)$. New data indicators based on local H^{-1} norms are introduced and then the AFEMs are based on a standard bulk chasing strategy (or Dörfler marking) combined with a procedure that adapts the mesh to reduce these new indicators. An analysis of our AFEM is given which establishes a contraction property and optimal convergence rates. In contrast to previous work, it is shown that it is not necessary to assume a compatible decay of the data estimator but rather that this is automatically guaranteed by the approximability assumptions on the solution by adaptive meshes, without additional assumptions on f. Computable surrogates for the data indicators are introduced and shown to also yield optimal convergence rates.

AMS subject classifications. 41A25, 41A65, 65N12, 65N15,65N30

1 Introduction

The theoretical understanding of the performance of Adaptive Finite Element Methods (AFEMs) for ellliptic problems has matured significantly during the last decade. This has led to the construction of AFEMs whose performance is, in a certain sense (described below), provably optimal when measured by error decay versus cardinality (and number of computations provided optimal iterative solvers and storage are used). However, even in the simplest settings, such as a Poisson problem on a polyhedral domain in $\mathbb{R}^2$, there remain important issues that need to be resolved in order to bring adaptivity theory to its most natural and complete form. The present paper centers on two of these issues:

[*] This research was supported by the Office of Naval Research Contracts ONR-N00014-08-1-1113 and ONR N00014-09-1-0107; the AFOSR Contract FA95500910500; the ARO/DoD Contract W911NF-07-1-0185; the NSF Grants DMS-0807811 and DMS-0915231; the Agence Nationale de la Recherche (ANR) project ECHANGE (ANR-08-EMER-006); the excellence chair of the Fondation "Sciences Mathématiques de Paris" held by Ronald DeVore. This publication is based on work supported by Award No. KUS-C1-016-04, made by King Abdullah University of Science and Technology (KAUST).

- The minimal conditions on the data f which are needed for building an AFEM and deriving convergence rates;
- The role of "data oscillation" assumptions in the analysis of convergence rates for AFEMs.

Since our main interest is to clearly put forward the new ideas necessary to properly handle these two issues, we shall only treat the model elliptic problem in two space dimensions:

$$-\mathrm{div}(A\nabla u) = f \quad \text{in} \quad \Omega, \qquad u|_{\partial\Omega} = 0, \tag{1.1}$$

where Ω is a polygonal domain in $\mathbb{R}^2$, and $x \mapsto A(x)$ is a matrix valued function such that $A(x)$ is symmetric positive definite for all $x \in \Omega$ and has eigenvalues $\lambda_i(x)$ satisfying

$$0 < a_{\min} \le \lambda_i(x) \le a_{\max}, \ \ x \in \Omega,$$

where $a_{\min}$ and $a_{\max}$ are fixed positive numbers. In addition, we assume that Ω can be decomposed into a (disjoint) partition

$$\Omega = \Omega_1 \cup \cdots \cup \Omega_j,$$

where each Ω_j is itself a polygonal subdomain and $A(x)$ equals a fixed matrix A_j for all $x \in \Omega_j$, i.e. $A(x)$ is piecewise constant over this partition. We further assume that the subdomains Ω_j are matched by the initial mesh $\mathcal{T}_0$ in the sense that each Ω_j is a union of triangles in $\mathcal{T}_0$. Our reason for working under these assumptions on A is that they allow for interface singularities [7],[9], which are much more extreme than corner singularities.

It is customary in the a posteriori error analysis of FEM to assume that $f \in L^2(\Omega)$. Then, the variational formulation of (1.1) consists in finding $u \in H_0^1(\Omega)$ such that

$$a(u,v) = L(v), \ \ v \in H_0^1(\Omega), \tag{1.2}$$

where the bilinear form a is given by

$$a(u,v) := \int_\Omega A\nabla u \cdot \nabla v,$$

and the linear form L is given by

$$L(v) := \int_\Omega fv.$$

Lax-Milgram theory ensures the existence and uniqueness of the solution u of (1.2) under the more general assumption that L is continuous on $H_0^1(\Omega)$ and a is coercive and continuous on $H_0^1(\Omega) \times H_0^1(\Omega)$. Therefore the well-posedness of (1.2) remains valid when $f \in H^{-1}(\Omega)$, the dual space of $H_0^1(\Omega)$, with L now defined by

$$L(v) := \langle f, v \rangle,$$

where $\langle \cdot, \cdot \rangle$ denotes the (H^{-1}, H_0^1) duality bracket.

Our interest in H^{-1} data is twofold: first it is the natural functional setting for (1.2) and the study of AFEM, and second there are important applications, some discussed below, for which $f \notin L^2(\Omega)$. With the exception of Nochetto [11] and Stevenson [13],[14], studies of AFEM always assume $f \in L^2(\Omega)$ and rely on a specific form of data oscillation which has significant implications on the structure of AFEM; see (3.21) below. Our approach is, however, different from [13],[14] in several respects: we do not approximate f by piecewise constants as in [13],[14], which is somewhat arbitrary for general $f \in H^{-1}(\Omega)$; we localize the global H^{-1} norm of f to stars ω_z and thus define the new local data indicators $\|f\|_{H^{-1}(\omega_z)}$; and we examine the relation between approximation classes for u and f, which require dealing with approximations of f other than piecewise constants. Despite these differences, we do have a procedure ADAPTDATA to reduce the data estimator which is similar to the inner loop in [13],[14].

AFEMs for approximating u generate a sequence of nested *conforming* (no hanging nodes) triangulations $\{\mathcal{T}_k\}_{k\geq 0}$ of Ω, starting from $\mathcal{T}_0$. For each $\mathcal{T}_k$ they find an approximation $U_{\mathcal{T}_k}$ to u by solving a Galerkin problem for a finite element space of piecewise polynomials subordinate to $\mathcal{T}_k$. Each AFEM is built on (i) a fixed rule for refining triangles, and (ii) a specified finite element space over triangulations generated by the particular refinement rule. Starting with an initial triangulation $\mathcal{T}_0$ of Ω, the subsequent triangulations $\mathcal{T}_k$ are generated by certain adaptive strategies. Typically, the AFEM computes the Galerkin solution $U_{\mathcal{T}_k}$ on the triangulation $\mathcal{T}_k$, estimates the residual $r_k := f + \operatorname{div}(AU_{\mathcal{T}_k})$ (which in our case is in $H^{-1}(\Omega)$), and uses this to compute an error indicator $e_T := e_T(U_{\mathcal{T}_k}, f)$ for each triangle $T \in \mathcal{T}_k$. These error indicators are then used to decide which triangles from $\mathcal{T}_k$ will be further refined in order to improve accuracy. This process is often referred to as *marking*. After refining the marked triangles and doing any additional refinements necessary to remove hanging nodes, the new triangulation $\mathcal{T}_{k+1}$ is obtained and the process is repeated. We refer the reader to Nochetto et al. [12] for an up to date survey of the current theory of AFEMs for elliptic problems.

To understand the issues raised in this paper, we briefly recall the main results of AFEM theory in the special case of newest vertex bisection and Lagrange P_m elements. The customary performance analysis of AFEMs measures the approximation error in the $H_0^1(\Omega)$ norm

$$\|v\|_{H_0^1(\Omega)} := \|\nabla v\|_{L^2(\Omega)},$$

and measures the complexity of a triangulation $\mathcal{T}$ by its cardinality $\#(\mathcal{T})$; this matches well the computational effort necessary to compute the Galerkin solution $U_\mathcal{T}$ provided optimal iterative solvers are available. We let $\mathfrak{T}_N$ denote the set of all conforming triangulations $\mathcal{T}$ with $\#(\mathcal{T}) \leq N$ that can be obtained from $\mathcal{T}_0$ by the newest vertex bisection process, and let

$$\sigma_N(u) := \inf_{\mathcal{T}\in\mathfrak{T}_N} \|u - U_\mathcal{T}\|_{H_0^1(\Omega)} \tag{1.3}$$

be the best error we could ever obtain by such triangulations (hereafter we assume $N \geq \#(\mathcal{T}_0)$). An ideal AFEM would be one that generates triangulations $\mathcal{T}_k$ such that

$$\|u - U_{\mathcal{T}_k}\|_{H_0^1(\Omega)} \leq C\sigma_{N_k}(u), \tag{1.4}$$

with $N_k := \#(\mathcal{T}_k)$ and C an absolute constant. There is no AFEM that is known to produce such an ideal performance.

On the other hand, there are AFEMs which come close to this performance. To describe these results, for $s > 0$ let $\mathcal{A}^s$ be the set of functions u such that

$$|u|_{\mathcal{A}^s} := \sup_{N \geq \#(\mathcal{T}_0)} N^s \sigma_N(u) < +\infty. \tag{1.5}$$

The set $\mathcal{A}^s$ is a quasi-Banach space, when equiped with the quasi-norm

$$\|u\|_{\mathcal{A}^s} := |u|_{\mathcal{A}^s} + \|u\|_{H^1_0(\Omega)}.$$

A more modest goal than (1.4) would be to construct AFEMs such that whenever $u \in \mathcal{A}^s$, the AFEM produces triangulations $\mathcal{T}_k$ such that

$$\|u - U_{\mathcal{T}_k}\|_{H^1_0(\Omega)} \leq C|u|_{\mathcal{A}^s}\#(\mathcal{T}_k)^{-s}, \tag{1.6}$$

with C an absolute constant. Starting with Binev, Dahmen, and DeVore [2], AFEMs have been constructed which exhibit this performance, except for one caveat: to guarantee this performance it is assumed that $f \in L^2(\Omega)$ [2],[5],[12], or alternatively $f \in H^{-1}(\Omega)$ can be approximated by piecewise constants over $\mathcal{T}_k$ [13],[14], properties which are *not* a consequence of $u \in \mathcal{A}^s$.

Let us elaborate further on this last issue. Given a triangulation $\mathcal{T}$ and the Galerkin solution $U_{\mathcal{T}}$ associated to $\mathcal{T}$, a typical AFEM uses the *residual-type* local error indicators

$$e_T^2 := h_T^2\|f + \operatorname{div}(A\nabla U_{\mathcal{T}})\|^2_{L^2(T)} + h_T\|J(U_{\mathcal{T}})\|^2_{L^2(\partial T)}, \tag{1.7}$$

on each triangle $T \in \mathcal{T}$. Here, h_T is the diameter of T and $J(U_{\mathcal{T}})$ is the jump of $A\nabla U_{\mathcal{T}} \cdot \nu_\sigma$ across a side (edge) σ of T, where ν_σ is a unit outward normal to σ. These error indicators do indeed control the global error from above in the sense that $\|u - U_{\mathcal{T}}\|_{H^1_0(\Omega)} \lesssim \sum_{T\in\mathcal{T}} e_T^2$. However, to control the error from below one needs to introduce an extra term $\operatorname{osc}(f, \mathcal{T})$, referred to as *data oscillation*, which then gives

$$\sum_{T\in\mathcal{T}} e_T^2 \lesssim \|u - U_{\mathcal{T}}\|^2_{H^1_0(\Omega)} + \operatorname{osc}(f, \mathcal{T})^2.$$

A typical form of this term for Lagrange finite elements of degree m is

$$\operatorname{osc}(f, \mathcal{T}) := \Big(\sum_{T\in\mathcal{T}} h_T^2\|f - a_T(f)\|^2_{L^2(T)}\Big)^{\frac{1}{2}}, \tag{1.8}$$

where $a_T(f)$ is the $L^2(T)$ orthogonal projection of f onto $\mathbb{P}_{m-1}$, the space of polynomials of total degree $\leq m - 1$. In the particular case $m = 1$ of piecewise linear elements, $a_T(f)$ is the meanvalue of f over T.

A typical AFEM computes the error indicators e_T for each $T \in \mathcal{T}$, and marks for refinement the triangles $T \in \mathcal{T}$ with largest error indicator. Among the many possible strategies on deciding which triangles to mark, the strongest analytical results are obtained when the marking is done using a bulk chasing criteria introduced by Dörfler [6], the so-called Dörfler marking. Recent work, see [5],[9],[12], has shown that AFEMs based on this marking strategy achieve a benchmark similar to (1.6) provided that $u \in \mathcal{A}^s$ and that in addition, f has enough smoothness so that the data oscillation term is controlled by the algorithm with the same rate of decay N^{-s}, for $s \leq m/2$. A glance at (1.7) and (1.8) reveals that a minimal requirement for the applicability of this approach is that $f \in L^2(\Omega)$. Moreover, this requirement on f is even more demanding when dealing with higher order Lagrange elements $m > 1$.

In order to remedy these issues, we need to develop new algorithms and theory for AFEMs that work on the minimal assumption that $f \in H^{-1}(\Omega)$. Moreover, such new algorithms could also be useful in practice since there are several relevant elliptic problems which give rise to a right hand side that belongs to $H^{-1}(\Omega)$ but not to $L^2(\Omega)$. Here are at least two of them:

- The function f belongs to $L^p(\Omega)$ for some $1 < p < 2$. Then standard imbedding theorems ensure that $f \in H^s(\Omega)$ for $s = 1 - \frac{2}{p} < 0$ and therefore $f \in H^{-1}(\Omega)$.
- The linear form L has the form $L(v) := \int_\Omega \nabla g \cdot \nabla v$, where g is a function in $H_0^1(\Omega)$ such that $\Delta g \notin L^2(\Omega)$. A typical example is when g is itself a $\mathbb{P}_m$ finite element function. Then $f = -\Delta g$ typically contains Dirac distributions supported on the edges σ of the mesh with densities the jumps of $\nabla g \cdot \nu_\sigma$. Such distributions are in $H^s(\Omega)$ only for $-1 \leq s < -\frac{1}{2}$.

The main accomplishment of this paper is the development of a new AFEM, based on new error and data indicators which apply to general data $f \in H^{-1}(\Omega)$. The algorithm will be shown to give the optimal convergence rate N^{-s} under the sole assumption that $u \in \mathcal{A}^s$, in contrast to earlier approaches such as in [2],[5],[6],[9],[10],[13],[14] where additional assumptions are made on f. Notice that the assumption $u \in \mathcal{A}^s$ only implies that $f \in H^{-1}(\Omega)$ (and not necessarily that $f \in L^2(\Omega)$). Since we are restricting our attention to piecewise linear elements ($m = 1$), the range of s is $0 < s \leq \frac{1}{2}$.

An outline of this paper is as follows. We begin in §2 by recalling the properties of newest vertex bisection that we shall need for our analysis. Then, we derive in §3 a posteriori error estimates for piecewise linear finite elements by using local H^{-1} norms for f. The error estimator consists of a jump estimator and a *data estimator* which is of the form

$$\mathcal{D}(f, \mathcal{T}) := \Big(\sum_{z \in \mathcal{N}(\mathcal{T})} \|f\|_{H^{-1}(\omega_z)}^2 \Big)^{\frac{1}{2}}, \tag{1.9}$$

where $\mathcal{N}(\mathcal{T})$ is the set of nodes in the triangulation $\mathcal{T}$ and ω_z the union of the triangles of $\mathcal{T}$ having z as a vertex (the so-called star or patch). The data estimator (1.9) is thus defined for any $f \in H^{-1}(\Omega)$, in contrast to the data terms involving L^2 norms in (1.7) and (1.8) or the replacement of $f \in H^{-1}(\Omega)$ by piecewise constants of [13],[14], which is a somewhat arbitrary.

We formulate in §4 an AFEM that applies to general $f \in H^{-1}(\Omega)$. We prove a contraction property for this AFEM that implies in particular convergence towards the exact solution. This algorithm combines the bulk chasing strategy based on the new error indicators, together with a generic refinement procedure ADAPTDATA, referred to as *data adaptation*, which reduces (1.9) to a prescribed tolerance. This procedure is similar to Stevenson's inner loop [13],[14], which was later eliminated by Cascón et al. [5] in the context of L^2 data. When treating general H^{-1} data, the reduction of $\mathcal{D}(f, \mathcal{T})$ must be explicitly enforced for the AFEM to converge and exhibit an optimal decay rate.

In §5, we begin our analysis of the convergence rates for our AFEM. We first show that the optimal convergence rate N^{-s} is achieved by our algorithm, provided that $u \in \mathcal{A}^s$ and that there is a subroutine $\mathsf{ADAPTDATA}$ that exhibits a similar convergence rate N^{-s} for $\mathcal{D}(f,\mathcal{T})$ in terms of $N = \#(\mathcal{T})$. We later build in §7 concrete realizations of $\mathsf{ADAPTDATA}$ which satisfy this property under various assumptions on f.

In §6, we prove that there exists an optimal data adaptation procedure for which the convergence rate N^{-s} of $\mathcal{D}(f,\mathcal{T})$ is ensured under the sole assumption that $u \in \mathcal{A}^s$, for $s < 1/2$; we also examine the borderline case $s = 1/2$. This optimal procedure is tied to evaluating the local $H^{-1}(\omega_z)$ norms in $\mathcal{D}(f,\mathcal{T})$ and finding optimal meshes for u. This is impractical because the computation of local $H^{-1}(\omega_z)$ norms in (1.9) is generally not viable. However, we introduce surrogate quantities in §7 that are computable, provided $f \in L^p(\Omega)$ for some $p > 1$ or f is a Dirac distribution on a 1-dimensional curve, thereby leading to a larger data estimator $\widetilde{\mathcal{D}}(f,\mathcal{T})$. In such cases, the AFEM can be built on these surrogate quantities, and $\mathsf{ADAPTDATA}$ can be implemented using a simple and practical *greedy algorithm* for data adaptation. In addition, the optimal decay rate N^{-s} is achieved under the assumption that $u \in \mathcal{A}^s$, for any $0 < s \leq 1/2$.

We end by some concluding remarks on possible extensions of our approach in §8.

2 Newest Vertex Bisection and Piecewise Linear Elements

In this section, we briefly recall newest vertex bisection and properties of the space of piecewise linear elements. The starting point is an initial conforming triangulation $\mathcal{T}_0$ of the polygonal domain Ω into a finite number of triangles. In newest vertex bisection, each edge of $\mathcal{T}_0$ is given an initial label of either 0 or 1. Such labels are required to depend only on the edge σ and not on the triangles to which σ belongs. The *initial labeling* of the edges of $\mathcal{T}_0$ is required to have the property that for each triangle $T \in \mathcal{T}_0$, exactly one edge of T has the label 0 and the other two have label 1. It is known that it is always possible to make such an initial assignment [8],[2].

Any triangle T created by the bisection procedure will have edges labeled $i, i+1, i+1$ with $i \geq 0$ integer. The vertex opposite the side labeled i is called *the newest vertex* [8],[2]. If T is to be bisected then this refinement is done by connecting the midpoint of the side marked i with the newest vertex. This leads to the creation of two new triangles (called the *children* of T), and three new edges. Each of these new edges is given the label $i+2$. The newest vertex of each children is the vertex opposite the side labeled $i+1$ (the side with the smallest label). Once $\mathcal{T}_0$ is equipped with the initial labeling, these rules guarantee that each edge has unique labeling regardless of the triangles sharing it and no ambiguity arises in this process; see [2],[8],[12] for details. The label i of an edge indicates its *generation*. Also the lowest labeled edge of a triangle T gives the generation of T, i.e. how many bisections were made to create T.

Unless implemented recursively [12], this *newest vertex bisection* procedure creates meshes $\mathcal{T}$ which are non-conforming. They can be represented by a (finite) binary forest whose roots are the elements in $\mathcal{T}_0$ and whose leaves are the elements of $\mathcal{T}$. Each such forest is contained in an (infinite) master forest which consists of all triangles that may be generated from $\mathcal{T}_0$ by the newest vertex bisection procedure. We are mainly interested in *conforming* meshes (no hanging nodes), which correspond to a restricted class of finite binary forests.

If $\mathcal{T}_k$ is a conforming mesh obtained from $\mathcal{T}_0$ by newest vertex bisection and $\mathcal{R}_k$ is a subset of $\mathcal{T}_k$ of triangles to be refined, the refinement procedure creates a non-conforming mesh $\overline{\mathcal{T}}_k$ by bisecting the elements in $\mathcal{R}_k$, and finds the *conforming refinement* of $\overline{\mathcal{T}}_k$ by adding the *smallest* number of additional newest vertex bisection steps so that the resulting mesh $\mathcal{T}_{k+1}$ is conforming. The second step (completion) is *nonlocal* and its complexity is rather intricate but critical for the overall complexity analysis. It was shown by Binev, Dahmen, and DeVore [2] that the cost of conforming refinement can be uniformly controlled. We express this crucial result as follows.

Lemma 2.1 (complexity of conforming bisection)**.** *There exists a constant $C_0 \geq 1$, depending only on $\mathcal{T}_0$, such that for all $k \geq 1$*

$$\#(\mathcal{T}_k) - \#(\mathcal{T}_0) \leq C_0 \sum_{j=0}^{k-1} \#(\mathcal{R}_j). \tag{2.1}$$

We now discuss two easy consequences of this result. Let $\widetilde{\mathcal{T}}$ be a non-conforming triangulation obtained from $\mathcal{T}_0$ by several newest vertex bisections. In other words, we do not assume that hanging nodes are removed after each set of bisections as is the case in the last Lemma. If $\mathcal{T}$ is the smallest conforming refinement of $\widetilde{\mathcal{T}}$, then

$$\#(\mathcal{T}) - \#(\mathcal{T}_0) \leq C_0\big(\#(\widetilde{\mathcal{T}}) - \#(\mathcal{T}_0)\big), \tag{2.2}$$

where C_0 is the same constant as in (2.1). This comes from the fact that the same $\mathcal{T}$ would also result from applying conforming refinement after each bisection step. A second scenario is that $\widetilde{\mathcal{T}}$ is a non-conforming refinement of an arbitrary conforming refinement $\mathcal{T}$ of $\mathcal{T}_0$. If $\overline{\mathcal{T}}$ is the smallest conforming refinement of $\widetilde{\mathcal{T}}$, then

$$\#(\overline{\mathcal{T}}) \leq \#(\mathcal{T}_0) + C_0\big((\#(\widetilde{\mathcal{T}}) - \#(\mathcal{T})) + (\#(\mathcal{T}) - \#(\mathcal{T}_0))\big) \leq C_0 \#(\widetilde{\mathcal{T}}). \tag{2.3}$$

For any two meshes $\mathcal{T}$ and $\mathcal{T}^*$ created from $\mathcal{T}_0$ by newest vertex bisection, we denote by

$$\mathcal{T} \oplus \mathcal{T}^*,$$

the *overlay* of the two meshes, consisting of the union of all triangles of $\mathcal{T}$ that do not contain smaller triangles of $\mathcal{T}^*$ and of all triangles of $\mathcal{T}^*$ that do not contain smaller triangles of $\mathcal{T}$. This overlay can be obtained by performing all bisections called for in the generation of $\mathcal{T}$ and $\mathcal{T}^*$ from $\mathcal{T}_0$. Hence, we clearly have

$$\#(\mathcal{T} \oplus \mathcal{T}^*) \leq \#(\mathcal{T}) + \#(\mathcal{T}^*) - \#(\mathcal{T}_0). \tag{2.4}$$

We sometimes write

$$\mathcal{T}^* \geq \mathcal{T},$$

to say that $\mathcal{T}^*$ is a *refinement* of $\mathcal{T}$, which means that it is obtained from $\mathcal{T}$ by applying additional steps of newest vertex bisection, or equivalently that $\mathcal{T} \oplus \mathcal{T}^* = \mathcal{T}^*$. Note that the overlay between two such conforming meshes is also conforming.

Remark 2.2. Most constants occuring in this paper depend only on the geometry of $\mathcal{T}_0$. This utilizes the property that any triangle created by newest vertex bisection is similar to one of a fixed number of equivalence classes dictated by $\mathcal{T}_0$ and its labeling; see [8]. Throughout this paper, C typically denotes a constant that only depends on the initial triangulation $\mathcal{T}_0$ and its labeling, unless stated otherwise. We shall also sometimes write

$$A(\cdot) \ \lesssim\ B(\cdot)$$

when $A(\cdot) \leq CB(\cdot)$ for such a constant independent of the arguments in A and B.

For a conforming mesh $\mathcal{T}$, we denote by $\mathcal{N}(\mathcal{T})$ the nodes (or vertices) of $\mathcal{T}$, and by $\mathcal{N}_0(\mathcal{T})$ the subset of nodes that are interior to Ω. For $z \in \mathcal{N}(\mathcal{T})$, we denote by ϕ_z the piecewise linear hat function such that $\phi_z(z') = \delta_{z,z'}$ for all $z' \in \mathcal{N}(\mathcal{T})$. The support sets

$$\omega_z := \mathrm{Supp}(\phi_z) = \cup\{T \in \mathcal{T}\ ;\ z \in T\}$$

are called *stars*. We denote by $\Gamma(\mathcal{T})$ the set of all inner edges σ of $\mathcal{T}$, i.e. edges which are not contained on the boundary $\partial\Omega$. For each $z \in \mathcal{N}(\mathcal{T})$ we denote by $\Gamma(\omega_z)$ the set of inner edges interior to ω_z, i.e. those inner edges which have z as an end point, and we define the skeleton of ω_z as

$$\gamma_z := \cup\{\sigma\ ;\ \sigma \in \Gamma(\omega_z)\}.$$

We denote by h_T the diameter of a triangle T and by h_z the diameter of ω_z. The triangulations built from newest vertex bisection are shape regular and graded in the sense that all possibly generated triangles satisfy a uniform *smallest angle* condition. From this it follows that h_z is uniformly equivalent to h_T: for all $z \in \mathcal{N}(\mathcal{T})$ and $T \in \mathcal{T}$ such that $T \subset \omega_z$,

$$h_z \lesssim h_T \leq h_z.$$

We denote by $\mathbb{V}(\mathcal{T})$ the space of piecewise linear functions subordinate to $\mathcal{T}$, and by $\mathbb{V}_0(\mathcal{T})$ those functions in $\mathbb{V}(\mathcal{T})$ which vanish on the boundary $\partial\Omega$ of Ω. The hat functions $\{\phi_z\}_{z\in\mathcal{N}(\mathcal{T})}$ are the canonical basis of $\mathbb{V}(\mathcal{T})$ and, likewise, $\{\phi_z\}_{z\in\mathcal{N}_0(\mathcal{T})}$ are the canonical basis of $\mathbb{V}_0(\mathcal{T})$. We recall the partition of unity property:

$$\sum_{z\in\mathcal{N}(\mathcal{T})} \phi_z(x) = 1, \quad x \in \Omega. \tag{2.5}$$

3 A Posteriori Error Analysis with H^{-1} Data

In this section, we introduce certain *local* $H^{-1}(\Omega)$ error indicators and derive some of their properties. We denote by

$$|||v|||_\Omega := a(v,v)^{\frac{1}{2}} = \Big(\int\limits_\Omega A\nabla v\cdot\nabla v\Big)^{\frac{1}{2}},$$

the energy norm associated to the problem (1.2) which is equivalent to the H_0^1 norm

$$\sqrt{a_{\min}}\|v\|_{H_0^1(\Omega)} \leq |||v|||_\Omega \leq \sqrt{a_{\max}}\|v\|_{H_0^1(\Omega)}, \tag{3.1}$$

where $\|v\|_{H_0^1(\Omega)} := \|\nabla v\|_{L^2(\Omega)}$. For any subdomain $\omega \subset \Omega$, we define the local energy seminorm

$$|||v|||_{\omega} := \Big(\int_{\omega} A\nabla v \cdot \nabla v\Big)^{\frac{1}{2}}.$$

3.1 An Error-Residual Equation

Let $f \in H^{-1}(\Omega)$, and let $u \in H_0^1(\Omega)$ be the exact solution of our model problem

$$a(u, v) = \langle f, v\rangle, \quad v \in H_0^1(\Omega). \tag{3.2}$$

Fix a conforming triangulation $\mathcal{T} \geq \mathcal{T}_0$ and denote by $U := U_{\mathcal{T}} \in \mathbb{V}_0(\mathcal{T})$ the Galerkin solution

$$a(U, V) = \langle f, V\rangle, \quad V \in \mathbb{V}_0(\mathcal{T}). \tag{3.3}$$

Equivalently, U is the orthogonal projection of u onto $\mathbb{V}_0(\mathcal{T})$ in the sense of the inner product $a(\cdot,\cdot)$. From the equivalence (3.1), U is a near best approximation to u in the H_0^1 norm:

$$\|u - U\|_{H_0^1(\Omega)} \leq \sqrt{\frac{a_{\max}}{a_{\min}}}\|u - V\|_{H_0^1(\Omega)}, \quad V \in \mathbb{V}_0(\mathcal{T}). \tag{3.4}$$

Integration by parts yields the following relation between the error $u - U$ and the residual

$$a(u - U, v) = \langle f + \operatorname{div}(A\nabla U), v\rangle = \langle f, v\rangle + \sum_{\sigma\in\Gamma(\mathcal{T})} \int_{\sigma} Jv, \quad v \in H_0^1(\Omega), \tag{3.5}$$

where on each edge σ, with normal $\nu := \nu_\sigma$,

$$J := J(U) := J_\sigma(U) := [A\nabla U] \cdot \nu,$$

is the jump residual. Since A is piecewise constant over $\mathcal{T} \geq \mathcal{T}_0$, J is constant on each $\sigma \in \Gamma(\mathcal{T})$. Using (2.5) in the error-residual relation, we end up with a similar relation localized to all stars ω_z:

$$a(u - U, v) = \sum_{z\in\mathcal{N}(\mathcal{T})} \Big(\langle f, v\phi_z\rangle + \int_{\gamma_z} Jv\phi_z\Big). \tag{3.6}$$

3.2 Reliability: Global Upper Bound

In order to derive an upper bound for the error, we first observe that *Galerkin orthogonality* implies

$$\langle f, \phi_z\rangle + \int_{\gamma_z} J\phi_z = 0, \quad z \in \mathcal{N}_0(\mathcal{T}). \tag{3.7}$$

We first exploit this to rewrite (3.6) as follows,

$$a(u - U, v) = \sum_{z\in\mathcal{N}(\mathcal{T})} \Big(\langle f, (v - \alpha_z(v))\phi_z\rangle + \int_{\gamma_z} J(v - \alpha_z(v))\phi_z\Big), \tag{3.8}$$

where $\alpha_z(v) \in \mathbb{R}$ is defined as the weighted meanvalue

$$\alpha_z(v) := \frac{\int\limits_{\omega_z} v\phi_z}{\int\limits_{\omega_z} \phi_z}, \quad z \in \mathcal{N}_0(\mathcal{T}),$$

and $\alpha_z(v) := 0$ if $z \in \mathcal{N}(\mathcal{T}) \setminus \mathcal{N}(\mathcal{T}_0)$ is a boundary node. We now estimate each term on the right-hand side of (3.8) separately assuming that $v \in H_0^1(\Omega)$. On the one hand, we have

$$\begin{aligned} \langle f, (v - \alpha_z(v))\phi_z \rangle &\leq \|f\|_{H^{-1}(\omega_z)} \|\nabla[(v - \alpha_z(v))\phi_z]\|_{L^2(\omega_z)} \\ &\leq \|f\|_{H^{-1}(\omega_z)} (\|\nabla v\|_{L^2(\omega_z)} + \|v - \alpha_z(v)\|_{L^2(\omega_z)} \|\nabla \phi_z]\|_{L^\infty(\omega_z)}) \\ &\lesssim \|f\|_{H^{-1}(\omega_z)} \|\nabla v\|_{L^2(\omega_z)}. \end{aligned}$$

Here we have used the fact that $\|\nabla \phi_z\|_{L^\infty(\omega_z)} \lesssim h_z^{-1}$, as well as the rescaled Poincaré type inequality

$$\|v - \alpha_z(v)\|_{L^2(\omega_z)} \lesssim h_z \|\nabla v\|_{L^2(\omega_z)}, \tag{3.9}$$

which is in its usual form for interior nodes but is also valid for boundary nodes (because v vanishes at least on one of the edges that constitute the boundary of ω_z; see [12]). On the other hand, from the rescaled trace theorem and (3.9), we have

$$\begin{aligned} \int\limits_{\gamma_z} J(v - \alpha_z(v))\phi_z &\leq \|J\|_{L^2(\gamma_z)} \|v - \alpha_z(v)\|_{L^2(\gamma_z)} \\ &\lesssim \|J\|_{L^2(\gamma_z)} \Big(h_z^{1/2} \|\nabla v\|_{L^2(\omega_z)} + h_z^{-1/2} \|v - \alpha_z(v)\|_{L^2(\omega_z)} \Big) \\ &\lesssim h_z^{1/2} \|J\|_{L^2(\gamma_z)} \|\nabla v\|_{L^2(\omega_z)} \\ &\lesssim \Big(\sum_{\sigma \in \Gamma(\omega_z)} |\sigma|^2 |J_\sigma|^2 \Big)^{1/2} \|\nabla v\|_{L^2(\omega_z)}. \end{aligned}$$

Since all points of Ω belong to at most 3 stars ω_z, except for a set of zero Lebesgue measure, this implies

$$a(u - U, v) \lesssim \Big(\sum_{z \in \mathcal{N}(\mathcal{T})} \Big(\|f\|^2_{H^{-1}(\omega_z)} + \sum_{\sigma \in \Gamma(\omega_z)} |\sigma|^2 |J_\sigma|^2 \Big) \Big)^{\frac{1}{2}} \|\nabla v\|_{L^2(\Omega)}. \tag{3.10}$$

Motivated by (3.10), we introduce the local jump residual and data indicators

$$j(z) := j(U, z, \mathcal{T}) := \Big(\sum_{\sigma \in \Gamma(\omega_z)} |\sigma|^2 |J_\sigma|^2 \Big)^{1/2} \text{ and } d(z) := d(f, z, \mathcal{T}) := \|f\|_{H^{-1}(\omega_z)}. \tag{3.11}$$

We also introduce their global counterparts

$$\mathcal{J}(U, \mathcal{T}) := \Big(\sum_{z \in \mathcal{N}(\mathcal{T})} j(z)^2 \Big)^{\frac{1}{2}} \text{ and } \mathcal{D}(f, \mathcal{T}) = \Big(\sum_{z \in \mathcal{N}(\mathcal{T})} d(z)^2 \Big)^{\frac{1}{2}}.$$

The local error indicator $e(z)$ and global error estimator $\mathcal{E}$ are then given by

$$e(z)^2 := e(U, f, z)^2 := j(z)^2 + d(z)^2 \text{ and } \mathcal{E}^2 := \mathcal{E}(U, f, \mathcal{T})^2 := \sum_{z \in \mathcal{N}(\mathcal{T})} e(z)^2. \tag{3.12}$$

Using (3.10), together with the norm equivalence (3.1), we reach an a posteriori global upper bound for H^{-1} data expressed as follows.

Lemma 3.1 (global upper bound)**.** *There exists a constant $C_G > 0$ that only depends on the initial mesh $\mathcal{T}_0$ and $a_{\min}$, such that*

$$|\!|\!| u - U |\!|\!|_\Omega \le C_G \; \mathcal{E}(U, f, \mathcal{T}). \tag{3.13}$$

If $\mathcal{M} \subset \mathcal{N}(T)$ is a set of nodes of $\mathcal{T}$, we use the notation $\mathcal{D}(f, \mathcal{M}; \mathcal{T})$, $\mathcal{J}(U, \mathcal{M}; \mathcal{T})$ and $\mathcal{E}(f, U, \mathcal{M}; \mathcal{T})$ for data, jump and error estimators *localized* to the nodes of $\mathcal{M}$. Namely, we define

$$\mathcal{E}(f, U, \mathcal{M}; \mathcal{T}) := \Big(\sum_{z \in \mathcal{M}} e(z)^2 \Big)^{\frac{1}{2}}, \tag{3.14}$$

and similarily define $\mathcal{D}(f, \mathcal{M}; \mathcal{T})$ and $\mathcal{J}(U, \mathcal{M}; \mathcal{T})$. It will be useful, for the discussion in §5, to have an upper bound for the energy error between two Galerkin solutions $U \in \mathbb{V}_0(\mathcal{T})$ and $U_* \in \mathbb{V}_0(\mathcal{T}_*)$ with $\mathcal{T}_*$ a conforming refinement of $\mathcal{T}$. The following result shows that this error is bounded by a localized estimator of the above type (3.14).

Lemma 3.2 (localized upper bound)**.** *Let $\mathcal{M} \in \mathcal{N}(\mathcal{T})$ be the set of all nodes $z \in \mathcal{N}(\mathcal{T})$ such that z is a vertex of a triangle $T \in \mathcal{T} \setminus \mathcal{T}^*$ which was refined in the process of constructing $\mathcal{T}^*$. Then,*

$$|\!|\!| U^* - U |\!|\!|_\Omega \le C_L \; \mathcal{E}(U, f, \mathcal{M}; \mathcal{T}), \tag{3.15}$$

where $C_L > C_G$ only depends on the initial mesh $\mathcal{T}_0$ as well as on $a_{\min}$ and C_G is the constant in (3.13).

Proof: We define $v := U^* - U \in \mathbb{V}_0(\mathcal{T}^*)$ and $w = v - V$ where $V \in \mathbb{V}_0(\mathcal{T})$ is an arbitrary function. In view of (3.6) and (3.7), we can write

$$\begin{aligned} |\!|\!| U^* - U |\!|\!|_\Omega^2 &= a(U^* - U, v) = a(u - U, v) = a(u - U, v - V) \\ &= \sum_{z \in \mathcal{N}(\mathcal{T})} \Big(\langle f, w\phi_z \rangle + \int\limits_{\gamma_z} J w \phi_z \Big) \\ &= \sum_{z \in \mathcal{N}(\mathcal{T})} \Big(\langle f, (w - \alpha_z(w))\phi_z \rangle + \int\limits_{\gamma_z} J (w - \alpha_z(w)) \phi_z \Big). \end{aligned}$$

We now take $V := \mathcal{P}_\mathcal{T} v$ where $\mathcal{P}_\mathcal{T}$ is a local Scott-Zhang projection operator [4] onto $\mathbb{V}_0(\mathcal{T})$ that we build as follows. For each $z \in \mathcal{N}_0(\mathcal{T})$ we pick a triangle $T_z \in \mathcal{T}$ in a such way that $T_z \subset \omega_z$ and the following property holds: if ω_z contains at least one triangle in $\mathcal{T} \cap \mathcal{T}^*$, i.e. a triangle of $\mathcal{T}$ that is not refined, we take for T_z such a triangle. For any $g \in L^2(\Omega)$, we then define $\pi_z g$ as its local $L^2(T_z)$-orthogonal projection onto Π_1 the space of affine polynomals, and $\beta_z(g) := \pi_z g(z)$ its value at z. We then set

$$\mathcal{P}_\mathcal{T} g := \sum_{z \in \mathcal{N}_0(\mathcal{T})} \beta_z(g) \phi_z.$$

It is easily seen that $\mathcal{P}_\mathcal{T}$ leaves $\mathbb{V}_0(\mathcal{T})$ invariant.

From the particular choice of T_z and the fact that $v \in \mathbb{V}_0(\mathcal{T}^*)$, we also find that $w = v - \mathcal{P}_\mathcal{T} v$ vanishes in all the triangles in $\mathcal{T} \cap \mathcal{T}^*$. Therefore,

$$|\!|\!| U^* - U |\!|\!|_\Omega^2 = \sum_{z \in \mathcal{M}} \Big(\langle f, w - \alpha_z(w)\phi_z \rangle + \int\limits_{\gamma_z} J (w - \alpha_z(w)) \phi_z \Big).$$

By the same arguments leading to Lemma 3.1, we thus obtain

$$|\!|\!|U^* - U|\!|\!|_\Omega^2 \lesssim \Big(\sum_{z\in\mathcal{M}} \Big(\|f\|_{H^{-1}(\omega_z)}^2 + \sum_{\sigma\in\Gamma(\omega_z)} |\sigma|^2 |J_\sigma|^2 \Big)\Big)^{\frac{1}{2}} \|\nabla w\|_{L^2(\Omega)}.$$

Observing that $\mathcal{P}_\mathcal{T}$ is uniformly H^1-stable in the sense that for all $g \in H_0^1(\Omega)$,

$$\|\nabla \mathcal{P}_\mathcal{T} g\|_{L^2(\Omega)} \;\lesssim\; \|\nabla g\|_{L^2(\Omega)},$$

we obtain that $\|\nabla w\|_{L^2(\Omega)} \le C\|\nabla v\|_{L^2(\Omega)} \le \frac{C}{\sqrt{a_{\min}}} |\!|\!|U - U^*|\!|\!|_\Omega$ with $C > 1$. This allows us to conclude the proof with $C_L > C_G$. □

3.3 Efficiency: Local Lower Bound

We want next to derive a local lower bound for the error $|\!|\!|u - U|\!|\!|_{\omega_z}$. We consider a star ω_z with $z \in \mathcal{N}(\mathcal{T})$ regardless of whether z is an interior or a boundary node. We construct a function of the form

$$\varphi = \sum_{\sigma\in\Gamma(\omega_z)} \alpha_\sigma \varphi_\sigma,$$

where the functions φ_σ are the canonical quadratic bubbles with value 1 at the mid-point of the side σ and zero at all other Lagrange quadratic nodes, and where

$$\alpha_\sigma := \frac{3}{2}|\sigma| J_\sigma, \;\; \sigma \in \Gamma(\omega_z).$$

Using φ as a test function in the error-residual relation yields (via Simpson's rule)

$$a(u - U, \varphi) = \langle f, \varphi\rangle + \sum_{\sigma\in\Gamma(\omega_z)} \alpha_\sigma J_\sigma \int_\sigma \varphi_\sigma = \langle f, \varphi\rangle + \frac{2}{3} \sum_{\sigma\in\Gamma(\omega_z)} |\sigma| \alpha_\sigma J_\sigma = \langle f, \varphi\rangle + j(z)^2.$$

We thus arrive at

$$\begin{aligned} j(z)^2 &= \int_{\omega_z} A\nabla(u - U)\cdot\nabla\varphi - \langle f, \varphi\rangle \\ &\le \Big(\sqrt{a_{\max}} |\!|\!|u - U|\!|\!|_{\omega_z} + \|f\|_{H^{-1}(\omega_z)}\Big) \|\nabla\varphi\|_{L^2(\omega_z)}. \end{aligned}$$

Since $\|\nabla\varphi_\sigma\|_{L^2(\omega_z)} \lesssim 1$, we have

$$\|\nabla\varphi\|_{L^2(\omega_z)} \lesssim \sum_{\sigma\in\Gamma(\omega_z)} |\sigma||J_\sigma| \lesssim j(z).$$

Therefore, we have proven the following local lower bound.

Lemma 3.3 (local lower bound)**.** *There exists a constant $c_1 > 0$, that only depends on $\mathcal{T}_0$ and on $a_{\max}$, such that*

$$c_1 j(z) \le |\!|\!|u - U|\!|\!|_{\omega_z} + d(z), \;\; z \in \mathcal{N}(\mathcal{T}). \tag{3.16}$$

Combining Lemmas 3.1 and 3.3, we immediately obtain the following result.

Corollary 3.4 (global lower and upper bound)**.** *There exist constants $0 < C_1 < C_2 \le 1 + C_G^2$, that only depend on $\mathcal{T}_0$ and on $a_{\min}$ and $a_{\max}$, such that*

$$C_1 \mathcal{E}(U, f, \mathcal{T})^2 \le |\!|\!|u - U|\!|\!|_\Omega^2 + \mathcal{D}(f, \mathcal{T})^2 \le C_2 \mathcal{E}(U, f, \mathcal{T})^2. \tag{3.17}$$

3.4 The Data Estimator

We discuss next some properties of the data estimator $\mathcal{D}(f,\mathcal{T})$. In particular, we compare the magnitude of this term with other data estimators previously used in the literature. If $f \in L^2(\Omega)$ and $U(H_0^1(\omega_z))$ is the unit ball of $H_0^1(\omega_z)$, then applying the Poincaré inequality to $v \in U(H_0^1(\omega_z))$ we get

$$d(z) = \|f\|_{H^{-1}(\omega_z)} = \sup_{v\in U(H_0^1(\omega_z))} \langle f, v\rangle \le \|f\|_{L^2(\omega_z)} \|v\|_{L^2(\omega_z)} \lesssim h_z \|f\|_{L^2(\omega_z)}. \tag{3.18}$$

The right side, which is the usual form of the interior residual for piecewise linear elements and piecewise constant coefficients A, can be much larger than $d(z)$ for an oscillatory function $f \in L^2(\Omega)$. In contrast, if f is constant in ω_z (polynomial suffices), then

$$\|f\|^2_{L^2(\omega_z)} = 3 \int\limits_{\omega_z} f^2 \phi_z \le 3 \|f\|_{H^{-1}(\omega_z)} \|f\phi_z\|_{H_0^1(\omega_z)} \lesssim h_z^{-1} \|f\|_{H^{-1}(\omega_z)} \|f\|_{L^2(\omega_z)},$$

and so in this case, $d(z)$ and the right side of (3.18) are of the same magnitude.

We can also shed some light on the relative sizes of $d(z)$ and $j(z)$: using (3.5) with $v \in U(H_0^1(\omega_z))$ leads to

$$d(z) = \|f\|_{H^{-1}(\omega_z)} \lesssim |\!|\!| u - U |\!|\!|_{\omega_z} + j(z). \tag{3.19}$$

This and Lemma 3.3 show that neither indicator $j(z), d(z)$ dominates the other for $f \notin L^2(\omega_z)$. On the other hand, the derivation of Lemmas 3.1 and 3.3 reveals that we could have replaced $\|f\|_{H^{-1}(\omega_z)}$ by $\|f - f_z\|_{H^{-1}(\omega_z)}$ for any constant f_z when $z \in \mathcal{N}_0(\mathcal{T})$. This is due to the fact that

$$\int\limits_{\omega_z} (v - \alpha_z(v))\phi_z = 0,$$

which allows us to remove f_z from f in (3.8) for all $z \in \mathcal{N}_0(\mathcal{T})$. With such a modification, instead of (3.16) we would obtain

$$j(z) \lesssim |\!|\!| u - U |\!|\!|_{\omega_z} + \|f - f_z\|_{H^{-1}(\omega_z)},$$

and in place of (3.19),

$$d(z) \lesssim |\!|\!| u - U |\!|\!|_{\omega_z} + \|f - f_z\|_{H^{-1}(\omega_z)}.$$

This in turn leads to a modified form of the lower bound in (3.17), namely

$$\mathcal{E}(U, f, \mathcal{T})^2 \lesssim |\!|\!| u - U |\!|\!|^2 + \sum_{z\in\mathcal{N}(\mathcal{T})} \|f - f_z\|^2_{H^{-1}(\omega_z)}. \tag{3.20}$$

The last term is an H^{-1}-version of the so-called *data oscillation* term. In fact, if $f \in L^2(\Omega)$, then the same argument employed in (3.18) yields the more familiar quantity [1],[5],[9],[10],[12]

$$\sum_{z\in\mathcal{N}(\mathcal{T})} \|f - \alpha_z(f)\|^2_{H^{-1}(\omega_z)} \lesssim \sum_{z\in\mathcal{N}(\mathcal{T})} h_z^2 \|f - \alpha_z(f)\|^2_{L^2(\omega_z)} = \operatorname{osc}(f, \mathcal{T})^2. \tag{3.21}$$

The decay of the right-hand side of (3.21) is strictly faster than that of (3.18) when $f = -\operatorname{div}(A\nabla u)$ is more regular than dictated by the regularity of u. For instance, reentrant corners of Ω and discontinuities of A may create singularities in u not reflected in f. In fact, u can be arbitrarily close to H^1 even for smooth data [7],[9],[12]. This explains the interest in data oscillation and the lower bound (3.20). On the other hand, in our setting of piecewise linear elements, the quasi-optimal convergence rates derived by Cascón et al. [5] hold for $f \in L^2(\Omega)$ without assuming any further regularity, and the meanvalue $\alpha_z(f)$ plays no role in making $d(z)$ smaller; one might as well take $\alpha_z(f) = 0$. In the present context of H^{-1} data, the role of f_z is similar in the sense that replacing $d(z)$ by $\|f - f_z\|_{H^{-1}(\omega_z)}$ does not yield faster asymptotic decay of the data estimator and setting $f_z = 0$ turns out to be sufficient for the optimal convergence of our AFEM algorithm of §5 and comparison of approximation classes of §6.

We therefore stick to the original definitions of *data indicators* and *data estimator* of §3.2

$$d(z) := \|f\|_{H^{-1}(\omega_z)} \ \text{ and } \ \mathcal{D}(f, \mathcal{T}) = \Big(\sum_{z \in \mathcal{N}(\mathcal{T})} d(z)^2 \Big)^{\frac{1}{2}}. \tag{3.22}$$

The following result gives some important properties of the data estimator $\mathcal{D}(f, \mathcal{T})$.

Lemma 3.5 (properties of data estimator)**.** *Let $\mathcal{T}$ be any conforming triangulation. Then*
(i) The application $f \mapsto \mathcal{D}(f, \mathcal{T})$ is a norm on $H^{-1}(\Omega)$.
(ii) For any $f \in H^{-1}(\Omega)$ we have

$$\mathcal{D}(f, \mathcal{T}) \le \sqrt{3} \|f\|_{H^{-1}(\Omega)}. \tag{3.23}$$

(iii) For any $f \in H^{-1}(\Omega)$ and for any conforming refinement $\mathcal{T}_*$ of $\mathcal{T}$, we have

$$\mathcal{D}(f, \mathcal{T}_*) \le \sqrt{3} \mathcal{D}(f, \mathcal{T}). \tag{3.24}$$

Proof: That $\mathcal{D}(\cdot, \mathcal{T})$ is a norm follows easily from the fact that $\mathcal{D}(f, \mathcal{T})$ is the $\ell_2(\mathcal{N}(\mathcal{T}))$ norm of the sequence $(\|f\|_{H^{-1}(\omega_z)})_{z \in \mathcal{N}(\mathcal{T})}$, and $f = 0$ in ω_z for all $z \in \mathcal{N}(\mathcal{T})$ implies $f = 0$ in Ω. To prove (3.23), note that for each $z \in \mathcal{N}(\mathcal{T})$, there exists $\lambda_z \in H_0^1(\omega_z)$ such that

$$\langle f, \lambda_z \rangle = \|f\|^2_{H^{-1}(\omega_z)}, \quad \text{and} \quad \|\lambda_z\|_{H_0^1(\omega_z)} = \|f\|_{H^{-1}(\omega_z)}.$$

If we define $\lambda := \sum_{z \in \mathcal{N}(\mathcal{T})} \lambda_z$, then using the fact that almost every point of Ω is interior to a triangle and thus contained in exactly 3 sets ω_z, we have

$$\|\lambda\|^2_{H_0^1(\Omega)} \le 3 \sum_{z \in \mathcal{N}(\mathcal{T})} \|\lambda_z\|^2_{H_0^1(\omega_z)} = 3 \sum_{z \in \mathcal{N}(\mathcal{T})} \|f\|^2_{H^{-1}(\omega_z)}. \tag{3.25}$$

On the other hand

$$\sum_{z \in \mathcal{N}(\mathcal{T})} \|f\|^2_{H^{-1}(\omega_z)} = \sum_{z \in \mathcal{N}(\mathcal{T})} \langle f, \lambda_z \rangle = \langle f, \lambda \rangle \le \|f\|_{H^{-1}(\Omega)} \|\lambda\|_{H_0^1(\Omega)}.$$

Replacing $\|\lambda\|_{H_0^1(\Omega)}$ by the bound in (3.25) gives (3.23). To prove (3.24), we observe that for each z_* in $\mathcal{N}(\mathcal{T}_*)$, there exists $z \in \mathcal{N}(\mathcal{T})$ such that $\omega_{z_*} \subset \omega_z$. Thus

$$\sum_{z_* \in \mathcal{N}(\mathcal{T}_*)} \|f\|^2_{H^{-1}(\omega_{z_*})} \le \sum_{z \in \mathcal{N}(\mathcal{T})} \sum_{\omega_{z_*} \subset \omega_z} \|f\|^2_{H^{-1}(\omega_{z_*})}. \tag{3.26}$$

Replacing Ω by ω_z, the previous derivation gives $\sum_{\omega_{z_*} \subset \omega_z} \|f\|^2_{H^{-1}(\omega_{z_*})} \le 3 \|f\|^2_{H^{-1}(\omega_z)}$. Inserting this into (3.26) we arrive at (3.24). □

4 An AFEM: Algorithm and Contraction Property

We now propose an AFEM that, starting from the initial mesh $\mathcal{T}_0$, iteratively constructs refined meshes and the corresponding Galerkin solutions. If $k \geq 0$ stands for the adaptive counter, we use a subscript k to indicate the corresponding mesh $\mathcal{T}_k$, the nodes $\mathcal{N}_k = \mathcal{N}(\mathcal{T}_k)$, the Galerkin solution U_k, the local indicators $j_k(z) := j(U_k, z),\ d_k(z) := d(f, z),\ e_k(z) := e(U, f, z)$, for $z \in \mathcal{N}_k$, and the global estimators $\mathcal{J}_k := \mathcal{J}(U_k, \mathcal{T}_k),\ \mathcal{D}_k := \mathcal{D}(f, \mathcal{T}_k),\ \mathcal{E}_k := \mathcal{E}(U_k, f, \mathcal{T}_k)$.

4.1 The Algorithm

Our adaptive algorithm takes the following form. We choose a parameter $0 < \theta < 1$ and an initial conforming mesh $\mathcal{T}_0$ satisfying the initial labeling of §2. Set $k = 0$ and iterate

```
$U_k = \mathsf{SOLVE}(\mathcal{T}_k)$;
$\{j_k(z), d_k(z)\}_{z\in\mathcal{N}_k} = \mathsf{ESTIMATE}(\mathcal{T}_k, U_k, f)$;
$\mathcal{M}_k = \mathsf{MARK}(\{e_k(z)\}_{z\in\mathcal{N}_k}, \mathcal{T}_k, \theta)$;
if $\mathcal{D}_k > \sigma_k := \frac{\theta}{3}\mathcal{E}_k$
    $\mathcal{T}_k^+ = \mathsf{ADAPTDATA}(\mathcal{T}_k, f, \frac{\sigma_k}{2\sqrt{3}})$;
else
    $\mathcal{T}_k^+ = \mathcal{T}_k$;
$\mathcal{T}_{k+1} = \mathsf{REFINE}(\mathcal{T}_k, \mathcal{M}_k) \oplus \mathcal{T}_k^+$;
$k \leftarrow k + 1$
```

This algorithm is based on the jump and data estimators $j_k(z)$ and $d_k(z)$; a more computational version will be discussed in §7. We now describe each subroutine appearing above in sufficient detail.

Procedure SOLVE. This module finds the Galerkin solution U_k of (3.3) *exactly*. We therefore assume that there is a way to evaluate (3.3), namely $\langle f, \phi_z\rangle$ for all $z \in \mathcal{N}_k$.

Procedure ESTIMATE. This module determines the jump indicator $j_k(z)$ and data indicator $d_k(z)$ for each $z \in \mathcal{N}_k$. We thus assume that we have access to $d_z = \|f\|_{H^{-1}(\omega_z)}$ for all $z \in \mathcal{N}_k$, even though these values are not immediately available. Later in §7 we replace d_z by surrogate quantities $\tilde{d}_z$, which are computable.

Procedure MARK. This module marks nodes $z \in \mathcal{N}_k$ with *largest* local indicators $e_k(z)$ according to the following bulk chasing strategy (Dörfler marking [6]): given a parameter $0 < \theta < 1$ determine a *smallest* marked set $\mathcal{M}_k \subset \mathcal{N}_k$ such that

$$\mathcal{E}(f, U_k, \mathcal{M}_k; \mathcal{T}_k) \geq \theta\mathcal{E}_k. \tag{4.1}$$

Note that the marking is driven by the total estimator $\mathcal{E}_k$ and not by any of its constituents $\mathcal{J}_k$ and $\mathcal{D}_k$. This is consistent with the fact that separate marking might not, in general, lead to optimal cardinality [5]. Bulk chasing ensures a *reduction property* of the estimator $\mathcal{E}_k$ provided

$f \in L^2(\Omega)$, which according with (3.18) gives the data indicator $h_z \|f\|_{L^2(\omega_z)}$ [5],[12]. This is not possible for data $f \in H^{-1}(\Omega)$ without further assumptions. Therefore we need to enforce a data indicator reduction separately, which is done in $\mathsf{ADAPTDATA}$.

Procedure $\mathsf{ADAPTDATA}$. This module is similar to Stevenson's inner loop to deal with H^{-1} data [13],[14]. It refines stars with relatively large data indicators $d_k(z)$ until the overall contribution in the conforming refinement $\mathcal{T}_k^+$ of $\mathcal{T}_k$ is smaller than a prescribed tolerance: $\mathcal{T}_k^+ = \mathsf{ADAPTDATA}(\mathcal{T}_k, f, \tau)$ should satisfy

$$\mathcal{D}_k^+ := \mathcal{D}(f, \mathcal{T}_k^+) \leq \tau.$$

Note that $\mathcal{D}_k^+ \leq \frac{\theta}{6\sqrt{3}} \mathcal{E}_k$. A contraction property of our AFEM is proved in §4.2 and is valid regardless of the complexity of the triangulation $\mathcal{T}_k^+$ produced by $\mathsf{ADAPTDATA}$. The latter, however, is crucial to examine the cardinality of AFEM. We will assume in §5 that this complexity is compatible with the rate of convergence permitted by the solution u, and use this assumption to derive optimal convergence rates of the AFEM. We will then show in §5 that this assumption can indeed be met by a certain version of $\mathsf{ADAPTDATA}$.

Procedure REFINE. This module performs one newest vertex bisection on each element $T \subset \omega_z$ for $z \in \mathcal{M}$ where $\mathcal{M} \subset \mathcal{N}(\mathcal{T})$ is a given set of nodes in a conforming triangulation $\mathcal{T}$. In addition it performs one newest vertex bisection on each of the resulting pairs of children (T', T'') so that each edge of T is bisected. This leads to a resulting non-conforming triangulation $\overline{\mathcal{T}}$ and $\mathcal{T}^* = \mathsf{REFINE}(\mathcal{T}, \mathcal{M})$ is its smallest conforming refinement.

Remark 4.1. In the above algorithm, the next mesh $\mathcal{T}_{k+1}$ is obtained from the current mesh $\mathcal{T}_k$ as the overlay of the two refinements of $\mathcal{T}_k$ by REFINE and $\mathsf{ADAPTDATA}$, which are done in parallel. An alternative, that we shall not further explore, would be to perform them sequencially.

4.2 Contraction Property

We begin our analysis of the AFEM with the following property, which is instrumental in the proof of a contraction property. A more general result is proved in [5].

Lemma 4.2 (jump residual reduction)**.** *Given a conforming refinement $\mathcal{T}$ of $\mathcal{T}_0$ and a set of nodes $\mathcal{M} \in \mathcal{N}(\mathcal{T})$, let $\mathcal{T}^* \geq \mathsf{REFINE}(\mathcal{T}, \mathcal{M})$ be any conforming refinement of $\mathsf{REFINE}(\mathcal{T}, \mathcal{M})$. Let $V \in \mathbb{V}_0(\mathcal{T})$ and $V^* \in \mathbb{V}_0(\mathcal{T}^*)$ be arbitrary. There exists a constant $C_3 > 0$ depending only on $\mathcal{T}_0$ and on $a_{\min}$ and $a_{\max}$ such that for all $\delta > 0$,*

$$\mathcal{J}(V^*, \mathcal{T}^*)^2 \leq (1+\delta)\Big(\mathcal{J}(V, \mathcal{T})^2 - \frac{1}{2}\,\mathcal{J}(V, \mathcal{M}; \mathcal{T})^2\Big) + (1+\delta^{-1})\, C_3\, |||V^* - V|||_{\Omega}^2.$$

Proof: As a first step, we compare the local quantities $j(V^*, z^*, \mathcal{T}^*)$ and $j(V, z^*, \mathcal{T}^*)$ for all $z^* \in \mathcal{N}(\mathcal{T}^*)$, where these quantities are defined as in (3.11) for the fine triangulation $\mathcal{T}^*$ and with U replaced by V^* or V. We remark that

$$\begin{aligned} j(V^*, z^*, \mathcal{T}^*) &\leq j(V, z^*, \mathcal{T}^*) + j(V^* - V, z^*, \mathcal{T}^*) \\ &\leq j(V, z^*, \mathcal{T}^*) + \Big(\textstyle\sum_{\sigma^* \in \Gamma(\omega_{z^*})} |\sigma^*|\, \|[A\nabla(V^* - V)]\|_{L^2(\sigma^*)}^2\Big)^{1/2}. \end{aligned}$$

Since A is piecewise constant over $\mathcal{T}^*$, a rescaled inverse inequality yields $\|[A\nabla W]\|_{L^2(\sigma^*)} \lesssim |\sigma^*|^{-1/2}\|\nabla W\|_{L^2(\omega_{z^*})}$ for all $\sigma^* \in \Gamma(\omega_{z^*})$ and all $W \in \mathbb{V}_0(\mathcal{T}^*)$. We therefore obtain

$$j(V^*, z^*, \mathcal{T}^*) \leq j(V, z^*, \mathcal{T}^*) + C\,\|\nabla(V^* - V)\|_{L^2(\omega_{z^*})},$$

where C only depends on $\mathcal{T}_0$ and $a_{\max}$. Squaring this inequality, applying the Young inequality

$$(a+b)^2 \leq (1+\delta)a^2 + (1+\delta^{-1})b^2 \quad \forall a, b \in \mathbb{R},$$

for any $\delta > 0$, and adding over $z \in \mathcal{N}(\mathcal{T}^*)$, we arrive at

$$\mathcal{J}(V^*, \mathcal{T}^*)^2 \leq (1+\delta)\,\mathcal{J}(V, \mathcal{T}^*)^2 + C_3\,(1+\delta^{-1})|\!|\!|V^* - V|\!|\!|_\Omega^2, \tag{4.2}$$

where C_3 only depends on $\mathcal{T}_0$ and on $a_{\min}$ and $a_{\max}$ (here we have used the finite overlapping property of stars ω_z). Since V exhibits jumps solely on interelement boundaries of $\mathcal{T}$, and the latter belong to exactly two stars, we can rewrite $\mathcal{J}(V, \mathcal{T}^*)^2$ as

$$\mathcal{J}(V, \mathcal{T}^*)^2 = 2 \sum_{\sigma \in \Gamma(\mathcal{T})} \Big(\sum_{\sigma^* \in \Gamma(\mathcal{T}^*), \sigma^* \subset \sigma} |\sigma^*|^2 |J_{\sigma^*}|^2 \Big) = 2 \sum_{\sigma \in \Gamma(\mathcal{T})} (\sum_{\sigma^* \in \Gamma(\mathcal{T}^*), \sigma^* \subset \sigma} |\sigma^*|^2) |J_\sigma|^2,$$

where $J_\sigma = J_{\sigma^*}$ is the jump of $A\nabla V$ across σ and thus across any $\sigma^* \subset \sigma$. On the other hand, we have

$$\mathcal{J}(V, \mathcal{T})^2 = 2 \sum_{\sigma \in \Gamma(\mathcal{T})} |\sigma|^2 |J_\sigma|^2.$$

We notice that we have $\sum_{\sigma^* \in \Gamma(\mathcal{T}^*), \sigma^* \subset \sigma} |\sigma^*|^2 \leq |\sigma|^2$ for all $\sigma \in \Gamma(\mathcal{T})$. In addition, if $\sigma \in \Gamma(\omega_z)$ for some $z \in \mathcal{M}$, then by definition of the procedure REFINE it has been split at least into two in the refinement process which leads to $\mathcal{T}^*$ and therefore for such edges we have

$$\sum_{\sigma^* \in \Gamma(\mathcal{T}^*), \sigma^* \subset \sigma} |\sigma^*|^2 \leq \frac{1}{2}|\sigma|^2.$$

From this it follows that

$$\mathcal{J}(V, \mathcal{T}^*)^2 \leq \frac{1}{2}\mathcal{J}(V, \mathcal{M}; \mathcal{T})^2 + \mathcal{J}(V, \mathcal{N}(\mathcal{T}) \setminus \mathcal{M}; \mathcal{T})^2 = \mathcal{J}(V, \mathcal{T})^2 - \frac{1}{2}\mathcal{J}(V, \mathcal{M}; \mathcal{T})^2.$$

Inserting this into (4.2), we conclude the proof. □

We next combine this result together with the fact that the procedure ADAPTDATA reduces the data estimator $\mathcal{D}_k$ strictly, in order to obtain a reduction property of the error estimator in *one* AFEM loop. The following result shows that such a reduction is ensured provided that the Galerkin solutions do not change much after such a loop.

Lemma 4.3 (estimator reduction)**.** *Let $0 < \theta \leq 1$ be the bulk parameter. If C_3 denotes the constant of Lemma 4.2, then we have for all $\delta > 0$*

$$\mathcal{E}_{k+1}^2 \leq (1+\delta)\Big(1 - \frac{\theta^2}{12}\Big)\mathcal{E}_k^2 + (1+\delta^{-1})C_3|\!|\!|U_{k+1} - U_k|\!|\!|_\Omega^2. \tag{4.3}$$

Proof: We distinguish two cases depending on whether AFEM calls $\mathsf{ADAPTDATA}$ or not.

If AFEM does not call $\mathsf{ADAPTDATA}$, then $\mathcal{D}_k \leq \frac{\theta}{3}\mathcal{E}_k$. The bulk property implies

$$\theta^2 \mathcal{E}_k^2 \leq \mathcal{E}(U_k, f, \mathcal{M}_k; \mathcal{T}_k)^2 \leq \mathcal{J}(U_k, \mathcal{M}_k; \mathcal{T}_k)^2 + \mathcal{D}_k^2 \leq \mathcal{J}(U_k, \mathcal{M}_k; \mathcal{T}_k)^2 + \frac{\theta^2}{9}\mathcal{E}_k^2,$$

whence

$$\mathcal{J}(U_k, \mathcal{M}_k; \mathcal{T}_k)^2 \geq \frac{8\theta^2}{9}\mathcal{E}_k^2. \tag{4.4}$$

We now examine the reduction of $\mathcal{E}_k^2$. In view of Lemma 4.2 and (3.24), we infer that

$$\begin{aligned} \mathcal{E}_{k+1}^2 &\leq (1+\delta)\Big(\mathcal{J}_k^2 - \tfrac{1}{2}\mathcal{J}(U_k, \mathcal{M}_k; \mathcal{T}_k)^2\Big) + (1+\delta^{-1})C_3 |\!|\!| U_{k+1} - U_k |\!|\!|_\Omega^2 + \mathcal{D}_{k+1}^2 \\ &\leq (1+\delta)\Big(\mathcal{J}_k^2 - \tfrac{1}{2}\mathcal{J}(U_k, \mathcal{M}_k; \mathcal{T}_k)^2 + 3\mathcal{D}_k^2\Big) + (1+\delta^{-1})C_3 |\!|\!| U_{k+1} - U_k |\!|\!|_\Omega^2. \end{aligned}$$

To estimate the first term on the right-hand side we use (4.4) and $\mathcal{E}_k^2 = \mathcal{J}_k^2 + \mathcal{D}_k^2$

$$\begin{aligned} \mathcal{J}_k^2 - \frac{1}{2}\mathcal{J}(U_k, \mathcal{M}_k; \mathcal{T}_k)^2 + 3\mathcal{D}_k^2 &\leq \mathcal{J}_k^2 - \frac{4\theta^2}{9}\mathcal{E}_k^2 + 3\mathcal{D}_k^2 \\ &= \Big(1 - \frac{2\theta^2}{9}\Big)\mathcal{J}_k^2 + \Big(3 - \frac{2\theta^2}{9}\Big)\mathcal{D}_k^2 - \frac{2\theta^2}{9}\mathcal{E}_k^2, \end{aligned} \tag{4.5}$$

followed by $\mathcal{D}_k^2 \leq \frac{\theta^2}{9}\mathcal{E}_k^2$ to finally derive

$$\mathcal{J}_k^2 - \frac{1}{2}\mathcal{J}(U_k, \mathcal{M}_k; \mathcal{T}_k)^2 + 3\mathcal{D}_k^2 \leq \Big(1 - \frac{2\theta^2}{9}\Big)(\mathcal{J}_k^2 + \mathcal{D}_k^2) = \Big(1 - \frac{2\theta^2}{9}\Big)\mathcal{E}_k^2.$$

Therefore, we obtain

$$\mathcal{E}_{k+1}^2 \leq (1+\delta)\Big(1 - \frac{2\theta^2}{9}\Big)\mathcal{E}_k^2 + (1+\delta^{-1})C_3 |\!|\!| U_{k+1} - U_k |\!|\!|_\Omega^2.$$

which implies the desired reduction property (4.3).

If AFEM calls $\mathsf{ADAPTDATA}$, which means that $\mathcal{D}_k > \frac{\theta}{3}\mathcal{E}_k$, then we are ensured by the definition of $\mathsf{ADAPTDATA}$ that

$$\mathcal{D}(f, \mathcal{T}_k^+) \leq \frac{\theta}{6\sqrt{3}}\mathcal{E}_k,$$

whence by (3.24), we have

$$\mathcal{D}_{k+1}^2 \leq \frac{\theta^2}{36}\mathcal{E}_k^2.$$

Moreover, $\mathcal{E}_k^2 > \mathcal{J}_k^2 + \frac{\theta^2}{9}\mathcal{E}_k^2$ yields

$$\mathcal{J}_k^2 < \Big(1 - \frac{\theta^2}{9}\Big)\mathcal{E}_k^2.$$

We again employ Lemma 4.2 to estimate $\mathcal{E}_{k+1}^2$ from above, now by

$$\mathcal{E}_{k+1}^2 \leq (1+\delta)(\mathcal{J}_k^2 + \mathcal{D}_{k+1}^2) + (1+\delta^{-1})C_3 |\!|\!| U_{k+1} - U_k |\!|\!|_\Omega^2.$$

Using the two above estimates for $\mathcal{D}_{k+1}^2$ and $\mathcal{J}_k^2$, we thus find that

$$\mathcal{E}_{k+1}^2 \leq (1+\delta)\Big(1 - \frac{\theta^2}{12}\Big)\mathcal{E}_k^2 + (1+\delta^{-1})C_3 |\!|\!| U_{k+1} - U_k |\!|\!|_\Omega^2,$$

which is the desired reduction property (4.3). □

We are now in a position to prove the main result of this section. This estimate is instrumental for the discussion of cardinality given in §5. Its proof is similar to that given in [5] for $f \in L^2(\Omega)$.

Theorem 4.4 (contraction property). *There exist constants $\gamma > 0$ and $0 < \alpha < 1$, depending on $\mathcal{T}_0$, $a_{\min}$, $a_{\max}$ and on the bulk parameter θ, such that for all $k \geq 0$,*

$$|\!|\!| u - U_{k+1} |\!|\!|_\Omega^2 + \gamma\, \mathcal{E}_{k+1}^2 \leq \alpha^2 \Big(|\!|\!| u - U_k |\!|\!|_\Omega^2 + \gamma\, \mathcal{E}_k^2 \Big). \tag{4.6}$$

Proof: For convenience, we use the notation

$$e_k = |\!|\!| u - U_k |\!|\!|_\Omega, \quad E_k = |\!|\!| U_{k+1} - U_k |\!|\!|_\Omega.$$

We invoke the Pythagoras equality for the energy norm

$$e_{k+1}^2 = e_k^2 - E_k^2,$$

along with (4.3), to arrive at

$$e_{k+1}^2 + \gamma\, \mathcal{E}_{k+1}^2 \leq e_k^2 + \big(\gamma\,(1+\delta^{-1})\, C_3 - 1\big)\, E_k^2 + (1+\delta)\,\gamma \Big(1 - \frac{\theta^2}{12}\Big) \mathcal{E}_k^2. \tag{4.7}$$

We now choose the parameters. We first select $\delta > 0$ so that

$$(1+\delta)\Big(1 - \frac{\theta^2}{24}\Big) = 1 - \frac{\theta^2}{48} =: \alpha_1$$

and next $\gamma > 0$ so that

$$\gamma\,(1+\delta^{-1})\, C_3 - 1 = 0 \quad \Rightarrow \quad \gamma\,(1+\delta) = \frac{\delta}{C_3}.$$

We invoke the upper a posteriori error bound (3.17), namely

$$e_k^2 \leq C_2\, \mathcal{E}_k^2,$$

to write

$$\Big(1 - \frac{\theta^2}{12}\Big) \mathcal{E}_k^2 \leq \Big(1 - \frac{\theta^2}{24}\Big) \mathcal{E}_k^2 - \frac{\theta^2}{24 C_2} e_k^2.$$

Inserting this back into (4.7), and setting $\alpha_2 := 1 - \frac{\delta\theta^2}{24 C_2 C_3}$, we get

$$e_{k+1}^2 + \gamma\, \mathcal{E}_{k+1}^2 \leq \alpha_2\, e_k^2 + \gamma\, \alpha_1\, \mathcal{E}_k^2.$$

The estimate (4.6) thus follows with $\alpha^2 = \max\{\alpha_1, \alpha_2\}$. □

5 Optimal Convergence Rates

In this section we study the asymptotic decay of the combined quantity

$$E(u, f, \mathcal{T})^2 := |\!|\!| u - U |\!|\!|_\Omega^2 + \mathcal{D}(f, \mathcal{T})^2.$$

5.1 Optimal Decay of $E(u, f, \mathcal{T})$

We first point out the following trivial consequence of (3.17)

$$\mathcal{E}(U, f, \mathcal{T})^2 \approx E(u, f, \mathcal{T})^2 \approx |||u - U|||_\Omega^2 + \gamma \mathcal{E}(U, f, \mathcal{T})^2, \tag{5.1}$$

along with the fact that the last quantity is contracted by AFEM according to Theorem 4.4. We assume that $u \in \mathcal{A}^s$ for $0 < s \leq 1/2$ which means that for all $N \geq \#(\mathcal{T}_0)$ there exist conforming meshes $\mathcal{T}_N \geq \mathcal{T}_0$ with $\#(\mathcal{T}_N) \leq N$ and

$$|||u - U_N|||_\Omega \leq |u|_{\mathcal{A}^s} N^{-s}. \tag{5.2}$$

Our AFEM can only meet this benchmark provided the data estimator exhibits a similar decay rate. We thus make a crucial assumption on the module $\mathsf{ADAPTDATA}$ and f.

Assumption A(s): For any $\tau > 0$, the output $\mathcal{T}^+ = \mathsf{ADAPTDATA}(\mathcal{T}, f, \tau)$ satisfies

$$\#(\mathcal{T}^+) - \#(\mathcal{T}) \leq \Big(\frac{F_s}{\tau}\Big)^{\frac{1}{s}}, \tag{5.3}$$

where F_s is a fixed constant.

We shall show below that we can construct subroutines $\mathsf{ADAPTDATA}$ for which this assumption is satisfied in a variety of settings. But for now, we continue on with our analysis assuming that we have such a subroutine in hand.

An immediate by-product of Assumption A(s) is that for all $N > \#(\mathcal{T}_0)$ there exist conforming meshes $\mathcal{T}_N \geq \mathcal{T}_0$ with $\#(\mathcal{T}_N) \leq N$ and

$$\mathcal{D}(f, \mathcal{T}_N) \lesssim F_s N^{-s}. \tag{5.4}$$

Lemma 5.1 (a priori asymptotic decay of E)**.** *Let assumption $A(s)$ on $\mathsf{ADAPTDATA}$ and f, and $u \in \mathcal{A}^s$ be valid. For all $N > \#(\mathcal{T}_0)$ there exist conforming meshes $\mathcal{T}_N \geq \mathcal{T}_0$ with $\#(\mathcal{T}_N) \leq N$ and*

$$E(u, f, \mathcal{T}_N) \lesssim \big(|u|_{\mathcal{A}^s} + F_s\big)\ N^{-s}. \tag{5.5}$$

Proof: Given $N > \#(\mathcal{T}_0)$ and the meshes $\mathcal{T}_N(u)$ and $\mathcal{T}_N(f)$ for u and f guaranteed by (5.2) and (5.4), respectively, we simply consider the overlay

$$\mathcal{T}(u, f) = \mathcal{T}_N(u) \oplus \mathcal{T}_N(f) \geq \mathcal{T}_0.$$

Invoking (2.4), we have

$$\#(\mathcal{T}(u, f)) \leq \#(\mathcal{T}_N(u)) + \#(\mathcal{T}_N(f)) - \#(\mathcal{T}_0) \leq 2N.$$

Moreover, from (3.24), we obtain

$$E(u, f, \mathcal{T}(u, f)) \leq 2^s\big(|u|_{\mathcal{A}^s} + \sqrt{3}F_s\big)(2N)^{-s}.$$

From this we immediately deduce (5.5). □

5.2 Quasi-Optimal Cardinality of AFEM

In order to prove the optimal convergence of our AFEM, we need to make the following assumption on the bulk parameter in MARK:

$$\textit{the bulk parameter satisfies } 0 < \theta < \theta_* \textit{ with } \theta_* := \sqrt{\tfrac{C_1}{1+C_L^2}}, \tag{5.6}$$

where C_1 and C_L are the constants appearing in (3.17) and (3.15). Since $C_1 < 1+C_G^2 < 1+C_L^2$, we deduce that $\theta_* < 1$ and that the larger the discrepancy between C_1 and C_L^2 (or between C_1 and C_2 in (3.17)) the smaller the value of the threshold θ_*.

We next prove that if a conforming refinement $\mathcal{T}^*$ of $\mathcal{T}$ reduces $E(u, f, \mathcal{T})$ substantially, then the refined set must capture the bulk of the error estimator. This property is what connects the AFEM with the best possible decay of $E(u, f, \mathcal{T})$ described in §5.1. This crucial insight is due to Stevenson [14] for the Laplacian and piecewise constant forcing f. The present formulation is closer to that of Cascón et al [5] for $f \in L^2(\Omega)$, but now the refined set $\mathcal{R}$ is indexed by nodes z instead of by triangles T.

Lemma 5.2 (bulk property)**.** *Let* $\xi := \sqrt{1 - \frac{\theta^2}{\theta_*^2}} > 0$. *Let* $\mathcal{T}^* \geq \mathcal{T}$ *be a refinement of* $\mathcal{T}$ *and let* $\mathcal{R} \subset \mathcal{N}(\mathcal{T})$ *designate the set of all nodes* $z \in \mathcal{N}(\mathcal{T})$ *such that* z *is the vertex of a triangle* $T \in \mathcal{T} \setminus \mathcal{T}^*$ *which was refined in the process of constructing* $\mathcal{T}^*$*. If*

$$E(u, f, \mathcal{T}^*) \leq \xi E(u, f, \mathcal{T}), \tag{5.7}$$

then the set $\mathcal{R}$ *satisfies the bulk property*

$$\mathcal{E}(U, f, \mathcal{R}; \mathcal{T}) \geq \theta \mathcal{E}(U, f, \mathcal{T}). \tag{5.8}$$

Proof: We use the lower bound in (3.17) and (5.7) to write

$$\begin{aligned}(1-\xi^2) C_1 \mathcal{E}(U, f, \mathcal{T})^2 &\leq E(u, f, \mathcal{T})^2 - E(u, f, \mathcal{T}^*)^2 \\ &= \| u - U \|_\Omega^2 - \| u - U_* \|_\Omega^2 + \mathcal{D}(f, \mathcal{T})^2 - \mathcal{D}(f, \mathcal{T}^*)^2.\end{aligned}$$

We observe that the Pythagoras equality in conjunction with (3.15) gives

$$\| u - U \|_\Omega^2 - \| u - U_* \|_\Omega^2 = \| U - U_* \|_\Omega^2 \leq C_L^2 \mathcal{E}(U, f, \mathcal{R}; \mathcal{T})^2,$$

Since $(1-\xi^2) C_1 = (1 + C_L^2)\theta^2$, we find that

$$(1 + C_L^2)\theta^2 \mathcal{E}(U, f, \mathcal{T})^2 \leq C_L^2 \mathcal{E}(U, f, \mathcal{R}; \mathcal{T})^2 + \mathcal{D}(f, \mathcal{T})^2 - \mathcal{D}(f, \mathcal{T}^*)^2.$$

On the other hand, we observe that

$$\mathcal{D}(f, \mathcal{T})^2 = \mathcal{D}(f, \mathcal{R}; \mathcal{T})^2 + \mathcal{D}(f, \mathcal{N}(\mathcal{T}) \setminus \mathcal{R}; \mathcal{T})^2 \leq \mathcal{E}(U, f, \mathcal{R}; \mathcal{T})^2 + \mathcal{D}(f, \mathcal{T}^*)^2,$$

and therefore

$$(1 + C_L^2)\theta^2 \mathcal{E}(U, f, \mathcal{T})^2 \leq (1 + C_L^2) \mathcal{E}(U, f, \mathcal{R}; \mathcal{T})^2,$$

which is the asserted estimate. □

We next show that the AFEM yields an estimate similar to (5.5) but with N replaced by $\#(\mathcal{M}_k)$.

Lemma 5.3 (cardinality of $\mathcal{M}_k$)**.** *Let $u \in \mathcal{A}^s$ and f and ADAPTDATA be such that Assumption A(s) is satisfied. Let the bulk parameter θ satisfy* (5.6)*. If $(\mathcal{T}_k, \mathcal{M}_k)$ are the k-th mesh and marked set generated by the AFEM from $\mathcal{T}_0$, then*

$$\#(\mathcal{M}_k) \lesssim \big(|u|_{\mathcal{A}^s} + F_s\big)^{1/s} E(u, f, \mathcal{T}_k)^{-1/s}. \tag{5.9}$$

Proof: Let $\varepsilon = \frac{\xi}{\sqrt{3}} E(u, f, \mathcal{T}_k)$ where $\xi := \sqrt{1 - \frac{\theta^2}{\theta_*^2}}$ was introduced in Lemma 5.2. In view of Lemma 5.1, there exists a conforming mesh $\mathcal{T}_\varepsilon \geq \mathcal{T}_0$ such that

$$E(u, f, \mathcal{T}_\varepsilon) \leq \varepsilon, \quad \text{and} \quad \#(\mathcal{T}_\varepsilon) \lesssim \big(|u|_{\mathcal{A}^s} + F_s\big)^{\frac{1}{s}} \varepsilon^{-\frac{1}{s}}.$$

We need to relate $\mathcal{T}_\varepsilon$ and $\mathcal{T}_k$. To do this, we introduce the overlay $\mathcal{T}^* = \mathcal{T}_\varepsilon \oplus \mathcal{T}_k$, which, according to (2.4), satisfies

$$\#(\mathcal{T}^*) \leq \#(\mathcal{T}_\varepsilon) + \#(\mathcal{T}_k) - \#(\mathcal{T}_0).$$

Since $\mathcal{T}^* \geq \mathcal{T}_\varepsilon$, we deduce that

$$E(u, f, \mathcal{T}^*)^2 = |\!|\!| u - U_* |\!|\!|_\Omega^2 + \mathcal{D}(f, \mathcal{T}^*)^2 \leq |\!|\!| u - U_\varepsilon |\!|\!|_\Omega^2 + 3\mathcal{D}(f, \mathcal{T}_\varepsilon)^2 \leq 3\varepsilon^2 = \xi^2 E(u, f, \mathcal{T}_k)^2.$$

Let $\mathcal{R} \subset \mathcal{N}(\mathcal{T}_k)$ be the set of all nodes $z \in \mathcal{N}(\mathcal{T}_k)$ such that z is the vertex of a triangle $T \in \mathcal{T}_k \setminus \mathcal{T}^*$ which was refined in the process of constructing $\mathcal{T}^*$. From Lemma 5.2, we conclude that this set satisfies the bulk property

$$\mathcal{E}(U_k, f, \mathcal{R}; \mathcal{T}_k) \geq \theta \mathcal{E}_k.$$

Since the set $\mathcal{M}_k$ is a minimal subset of $\mathcal{N}(\mathcal{T}_k)$ that satisfies the same property, we infer that

$$\begin{aligned}\#(\mathcal{M}_k) &\leq \#(\mathcal{R}) \leq \#(\mathcal{N}^*) - \#(\mathcal{N}_k) \lesssim \#(\mathcal{T}^*) - \#(\mathcal{T}_k)\\ &\leq \#(\mathcal{T}_\varepsilon) - \#(\mathcal{T}_0) \lesssim \big(|u|_{\mathcal{A}^s} + F_s\big) \varepsilon^{-\frac{1}{s}} \lesssim \big(|u|_{\mathcal{A}^s} + F_s\big)^{\frac{1}{s}} E(u, f, \mathcal{T}_k)^{-\frac{1}{s}},\end{aligned}$$

as asserted. □

We are now ready to prove the main result of this section, namely that the AFEM achieves a performance comparable with the benchmark (5.5).

Theorem 5.4 (quasi-optimal cardinality of AFEM)**.** *Let $u \in \mathcal{A}^s$ and f and ADAPTDATA be such that Assumption A(s) is satisfied. Let the bulk parameter satisfy* (5.6)*. If $(\mathcal{T}_k, \mathbb{V}_k, U_k)_{k \geq 0}$ is a sequence of conforming meshes, nested spaces $\mathbb{V}_k$ and Galerkin solutions $U_k \in \mathbb{V}_k$ generated by our AFEM, then*

$$E(u, f, \mathcal{T}_k) \lesssim \big(|u|_{\mathcal{A}^s} + F_s\big) \#(\mathcal{T}_k)^{-s}. \tag{5.10}$$

Proof: At each iteration j of our AFEM, there are two instances where elements are added. The first one is in the subroutine MARK. Lemma 5.3 shows that the set $\mathcal{M}_j$ of marked nodes satisfies

$$\#(\mathcal{M}_j) \lesssim \big(|u|_{\mathcal{A}^s} + F_s\big)^{\frac{1}{s}} E(u, f, \mathcal{T}_j)^{-\frac{1}{s}}.$$

The second instance is due to data adaptation within ADAPTDATA. For the j-th iterate, we may apply (5.3) with $\tau = \frac{\theta}{6\sqrt{3}}\mathcal{E}(U_j, f, \mathcal{T}_j)$ and obtain

$$\#(\mathcal{T}_j^+) - \#(\mathcal{T}_j) \le F_s^{\frac{1}{s}} \Big(\frac{\theta}{6\sqrt{3}} \mathcal{E}(U_j, f, \mathcal{T}_j) \Big)^{-\frac{1}{s}} \approx F_s^{\frac{1}{s}} E(u, f, \mathcal{T}_j)^{-\frac{1}{s}}$$

because of the equivalence (5.1). In light of Lemma 2.1, we deduce

$$\#(\mathcal{T}_k) - \#(\mathcal{T}_0) \lesssim \sum_{j=0}^{k-1} \big(\#(\mathcal{M}_j) + \#(\mathcal{T}_j^+) - \#(\mathcal{T}_j) \big) \lesssim \big(|u|_{\mathcal{A}^s} + F_s \big)^{\frac{1}{s}} \sum_{j=0}^{k-1} E(u, f, \mathcal{T}_j)^{-\frac{1}{s}}. \quad (5.11)$$

We now recall the contraction property (4.6) for $0 \le j < k$

$$|\!|\!| u - U_k |\!|\!|_\Omega^2 + \gamma \mathcal{E}_k^2 \le \alpha^{2(k-j)} \big(|\!|\!| u - U_j |\!|\!|_\Omega^2 + \gamma \mathcal{E}_j^2 \big),$$

which can be written equivalently as follows, upon employing (5.1),

$$E(u, f, \mathcal{T}_k)^{\frac{1}{s}} \lesssim \alpha^{\frac{k-j}{s}} E(u, f, \mathcal{T}_j)^{\frac{1}{s}}.$$

Inserting this into (5.11), we obtain

$$\#(\mathcal{T}_k) - \#(\mathcal{T}_0) \lesssim \big(|u|_{\mathcal{A}^s} + F_s \big)^{\frac{1}{s}} E(u, f, \mathcal{T}_k)^{-\frac{1}{s}} \sum_{j=0}^{k-1} \alpha^{\frac{k-j}{s}} \lesssim \big(|u|_{\mathcal{A}^s} + F_s \big)^{\frac{1}{s}} E(u, f, \mathcal{T}_k)^{-\frac{1}{s}}$$

because $\alpha < 1$ and so the geometric series $\sum_{j=0}^{\infty} \alpha^{\frac{j}{s}}$ converges. This gives (5.10) provided $\#(\mathcal{T}_k) \ge 2\#(\mathcal{T}_0)$. If instead $\#(\mathcal{T}_0) < \#(\mathcal{T}_k) < 2\#(\mathcal{T}_0)$, then $\#(\mathcal{T}_k) - \#(\mathcal{T}_0) \ge 1 \ge \frac{\#(\mathcal{T}_k)}{2\#(\mathcal{T}_0)}$, which also yields (5.10). This concludes the proof. □

6 Approximation Classes for the Data

The results of the previous sections show that if $u \in \mathcal{A}^s$, the optimal convergence rate N^{-s} is met by our AFEM algorithm, provided that for the given data f, the procedure ADAPTDATA satisfies the property (5.3) that defines Assumption A(s). The goal of this section is to discuss under which circumstances such a property may hold. In view of (5.4), it is thus natural to introduce approximation classes $\mathcal{B}^s$ for the data, whose definition mimics that of the classes $\mathcal{A}^s$ for the solution u. Accordingly, we define for all $N \ge \#(\mathcal{T}_0)$

$$\delta_N(f) := \min_{\mathcal{T} \in \mathfrak{T}_N} \mathcal{D}(f, \mathcal{T}),$$

and denote by $\mathcal{B}^s$ the set of all $f \in H^{-1}(\Omega)$ such that

$$|f|_{\mathcal{B}^s} := \sup_{N \ge \#(\mathcal{T}_0)} N^s \delta_N(f) < +\infty.$$

This is a (quasi) semi-norm, and a quasi-norm for the space $\mathcal{B}^s$ can be defined by

$$\|f\|_{\mathcal{B}^s} := \|f\|_{H^{-1}(\Omega)} + |f|_{\mathcal{B}^s}.$$

The main result of this section shows that the condition $u \in \mathcal{A}^s$ implies that $f \in \mathcal{B}^s$, $0 < s < 1/2$.

Theorem 6.1 (relation between $\mathcal{A}_s$ and $\mathcal{B}_s$)**.** *If $u \in \mathcal{A}^s$ for some $0 < s < 1/2$, then $f \in \mathcal{B}^s$ and*

$$\|f\|_{\mathcal{B}^s} \leq C_4 \|u\|_{\mathcal{A}^s}, \tag{6.1}$$

where C_4 depends only on s, $a_{\min}, a_{\max}$ and on the initial triangulation $\mathcal{T}_0$.

Since the class $\mathcal{A}^s$ is defined via approximation by piecewise linear functions in $\mathbb{V}_0(\mathcal{T})$, a natural approach to proving this theorem is to start with (5.2) and then approximate $f = -\mathrm{div}(A\nabla u)$ by $S = -\mathrm{div}(A\nabla V)$ for some $V \in \mathbb{V}_0(\mathcal{T})$. Each such S is a Dirac distribution supported on the interior edges $\Gamma(\mathcal{T})$ of $\mathcal{T}$. With this in mind, the heuristic argument to prove Theorem 6.1 is as follows: we let $N \approx 2^n \#(\mathcal{T}_0)$ for $n \geq 1$ integer, and let $\mathcal{T}_j \geq \mathcal{T}_0$ and $V_j \in \mathbb{V}(\mathcal{T}_j)$ satisfy

$$\#(\mathcal{T}_j) \leq 2^j \#(\mathcal{T}_0), \qquad \|u - V_j\|_{H_0^1(\Omega)} \lesssim |u|_{\mathcal{A}^s} \big(2^j \#(\mathcal{T}_0)\big)^{-s}.$$

We realize that $u = u - V_n + \sum_{j=1}^n V_j - V_{j-1}$ with $V_0 = 0$ induces the representation of f

$$f = f - S_n + \sum_{j=1}^n S_j - S_{j-1},$$

with $S_j = -\mathrm{div}(A\nabla V_j)$. We next introduce a mesh $\mathcal{T}^* \geq \mathcal{T}_j$, which is a common refinement of all $\mathcal{T}_j$ for $0 \leq j \leq n$. We achieve this by further refining each interior edge of $\mathcal{T}_j$ at least m_j times and so creating an admissible mesh $\mathcal{R}_{m_j}(\mathcal{T}_j) \geq \mathcal{T}_j$ with a level of resolution comparable with $\mathcal{T}_n$. We finally set $\mathcal{T}^* := \oplus_{j=1}^n \mathcal{R}_{m_j}(\mathcal{T}_j)$ and evaluate each term in

$$\mathcal{D}(f, \mathcal{T}^*) \leq \mathcal{D}(f - S_n, \mathcal{T}^*) + \sum_{j=1}^n \mathcal{D}(S_j - S_{j-1}, \mathcal{T}^*).$$

To accomplish this program, we first examine in §6.1 the effect of m edge refinements and completion to create $\mathcal{R}_m(\mathcal{T})$ from $\mathcal{T}$. Then, we characterize $\mathcal{D}(S, \mathcal{T})$ and derive crucial properties of $\mathcal{D}(S, \mathcal{T})$ in §6.2. We finally prove Theorem 6.1 in §6.3. Inspection of this proof reveals that in the limit case $s = \frac{1}{2}$, we have a logarithmic loss in the sense that $u \in \mathcal{A}^s$ only implies that $\delta_N(f) \leq CN^{-s} \log N$. We show in §6.4 that this loss cannot be avoided.

Theorem 6.1 suggests that one should be able to design the ADAPTDATA procedure so that the property (5.3) holds for any $f \in H^{-1}$ whenever $u \in \mathcal{A}^s$. We discuss in §6.5 how this may be achieved by using a specific refinement procedure. However, this procedure requires the evaluation of the data indicators $d(z)$ which are not easily computable in practice since they involve the $H^{-1}(\omega_z)$ norms. We present in §7 simpler greedy strategies based on more computable surrogate quantities $\widetilde{d}(z)$, under some additional assumptions on f.

6.1 Edge Refinement

In this subsection, we introduce certain conforming refinements $\mathcal{R}_m(\mathcal{T})$ of a given admissible triangulation $\mathcal{T}$ obtained by successively bisecting the *inner* edges of $\mathcal{T}$ at least m times. Given an edge σ of length $|\sigma|$, its bisection results in two edges of length $\frac{|\sigma|}{2}$. If each of these two edges is again bisected we obtain four edges of length $\frac{|\sigma|}{4}$. Notice that an application of newest vertex bisection to a triangle results in bisecting one of its edges. If a refinement is applied to each of the resulting children, then the remaining two edges will also be bisected. So we can always refine an edge σ in the original triangulation $\mathcal{T}$ into arbitrarily fine edges through a sequence of newest vertex bisections. We are interested in doing this with a minimal number of refinements, and we only want to refine the inner edges of $\mathcal{T}$.

Given $\mathcal{T}$, we let $\mathcal{R}_m(\mathcal{T})$ be a triangulation with the following properties

(i) $\mathcal{R}_m(\mathcal{T})$ is a refinement of $\mathcal{T}$ which is admissible;

(ii) Any edge of $\mathcal{R}_m(\mathcal{T})$ contained in an inner edge σ of $\mathcal{T}$ has length at most $2^{-m}|\sigma|$.

(iii) The cardinality of $\mathcal{R}_m(\mathcal{T})$ is minimal among all triangulations with these two properties.

Lemma 6.2 (the cardinality of $\mathcal{R}_m(\mathcal{T})$)**.** *Given any admissible $\mathcal{T}$, we have*

$$\#(\mathcal{R}_m(\mathcal{T})) \leq 7C_0 2^m \#(\mathcal{T}), \quad m > 0, \tag{6.2}$$

with C_0 the constant in (2.1).

Proof: Let us first observe that given any triangle T that occurs in newest vertex bisection and given one of its edges σ we can always perform a bisection of σ with at most 2 newest vertex bisections. Namely, the first subdivision of $\mathcal{T}$ will bisect one of the edges of T. If this is not σ then σ is an edge of one of the children of T and the subdivision of that child will bisect σ. We shall call this sequence (of one or two bisections) a *bisection of* σ. If we perform three subdivisions then each edge of T is bisected. We call the latter a *full bisection* of T.

Now given our original triangulation $\mathcal{T} =: \widetilde{\mathcal{T}}_0$, we denote by $\Gamma_0 = \Gamma(\widetilde{\mathcal{T}}_0)$ its collection of inner edges. We perform a full bisection on each of the triangles $T \in \widetilde{\mathcal{T}}_0$ and denote the resulting nonconforming triangulation by $\widetilde{\mathcal{T}}_1$. Thus each edge of $\mathcal{T}$ has been bisected once in $\widetilde{\mathcal{T}}_1$. We denote by Γ_1 the inner edges of $\widetilde{\mathcal{T}}_1$ which are contained in an edge of Γ_0.

Given that $\widetilde{\mathcal{T}}_m$, $m \geq 1$, has already been constructed and Γ_m are the edges of $\widetilde{\mathcal{T}}_m$ that are contained in edges from Γ_0, we perform a bisection of each $\sigma \in \Gamma_m$. We denote by $\widetilde{\mathcal{T}}_{m+1}$ the resulting non-conforming triangulation after each edge $\sigma \in \Gamma_m$ has been bisected. Note that $\widetilde{\mathcal{T}}_m$ satisfies the property (ii) required for $\mathcal{R}_m(\mathcal{T})$.

It is easy to bound the cardinality of each $\widetilde{\mathcal{T}}_m$. Let $N_0 = \#(\Gamma_0)$ be the number of inner edges in $\widetilde{\mathcal{T}}_0 = \mathcal{T}$. Since each interior edge requires bisecting two triangles, we have

$$\#(\widetilde{\mathcal{T}}_1) \leq 2N_0 + \#(\widetilde{\mathcal{T}}_0).$$

In the general case of $m \geq 1$, we have $\#(\Gamma_m) = 2^m N_0$ and

$$\#(\widetilde{\mathcal{T}}_m) \leq \#(\widetilde{\mathcal{T}}_{m-1}) + 4\#(\Gamma_{m-1}) = \#(\widetilde{\mathcal{T}}_{m-1}) + 4 \cdot 2^{m-1} N_0.$$

The factor 4 arises because to bisect σ we need to bisect the two triangles containing σ either once or twice. It follows by induction that

$$\#(\widetilde{\mathcal{T}}_m) \leq \#(\widetilde{\mathcal{T}}_0) + \{2 + 4(2 + \cdots + 2^{m-1})\}N_0 \leq \#(\widetilde{\mathcal{T}}_0) + 4 \cdot 2^m N_0.$$

Observing that $N_0 \leq \frac{3}{2}\#(\widetilde{\mathcal{T}}_0)$, we thus find that

$$\#(\widetilde{\mathcal{T}}_m) \leq (6 \cdot 2^m + 1)\#(\mathcal{T}) \leq 7 \cdot 2^m \#(\mathcal{T}).$$

Given $\widetilde{\mathcal{T}}_m$ as described above, we define $\overline{\mathcal{T}}_m$ as the minimal completion of $\widetilde{\mathcal{T}}_m$ into a conforming triangulation. Then, $\overline{\mathcal{T}}_m$ satisfies properties (i) and (ii) in the definition of $\mathcal{R}_m(\mathcal{T})$. From (2.3), we derive that $\#(\overline{\mathcal{T}}_m) \leq C2^m\#(\mathcal{T})$, with $C = 7C_0$, which concludes the proof since $\#(\mathcal{R}_m(\mathcal{T})) \leq \#(\overline{\mathcal{T}}_m)$. □

6.2 The Data Estimator for Dirac Distributions

We next derive properties of the data estimator $\mathcal{D}(S, \mathcal{T})$ provided S is a linear combination of the Dirac distributions of the inner edges $\sigma \in \Gamma(\mathcal{T})$ of $\mathcal{T}$, i.e. S is of the form

$$S := \sum_{\sigma\in\Gamma(\mathcal{T})} c_\sigma \delta_\sigma. \tag{6.3}$$

Note that $S = -\mathrm{div}(A\nabla V)$ for some $V \in \mathbb{V}_0(\mathcal{T})$ is of this form with $c_\sigma := J_\sigma(V)$ the jump of the normal component of $A\nabla V$ across σ.

Lemma 6.3 (a characterization of $\mathcal{D}(S, \mathcal{T})$)**.** *For all $\mathcal{T}$ and all S of the form* (6.3)*, we have*

$$\sum_{\sigma\in\Gamma(\mathcal{T})} |c_\sigma|^2|\sigma|^2 \lesssim \mathcal{D}(S, \mathcal{T})^2 \lesssim \sum_{\sigma\in\Gamma(\mathcal{T})} |c_\sigma|^2|\sigma|^2. \tag{6.4}$$

Proof: We shall prove that for each $z \in \mathcal{N}(\mathcal{T})$, we have

$$\sum_{\sigma\in\Gamma(\omega_z)} |c_\sigma|^2|\sigma|^2 \lesssim \|S\|^2_{H^{-1}(\omega_z)} \lesssim \sum_{\sigma\in\Gamma(\omega_z)} |c_\sigma|^2|\sigma|^2, \tag{6.5}$$

where $\Gamma(\omega_z)$ are the inner edges of $\mathcal{T}$ which admit z as an end point. The lemma then follows by adding these estimates over all $z \in \mathcal{N}(\mathcal{T})$.

To prove the right inequality in (6.5), we let $v \in H^1_0(\omega_z)$ with $\|v\|_{H^1_0(\omega_z)} = 1$. Then

$$\langle S, v\rangle_{H^{-1}(\omega_z),H^1_0(\omega_z)} = \sum_{\sigma\in\Gamma(\omega_z)} c_\sigma \int_\sigma v \leq \sum_{\sigma\in\Gamma(\omega_z)} |c_\sigma|\, |\sigma|^{1/2}\|v\|_{L^2(\sigma)}. \tag{6.6}$$

From the trace theorem (see e.g. [4]), Poincaré's inequality and standard scaling arguments we have for any triangle T of ω_z and any edge σ of T,

$$\|v\|_{L^2(\sigma)} \leq \|v\|_{L^2(\partial T)} \lesssim |\sigma|^{\frac{1}{2}}\|v\|_{H^1_0(\omega_z)} \lesssim |\sigma|^{\frac{1}{2}}. \tag{6.7}$$

Combining (6.6) and (6.7), and taking the supremum over all v so that $\|v\|_{H^1_0(\omega_z)} = 1$, we get

$$\|S\|_{H^{-1}(\omega_z)} \lesssim \sum_{\sigma\in\Gamma(\omega_z)} |c_\sigma|\, |\sigma|,$$

which implies the right inequality in (6.5) since the number of terms in the sum is bounded by a fixed integer that only depends on the initial triangulation $\mathcal{T}_0$.

To prove the left inequality in (6.5), we choose an edge $\sigma \in \Gamma(\omega_z)$ and let φ_σ be the canonical quadratic bubble function with value 1 at the mid-point of the side σ and zero at all other Lagrange quadratic nodes. It is easily checked that

$$\|\varphi_\sigma\|_{H^1_0(\omega_z)} \lesssim 1.$$

Moreover, we have

$$\langle S, \varphi_\sigma \rangle_{H^{-1}(\omega_z), H_0^1(\omega_z)} = c_\sigma \int\limits_\sigma \varphi_\sigma = \frac{2}{3}|\sigma| c_\sigma,$$

whence

$$\|S\|_{H^{-1}(\omega_z)} \gtrsim |c_\sigma||\sigma|.$$

Squaring and adding over $\sigma \in \Gamma(z)$ we derive the left inequality in (6.5). □

A first consequence of Lemma 6.3 is that edge refinement provides a decrease in the data estimator when S is of the form (6.3), as expressed in the following.

Lemma 6.4 (decrease of $\mathcal{D}(S, \mathcal{T})$)**.** *If S is of the form* (6.3) *and $\mathcal{R}_k(\mathcal{T})$ is the edge bisection refinement of Lemma 6.2, then*

$$\mathcal{D}(S, \mathcal{R}_k(\mathcal{T}))^2 \lesssim 2^{-k} \mathcal{D}(S, \mathcal{T})^2. \tag{6.8}$$

Proof: We have

$$\mathcal{D}(S, \mathcal{R}_k(\mathcal{T}))^2 \lesssim \sum_{\sigma \in \Gamma(\mathcal{R}_k(\mathcal{T}))} |c_\sigma|^2 |\sigma|^2 \le 2^{-k} \sum_{\sigma \in \Gamma(\mathcal{T})} |c_\sigma|^2 |\sigma|^2 \lesssim 2^{-k} \mathcal{D}(S, \mathcal{T})^2, \tag{6.9}$$

where we have used both inequalities in Lemma 6.3 and the fact that the edges of $\mathcal{R}_k(\mathcal{T})$ are obtained by k successive refinements from those of $\mathcal{T}$. □

Combining Lemmas 6.4 and 6.2, we give an estimate of $\|S\|_{\mathcal{B}^{1/2}}$ when S is of the form (6.3).

Lemma 6.5 (estimate of $\|S\|_{\mathcal{B}^{1/2}}$)**.** *If S is of the form* (6.3) *over a mesh $\mathcal{T} \ge \mathcal{T}_0$, then*

$$\|S\|_{\mathcal{B}^{1/2}} \lesssim \left(\#(\mathcal{T})\right)^{\frac{1}{2}} \|S\|_{H^{-1}(\Omega)}. \tag{6.10}$$

Proof: Let $N = \#(\mathcal{T})$ and $m \le 7C_0 N$. From Lemma 3.5, we have

$$\delta_m(S) \le \sqrt{3} \|S\|_{H^{-1}(\Omega)}, \tag{6.11}$$

whence

$$m^{1/2} \delta_m(S) \lesssim N^{1/2} \|S\|_{H^{-1}(\Omega)}. \tag{6.12}$$

To bound $\delta_m(f)$ for $m > 7C_0 N$, we use the edge refinements $\mathcal{T}_k := \mathcal{R}_k(\mathcal{T})$ of Lemma 6.2, for which we have $\#(\mathcal{T}_k) \le 7C_0 2^k N$. For $m = \#(\mathcal{T}_k)$, we have from Lemma 6.4 and Lemma 3.5

$$\delta_m(S)^2 \le \mathcal{D}(S, \mathcal{T}_k)^2 \lesssim 2^{-k} \mathcal{D}(S, \mathcal{T})^2 \lesssim 2^{-k} \|S\|^2_{H^{-1}(\Omega)}, \tag{6.13}$$

whence

$$m^{1/2} \delta_m(S) \lesssim N^{1/2} \|S\|_{H^{-1}(\Omega)}. \tag{6.14}$$

For a general m we find k so that $\#(\mathcal{T}_k) < m \le \#(\mathcal{T}_{k+1})$ and use the fact that $\delta_m(S)$ is monotone decreasing with increasing m, to derive (6.14) for all $m \ge N$. We thus obtain (6.10). □

6.3 Proof of Theorem 6.1

We fix a value of n and estimate $\delta_N(f)$ for $N \approx 2^n$, assuming that $u \in \mathcal{A}^s$. According to (1.5), under such an assumption there is a sequence of admissible triangulations $\{\mathcal{T}_j\}_{j=1}^n$ and best approximations $V_j \in \mathbb{V}(\mathcal{T}_j)$ of u in $H_0^1(\Omega)$ such that $\#(\mathcal{T}_j) \le 2^j \#(\mathcal{T}_0)$ and

$$\|u - V_j\|_{H_0^1(\Omega)} \le M 2^{-js}, \quad 1 \le j \le n \tag{6.15}$$

with $M := |u|_{\mathcal{A}^s} \#(\mathcal{T}_0)^{-s}$. For each $\mathcal{T}_j$, we construct the edge refinement $\mathcal{R}_{m_j}(\mathcal{T}_j)$ upon applying Lemma 6.2 with the choice

$$m_j := n - j - \lceil \log_2 j^2 \rceil, \quad 1 \le j \le n,$$

where $\lceil \cdot \rceil$ is the ceiling function. Let $\mathcal{T}^* = \oplus_{j=1}^n \mathcal{R}_{m_j}(\mathcal{T}_j)$ be the overlay of all meshes $\mathcal{R}_{m_j}(\mathcal{T}_j)$; note that $V_j \in \mathbb{V}(\mathcal{T}^*)$ for all $1 \le j \le n$. In view of (2.4) and (6.2), we have

$$\#(\mathcal{T}^*) \le \sum_{j=1}^n \#(\mathcal{R}_{m_j}(\mathcal{T}_j)) \le 7C_0 \#(\mathcal{T}_0) \sum_{j=1}^n 2^{m_j + j} \le 7C_0 \#(\mathcal{T}_0) 2^n \sum_{j=1}^\infty j^{-2},$$

where C_0 is the constant in (2.1). If $C^* = \frac{7\pi^2}{6} C_0$, then we infer that

$$\#(\mathcal{T}^*) \le C^* \#(\mathcal{T}_0) 2^n. \tag{6.16}$$

We now give a bound for $\mathcal{D}(S, \mathcal{T}^*)$. If $V_0 := 0$, we have the decomposition $u = u - V_n + \sum_{j=1}^n V_j - V_{j-1}$, which induces the corresponding decomposition of f

$$f = f - S_n + \sum_{j=1}^n S_j - S_{j-1},$$

with $S_j = -\mathrm{div}(A \nabla V_j)$. Since $f \mapsto \mathcal{D}(f, \mathcal{T}^*)$ is a norm (see Lemma 3.5) we obtain

$$\mathcal{D}(f, \mathcal{T}^*) \le \mathcal{D}(f - S_n, \mathcal{T}^*) + \sum_{j=1}^n \mathcal{D}(S_j - S_{j-1}, \mathcal{T}^*).$$

Using (3.23) and the continuity of the mapping $v \mapsto \mathrm{div}(A \nabla v)$ from $H_0^1(\Omega)$ to $H^{-1}(\Omega)$, we can bound the first term in the right hand side by

$$\mathcal{D}(f - S_n, \mathcal{T}^*) \le \sqrt{3} \|f - S_n\|_{H^{-1}(\Omega)} \le \sqrt{3}\, a_{\max} \|u - V_n\|_{H_0^1(\Omega)}.$$

Similarly, the terms in the sum are each bounded by the following argument

$$\begin{aligned} \mathcal{D}(S_j - S_{j-1}, \mathcal{T}^*) &\lesssim \mathcal{D}(S_j - S_{j-1}, \mathcal{R}_{m_j}(\mathcal{T}_j)) \\ &\lesssim 2^{-m_j/2} \mathcal{D}(S_j - S_{j-1}, \mathcal{T}_j) \\ &\lesssim 2^{-m_j/2} \|S_j - S_{j-1}\|_{H^{-1}(\Omega)} \\ &\lesssim 2^{-m_j/2} a_{\max} \|V_j - V_{j-1}\|_{H_0^1(\Omega)}, \end{aligned}$$

where we have used Lemma 6.4 and both inequalities of Lemma 3.5. It follows that

$$\mathcal{D}(f, \mathcal{T}^*) \le C\Big(\|u - V_n\|_{H_0^1(\Omega)} + \sum_{j=1}^n 2^{-m_j/2} \|V_j - V_{j-1}\|_{H_0^1(\Omega)} \Big), \tag{6.17}$$

where the constant C only depends on $a_{\max}$ and on the initial triangulation $\mathcal{T}_0$.

We next invoke (6.15) to deduce that $\|u - V_n\|_{H_0^1(\Omega)} \le M2^{-ns}$,

$$\|V_j - V_{j-1}\|_{H_0^1(\Omega)} \le \|u - V_j\|_{H_0^1(\Omega)} + \|u - V_{j-1}\|_{H_0^1(\Omega)} \le (1+2^s)M2^{-js}, \quad 1 < j < n,$$

and, since V_1 is the best approximation of u within $\mathbb{V}(\mathcal{T}_1)$,

$$\|V_1\|_{H_0^1(\Omega)} \le \|u\|_{H_0^1(\Omega)}.$$

Replacing this into (6.17) gives

$$\mathcal{D}(f, \mathcal{T}^*) \le CM\Big(2^{-ns} + \sum_{j=2}^{n-1} 2^{-m_j/2-js}\Big) + C2^{-n/2}\|u\|_{H_0^1(\Omega)}.$$

We observe that $s < 1/2$ implies

$$\sum_{j=2}^{n-1} 2^{-m_j/2-js} \le 2^{-ns}\sum_{j=2}^{n-1} j2^{(n-j)(s-1/2)} \le C2^{-ns}.$$

Thus, for $N = C^*\#(\mathcal{T}_0)2^n$ with C^* the constant in (6.16), we thus find that

$$\delta_N(f) \le \mathcal{D}(f, \mathcal{T}^*) \le C\big(M + \|u\|_{H_0^1(\Omega)}\big)2^{-sn} \le C\|u\|_{\mathcal{A}^s}N^{-s}, \tag{6.18}$$

where C only depends on s, $a_{\max}$ and the initial triangulation $\mathcal{T}_0$. Taking into account the monotonicity of $\delta_N(f)$ we complete the proof of the theorem. □

6.4 The Case $s = \frac{1}{2}$: Counterexample

The limit case $s = 1/2$ is not covered by Theorem 6.1. In fact, (6.18) becomes

$$\delta_N(f) \le C\|u\|_{\mathcal{A}^s}N^{-1/2}\log N.$$

We will now give a counterexample that shows that (6.18) cannot possibly hold for $s = 1/2$. To this end, we use the unit square domain $\Omega = [0,1]^2$ and the matrix coefficient $A(x) = I$, therefore the equation is simply $-\Delta u = f$.

We take as an initial triangulation $\mathcal{T}_0$ the four triangles obtained by inserting the two diagonals connecting $(0,0)$ to $(1,1)$ and $(1,0)$ to $(0,1)$ respectively. We shall refer to this triangulation of a square as *the base pattern*. We label the sides of the square Ω by 0 and the other four edges in $\mathcal{T}_0$ by 1, as an initial labeling for newest vertex bisection. Let Q be one of the 4 squares obtained by spliting Ω by the mid-point of its sides. If we refine only the triangles in $\mathcal{T}_0$ which intersect Q, then applying four bisections, we arrive at a triangulation with the base pattern on Q. This can be repeated: if Q is a dyadic square of side length 2^{-n} then applying $4n$ bisections from $\mathcal{T}_0$ we reach a (non-conforming) triangulation which contains the base pattern for Q.

We let ϕ denote the nodal basis function for $\mathcal{T}_0$ corresponding to the center vertex $(1/2, 1/2)$. More generally, for any dyadic square Q of Ω, we denote by ϕ_Q the nodal basis function subordinate to the base pattern for Q associated to the vertex which is the center of this base pattern. We have $\|\phi_Q\|_{L_\infty(\Omega)} = 1$ and $\|\phi_Q\|_{H_0^1(\Omega)} = \|\phi\|_{H_0^1(\Omega)} = 2$.

We now choose a sequence $(Q_j)_{j>0}$ of *disjoint* dyadic squares of Ω with $|Q_j| = 4^{-j}$; we thus select one dyadic square from each scale j. To each Q_j we associate the 4^j dyadic squares $Q_{i,j} \subset Q_j$ with $|Q_{i,j}| = 4^{-2j}$, for $i = 1, \cdots 4^j$. We consider the function

$$u := \sum_{j=1}^{\infty} 2^{-j}\psi_j, \quad \text{where } \psi_j := 2^{-j}\sum_{i=1}^{4^j}\phi_{Q_{i,j}}. \tag{6.19}$$

From disjointness of the supports of the basis functions involved in this definition, we find that $\|\psi_j\|_{H_0^1(\Omega)} = 2$ and $\|u\|_{H_0^1(\Omega)} = \frac{2}{\sqrt{3}}$. The associate data f for the equation $-\Delta u = f$ is thus in $H^{-1}(\Omega)$ with $\|f\|_{H^{-1}(\Omega)} = \frac{2}{\sqrt{3}}$.

It is easy to see that $u \in \mathcal{A}^{1/2}$. Indeed, for each j, we can start from Ω and apply $4j$ refinements and arrive at the base pattern for Q_j. Then, on Q_j we apply an additional $3 \cdot 4^j$ (uniform) refinements and arrive at a triangulation which contains the base pattern for each $Q_{i,j}$ for $i = 1, \cdots, 4^j$. Thus, using at most $\sum_{j=1}^{n}(4j+3\cdot 4^j)$ refinements we arrive at a triangulation $\bar{\mathcal{T}}_n$ that contains the base pattern for each $Q_{i,j}$ for $i = 1, \cdots, 4^j$ and $j = 1, \cdots, n$. This triangulation is not necessarily conforming and we define $\mathcal{T}_n$ its smallest conforming refinement, so that all functions ψ_j for $j = 1, \cdots, n$ belong to $\mathbb{V}_0(\mathcal{T}_n)$, and therefore so does the function

$$V_n := \sum_{j=1}^{n} 2^{-j}\psi_j.$$

According to (2.1), we have

$$\#(\mathcal{T}_n) \le C_0 \sum_{j=1}^{n}(4j + 3 \cdot 4^j) \le C4^n.$$

We thus find that

$$\sigma_{C\cdot 4^n}(u) \le \|u - V_n\|_{H_0^1(\Omega)} \le \sum_{j=n+1}^{\infty} 2^{-j}\|\psi_j\|_{H_0^1(\Omega)} \le 2 \cdot 2^{-n},$$

which shows that $u \in \mathcal{A}^{1/2}$.

We now show that f is not in $\mathcal{B}^{1/2}$. Let $\mathcal{T}$ be any conforming triangulation obtained from $\mathcal{T}_0$ by using 4^n bisections. Writing $S_n = -\Delta V_n$ and using Lemma 3.5, we see that

$$\begin{aligned}
\mathcal{D}(f, \mathcal{T}) &\ge \mathcal{D}(S_n, \mathcal{T}) - \sqrt{3}\|f - S_n\|_{H^{-1}(\Omega)} \\
&= \mathcal{D}(S_n, \mathcal{T}) - \sqrt{3}\|u - V_n\|_{H_0^1(\Omega)} \\
&\ge \mathcal{D}(S_n, \mathcal{T}) - 2\sqrt{3}2^{-n}.
\end{aligned}$$

Thus, it is sufficient to show that

$$\mathcal{D}(S_n, \mathcal{T}) \ge M_n 2^{-n}, \tag{6.20}$$

where $M_n \to \infty$ as $n \to \infty$. To show this, we first let $\mathcal{T}^* = \mathcal{T} \oplus \mathcal{T}_n$ be the overlay of $\mathcal{T}_n$ and $\mathcal{T}$, so that $\#(\mathcal{T}^*) \le C4^n$. Using Lemmas 6.3 and 3.5 gives

$$\mathcal{D}(S_n, \mathcal{T})^2 \ge \frac{1}{3}\mathcal{D}(S_n, \mathcal{T}^*)^2 \ge C \sum_{\sigma \in \Gamma(\mathcal{T}^*)} J_\sigma^2 |\sigma|^2, \tag{6.21}$$

where $J_\sigma = J_\sigma(V_n)$. Notice that $J_\sigma = 0$ unless σ is contained in an edge of $\mathcal{T}_n$ and in any case J_σ is constant on σ.

Let us fix one of the edges σ from $\mathcal{T}_n$ and let σ_i, $i = 1, 2, \cdots, m_\sigma$, denote the edges in $\mathcal{T}^*$ contained in σ. Then,

$$|\sigma| = \sum_{i=1}^{m_\sigma} |\sigma_i| \leq m_\sigma^{1/2} \Big(\sum_{i=1}^{m_\sigma} |\sigma_i|^2 \Big)^{1/2}, \tag{6.22}$$

and therefore

$$\sum_{i=1}^{m_\sigma} J_{\sigma_i}^2 |\sigma_i|^2 \geq m_\sigma^{-1} J_\sigma^2 |\sigma|^2, \tag{6.23}$$

where we have used the fact that $J_{\sigma_i} = J_\sigma$ for each $\sigma_i \subset \sigma$. This gives

$$\mathcal{D}(S_n, \mathcal{T}^*)^2 \geq \sum_{\sigma \in \Gamma(\mathcal{T}_n)} m_\sigma^{-1} J_\sigma^2 |\sigma|^2. \tag{6.24}$$

Notice that the only $\sigma \in \Gamma(\mathcal{T}_n)$ for which $J_\sigma \neq 0$ are the edges contained in the interior of one of the $Q_{i,j}$. For each fixed i, j, there are only four such edges. Also the jump in $\nabla \phi_{Q_{i,j}}$ across each of these edges is the same and equal to $2\sqrt{2}\, 4^j$, and so the jump in $\nabla \psi_j$ across each such edge is $\sqrt{2}\, 2^{j+1}$. Finally the jump $J_\sigma(V_n)$ across each such edges is $2\sqrt{2}$. Since $|\sigma| \geq C 4^{-j}$ for each such edge, we obtain

$$\mathcal{D}(S_n, \mathcal{T}^*)^2 \geq \sum_{j=1}^{n} \sum_{i=1}^{4^j} N_{i,j}^{-1} 4^{-2j} \tag{6.25}$$

where $N_{i,j}$ is the largest of the m_σ for σ in $Q_{i,j}$.

On the other hand, we know that $\sum_{j=1}^n \sum_{i=1}^{4^j} N_{i,j} \leq 4^n$ because there are at most 4^n refinements in creating $\mathcal{T}$ and any refinement of an edge from $\mathcal{T}_n$ must come from the refinements used to create $\mathcal{T}$. Therefore, for at least half of the $j \in \{1, \cdots, n\}$, we have $\sum_{i=1}^{4^j} N_{i,j} \leq 2 \cdot 4^n/n$ and for each of these j for at least half of the $i \in \{1, \cdots, 4^j\}$, we have $N_{i,j} \leq 4 \cdot 4^{n-j}/n$, i.e. $N_{i,j}^{-1} \geq 4^{-n+j-1} n$. Summing over just these i, j, we obtain that the right side of (6.25) is larger than $n 4^{-n-2}$. Thus, we have verified (6.20) as desired.

6.5 An Optimal Data Adaptation Procedure

Theorem 6.1 implies that whenever $u \in \mathcal{A}^s$ for some $0 < s < \frac{1}{2}$, then for all $\tau > 0$, the triangulation $\mathcal{T}_\tau \geq \mathcal{T}_0$ of minimal cardinality such that

$$\mathcal{D}(f, \mathcal{T}_\tau) \leq \tau,$$

satisfies

$$\#(\mathcal{T}_\tau) - \#(\mathcal{T}_0) \leq \|f\|_{\mathcal{B}^s}^{1/s} \tau^{-1/s} \leq C_4^{1/s} \|u\|_{\mathcal{A}^s}^{1/s} \tau^{-1/s}.$$

Therefore, if we define $\mathcal{T}^+ = \mathsf{ADAPTDATA}(f, \mathcal{T}, \tau)$ by

$$\mathcal{T}^+ := \mathcal{T} \oplus \mathcal{T}_{\frac{\tau}{\sqrt{3}}},$$

and recall (2.4), then we find that

$$\#(\mathcal{T}^+) - \#(\mathcal{T}) \le \#(\mathcal{T}_{\frac{\tau}{\sqrt{3}}}) - \#(\mathcal{T}_0) \lesssim \|u\|_{\mathcal{A}^s}^{1/s} \tau^{-1/s}.$$

In addition, using (3.24), we find that

$$\mathcal{D}(f, \mathcal{T}^+) \le \tau.$$

Therefore, we realize that Assumption A(s) holds with $F_s \lesssim \|u\|_{\mathcal{A}^s}$.

This ADAPTDATA procedure is not realistic for several reasons. First, it assumes that we are able to identify the triangulation of minimal cardinality such that the data estimator is controlled by some prescribed tolerance. Even if we have exact knowledge of the data indicators $d(z)$, this task may not be achievable in reasonable computational time. A possibility is to try to compute *near optimal* triangulations, that would still retain the property

$$\#(\mathcal{T}_\tau) - \#(\mathcal{T}_0) \lesssim \|u\|_{\mathcal{A}^s}^{1/s} \tau^{-1/s},$$

with a constant larger than $C_4^{1/s}$ but still fixed; see (6.1). This may be achieved by properly adapting the *near best tree algorithm* of Binev and DeVore [3]; however we shall not engage into this discussion, due to the fact that we face a more severe obstruction, namely the fact that we may not have practical access to the quantities $d(z)$. We address this important issue next.

7 Computable Data Estimators and AFEM

In §7.1 we replace the data indicators $d(z)$ with computable surrogates $\widetilde{d}(z)$, provided more information is known on f. We then adapt our optimal convergence analysis in §7.2 to this setting and show in §7.4 that the procedure ADAPTDATA can then be implemented as a simple *greedy algorithm.*

7.1 Computable Data Estimators

The quantities $d(z)$ require minimal $H^{-1}(\Omega)$ regularity of f but are not easy to evaluate in practice since they involve local $H^{-1}(\omega_z)$ norms. One may circumvent this obstruction provided more information is known on f, by introducing computable surrogate quantities $\widetilde{d}(z)$ that satisfy

$$d(z) \lesssim \widetilde{d}(z),$$

for all $\mathcal{T}$ and $z \in \mathcal{N}(\mathcal{T})$. Two important examples are when $f \in L^p(\Omega)$ for some $1 < p \le 2$ or when f is a Dirac distribution on a 1-dimensional Lipschitz curve $\mathcal{C}$. Note that in such examples f is generally *not* in $L^2(\Omega)$.

Since $\Omega \subset \mathbb{R}^2$, it is well known that $H_0^1(\Omega)$ embeds into $L^q(\Omega)$ for all $q < \infty$ and therefore that $L^p(\Omega)$ continuously embeds in $H^{-1}(\Omega)$ for all $p > 1$. Using the Poincaré inequality and scaling arguments we obtain that for all $g \in H_0^1(\omega_z)$ and $q < \infty$

$$\|g\|_{L^q(\omega_z)} \lesssim h_z^{2/q} \|\nabla g\|_{L^2(\omega_z)}.$$

By duality, we obtain that for all $f \in L^p(\Omega)$ and $q = p/(p-1)$,

$$\|f\|_{H^{-1}(\omega_z)} \lesssim h_z^{2/q} \|f\|_{L^p(\omega_z)}.$$

Therefore, if $|\omega_z| \approx h_z^2$ denotes the measure of ω_z, we obtain

$$d(z) \lesssim \widetilde{d}(z) := |\omega_z|^{1/q} \|f\|_{L^p(\omega_z)}. \tag{7.1}$$

Likewise, assume that f is of the form $f := v\delta_{\mathcal{C}}$, where $\mathcal{C}$ is a Lipschitz curve contained in Ω and $v \in L^p(\mathcal{C})$ for $p > 1$. We then have for any $g \in H_0^1(\Omega)$ and $q = p/(p-1)$

$$\langle f, g \rangle = \int_{\mathcal{C}} v(s) g(s) ds \leq \|v\|_{L^p(\mathcal{C})} \|g\|_{L^q(\mathcal{C})} \lesssim \|v\|_{L^p(\mathcal{C})} \|g\|_{H_0^1(\Omega)},$$

where we have used the Sobolev embedding $H_0^1(\Omega) \subset W_q^r(\Omega)$ with $r = \min\{1, 2/q\}$ and trace theorem (the trace operator maps continuously $W_q^r(\Omega)$ into $W_q^{r-1/q}(\mathcal{C})$); hence $f \in H^{-1}(\Omega)$. Localizing this estimate near the portion of $\mathcal{C}$ which intersects a star ω_z, we find that

$$\|f\|_{H^{-1}(\omega_z)} \lesssim |\mathcal{C} \cap \omega_z|^{1/q} \ \|v\|_{L^p(\mathcal{C} \cap \omega_z)},$$

where $|\mathcal{C} \cap \omega_z|$ denotes the length of $\mathcal{C} \cap \omega_z$. It follows that

$$d(z) \lesssim \widetilde{d}(z) := |\mathcal{C} \cap \omega_z|^{1/q} \ \|v\|_{L^p(\mathcal{C} \cap \omega_z)}. \tag{7.2}$$

Such surrogate data indicators $\widetilde{d}(z)$ lead to a computable data estimator

$$\widetilde{\mathcal{D}}(f, \mathcal{T}) := \Big(\sum_{z \in \mathcal{N}(\mathcal{T})} \widetilde{d}_z(f)^2 \Big)^{\frac{1}{2}},$$

which obviously satisfies $\mathcal{D}(f, \mathcal{T}) \lesssim \widetilde{\mathcal{D}}(f, \mathcal{T})$, along with computable error indicators $e(z)$ and estimator $\widetilde{\mathcal{E}}$

$$\widetilde{e}(z)^2 := j(z)^2 + \widetilde{d}(z)^2 \ \text{ and } \ \widetilde{\mathcal{E}}^2 := \widetilde{\mathcal{E}}(U, f, \mathcal{T})^2 := \sum_{z \in \mathcal{N}(\mathcal{T})} \widetilde{e}(z)^2.$$

It is immediate to check that Lemmas 3.1, 3.2 and 3.3, as well as Corollary 3.4 remain valid when $\mathcal{D}$ and $\mathcal{E}$ are replaced by $\widetilde{\mathcal{D}}$ and $\widetilde{\mathcal{E}}$, with different multiplicative constants $\widetilde{C}_G$, $\widetilde{C}_L$, $\widetilde{c}_1$, $\widetilde{C}_1$ and $\widetilde{C}_2$ involved in the corresponding estimates.

The following result shows that quasi-monotonicity property (iii) in Lemma (3.5) also remains valid up to a modification of the constant $\sqrt{3}$.

Lemma 7.1 (quasi-monotonicity). *Let $\mathcal{T}^*$ be a refinement of $\mathcal{T}$ and $1 < p \leq \infty$. We then have*

$$\widetilde{\mathcal{D}}(f, \mathcal{T}^*) \leq A \widetilde{\mathcal{D}}(f, \mathcal{T}), \tag{7.3}$$

with $A = 3^{1/r}$ and $r = \min\{p, 2\}$ for both surrogate indicators (7.1) *and* (7.2).

Proof: Let $\widetilde{d}(z)$ be given by (7.1). If $1 < p \leq 2$, then

$$\begin{aligned}\widetilde{\mathcal{D}}(f,\mathcal{T}^*)^2 &= \textstyle\sum_{z^*\in\mathcal{N}(\mathcal{T}^*)} \big(|\omega_{z^*}|^{1/q}\|f\|_{L^p(\omega_{z^*})}\big)^2 \\ &\leq \textstyle\sum_{z\in\mathcal{N}(\mathcal{T})}\sum_{\omega_{z^*}\subset\omega_z} |\omega_{z^*}|^{2/q}\|f\|^2_{L^p(\omega_{z^*})} \\ &\leq \textstyle\sum_{z\in\mathcal{N}(\mathcal{T})} |\omega_z|^{2/q}\sum_{\omega_{z^*}\subset\omega_z} \|f\|^2_{L^p(\omega_{z^*})} \\ &\leq \textstyle\sum_{z\in\mathcal{N}(\mathcal{T})} |\omega_z|^{2/q}\Big(\sum_{\omega_{z^*}\subset\omega_z} \|f\|^p_{L^p(\omega_{z^*})}\Big)^{2/p}.\end{aligned}$$

If $p > 2$, instead, we resort to Hölder inequality with exponents $p/2$ and $p/(p-2)$ to bound the above inner sum as follows:

$$\sum_{\omega_{z^*}\subset\omega_z} |\omega_{z^*}|^{2/q}\|f\|^2_{L^p(\omega_{z^*})} \leq |\omega_z|\Big(\sum_{\omega_{z^*}\subset\omega_z}|\omega_{z^*}|\Big)^{(p-2)/p}\Big(\sum_{\omega_{z^*}\subset\omega_z}\|f\|^p_{L^p(\omega_{z^*})}\Big)^{2/p}.$$

Since every $T^* \in \mathcal{T}^*$ belongs to exactly three stars ω_{z^*}, this immediately gives (7.3) with constant $A = 3^{1/p}$ for $1 < p \leq 2$ and $A = 3^{1/2}$ for $p > 2$.

It remains to consider $\widetilde{d}(z)$ given by (7.2). Since the argument is identical to that above, with $|\omega_{z^*}|$ replaced by $|\mathcal{C}\cap\omega_{z^*}|$, we omit the proof. □

7.2 A Modified AFEM

We may now consider our AFEM based on the surrogate data indicators $\widetilde{d}(z)$. The algorithm has some slight changes compared to the version proposed in §3.1, which we now describe. Choose parameters $0 < \theta < 1$, and an initial conforming mesh $\mathcal{T}_0$ satisfying the initial labeling of §2. Set $k = 0$ and $\mathcal{T}^+_{-1} := \mathcal{T}_0$, and iterate

```
$U_k = \mathsf{SOLVE}(\mathcal{T}_k)$;
$\Big\{j_k(z), \widetilde{d}_k(z)\Big\}_{z\in\mathcal{N}_k} = \mathsf{ESTIMATE}(\mathcal{T}_k, U_k, f)$;
$\mathcal{M}_k = \mathsf{MARK}(\{\widetilde{e}_k(z)\}_{z\in\mathcal{N}_k}, \mathcal{T}_k, \theta)$;
if $\widetilde{\mathcal{D}}_k > \sigma_k := \frac{\theta}{3}\widetilde{\mathcal{E}}_k$
    $\mathcal{T}^+_k = \mathsf{ADAPTDATA}(\mathcal{T}^+_{k-1}, f, \frac{\sigma_k}{2A})$;
else
    $\mathcal{T}^+_k = \mathcal{T}^+_{k-1}$;
$\mathcal{T}_{k+1} = \mathsf{REFINE}(\mathcal{T}_k, \mathcal{M}_k) \oplus \mathcal{T}^+_k$;
$k \leftarrow k + 1$,
```

where $A = 3^{1/\min\{p,2\}}$ is the constant of Lemma 7.1. Note that the modified AFEM computes two sequences of meshes $\{\mathcal{T}_{k+1}, \mathcal{T}^+_k\}_{k=-1}^{\infty}$ with $\mathcal{T}_{k+1} \geq \mathcal{T}^+_k$; $\mathcal{T}_k$ controls the error whereas $\mathcal{T}^+_k$ deals with data adaptation. This is due to the structure of ADAPTDATA discussed in §7.4.

The modules SOLVE and REFINE are left unchanged. The module ESTIMATE now determines the jump indicators $j_k(z)$ and surrogate data indicators $\widetilde{d}_k(z)$ for each $z \in \mathcal{N}_k$. The module MARK is now based on the bulk criterion

$$\widetilde{\mathcal{E}}(f, U_k, \mathcal{M}_k; \mathcal{T}_k) \geq \theta\widetilde{\mathcal{E}}_k.$$

The procedure $\mathsf{ADAPTDATA}$ builds a conforming refinement of $\mathcal{T}_{k-1}^+$ such that the new data estimator is smaller than a prescribed tolerance: $\mathcal{T}_k^+ = \mathsf{ADAPTDATA}(\mathcal{T}_{k-1}^+, f, \tau)$ should satisfy

$$\widetilde{\mathcal{D}}_k^+ := \widetilde{\mathcal{D}}(f, \mathcal{T}_k^+) \leq \tau.$$

Note that the tolerance parameter of AFEM has been modified from $\frac{\sigma_k}{2\sqrt{3}}$ to $\frac{\sigma_k}{2A}$.

With such modifications, similar results as those of §3 can be obtained with exactly the same proofs. In particular, we reach the contraction property

$$|||u - U_{k+1}|||_\Omega^2 + \gamma\, \widetilde{\mathcal{E}}_{k+1}^2 \leq \alpha^2 \Big(|||u - U_k|||_\Omega^2 + \gamma\, \widetilde{\mathcal{E}_k}^2 \Big). \tag{7.4}$$

Likewise, similar results to those of §4 can be obtained under the Assumption A(s) for the modified module $\mathsf{ADAPTDATA}$, with constant $\widetilde{F}_s$. We denote the constants appearing in the modified estimates by $\widetilde{C}_1, \widetilde{C}_2$, etc. If we assume

$$\textit{the bulk parameter satisfies } 0 < \theta < \widetilde{\theta}_* \textit{ with } \widetilde{\theta}_* := \sqrt{\tfrac{\widetilde{C}_1}{1+\widetilde{C}_L^2}}, \tag{7.5}$$

then we see that $\widetilde{\theta}_* < 1$ and we obtain the following optimal convergence result.

Theorem 7.2 (optimality of the modified AFEM)**.** *Let $u \in \mathcal{A}^s$ and f and ADAPTDATA be such that Assumption A(s) is satisfied with constant $\widetilde{F}_s$. Assume that the bulk parameter satisfies* (7.5)*. If $(\mathcal{T}_k, \mathbb{V}_k, U_k)_{k\geq 0}$ is a sequence of conforming meshes, nested spaces $\mathbb{V}_k$ and Galerkin solutions $U_k \in \mathbb{V}_k$ generated by our AFEM, then*

$$\widetilde{E}(u, f, \mathcal{T}_k) \lesssim \big(|u|_{\mathcal{A}^s} + \widetilde{F}_s \big) \#(\mathcal{T}_k)^{-s}. \tag{7.6}$$

Proof: Upon replacing $\#(\mathcal{T}_j^+) - \#(\mathcal{T}_j)$ by $\#(\mathcal{M}_j^+)$ in (5.11), this proof is identical to that of Theorem 5.4. □

7.3 Membership in $\mathcal{B}^{\frac{1}{2}}$: Constructive Proof

We now give a constructive proof that both forcing functions f of §7.1 satisfy $f \in \mathcal{B}^{\frac{1}{2}}$. To create a suitable approximation to f, we use the following *greedy algorithm* which, starting from the initial mesh $\mathcal{T}_0$, iteratively marks the node z corresponding to the largest data indicator $\widetilde{d}(z)$ (with ties handled in an arbitrary way) and refines the corresponding star ω_z, until the data estimator is below the prescribed tolerance τ:

```
$\mathcal{T} = \mathsf{GREEDY}(\mathcal{T}_0, f, \tau)$
$\mathcal{T} = \mathcal{T}_0$;
do while $\widetilde{\mathcal{D}}(f, \mathcal{T}) > \tau$
    $z := \text{argmax}\{\widetilde{d}(z) = \widetilde{d}(f, z, \mathcal{T}) \ : \ z \in \mathcal{N}(\mathcal{T})\}$;
    $\mathcal{T} = \mathsf{REFINE}(\mathcal{T}, \{z\})$;
end do
```

Theorem 7.3 ($L^p(\Omega)$ is contained in $\mathcal{B}^{\frac{1}{2}}$)**.** *Assume that $f \in L^p(\Omega)$ for $1 < p \le \infty$ and that $\widetilde{d}(z)$ is given by (7.1). Then, for any $\tau > 0$, the cardinality of $\mathcal{T} = \mathsf{GREEDY}(\mathcal{T}_0, f, \tau)$ is controlled by*

$$\#(\mathcal{T}) - \#(\mathcal{T}_0) \le K_p^2 \|f\|_{L^p(\Omega)}^2 \tau^{-2}, \tag{7.7}$$

where K_p depends only on $\mathcal{T}_0$ and on p. In particular, $f \in \mathcal{B}^{\frac{1}{2}}$ with $\|f\|_{\mathcal{B}^{\frac{1}{2}}} \lesssim \widetilde{F}_{\frac{1}{2}} := K_p \|f\|_{L^p(\Omega)}$.

Proof: Let N be the number of iterations of the GREEDY algorithm for the prescribed tolerance τ, and $\{z^i, \mathcal{T}^i\}_{i=0}^N$ be the nodes marked and meshes generated by GREEDY with $\mathcal{T}^0 = \mathcal{T}_0$. Let us define $\mathcal{N}^i = \mathcal{N}(\mathcal{T}^i)$ for $0 \le i \le N-1$ and

$$\delta := \widetilde{d}(f, z^{N-1}, \mathcal{T}^{N-1}) = \max\{\widetilde{d}(f, z, \mathcal{T}^{N-1}) \ : \ z \in \mathcal{N}^{N-1}\}.$$

We thus have

$$\tau \le \widetilde{D}(f, \mathcal{T}^{N-1}) \le \sqrt{\#(\mathcal{N}^{N-1})}\delta. \tag{7.8}$$

On the other hand, it is easily seen that for $0 \le i \le N-1$ we have

$$\widetilde{d}(f, z^i, \mathcal{T}^i) \ge \delta. \tag{7.9}$$

Indeed any star ω_z of $\mathcal{T}^{N-1}$ is contained in a star $\omega_{z'}$ of $\mathcal{T}^i$ which has larger diameter, and in view of the particular form (7.1) of the data indicator, we have

$$\widetilde{d}(f, z, \mathcal{T}^{N-1}) \le \widetilde{d}(f, z', \mathcal{T}^i),$$

whence

$$\max\{\widetilde{d}(f, z, \mathcal{T}^i) \ : \ z \in \mathcal{N}^i\} \ge \max\{\widetilde{d}(f, z, \mathcal{T}^{N-1}) \ : \ z \in \mathcal{N}^{N-1}\} = \delta,$$

which proves (7.9). We thus have

$$|\omega_{z^i}|^{1/q} \|f\|_{L^p(\omega_{z^i})} \ge \delta,$$

and this implies that for each i, we can find at least one triangle $T^i \in \mathcal{T}^i$ contained in the star ω_{z^i} with measure $|T^i|$ and such that

$$|T^i|^{1/q} \|f\|_{L^p(T^i)} \gtrsim \delta.$$

The triangles $T^0, \cdots, T^{N-1}$ are distinct from each other because each T^i is bisected in the refinement process from $\mathcal{T}^i$ to $\mathcal{T}^{i+1}$. We denote by $\mathcal{B} = \{T^0, \cdots, T^{N-1}\}$ this collection of triangles, and by $\mathcal{B}_j$ the set of T^i's satisfying

$$2^{-(j+1)}|\Omega| < |T^i| \le 2^{-j}|\Omega|, \quad j \ge 0.$$

We observe that $|\Omega| 2^{-(j+1)} \#(\mathcal{B}_j) < |\Omega|$, whence $\#(\mathcal{B}_j) < 2^{j+1}$. Moreover, if $T^i \in \mathcal{B}_j$, we have

$$\delta \ \lesssim \ |T^i|^{1/q} \|f\|_{L^p(T^i)} \ \lesssim \ 2^{-j/q} |\Omega|^{1/q} \|f\|_{L^p(T^i)}.$$

Assume now $1 < p < \infty$. Raising the last inequality to the power p and summing up on the triangles $T^i \in \mathcal{B}_j$, which are pairwise disjoint, we find that

$$\#(\mathcal{B}_j) \ \lesssim \ \delta^{-p} 2^{-jp/q} |\Omega|^{p/q} \|f\|_{L^p(\Omega)}^p.$$

Therefore

$$N = \#(\mathcal{B}) = \sum_{j\geq 0} \#(\mathcal{B}_j) \leq \sum_{j\geq 0} \min\left\{C2^{-jp/q}\delta^{-p}\|f\|^p_{L^p(\Omega)}, 2^{j+1}\right\},$$

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

Let us denote by j_0 the smallest j such that the second term dominates the first in the minimum, and assume first that $j_0 > 0$. By definition of j_0 we have

$$2^{j_0} \leq C2^{-(j_0-1)p/q}\delta^{-p}\|f\|^p_{L^p(\Omega)},$$

whence

$$2^{j_0} \leq A_p\delta^{-\frac{pq}{q+p}}\|f\|^{\frac{pq}{q+p}}_{L^p(\Omega)} = A_p\delta^{-1}\|f\|_{L^p(\Omega)}$$

with A_p a constant which only depends on $\mathcal{T}_0, |\Omega|$ and p. We thus have

$$\begin{aligned} N &\leq \textstyle\sum_{j<j_0} 2^{j+1} + C\delta^{-p}\|f\|^p_{L^p(\Omega)} \sum_{j\geq j_0} 2^{-jp/q} \\ &\leq 2^{j_0+1} + \frac{C}{2^{p/q}-1}\delta^{-p}\|f\|^p_{L^p(\Omega)}2^{-(j_0-1)p/q} \\ &\leq B_p\delta^{-1}\|f\|_{L^p(\Omega)}, \end{aligned}$$

with B_p a constant which only depends on $\mathcal{T}_0, |\Omega|$, and p. Combining this bound on N with (7.8), and the following relation ensuing from (2.1)

$$\#(\mathcal{N}^{N-1}) - \#(\mathcal{N}^0) \leq C^0 N$$

due to the proportionality between $\#(\mathcal{T}^i)$ and $\#(\mathcal{N}^i)$,we obtain

$$\tau \leq \sqrt{\#(\mathcal{N}^{N-1})}\delta \leq B_p\frac{\sqrt{\#(\mathcal{N}^0) + C^0N}}{N}\|f\|_{L^p(\Omega)} \lesssim B_p\frac{1}{\sqrt{N}}\|f\|_{L^p(\Omega)}.$$

This is the asserted estimate (7.7) for $1 < p < \infty$ and $j_0 > 0$. If $1 < p < \infty$ and $j_0 = 0$ instead, then we infer that $\delta^{-p}\|f\|^p_{L^p(\Omega)} \lesssim 1$. This implies

$$N \lesssim \delta^{-p}\|f\|^p_{L^p(\Omega)} \lesssim \delta^{-1}\|f\|_{L^p(\Omega)},$$

and (7.7) follows as before. It remains to examine the case $p = \infty$, for which we have $\delta \lesssim |T^i|\|f\|_{L^\infty(T^i)}$. Consequently, squaring and summing over all $T^i \in \mathcal{B}_j$, we deduce

$$\delta^2\#(\mathcal{B}_j) \lesssim 2^{-j}|\Omega|\sum_i |T^i|\|f\|^2_{L^\infty(T^i)} \leq 2^{-j}|\Omega|^2\|f\|^2_{L^\infty(\Omega)}.$$

The argument from now on proceeds as before. This concludes the proof. □.

Theorem 7.4 (line Dirac masses belong to $\mathcal{B}^{\frac{1}{2}}$)**.** *Assume that $f := v\delta_{\mathcal{C}}$ where $\mathcal{C}$ is a Lipschitz curve and $v \in L^p(\mathcal{C})$ with $1 < p \leq \infty$, and that $\widetilde{d}(z)$ is given by (7.2). Then, for any $\tau > 0$, the cardinality of $\mathcal{T} =$ GREEDY$(\mathcal{T}_0, f, \tau)$ is controlled by*

$$\#(\mathcal{T}) - \#(\mathcal{T}_0) \leq K^2_{\mathcal{C}}\|v\|^2_{L^p(\mathcal{C})}\tau^{-2}, \tag{7.10}$$

where the constant $K_{\mathcal{C}}$ depends on $\mathcal{T}_0$ and on the length of curve $\mathcal{C}$. In particular, $f \in \mathcal{B}^{\frac{1}{2}}$ with $\|f\|_{\mathcal{B}^{\frac{1}{2}}} \lesssim \widetilde{F}_s := K_{\mathcal{C}}\|v\|_{L^p(\mathcal{C})}$.

Proof: The proof is essentially the same as the previous one so we just sketch it. We define the sets $\mathcal{B}_j$ in a similar way, and we now obtain

$$\delta \lesssim |\mathcal{C} \cap T^i|^{1/q} \|v\|_{L^p(\mathcal{C}\cap T^i)} \leq A_{\mathcal{C}} 2^{-j/2q} \|v\|_{L^p(\mathcal{C}\cap T^i)},$$

for all $T^i \in \mathcal{B}_j$, where $A_{\mathcal{C}}$ is a constant that depends on $\mathcal{T}_0$ and on the ratio $|\mathcal{C} \cap T^i|/|T^i|^{1/2}$, which is uniformly bounded for $\mathcal{C}$ Lipschitz. Raising this bound for δ to the power p and summing up on the triangles $T^i \in \mathcal{B}_j$, we find that

$$\#(\mathcal{B}_j) \lesssim \delta^{-p} A_{\mathcal{C}}^p 2^{-jp/2q} \|f\|_{L^p(\mathcal{C})}^p.$$

On the other hand, we also have that

$$2^{-(j+1)/2} |\Omega|^{1/2} \#(\mathcal{B}_j) \lesssim |\mathcal{C}|$$

and therefore

$$\#(\mathcal{B}_j) \leq B_{\mathcal{C}} 2^{j/2},$$

where $B_{\mathcal{C}}$ is a constant that depends on $\mathcal{T}_0$ and on the length $|\mathcal{C}|$ of $\mathcal{C}$. This is due to the fact that the triangles of $\mathcal{B}_j$ should have a vertex z such that ω_z captures a substantial portion of $\mathcal{C}$ on ω_z. The rest of the proof is similar to the previous one and is thus omitted. □

7.4 Optimal Data Adaptation: Greedy Algorithm

We finally show that Assumption A(s) is met for the optimal rate $s = \frac{1}{2}$ by a simple concrete realization of the new $\mathsf{ADAPTDATA}$ procedure, via the above *greedy algorithm.* If $\mathcal{T}_{-1}^+ = \mathcal{T}_0$, the following algorithm generates the meshes $\mathcal{T}_k^+$ for $k \geq 0$

$$\mathcal{T}_k^+ = \mathsf{GREEDY}(\mathcal{T}_{k-1}^+, f, \tau_k).$$

Corollary 7.5 (assumption A($\frac{1}{2}$))**.** *Let f be any one of the distributions from §7.1 and let $\widetilde{F}_{\frac{1}{2}}$ be as defined in Theorems 7.3 and 7.4. Then $\mathcal{T}_k^+ = \mathsf{GREEDY}(\mathcal{T}_{k-1}^+, f, \tau_k)$ satisfies Assumption A($\frac{1}{2}$) with constant $\widetilde{F}_{\frac{1}{2}}$.*

Proof: We note that $\widetilde{\mathcal{D}}(f, \mathcal{T}_{k-1}^+) \leq \tau_{k-1}$ for $k \geq 1$. If $\tau_k \geq \tau_{k-1}$, then $\mathcal{T}_k^+ = \mathcal{T}_{k-1}^+$ and there is nothing to prove. On the contrary, if $\tau_k < \tau_{k-1}$ then we observe that the concatenation $\mathcal{T}_k^+ = \mathsf{GREEDY}(\mathsf{GREEDY}(\mathcal{T}_0, f, \tau_{k-1}), f, \tau_k)$ is equivalent $\mathsf{GREEDY}(\mathcal{T}_0, f, \tau_k)$, because the decisions within GREEDY are independent of the tolerance. Consequently, using Theorems 7.3 and 7.4, we deduce

$$\#(\mathcal{T}_k^+) - \#(\mathcal{T}_{k-1}^+) \leq \#(\mathcal{T}_k^+) - \#(\mathcal{T}_0^+) \leq \widetilde{F}_{\frac{1}{2}} \tau_k^{-2}.$$

This is Assumption A($\frac{1}{2}$) with constant $\widetilde{F}_{\frac{1}{2}}$. □

8 Conclusions and Extensions

We finally summarize our rather technical results and discuss possible extensions.

- *H^{-1} data*: We provided a natural framework for a posteriori error estimation with rough data f. This framework consists of localization of f to stars ω_z via a partition of unity and a corresponding localization of the H^{-1} norms. We discussed two relevant and practical examples in detail.

- *Data estimator*: We introduced data indicators $d(z) = \|f\|_{H^{-1}(\omega_z)}$ which measure local data resolution in H^{-1}, and corresponding global data estimators $\mathcal{D}(f,\mathcal{T})$. We explored their connection to the often used data oscillation $\operatorname{osc}(f,\mathcal{T})$ which is defined only for data $f \in L^2(\Omega)$.

- *Contraction property and optimality of our AFEM*: We presented an AFEM with an inner loop to reduce the data estimator, and showed that the main iterative step of the AFEM induces a contraction for the scaled sum of energy error and error estimator. We proved that the AFEM exhibits optimal performance relative to the best decay rates allowed by the solution and data.

- *Approximation classes for the solution and the data*: We proved that, with no further assumptions on f beyond $H^{-1}(\Omega)$, the decay rate N^{-s} of $\mathcal{D}(f,\mathcal{T})$ is ensured if $u \in \mathcal{A}^s$ provided $s < 1/2$. We explored the exceptional case $s = 1/2$ and construct a counterexample. We also discussed two important examples that lead to surrogate estimators $\widetilde{\mathcal{D}}(f,\mathcal{T})$, which are computable, and for which the decay rate $N^{-1/2}$ is ensured for this data estimator.

- *Variable coefficients*: If the diffusion matrix $A(x)$ is no longer piecewise constant over $\mathcal{T}_0$, then the terms $\|f\|_{H^{-1}(\omega_z)}$ that constitute $\mathcal{D}(f,\mathcal{T})$ are replaced by the local residuals $\|f+\operatorname{div}_{\mathcal{T}}(A\nabla U)\|_{H^{-1}(\omega_z)}$ and therefore depend on the discrete solution U; here $\operatorname{div}_{\mathcal{T}}$ stands for the divergence computed elementwise. This setting has been studied by Cascón et al in [5] under the restriction $f \in L^2(\Omega)$.

- *Higher dimensions*: The restriction to $d = 2$ is made for simplicity. Most results of this paper are valid for $d > 2$ with minor modifications.

- *Higher polynomial degree*: If $V \in \mathbb{V}(\mathcal{T})$ is piecewise polynomial of degree $m > 1$, then the quantity $\operatorname{div}(A\nabla V)$ consists of two terms, the usual jump residual $J(V)$ on the skeleton of $\mathcal{T}$ and a polynomial of degree less or equal to $m-2$ inside each element $T \in \mathcal{T}$. This suggests how to modify the definition (3.11) of the data indicators $d(z)$, or the definitions (7.1) and (7.2) of the surrogate quantities $\widetilde{d}(z)$, to reach the larger range of convergence rates $s \le m/2$. Concerning the surrogate definitions, there are several posibilities depending on m and f. For instance, take $m = 2$ and $f \in B^1_r(L^r(\Omega))$, the Besov space with differentiability order 1 and integrability $\frac{2}{3} < r \le 2$; note that $B^1_r(L^r(\Omega)) \subset L^p(\Omega) \subset H^{-1}(\Omega)$ for

$p = \frac{2r}{2-r}$. If we let $d(z) := \|f - f_z\|_{H^{-1}(\omega_z)}$, as discussed in §3.4, then it can be proved that

$$d(z) \lesssim h_z^{3-\frac{2}{r}} \|f\|_{B^1_r(L^r(\omega_z))} =: \widetilde{d}(z),$$

The proof of Theorem 7.3 extends, thereby showing that $f \in \mathcal{B}^1$ with $\|f\|_{\mathcal{B}^1} \lesssim \|f\|_{B^1_r(L^r(\Omega)}$. The proof of Theorem 7.2 is also valid and the modified AFEM exhibits optimal complexity

$$\widetilde{E}(u, f, \mathcal{T}_k) \lesssim \big(|u|_{\mathcal{A}^1} + \|f\|_{B^1_r(L^r(\omega_z))}\big)\#(\mathcal{T}_k)^{-1}.$$

If f is a line Dirac mass, instead, then $u \in \mathcal{A}^1$ but the definition above of $d(z)$ is inadequate for the decay rate $s = 1$. We could replace f_z by a linear combination of Dirac masses $\sum_{\sigma\in\Gamma_z} c_\sigma \delta_\sigma$, supported on the skeleton γ_z of ω_z, or incorporate the line integral somehow into the jump residual as is proposed in [11]. The situation gets even more involved for $m > 2$. Finally, we point out that in order to relate the approximation classes $\mathcal{A}^s$ and $\mathcal{B}^s$, namely to extend Theorem 6.1 to $m > 1$, we would need a definition of data indicator $d(z)$ that includes both a bulk correction f_z and a suitable Dirac distribution supported on γ_z. We leave these issues open.

References

[1] I. Babuška and A. Miller, *A feedback finite element method with a posteriori error estimation. I. The finite element method and some basic properties of the a posteriori error estimator*, Comput. Methods Appl. Mech. Engrg. 61 (1) (1987), 1–40.

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

[3] P. Binev and R. DeVore, *Fast computation in adaptive tree approximation*, Numer. Math. 97 (2004), 193–217.

[4] S. Brenner and L.R. Scott, *The Mathematical Theory of Finite Element Methods*, Springer Texts in Applied Mathematics 15 (2008).

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

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

[7] R.B. Kellogg, *On the Poisson equation with intersecting interfaces*, Applicable Anal., 4 (1974/75), 101–129.

[8] W.F. Mitchell, *A comparison of adaptive refinement techniques for elliptic problems*, ACM Trans. Math Softw., 15 (1989), 326–347.

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

[10] P. Morin, R.H. Nochetto, and K.G.Siebert, *Local problems on stars: A posteriori error estimators, convergence, and performance*, Math. Comp. 72 (2003), 1067–1097.

[11] R.H. Nochetto, *Pointwise a posteriori error estimates for elliptic problems on highly graded meshes*, Math. Comp., 64 (1995), 1–22.

[12] R.H. Nochetto, K.G. Siebert, and A. Veeser, *Theory of adaptive finite element methods: an introduction*, in: Multiscale, Nonlinear, and Adaptive Approximation, R. DeVore and A. Kunoth eds, Springer, 2009, pages 409–542.

[13] R. Stevenson, *An optimal adaptive finite element method*, SIAM J. Numer. Anal. 42 (2005), 2188–2217.

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

Albert Cohen
UPMC Univ Paris 06, UMR 7598, Laboratoire Jacques-Louis Lions, F-75005, Paris, France
CNRS, UMR 7598, Laboratoire Jacques-Louis Lions, F-75005, Paris, France
cohen@ann.jussieu.fr

Ronald DeVore
Department of Mathematics
Texas A& M University
College Station, TX 77843
rdevore@math.tamu.edu

Ricardo H. Nochetto
Department of Mathematics and Institute for Physical Science and Technology
University of Maryland
College Park, MD 20742
rhn@math.umd.edu