HIGH-ORDER AFEM FOR THE LAPLACE-BELTRAMI OPERATOR: CONVERGENCE RATES

ANDREA BONITO [*], J. MANUEL CASCÓN [†], KHAMRON MEKCHAY [‡],
PEDRO MORIN [§], AND RICARDO H. NOCHETTO [¶]

Abstract. We present a new AFEM for the Laplace-Beltrami operator with arbitrary polynomial degree on parametric surfaces, which are globally W^1_∞ and piecewise in a suitable Besov class embedded in $C^{1,\alpha}$ with $\alpha \in (0,1]$. The idea is to have the surface sufficiently well resolved in W^1_∞ relative to the current resolution of the PDE in H^1. This gives rise to a conditional contraction property of the PDE module. We present a suitable approximation class and discuss its relation to Besov regularity of the surface, solution, and forcing. We prove optimal convergence rates for AFEM which are dictated by the worst decay rate of the surface error in W^1_∞ and PDE error in H^1.

1. Introduction. Let γ be a d dimensional surface in $\mathbb{R}^{d+1}$ ($d \geq 1$) either with or without boundary, which is globally Lipschitz and piecewise in a suitable Besov class embedded in $C^{1,\alpha}$ with $\alpha \in (0,1]$. We design and study a quasi-optimal adaptive finite element method (AFEM) to approximate the solution of

$$-\Delta_\gamma u = f \quad \text{on} \quad \gamma, \tag{1.1}$$

where $f \in L_2(\gamma)$ and $-\Delta_\gamma$ is the Laplace-Beltrami operator (or surface Laplacian) on γ. In addition, we impose that $u = 0$ on $\partial\gamma$ or require that $\int_\gamma u = 0$ if $\partial\gamma = \emptyset$ (with $\int_\gamma f = 0$ for compatibility). To represent Δ_γ, one needs to describe γ mathematically using, for example, parametric representations on charts, level sets, distance functions, graphs of functions, etc. Moreover, one usually obtains approximate solutions (finite element solutions) by solving the problem on approximate polyhedral surfaces rather than the surface γ itself. Exploiting the variational structure of the Laplace-Beltrami operator, [20] gives an a priori error analysis whereas [17, 16, 25, 6] provide a posteriori counterparts. Our present objective is to continue our research on AFEM for elliptic PDEs on surfaces initiated in [25] for graphs and extended in [6] to parametric surfaces, the latter with polynomial degree $n = 1$. We design herein an AFEM for parametric surfaces using C^0 finite elements of degree $n \geq 1$, prove optimal convergence rates and workload estimates, and study suitable approximation classes for the triple (u, f, γ).

High-order finite elements are superior to linears for geometric problems: they provide better approximation of important geometric quantities such as curvature,

and they are less sensitive to mesh tangling due to tangential node motion for time dependent problems; we refer to [6] for a discussion of several applications. The advantage of high-order methods is even more pronounced when they are combined with adaptivity. AFEMs are known to exploit the nonlinear Besov regularity scale, instead of the linear Sobolev scale, and to deliver optimal convergence rates $N^{-n/d}$ in terms of degrees of freedom N for singular elliptic problems on flat domains with limited Sobolev regularity [30, 12], [11, 13, 21, 22, 28]. The study of AFEM for the Laplace-Beltrami operator on parametric surfaces is, however, restricted to $n = 1$ because the first fundamental form, area element, and normal vector to the discrete surface as well as the surface gradient of discrete functions are piecewise constant, which greatly simplifies the analysis [6]. This paper bridges this gap and provides a comprehensive approach to high-order AFEM on parametric surfaces.

It is standard practice to pose the discrete problem on a piecewise polynomial approximation Γ of the exact surface γ. This is unavoidable when dealing with evolving surfaces, such as time dependent free boundary problems, for which γ is unknown [6]. This surface discrepancy is responsible for a geometric consistency error not present in the flat case, which makes this setting mathematically challenging and intriguing. In fact, there is a non-linear interplay between the approximate surface Γ and the approximate solution U defined on Γ. To elucidate this issue, one might think of the Laplace-Beltrami operator as a linear elliptic operator with variable coefficients in a flat parametric domain, except that the approximate coefficients are not piecewise polynomials as in [7] but rather some rational functions when $n > 1$. The multiplicative structure of the solution-coefficient interaction is an essential new difficulty we must cope with to develop high-order AFEM and study their performance.

To handle this nonlinear interaction, we propose an AFEM which successively applies two different modules: ADAPT_SURFACE approximates the surface γ in W^1_∞ and ADAPT_PDE approximates the solution u in H^1. The former is a greedy algorithm which monitors the geometric estimator whereas the latter deals with a residual estimator. If $\{\mathcal{T}_k, U_k\}_{k=0}^\infty$ denotes the sequence of meshes and Galerkin solutions generated by AFEM in step k using a discrete forcing function F_k, the method reads as follows:

AFEM: Given $\mathcal{T}_0$ and parameters $\varepsilon_0 > 0$, $0 < \rho < 1$, and $\omega > 0$, set $k = 0$.
1. $\mathcal{T}_k^+ = \mathsf{ADAPT\_SURFACE}(\mathcal{T}_k, \omega \varepsilon_k)$
2. $[U_{k+1}, \mathcal{T}_{k+1}] = \mathsf{ADAPT\_PDE}(\mathcal{T}_k^+, \varepsilon_k)$
3. $\varepsilon_{k+1} = \rho \varepsilon_k$; $k = k + 1$
4. go to 1.

This strategy bears similarities with the algorithms proposed in [7] targeting diffusion problems with partial information on the coefficients. This concept relates directly to surface approximation in the present context but it is intrinsically different than piecewise polynomial approximation of coefficients. We develop herein new techniques to handle such differences upon insisting on the geometric nature of the approximation.

We now describe AFEM. For the purpose of this introduction, we assume that γ can be parameterized by a single map $\chi : \Omega \to \gamma$, where $\Omega \subset \mathbb{R}^d$ denotes the corresponding parametric domain and refer to § 2.1 for the more general case. If $\widehat{\mathcal{T}} := \widehat{\mathcal{T}}(\Omega)$ is a generic triangulation in Ω, then $\mathbb{V}(\widehat{\mathcal{T}})$ denotes the space of continuous piecewise polynomial functions of degree $\leq n$ subordinate to $\widehat{\mathcal{T}}$. Let $I_{\mathcal{T}} : C(\overline{\Omega}) \to \mathbb{V}(\widehat{\mathcal{T}})$ be the Lagrange interpolation operator and $X_{\mathcal{T}} = I_{\mathcal{T}} \chi$ be the interpolant of the parametrization χ. The map $X_{\mathcal{T}}$ induces the discrete (piecewise polynomial)

surface $\Gamma := X_{\mathcal{T}}(\Omega)$ and a curved mesh $\mathcal{T} := \{T = X_{\mathcal{T}}(\widehat{T}) \mid \widehat{T} \in \widehat{\mathcal{T}}\}$; note the correspondence between a flat element $\widehat{T} \in \widehat{\mathcal{T}}$ and a curved element $T \in \mathcal{T}$. This one-to-one correspondence between $\widehat{\mathcal{T}}$ and $\mathcal{T}$ justifies the use of subscript $\mathcal{T}$ instead of $\widehat{\mathcal{T}}$ in the notation of $I_{\mathcal{T}}$ and $X_{\mathcal{T}}$; the same slight abuse of notation will be employed throughout the paper with other quantities defined in Ω or $\widehat{\mathcal{T}}$. In this vein, we next define the *geometric estimator* $\lambda_{\mathcal{T}}(\gamma) := \max_{T \in \mathcal{T}} \lambda_{\mathcal{T}}(\gamma, T)$ in terms of the *geometric element indicator*

$$\lambda_{\mathcal{T}}(\gamma, T) := \|\widehat{\nabla}(\chi - X_{\mathcal{T}})\|_{L_{\infty}(\widehat{T})} \qquad \forall T \in \mathcal{T}; \tag{1.2}$$

we observe that $\lambda_{\mathcal{T}}(\gamma, T)$ is evaluated in the flat element $\widehat{T} \in \widehat{\mathcal{T}}$. Given a tolerance $\varepsilon > 0$ and a mesh $\mathcal{T}$, the procedure

$$\mathcal{T}_* = \mathsf{ADAPT\_SURFACE}(\mathcal{T}, \varepsilon),$$

finds adaptively a refinement $\mathcal{T}_*$ of $\mathcal{T}$, denoted $\mathcal{T}_* \geq \mathcal{T}$, and its corresponding piecewise polynomial approximation Γ_* of γ, such that

$$\lambda_{\mathcal{T}_*}(\gamma) \leq \varepsilon. \tag{1.3}$$

We say that this module is $t-optimal$ provided the number of marked elements $\#\mathcal{M}$ to achieve (1.3) satisfies

$$\#\mathcal{M} \preccurlyeq C(\gamma) \, \varepsilon^{-1/t}. \tag{1.4}$$

The largest value of $t \leq n/d$ depends on the dimension d and the polynomial degree $n \geq 1$. In §7.3, we show that (1.4) holds if χ belongs to a suitable Besov space.

Throughout this paper we use the notation $A \preccurlyeq B$ to denote $A \leq CB$ with a constant C independent of A and B, and write $A \cong B$ to mean $B \preccurlyeq A \preccurlyeq B$. We shall indicate if appropriate on which quantities the constant C depends on.

Since the exact and approximate solutions u and U are defined on different surfaces γ and Γ, we have to decide how to compare them. We lift U to γ via the map $X_{\mathcal{T}} \circ \chi^{-1}$, but keep the symbol U, and define the *energy error* to be

$$e_{\mathcal{T}}(U) := \|\nabla_{\gamma}(u - U)\|_{L_2(\gamma)}, \tag{1.5}$$

where ∇_{γ} denotes the *surface gradient* on γ defined below in §3.1. We further denote by $\eta_{\mathcal{T}}(U, F)$ the *residual estimator* of $e_{\mathcal{T}}(U)$, defined later in §4.3. If ε stands for a tolerance, the procedure

$$[U_*, \mathcal{T}_*] = \mathsf{ADAPT\_PDE}(\mathcal{T}, \varepsilon)$$

finds adaptively a refinement $\mathcal{T}_*$ of $\mathcal{T}$ such that the Galerkin solution $U_* \in \mathbb{V}(\mathcal{T}_*)$ on Γ_*—the approximate surface corresponding to $\mathcal{T}_*$—satisfies the prescribed bound $\eta_{\mathcal{T}_*}(U_*, F_*) \leq \varepsilon$. This is the usual loop for linear elliptic PDE [19, 26], [6, 8, 11, 12, 13, 22, 25, 28]

$$\mathsf{SOLVE} \to \mathsf{ESTIMATE} \to \mathsf{MARK} \to \mathsf{REFINE}, \tag{1.6}$$

except that the approximate surface Γ is updated after each REFINE call and therefore changes within $\mathsf{ADAPT\_PDE}$.

Note that there is a tolerance ε_k being reduced geometrically in every outer loop of AFEM, and a small parameter ω (to be determined explicitly) that relates the

tolerances for both procedures. The role of ω is critical to derive convergence rates for AFEM, and is explored computationally in [6, Section 2] for dimension $d = 2$ and polynomial degree $n = 1$. The presence of ω is in the spirit of the inner loop of [30] to handle data $f \in H^{-1}$ and of [7] to deal with discontinuous coefficients in the flat case. It means that the surface must be resolved slightly better than the solution for $\eta_{\mathcal{T}}(U, F)$ to provide reliable information about $e_{\mathcal{T}}(U)$.

Our first main result is a *conditional contraction property* of ADAPT_PDE, which reads as follows and is shown in §6:

> *If the parameter $\omega > 0$ is small enough, there exist constants $0 < \alpha < 1$ and $\beta > 0$ such that, if the geometric estimator $\lambda_{\mathcal{T}_k}(\gamma) \leq \omega \varepsilon_k$ and the error estimator $\eta_k = \eta_{\mathcal{T}_k}(U_k, F_k) \geq \varepsilon_k$, then the inner iterates $\{\Gamma_j, \mathcal{T}_j, U_j, \eta_j\}_{j=0}^J$ of ADAPT_PDE($\mathcal{T}_k^+, \varepsilon_k$) satisfy*
>
> $$e_{j+1}^2 + \beta \eta_{j+1}^2 \leq \alpha^2 \left(e_j^2 + \beta \eta_j^2 \right) \qquad \forall\, 0 \leq j < J,$$
>
> *where $e_j := e_{\mathcal{T}_j}(U_j)$ and J is uniformly bounded with respect to k.*

To derive convergence rates we need to seek a suitable error quantity and associated approximation class; this is fully discussed in §7. Since all decisions of the AFEM are based on the estimators $\{\eta_{\mathcal{T}}(U, F), \lambda_{\mathcal{T}}(\gamma)\}$, the convergence rate of AFEM is dictated by these quantities. We will show in Lemma 4.8 that for all the inner iterates $(\mathcal{T}, U)$ within ADAPT_PDE

$$\|\nabla_\gamma (u - U)\|_{L_2(\gamma)}^2 + \mathrm{osc}_{\mathcal{T}}(U, f)^2 \approx \eta_{\mathcal{T}}(U, F)^2. \tag{1.7}$$

where the oscillation $\mathrm{osc}_{\mathcal{T}}(U, f)$, a quantity evaluated in the parametric domain, can be bounded separately in terms of U and f as follows:

$$\mathrm{osc}_{\mathcal{T}}(U, f)^2 \leq \mathrm{osc}_{\mathcal{T}}(U)^2 + \mathrm{osc}_{\mathcal{T}}(f)^2. \tag{1.8}$$

The presence of the first term is a feature inherent to polynomial degree $n > 1$ which is absent in [6]. This justifies the following notion of *total error*

$$E_{\mathcal{T}}(U; u, f, \gamma) := \left(\|\nabla_\gamma (u - U)\|_{L_2(\gamma)}^2 + \mathrm{osc}_{\mathcal{T}}(U, f)^2 + \omega^{-1} \lambda_{\mathcal{T}}(\gamma)^2 \right)^{1/2}, \tag{1.9}$$

where the scaling ω^{-1} brings the geometric estimator $\lambda_{\mathcal{T}}(\gamma)$ to a size comparable with $\eta_{\mathcal{T}}(U, F)$. Then the quality of the best approximation of (u, f, γ) with N degrees of freedom can be assessed in terms of the following best approximation error:

$$\sigma(N; u, f, \gamma) := \inf_{\mathcal{T} \in \mathbb{T}_N} \inf_{V \in \mathbb{V}(\mathcal{T})} E_{\mathcal{T}}(V; u, f, \gamma),$$

where $\mathbb{V}(\mathcal{T})$ denotes the approximation space on the discrete surface Γ, and $\mathbb{T}_N$ is the set of conforming triangulations obtained after N bisections from $\mathcal{T}_0$. We say that the triple (u, f, γ) belongs to the approximation class $\mathbb{A}_s$, with $0 < s \leq n/d$, if

$$\sigma(N; u, f, \gamma) \preccurlyeq N^{-s}; \tag{1.10}$$

equivalently, for any natural number $N \geq 1$, there is a conforming mesh refinement $\mathcal{T}_N$ of the initial mesh $\mathcal{T}_0$ satisfying $\#\mathcal{T}_N - \#\mathcal{T}_0 \leq N$ and such that $E_{\mathcal{T}_N}(V; u, f, \gamma) \preccurlyeq N^{-s}$ for some $V \in \mathbb{V}(\mathcal{T}_N)$. We observe that if $(u, f, \gamma) \in \mathbb{A}_s$ then the module ADAPT_SURFACE is s-optimal, namely (1.3)–(1.4) are valid with $t = s$.

The algebraic error decay (1.10) relates to Besov regularity for flat domains [4, 23, 21]. The situation for surfaces is much more intricate due to the nonlinear surface-PDE interaction. We wonder whether regularity of γ enabling an error decay N^{-s} in W_∞^1 is compatible with a similar decay rate for $e_{\mathcal{T}}(U)$ and $\mathrm{osc}_{\mathcal{T}}(U, f)$, which depend on the approximate surface Γ. Exploring this question is a fundamental contribution of this paper and entails the study of Besov regularity of products and composition of functions, which we carry out in §9 and is of independent interest. We apply our findings in §7 to quantify the effect of surface approximation in the decay of both $e_{\mathcal{T}}(U)$ and $\mathrm{osc}_{\mathcal{T}}(U, f)$. This leads to our second main contribution:

> Let $0 < p, q \leq \infty$, $0 < s \leq n/d$ such that $s > \frac{1}{p} - \frac{1}{2}$, $s > \frac{1}{q}$. If the triple (u, f, γ) satisfies
>
> $$u \in B_p^{1+sd}(L_p(\Omega)), \quad f \in B_p^{sd}(L_p(\Omega)), \quad \chi \in B_q^{1+sd}(L_q(\Omega)),$$
>
> then (1.10) holds, i.e., $(u, f, \gamma) \in \mathbb{A}_s$. Moreover, $\mathrm{osc}_{\mathcal{T}}(f)$ exhibits a faster decay $s + 1/d$.

We observe that $s > \frac{1}{p} - \frac{1}{2}$ and $s > \frac{1}{q}$ guarantee that $B_p^{1+sd}(L_p(\Omega)) \subset H^1(\Omega)$ and $B_q^{1+sd}(L_q(\Omega)) \subset W_\infty^1(\Omega)$, whence the additional regularity is just above the nonlinear Sobolev scale for both u and γ. This shows that the two scales are indeed compatible, and that if $s = \frac{n}{d}$, then $p > \frac{2d}{2n+d}$ and $q > \frac{d}{n}$ may be smaller than 1 for $n > \frac{d}{2}$. The latter does not happen for $n = 1$ and represents a striking difference with [6].

Our third main contribution is a quasi-optimal decay rate for the AFEM in terms of degrees of freedom under natural restrictions on the initial triangulation $\mathcal{T}_0$, marking parameter θ of MARK and parameter ω of AFEM. This is developed in §8 and reads:

> Let the initial mesh $\mathcal{T}_0$ have an admissible labeling for refinement, and $\theta \in (0, \theta_*)$, $\omega \in (0, \omega_*)$ for θ_*, ω_* sufficiently small. If $(u, f, \gamma) \in \mathbb{A}_s$ then the sequence $\{\Gamma_k, \mathcal{T}_k, U_k\}$ generated by AFEM verifies
>
> $$E_{\mathcal{T}_k}(U_k; u, f, \gamma) \preccurlyeq (\#\mathcal{T}_k - \#\mathcal{T}_0)^{-s}.$$
>
> Moreover, the workload up to step k of AFEM is proportional to $\varepsilon_k^{-1/s}$ provided each inner loop of ADAPT_PDE has linear complexity.

The rest of the paper is organized as follows. We discuss the representation and interpolation of γ in § 2, and basic differential geometry leading to the Laplace-Beltrami operator in § 3. In § 4 we obtain a posteriori error estimates for the energy error and derive several properties of the estimator and oscillation. In § 5 we examine the various modules of AFEM and in § 6 establish the conditional contraction property of ADAPT_PDE (first main result). In § 7 we show that suitable Besov regularity of the triple (u, f, γ) implies $(u, f, \gamma) \in \mathbb{A}_s$, which is our second main result. We next prove quasi-optimal convergence rates in § 8 – our third main result. After recalling a definition of Besov spaces, we establish in § 9 scale independent estimates for the products and compositions of functions in Besov spaces. These results are instrumental in § 7 and do not seem to be available in the literature. Glossaries of notations, relevant constants, definitions and algorithms are provided at the end.

2. Parametric Surfaces. In this section we discuss how to represent a parametric surface γ by non-overlapping charts and its discretization Γ by interpolation of γ, which is instrumental for the design, analysis, and implementation of our AFEM. Dealing with overlapping charts is not practical computationally.

2.1. Representation of Parametric Surfaces. We assume that the surface γ is described as the deformation of a d dimensional polyhedral surface $\overline{\Gamma}_0$ by a globally Lipschitz *homeomorphism* $P_0 : \overline{\Gamma}_0 \to \gamma \subset \mathbb{R}^{d+1}$. The overline notation is to emphasize that $\overline{\Gamma}_0$ is piecewise affine. Moreover, if $\overline{\Gamma}_0 = \bigcup_{i=1}^{M} \overline{\Gamma}_0^i$ is made up of M (closed) faces $\overline{\Gamma}_0^i, i = 1, \ldots, M$, we denote by $P_0^i : \overline{\Gamma}_0^i \to \mathbb{R}^{d+1}$ the restriction of P_0 to $\overline{\Gamma}_0^i$. We refer to $\overline{\Gamma}_0^i$ as a *macro-element* which induces the partition $\{\gamma^i\}_{i=1}^M$ of γ upon setting

$$\gamma^i := P_0^i(\overline{\Gamma}_0^i).$$

Note that this *non-overlapping* parametrization allows for piecewise smooth surfaces γ with possible kinks matched by the decomposition $\{\gamma^i\}_{i=1}^M$.

In order to avoid technicalities, we assume that all the macroelements are simplices, i.e. there is a (closed) reference simplex $\Omega \subset \mathbb{R}^d$, from now on called the *local parametric domain*, and an affine map $X_0^i : \mathbb{R}^d \to \mathbb{R}^{d+1}$ such that $\overline{\Gamma}_0^i = X_0^i(\Omega)$; Figure 2.1 sketches the situation when $d = 2$. We thus let $\chi^i := P_0^i \circ X_0^i : \Omega \to \gamma^i$ be a local parametrization of γ which is bi-Lipschitz, namely there exists a universal constant $L \geq 1$ such that for all $1 \leq i \leq M$

$$L^{-1}|\hat{\mathbf{x}} - \hat{\mathbf{y}}| \leq |\chi^i(\hat{\mathbf{x}}) - \chi^i(\hat{\mathbf{y}})| \leq L|\hat{\mathbf{x}} - \hat{\mathbf{y}}|, \qquad \forall \hat{\mathbf{x}}, \hat{\mathbf{y}} \in \Omega. \tag{2.1}$$

This *minimal regularity* of γ, to be soon strengthened out locally in each macro-element, implies the more familiar condition, valid for a.e. $\hat{\mathbf{x}} \in \Omega$,

$$L^{-1}|\mathbf{w}| \leq |\widehat{\nabla}\chi^i(\hat{\mathbf{x}})\mathbf{w}| \leq L|\mathbf{w}| \qquad \forall \, \mathbf{w} \in \mathbb{R}^d. \tag{2.2}$$

We use bold notation to denote the collection of these parametrizations, i.e. $\boldsymbol{\chi} := \{\chi^i\}_{i=1}^M$. We further assume that $P_0(\mathbf{v}) = \mathbf{v}$ for all vertices $\mathbf{v}$ of $\overline{\Gamma}_0$, so that X_0^i is the nodal interpolant of χ^i into linears.
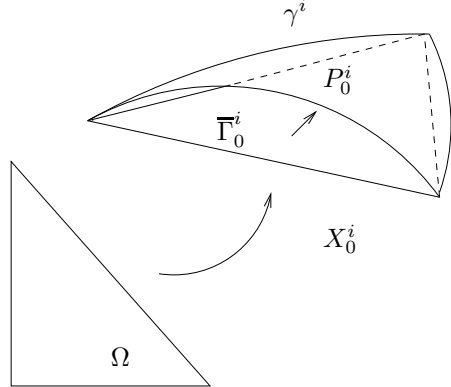
FIG. 2.1. *Representation of each component γ^i when $d = 2$ as a parametrization from a flat triangle $\overline{\Gamma}_0^i \subset \mathbb{R}^3$ as well as from the reference simplex $\Omega \subset \mathbb{R}^2$. The map $X_0^i : \Omega \to \overline{\Gamma}_0^i$ is affine.*

The structure of the map P_0 depends on the application. For instance, if γ^i is described on Γ_0^i via the *distance function* $\text{dist}(\mathbf{x})$ to γ, then

$$\gamma^i \ni \tilde{\mathbf{x}} = \mathbf{x} - \text{dist}(\mathbf{x})\nabla \text{dist}(\mathbf{x}) = P_0(\mathbf{x}) \qquad \forall \, \mathbf{x} \in \overline{\Gamma}_0^i,$$

provided dist($\mathbf{x}$) is sufficiently small so that the distance is uniquely defined. If, instead, γ^i is the *zero level set* $\phi(\mathbf{x}) = 0$ of a function ϕ, then

$$\overline{\Gamma}_0^i \ni \mathbf{x} = \tilde{\mathbf{x}} + \frac{\nabla\phi(\tilde{\mathbf{x}})}{|\nabla\phi(\tilde{\mathbf{x}})|}\,|\mathbf{x} - \tilde{\mathbf{x}}| = P_0^{-1}(\tilde{\mathbf{x}}) \qquad \forall\,\tilde{\mathbf{x}} \in \gamma^i$$

is the inverse map of P_0. In both cases, dist and ϕ must be C^2 for P_0^i to be $C^1(\overline{\Gamma}_0^i)$. Yet another option is to view γ^i as a graph on $\overline{\Gamma}_0^i$, in which case P_0^i is a lift in the normal direction to $\overline{\Gamma}_0^i$ and P_0 is $C^1(\overline{\Gamma}_0^i)$ if and only if γ^i is $C^1(\overline{\Omega})$; we refer to [25]. Notice that the inverse mapping theorem implies $(P_0^i)^{-1} \in C^1(\gamma^i)$.

The *regularity* of γ is expressed in terms of the regularity of the maps χ^i. If $s > 0$, $0 < p, q \leq \infty$, we say that γ is piecewise $B_q^{1+s}(L_p(\Omega))$ whenever $\chi^i \in [B_q^{1+s}(L_p(\Omega))]^{d+1}$, $i = 1, \ldots, M$; or shortly $\boldsymbol{\chi} \in [B_q^{1+s}(L_p(\Omega))]^{(d+1)\times M}$. We refer to § 9 for the definition of Besov norms and spaces.

We observe that a function $v^i : \gamma^i \to \mathbb{R}$ defines uniquely two functions $\widehat{v}^i : \Omega \to \mathbb{R}$ and $\bar{v}^i : \overline{\Gamma}_0^i \to \mathbb{R}$ via the maps χ^i and P_0, namely

$$\widehat{v}^i(\hat{\mathbf{x}}) := v^i(\chi^i(\hat{\mathbf{x}})) \quad \forall\,\hat{\mathbf{x}} \in \Omega, \qquad \bar{v}^i(\bar{\mathbf{x}}) := v^i(P_0(\bar{\mathbf{x}})) \quad \forall\,\bar{\mathbf{x}} \in \overline{\Gamma}_0^i. \qquad (2.3)$$

Conversely, a function $\widehat{v}^i : \Omega \to \mathbb{R}$ (respectively, $\bar{v}^i : \overline{\Gamma}_0^i \to \mathbb{R}$) defines uniquely the two functions $v^i : \gamma^i \to \mathbb{R}$ and $\bar{v}^i : \overline{\Gamma}_0^i \to \mathbb{R}$ (respectively, $v^i : \gamma^i \to \mathbb{R}$ and $\widehat{v}^i : \Omega \to \mathbb{R}$). When no confusion is possible, we will denote by v^i the three functions $v^i, \bar{v}^i$ and $\widehat{v}^i$ and set $\tilde{\mathbf{x}}^i := \chi^i(\hat{\mathbf{x}})$ for all $\hat{\mathbf{x}} \in \Omega$, $i = 1, \ldots, M$. Moreover, we will use vector notation

$$\mathbf{v} := \{v^i\}_{i=1}^M, \qquad (2.4)$$

along with the convention

$$\|\mathbf{v}\|_{B(\Omega)} := \max_{i=1,\ldots,M} \|v^i\|_{B(\Omega)}, \qquad |\mathbf{v}|_{B(\Omega)} := \max_{i=1,\ldots,M} |v^i|_{B(\Omega)}. \qquad (2.5)$$

for (quasi) norms and semi-norms defined on a quasi-normed linear space $B(\Omega)$; typically $B(\Omega)$ will be a Lebesgue, Sobolev, or Besov space. Moreover, we will write

$$\|v\|_{B(\widehat{T})}, \qquad |v|_{B(\widehat{T})} \qquad \forall\,\widehat{T} \in \widehat{\mathcal{T}} \qquad (2.6)$$

to indicate the *local* (quasi) norms and semi-norms over a generic element $\widehat{T} \in \widehat{\mathcal{T}}^i$ without specifying the superscript i in either function v or mesh $\widehat{\mathcal{T}}$.

Before proceeding further, we note that as a general rule, we use hat symbols to denote quantities related to Ω, an overline to refer to quantities on $\overline{\Gamma}_0$, tilde to characterize quantities in γ and bold to indicate vector quantities.

2.2. Interpolation of Parametric Surfaces and Finite Element Spaces. The partition of the initial polyhedral surface $\overline{\Gamma}_0$ in macro-elements (or faces) induces a conforming triangulation of $\overline{\Gamma}_0$; we call this set $\overline{\mathcal{T}}_0$. We only discuss the class of conforming meshes $\overline{\mathbb{T}} := \mathbb{T}(\overline{\mathcal{T}}_0)$ created by successive bisections of this initial mesh $\overline{\mathcal{T}}_0$. However, our results remain valid for quad-refinements and red-refinements all with hanging nodes; we refer to Remark 5.3 for the notion of admissible refinements. A triangulation $\overline{\mathcal{T}} \in \overline{\mathbb{T}}$ yields triangulations of M copies of Ω and a piecewise polynomial approximation Γ of γ defined below.

2.2.1. Finite Element Spaces and Surface Approximations. Any number of conforming graded bisections of each macro-element $\overline{\Gamma}_0^i$ generate via $(X_0^i)^{-1}$ a conforming partition of the local parametric domain $\Omega \subset \mathbb{R}^d$ denoted $\widehat{\mathcal{T}}^i(\Omega)$ or simply $\widehat{\mathcal{T}}^i$. For $n \geq 1$, let $\widehat{\mathbb{V}}^i := \mathbb{V}(\widehat{\mathcal{T}}^i)$ be the finite element space of globally continuous piecewise polynomials of degree $\leq n$ on Ω subordinate to the partition $\widehat{\mathcal{T}}^i$, and let $I_{\mathcal{T}^i} : C^0(\overline{\Omega}) \to \widehat{\mathbb{V}}^i$ (resp. $I_{\mathcal{T}^i} : C^0(\overline{\Omega})^d \to (\widehat{\mathbb{V}}^i)^d$) be the Lagrange interpolation operator of scalar functions (resp. of vector-valued functions). We next define

$$X_{\mathcal{T}^i}^i := I_{\mathcal{T}^i} \chi^i, \qquad \Gamma^i := X_{\mathcal{T}^i}^i(\Omega), \qquad \mathcal{T}^i := \left\{ T := X_{\mathcal{T}^i}^i(\widehat{T}) : \widehat{T} \in \widehat{\mathcal{T}}^i \right\}$$

to be the piecewise polynomial interpolation of χ^i and γ^i, and their associated mesh.

We now define the corresponding global quantities. The *global parametric space* Ω^M consists of M identical copies of the local parametric space Ω. Its subdivision is denoted $\widehat{\mathcal{T}}$ and is defined as

$$\widehat{\mathcal{T}} := \cup_{i=1}^M \widehat{\mathcal{T}}^i.$$

Each triangulation $\overline{\mathcal{T}} \in \overline{\mathbb{T}}$ uniquely determines $\widehat{\mathcal{T}}$, so we can define the forest

$$\widehat{\mathbb{T}} := \mathbb{T}(\widehat{\mathcal{T}_0}) := \{\widehat{\mathcal{T}} : \overline{\mathcal{T}} \in \overline{\mathbb{T}}\}.$$

Notice that $\widehat{\mathbb{T}}$ does not correspond necessarily to M copies of the same forest, it is rather a set of M different but compatible forests. Indeed, the bisection rule is governed by the topology of $\overline{\mathcal{T}}_0$ and dictates which initial bisection of each separate Ω is performed. Moreover, refinement of a macroelement in $\overline{\mathcal{T}}_0$ induces a partition of its boundary which must be compatible with refinements of adjacent macroelements. Similarly, the global subdivision $\mathcal{T}$ is given by

$$\mathcal{T} := \cup_{i=1}^M \mathcal{T}^i$$

and

$$\mathbb{T} := \mathbb{T}(\mathcal{T}_0) := \{\mathcal{T} : \overline{\mathcal{T}} \in \overline{\mathbb{T}}\};$$

note that $\mathcal{T}_0 = \overline{\mathcal{T}}_0$ only for polynomial degree $n = 1$. The global piecewise polynomial surface Γ and parametrization $\mathbf{X}_{\mathcal{T}}$ of Γ are then given by

$$\Gamma := \Gamma_{\mathcal{T}} := \cup_{i=1}^M \Gamma^i, \qquad \mathbf{X}_{\mathcal{T}} := \left\{ X_{\mathcal{T}^i}^i \right\}_{i=1}^M.$$

At this point, we remark that $\mathbf{X}_0$ and $\mathbf{X}_{\mathcal{T}_0}$ are, in general, different maps; the first one is the nodal interpolant of χ into linears whereas the last one is the parametrization of $\Gamma_{\mathcal{T}_0}$. Moreover, we say that $(\mathcal{T}, \Gamma)$ is a pair of mesh-surface approximation when $\mathcal{T} \in \mathbb{T}$ and $\Gamma = \Gamma_{\mathcal{T}}$. Also, for a subdivision $\mathcal{T} \in \mathbb{T}$, we denote by $S_{\mathcal{T}}$ the set of interior faces (edges if $d = 2$). Finally, we define the finite element space over $\mathcal{T}$

$$\mathbb{V}(\mathcal{T}) := \{V \in C^0(\Gamma) : V|_{\Gamma^i} \text{ is the lift of some } \widehat{V}^i \in \widehat{\mathbb{V}}^i \text{ via } X_{\mathcal{T}}^i,$$
$$\text{with } V = 0 \text{ on } \partial\Gamma \text{ or } \int_\Gamma V = 0 \text{ if } \partial\Gamma = \emptyset\}, \qquad (2.7)$$

and observe that functions in $\mathbb{V}(\mathcal{T})$ are not piecewise polynomials.

The refinement procedure consists of bisecting elements in $\overline{\mathcal{T}}_0$ and propagating its effects on $\widehat{\mathcal{T}}$ and $\mathcal{T}$ via the mappings $\mathbf{X}_0^{-1}$ and $\mathbf{X}_{\mathcal{T}} \circ \mathbf{X}_0^{-1}$, respectively. Figure 2.2
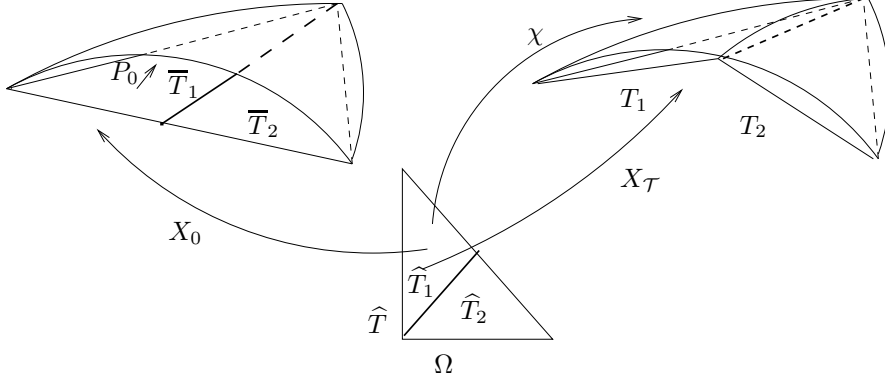
FIG. 2.2. *Effect of one bisection of the macro-element $X_0(\Omega)$ when $d = 2$ and $n = 1$; the superscript i is omitted for simplicity. (Left) A triangle $\overline{T} \in \overline{\mathcal{T}}_0 = \mathcal{T}_0$ is split into two triangles $\overline{T}_1$, $\overline{T}_2 \subset \mathbb{R}^3$. (Bottom) Equivalently, via the affine map X_0^{-1}, the corresponding triangle $\widehat{T} \in \widehat{\mathcal{T}}$ is split into two triangles $\widehat{T}_1$, $\widehat{T}_2 \subset \mathbb{R}^2$, whereas (Right) γ is interpolated by a new piecewise linear surface $\Gamma := X_\mathcal{T}(\Omega)$, with $X_\mathcal{T} = \mathcal{I}_\mathcal{T}\chi$ the piecewise linear interpolant of the parametrization χ defined in Ω and subordinate to the new triangulation $\widehat{\mathcal{T}}$. The images via $X_\mathcal{T}$ of $\widehat{T}_1$ and $\widehat{T}_2$ are denoted T_1 and T_2 respectively; they are affine when $n = 1$.*

depicts one bisection refinement for $d = 2$. For $\overline{\mathcal{T}}, \overline{\mathcal{T}}_* \in \overline{\mathbb{T}}$, we use the notation $\overline{\mathcal{T}}_* \geq \overline{\mathcal{T}}$ to indicate that $\overline{\mathcal{T}}_*$ is a conforming refinement of $\overline{\mathcal{T}}$. In addition, slightly abusing the notation, given two subdivisions $\mathcal{T}, \mathcal{T}_* \in \mathbb{T}$, we write $\mathcal{T}_* \geq \mathcal{T}$ to indicate that $\overline{\mathcal{T}}_* \geq \overline{\mathcal{T}}$. Notice that given $\mathcal{T}, \mathcal{T}_* \in \mathbb{T}$, with $\mathcal{T}_* \geq \mathcal{T}$, the finite element space $\mathbb{V}(\mathcal{T})$ is not a subspace of $\mathbb{V}(\mathcal{T}_*)$ since the associated surface approximation Γ and Γ_* do not match. This lack of consistency is accounted for in the discussion below taking advantage of the nested property $\mathbb{V}(\widehat{\mathcal{T}}^i) \subset \mathbb{V}(\widehat{\mathcal{T}}_*^i)$ for all $1 \leq i \leq M$.

At this point, the three different subdivisions $\overline{\mathcal{T}}, \widehat{\mathcal{T}}$ and $\mathcal{T}$ are defined. Notice that any of these three subdivisions uniquely determines the other two, which is repeatedly used in this work. In practice only $\overline{\mathcal{T}}$ is required and the other two are recovered using the mappings $\mathbf{X}_0^{-1}$ and $\mathbf{X}_\mathcal{T} \circ \mathbf{X}_0^{-1}$. However, they have theoretical different purposes: the subdivision $\overline{\mathcal{T}}$ is made of flat faces obtained as refinement of the initial polyhedral surface and drives the refinement procedure; $\widehat{\mathcal{T}}$ is the triangulation on the parametric space and it is used to evaluate some quantities (geometric estimator, oscillation, etc.) associated to the AFEM in a nested framework; $\mathcal{T}$ is made of curved faces and is the subdivision defining $\Gamma = \Gamma_\mathcal{T}$ where the approximate PDE is solved.

2.2.2. Stability of the Lagrange Interpolation Operator. The Lagrange interpolation operator is instrumental to define the approximate surface and will be central in the definition of the geometric estimator in § 2.2.3. The following lemma discusses its local stability in Besov space $B_p^s(L_q)$ and Sobolev space W_∞^1. We refer to § 9 for a definition of the Besov seminorms.

LEMMA 2.1 (Local stability of Lagrange interpolation). *Let $\widehat{\mathcal{T}}$ be any conforming refinement of $\widehat{\mathcal{T}}_0$. If $s > 0$ and $0 < p, q \leq \infty$ satisfy $s > d/q$ and $s \leq n + 1$, then the Lagrange interpolation operator $I_\mathcal{T}$ (with polynomial degree $n \geq 1$) is stable in $B_p^s(L_q(\widehat{T}))$, namely there exists a constant C depending on s, d, p, q and n such that*

$$|I_\mathcal{T} v|_{B_p^s(L_q(\widehat{T}))} \leq C |v|_{B_p^s(L_q(\widehat{T}))} \qquad \forall v \in B_p^s(L_q(\widehat{T})), \tag{2.8}$$

9

for any $\widehat{T} \in \widehat{\mathcal{T}}$ *and* $\widehat{\mathcal{T}} \in \widehat{\mathbb{T}}$. *The same bound is valid in* $W^1_\infty(\widehat{T})$, *i.e.*

$$\|\widehat{\nabla} I_{\mathcal{T}} v\|_{L_\infty(\widehat{T})} \leq C\|\widehat{\nabla} v\|_{L_\infty(\widehat{T})} \qquad \forall v \in W^1_\infty(\widehat{T}). \tag{2.9}$$

Proof. We consider an arbitrary element $\widehat{T} \in \widehat{\mathcal{T}}$. If $m < s \leq m+1$, with $0 \leq m \leq n$, then for any $P \in \mathbb{P}_m(\widehat{T})$ (the space of polynomials of degree $\leq m$ over $\widehat{T}$) we have $P = I_{\mathcal{T}} P$ and the following holds

$$|I_{\mathcal{T}} v|_{B^s_p(L_q(\widehat{T}))} = |I_{\mathcal{T}}(v-P)|_{B^s_p(L_q(\widehat{T}))} \precsim |\widehat{T}|^{-s/d}\|I_{\mathcal{T}}(v-P)\|_{L_q(\widehat{T})},$$

where in the last inequality we use an inverse estimate for Besov semi-norms, which is obtained by scaling. We now introduce $w := v - P$ and estimate $\|I_{\mathcal{T}} w\|_{L_q(\widehat{T})}$. First, scaling to the reference simplex $\widehat{T}_R$ we get

$$\|I_{\mathcal{T}} w\|_{L_q(\widehat{T})} \precsim |\widehat{T}|^{1/q}\|w\|_{L_\infty(\widehat{T}_R)} \precsim |\widehat{T}|^{1/q}\|w\|_{B^s_p(L_q(\widehat{T}_R))}$$

because $B^s_p(L_q(\widehat{T}_R))$ is embedded in $L_\infty(\widehat{T}_R)$ in view of $sq > d$. Notice that we do not distinguish between the function w defined on $\widehat{T}$ and the corresponding one defined on the reference element $\widehat{T}_R$. We recall that $\|w\|_{B^s_p(L_q(\widehat{T}_R))} = \|w\|_{L_q(\widehat{T}_R)} + |v|_{B^s_p(L_q(\widehat{T}_R))}$, according to the definitions of § 9, and scale back to $\widehat{T}$:

$$|\widehat{T}|^{1/q}\|w\|_{B^s_p(L_q(\widehat{T}_R))} \precsim \|w\|_{L_q(\widehat{T})} + |\widehat{T}|^{s/d}|w|_{B^s_p(L_q(\widehat{T}))}.$$

Combining previous estimates with the immediate generalization of [23, Lemma 4.15]

$$\inf_{P \in \mathbb{P}_m}\|v-P\|_{L_q(\widehat{T})} \precsim |\widehat{T}|^{s/d}|v|_{B^s_p(L_q(\widehat{T}))},$$

and the property $|P|_{B^s_p(L_q(\widehat{T}))} = 0$, we conclude the desired result (2.8)

$$|I_{\mathcal{T}} v|_{B^s_p(L_q(\widehat{T}))} \precsim |\widehat{T}|^{-s/d}\inf_{P \in \mathbb{P}_m}\|v-P\|_{L_q(\widehat{T})} + |v|_{B^s_p(L_q(\widehat{T}))} \precsim |v|_{B^s_p(L_q(\widehat{T}))}.$$

To prove the stability bound (2.9), we take advantage of the representation $I_{\mathcal{T}} v = \sum_{j=1}^{d+1} v(\mathbf{z}_j)\phi_{\mathbf{z}_j}$ in terms of the canonical basis functions $\phi_{\mathbf{z}_j} \in \mathbb{P}_n(\widehat{T})$. Then

$$\widehat{\nabla} I_{\mathcal{T}} v = \sum_{j=1}^{d+1}\left(v(\mathbf{z}_j) - v(\mathbf{z}_l)\right)\widehat{\nabla}\phi_{\mathbf{z}_j} \qquad 1 \leq l \leq d+1,$$

where we exploit that $\{\phi_{\mathbf{z}_j}\}_j$ is a partition of unity over $\widehat{T}$, ie. $\sum_{j=1}^{d+1}\phi_{\mathbf{z}_j} = 1$. Since any sequence of meshes in the flat parametric domain Ω obtained by successive bisections is shape regular, using inverse estimates in $\widehat{T}$ and interpolation in $L_\infty(\widehat{T})$ yields

$$\|\widehat{\nabla} I_{\mathcal{T}} v\|_{L_\infty(\widehat{T})} \leq \max_{1 \leq j \leq d+1}|v(\mathbf{z}_j) - v(\mathbf{z}_l)|\sum_{j=1}^{d+1}\|\widehat{\nabla}\phi_{\mathbf{z}_j}\|_{L_\infty(\widehat{T})} \leq C\|\widehat{\nabla} v\|_{L_\infty(\widehat{T})},$$

where $C > 0$ is a geometric constant independent of γ and proportional to the sum $\sum_{j=1}^{d+1}\|\widehat{\nabla}\phi_{\mathbf{z}_j}\|_{L_\infty(\widehat{T})}$ which depends only on n and d. This concludes the proof. $\square$

2.2.3. Shape Regularity and Geometric Estimators. The proof of Lemma 2.1 utilizes direct and inverse estimates that rely on the shape regularity of elements $\widehat{T} \in \widehat{\mathcal{T}} \in \widehat{\mathbb{T}}$. A discussion about shape regularity of the forests $\overline{\mathbb{T}}$, $\widehat{\mathbb{T}}$ and $\mathbb{T}$ is in order.

The forest $\overline{\mathbb{T}}$ induced by bisection on the flat faces of the initial subdivision $\overline{\mathcal{T}}_0$ is shape regular [3, 28, 31] and so is its counterpart $\widehat{\mathbb{T}}$ on the parametric domain. Regarding the forest $\mathbb{T}$, the question is more subtle and we start with a definition.

DEFINITION 2.2 (Shape regularity). *We say that the class of conforming meshes $\mathbb{T}$ is shape regular if there is a constant C_0 only depending on $\overline{\mathcal{T}}_0$, such that for all $\widehat{\mathcal{T}} \in \widehat{\mathbb{T}}$, and all $i = 1, ..., M$,*

$$C_0^{-1}|\hat{\mathbf{x}} - \hat{\mathbf{y}}| \leq |X_{\mathcal{T}^i}^i(\hat{\mathbf{x}}) - X_{\mathcal{T}^i}^i(\hat{\mathbf{y}})| \leq C_0|\hat{\mathbf{x}} - \hat{\mathbf{y}}| \qquad \forall \hat{\mathbf{x}}, \hat{\mathbf{y}} \in \widehat{T}, \quad \forall \widehat{T} \in \widehat{\mathcal{T}}^i. \quad (2.10)$$

We have already noted that $\widehat{\mathbb{T}}$ is shape regular and observe that (2.10) states that the deformation of $\widehat{T} \in \widehat{\mathcal{T}}^i$ leading to $T = X_{\mathcal{T}^i}^i(\widehat{T}) \in \mathcal{T}^i$ does not degenerate. We also point out that (2.10) implies the usual condition on the Jacobian $\widehat{\nabla} X_{\mathcal{T}^i}^i$, valid for a.e. $\hat{\mathbf{x}} \in \Omega$,

$$C_0^{-1}|\mathbf{w}| \leq |\widehat{\nabla} X_{\mathcal{T}^i}^i(\hat{\mathbf{x}})\,\mathbf{w}| \leq C_0|\mathbf{w}| \qquad \forall \mathbf{w} \in \mathbb{R}^d, \quad (2.11)$$

and that $\widehat{\nabla} X_{\mathcal{T}^i}^i$ happens to be constant on $\widehat{T}$ for an affine map $X_{\mathcal{T}^i}^i$ [14].

We stress that a bi-Lipschitz parametrization satisfying (2.1) does *not* guarantee that $\mathbb{T}$ is shape regular. This issue has been tackled in [9] assuming that the surface γ is W_∞^2 and $\overline{\mathcal{T}}_0$ is sufficiently fine. We present a similar result in Lemma 2.4, invoking piecewise C^1-regularity of γ, which hinges on the quasi-monotonicity of the geometric estimator $\lambda_{\mathcal{T}}(\gamma)$, which we prove first in Lemma 2.3. We start with the definition of $\lambda_{\mathcal{T}}(\gamma)$. Since there is a one-to-one correspondence between subdivisions $\mathcal{T} \in \mathbb{T}$ defining the surface interpolant $\Gamma = \Gamma_{\mathcal{T}}$ and subdivisions $\widehat{\mathcal{T}} \in \widehat{\mathbb{T}}$ of M copies of the parametric domain Ω, we define $\lambda_{\mathcal{T}}(\gamma)$ as follows. For $1 \leq i \leq M$ and $T \in \mathcal{T}^i$ ($\widehat{T} \in \widehat{\mathcal{T}}^i$), let the *geometric element indicator* be

$$\lambda_{\mathcal{T}^i}(\gamma, T) := \|\widehat{\nabla}(\chi^i - X_{\mathcal{T}^i}^i)\|_{L_\infty(\widehat{T})} = \|\widehat{\nabla}(\chi^i - I_{\mathcal{T}^i}\chi^i)\|_{L_\infty(\widehat{T})}, \quad (2.12)$$

and the corresponding *geometric estimator* be

$$\lambda_{\mathcal{T}}(\gamma) := \max_{i=1,...,M} \max_{T \in \mathcal{T}^i} \lambda_{\mathcal{T}^i}(\gamma, T). \quad (2.13)$$

It is worth mentioning that this quantity could increase upon refinement, especially in the pre-asymptotic regime, but the following *quasi-monotonicity* property is valid instead.

LEMMA 2.3 (Quasi-monotonicity of the geometric estimator). *There exists a constant $\Lambda_0 > 1$, solely depending on $\overline{\mathcal{T}}_0$, the polynomial degree n, and dimension d, such that*

$$\lambda_{\mathcal{T}_*}(\gamma) \leq \Lambda_0 \lambda_{\mathcal{T}}(\gamma) \quad (2.14)$$

for any $\mathcal{T}, \mathcal{T}_* \in \mathbb{T}$ with $\mathcal{T}_* \geq \mathcal{T}$. This bound holds elementwise as well.

Proof. We consider an arbitrary element $\widehat{T} \in \widehat{\mathcal{T}}$, i.e. $\widehat{T} \in \widehat{\mathcal{T}}^i$ for some i, but we do not write explicitly the superscript i. We further observe that the Lagrange interpolation operator $I_{\mathcal{T}_*}$ is invariant on polynomials of degree $\leq n$ over $\widehat{T}$, whence

$$\|\widehat{\nabla}(\chi - I_{\mathcal{T}_*}\chi)\|_{L_\infty(\widehat{T})} = \|\widehat{\nabla}(\chi - I_{\mathcal{T}}\chi) - \widehat{\nabla} I_{\mathcal{T}_*}(\chi - I_{\mathcal{T}}\chi)\|_{L_\infty(\widehat{T})}.$$

11

From the local W^1_∞ stability bound (2.9) of $I_{\mathcal{T}_*}$, we deduce the existence of $C > 0$, solely depending on $\overline{\mathcal{T}}_0$, d and n, such that

$$\|\widehat{\nabla} I_{\mathcal{T}_*}(\chi - I_{\mathcal{T}}\chi)\|_{L_\infty(\widehat{T})} \leq C\|\widehat{\nabla}(\chi - I_{\mathcal{T}}\chi)\|_{L_\infty(\widehat{T})}.$$

The desired estimate (2.14) thus follows with $\Lambda_0 = 1 + C$. $\square$

This result turns out to be critical not only for Lemma 2.4 below, which guarantees the shape regularity of $\mathbb{T}$, but also to control the possible increase of the geometric estimator due to the ADAPT_PDE calls. We reiterate that bi-Lipschitz parametrizations satisfying (2.1) do *not* guarantee that $\mathbb{T}$ is shape regular.

LEMMA 2.4 (Shape regularity). *The forest* $\mathbb{T} = \mathbb{T}(\mathcal{T}_0)$ *is shape-regular with constant* $C_0 = 2L$ *provided*

$$\lambda_{\mathcal{T}_0}(\gamma) \leq \frac{1}{2\Lambda_0 L}, \tag{2.15}$$

with $L \geq 1$ *the non-degeneracy constant in (2.1) and* $\Lambda_0 > 0$ *the constant in (2.14).*

Proof. Let $\mathcal{T} \in \mathbb{T}$ be an arbitrary mesh. For any $T \in \mathcal{T}$, we recall that T belongs to a mesh patch $\mathcal{T}^i$ for some i, which we do not write explicitly. Let $\widehat{T}$ be the corresponding element in $\widehat{\mathcal{T}}$. Since for $\widehat{\mathbf{x}}, \widehat{\mathbf{y}} \in \widehat{T}$

$$|(\chi - X_{\mathcal{T}})(\widehat{\mathbf{x}}) - (\chi - X_{\mathcal{T}})(\widehat{\mathbf{y}})| \leq |\widehat{\mathbf{x}} - \widehat{\mathbf{y}}|\|\widehat{\nabla}(\chi - X_{\mathcal{T}})\|_{L_\infty(\widehat{T})} = |\widehat{\mathbf{x}} - \widehat{\mathbf{y}}|\lambda_{\mathcal{T}}(\gamma, T),$$

the shape-regularity assertion is a consequence of (2.1) and (2.14). $\square$

Note that once shape-regularity of the forest $\mathbb{T}$ is established, the triangulations $\mathcal{T}_0$ and $\overline{\mathcal{T}_0}$ are equivalent. Therefore, with a slight abuse of notation, hereafter we regard $\mathcal{T}_0$ as the initial triangulation of the AFEM.

We refer to [6, Figure 11] for an intermediate degenerate situation in which $\lambda_{\mathcal{T}_1}(\gamma) > (2\Lambda_0 L)^{-1}$ and (2.15) is violated for polynomial degree $n = 1$.

3. The Laplace-Beltrami Operator. In this section, we start the discussion with basic differential geometry properties leading to the definition of the Laplace-Beltrami operator Δ_γ together with other relevant geometric operators. We then derive a weak formulation of $-\Delta_\gamma u = f$ as well as its finite element counterpart. We assume γ to be globally W^1_∞ and piecewise C^1, i.e., $\chi^i \in C^1(\overline{\Omega})^{d+1}$ for all $1 \leq i \leq M$, and Γ denote its piecewise polynomial approximation. In the discussion below we often remove the superscript i, because no confusion is possible. We also note that the hat on a function such as $\hat{v}$ will be omitted when the domain is explicit in the formula or when it appears on an operator being applied to the function.

3.1. Basic Differential Geometry. In this subsection we recall a matrix formulation of some basic differential geometry facts and refer to [6] for details. Our first task is to relate the gradient $\widehat{\nabla}$ in the parametric domain Ω with the tangential gradient ∇_γ on γ. To this end, let $\mathbf{T} \in \mathbb{R}^{(d+1)\times d}$ be the matrix

$$\mathbf{T} := \mathbf{T}_\gamma := [\widehat{\partial}_1\chi, \ldots, \widehat{\partial}_d\chi],$$

whose j-th column $\widehat{\partial}_j\chi \in \mathbb{R}^{d+1}$ is the vector of partial derivatives of χ with respect to the jth coordinate of Ω. Since χ is a diffeomorphism, the set $\{\widehat{\partial}_j\chi\}_{j=1}^d$ of tangent vectors to γ is well defined, linearly independent, and expands the tangent hyperplane to each γ^i at interior points for all $1 \leq i \leq M$. The *first fundamental form* of γ is the symmetric and positive definite matrix $\mathbf{G} \in \mathbb{R}^{d\times d}$ defined by

$$\mathbf{G} := (\widehat{\partial}_i\chi^T\widehat{\partial}_j\chi)_{1 \leq i,j \leq d} = \mathbf{T}^T\mathbf{T}. \tag{3.1}$$

12

Given $v : \gamma \to \mathbb{R}$, the tangent gradient $\nabla_\gamma v(\tilde{\mathbf{x}}) = \sum_{i=1}^d \alpha_i(\hat{\mathbf{x}})\hat{\partial}_i\chi(\hat{\mathbf{x}})^T$ is a row vector that satisfies the chain rule

$$\widehat{\nabla} v = \nabla_\gamma v\, \mathbf{T}. \tag{3.2}$$

To get the reverse relation, we augment $\mathbf{T}$ to the matrix $\widetilde{\mathbf{T}} \in \mathbb{R}^{(d+1)\times(d+1)}$ by adding the (outer) unit normal $\boldsymbol{\nu} = (\nu_1, \cdots, \nu_{d+1})^T \in \mathbb{R}^{(d+1)}$ to the tangent hyperplane $\mathrm{span}\{\hat{\partial}_i\chi\}_{i=1}^d$ to γ as the last column, namely

$$\widetilde{\mathbf{T}} := [\mathbf{T}, \boldsymbol{\nu}] = [\hat{\partial}_1\chi, \ldots, \hat{\partial}_d\chi, \boldsymbol{\nu}].$$

Since $\widetilde{\mathbf{T}}$ is invertible, we let $\widetilde{\mathbf{D}} = \widetilde{\mathbf{T}}^{-1}$ and use (3.2) to realize that

$$\nabla_\gamma v = \nabla_\gamma v\, \widetilde{\mathbf{T}}\,\widetilde{\mathbf{D}} = [\widehat{\nabla} v, 0]\widetilde{\mathbf{D}} = \widehat{\nabla} v\, \mathbf{D}, \tag{3.3}$$

where $\mathbf{D} \in \mathbb{R}^{d\times(d+1)}$ results from $\widetilde{\mathbf{D}}$ by suppressing its last row. Moreover, the first fundamental form $\mathbf{G}$ has inverse $\mathbf{G}^{-1} = \mathbf{D}\mathbf{D}^T$. We let

$$q := \sqrt{\det \mathbf{G}} \tag{3.4}$$

be the area element of γ and point out the change of variables formula

$$\int_\omega vq = \int_{\chi(\omega)} v, \tag{3.5}$$

for any $\omega \subset \Omega$ measurable. When χ is C^2 and $v \in H^2(\gamma)$ ($\hat{v} \in H^2(\Omega)$), we have the compact expression for the Laplace-Beltrami operator on γ

$$\Delta_\gamma v = \frac{1}{q}\widehat{\mathrm{div}}\big(q\widehat{\nabla} v\mathbf{G}^{-1}\big).$$

The above representation is instrumental to derive the following integration by parts formula on surfaces

$$\int_\gamma \nabla_\gamma w\nabla_\gamma^T v = \int_\gamma -\Delta_\gamma w\, v + \int_{\partial\gamma} \nabla_\gamma w\, \mathbf{n}\, v \qquad \forall v, w \in H^2(\gamma), \tag{3.6}$$

where $\mathbf{n}$ is the unit co-normal on $\partial\gamma$ pointing outside γ.

The discussion above applies as well to the piecewise polynomial surface Γ (recall that we dropped the index i specifying the underlying patch). We denote the corresponding matrices $\mathbf{T}_\Gamma = \widehat{\nabla} X_{\mathcal{T}}$ and $\mathbf{D}_\Gamma$ associated with $X_{\mathcal{T}} : \Omega \to \Gamma$, and get

$$\nabla_\Gamma v = \widehat{\nabla} v\, \mathbf{D}_\Gamma. \tag{3.7}$$

The first fundamental form $\mathbf{G}_\Gamma$ of Γ and its elementary area q_Γ are defined by

$$\mathbf{G}_\Gamma := \mathbf{T}_\Gamma^T\, \mathbf{T}_\Gamma, \qquad q_\Gamma := \sqrt{\det \mathbf{G}_\Gamma}, \tag{3.8}$$

and the (outer) unit normal to Γ is denoted by $\boldsymbol{\nu}_\Gamma$. The corresponding expression of the Laplace-Beltrami operator is

$$\Delta_\Gamma V = \frac{1}{q_\Gamma}\widehat{\mathrm{div}}\big(q_\Gamma\widehat{\nabla} V\mathbf{G}_\Gamma^{-1}\big), \tag{3.9}$$

13

and only makes sense elementwise. In addition, we recall that for $T \in \mathcal{T}$ and S a side of T, the unit co-normal $\mathbf{n}_\Gamma$ on S pointing outside T satisfies

$$\widehat{\mathbf{n}} = \frac{r_\Gamma}{q_\Gamma} \mathbf{T}_\Gamma^T \mathbf{n}_\Gamma, \qquad \mathbf{n}_\Gamma = \frac{q_\Gamma}{r_\Gamma} \mathbf{D}_\Gamma^T \widehat{\mathbf{n}} \tag{3.10}$$

where r_Γ is the area element associated with the subsimplex $\widehat{S} := X_\mathcal{T}^{-1}(S)$ (see [6] for a detailed expression). Hence, (3.10) and (3.7) give the following local expression for the tangential derivative of v in the direction $\mathbf{n}_\Gamma$ on S

$$\nabla_\Gamma v \, \mathbf{n}_\Gamma = \frac{q_\Gamma}{r_\Gamma} \widehat{\nabla} v \mathbf{G}_\Gamma^{-1}|_{\widehat{S}} \, \widehat{\mathbf{n}}. \tag{3.11}$$

This is of particular importance when considering residual type estimators as in the present work; see § 4.

3.2. Variational Formulation and Galerkin Method. We start by introducing relevant Lebesgue and Sobolev spaces. Let

$$L_{2,\#}(\gamma) := \left\{ v \in L_2(\gamma) \; : \; \int_\gamma v = 0 \quad \text{if} \quad \partial\gamma = \emptyset \right\}$$

be the subspace of $L_2(\gamma)$ of functions with vanishing meanvalue whenever the surface γ is closed, and let $H_\#^1(\gamma)$ be the subspace of $H^1(\gamma)$ given by

$$H_\#^1(\gamma) := \left\{ v \in L_{2,\#}(\gamma) \; : \nabla_\gamma v^i \in [L_2(\gamma^i)]^{d+1}, \right.$$
$$\left. v^i = v^j \text{ on } \gamma^i \cap \gamma^j \; 1 \leq i, j \leq M, \, v = 0 \text{ on } \partial\gamma \right\},$$

where ∇_γ and traces of $v^i = v_{|\gamma^i}$ are well defined in each component γ^i due to (3.3). Let the weak form of the *Laplace-Beltrami operator* $\Delta_\gamma v$ for any function $v \in H_\#^1(\gamma)$ be

$$\langle -\Delta_\gamma v, \varphi \rangle := \sum_{i=1}^M \int_{\gamma^i} \nabla_\gamma v^i \, \nabla_\gamma^T \varphi^i \qquad \forall \varphi \in H_\#^1(\gamma), \tag{3.12}$$

where $\langle \cdot, \cdot \rangle$ denotes the $(H_\#^1(\gamma))^*$-$H_\#^1(\gamma)$ duality product.

We now build on (3.12) and write the weak formulation of $-\Delta_\gamma u = f$ as follows: given $f \in L_{2,\#}(\gamma)$, we seek $u \in H_\#^1(\gamma)$ satisfying

$$\int_\gamma \nabla_\gamma u \nabla_\gamma^T \varphi = \int_\gamma f \varphi, \qquad \forall \, \varphi \in H_\#^1(\gamma), \tag{3.13}$$

where we have written $\int_\gamma \nabla_\gamma u \nabla_\gamma^T \varphi$ to denote $\sum_{i=1}^M \int_{\gamma^i} \nabla_\gamma u^i \nabla_\gamma^T \varphi^i$. Existence and uniqueness of a solution $u \in H_\#^1(\gamma)$ is a consequence of the Lax-Milgram theorem provided γ is Lipschitz.

When χ^i is C^2 and $u \in H^2(\gamma^i)$ for each $1 \leq i \leq M$, we showed in [6] that in the interior of each component γ^i, namely $\chi^i(\text{int}(\Omega))$, we have

$$-\Delta_\gamma u^i = f^i, \quad 1 \leq i \leq M, \tag{3.14}$$

together with vanishing jump conditions at the interfaces $\gamma^i \cap \gamma^j$

$$\mathcal{J}(u)|_{\gamma^i \cap \gamma^j} := \nabla_{\gamma^i} u \, \mathbf{n}^i + \nabla_{\gamma^j} u \, \mathbf{n}^j = 0 \qquad \forall \, 1 \leq i, j \leq M, \tag{3.15}$$

14

where $\mathbf{n}^i$ is the unit outer normal to $\partial\gamma^i$ in the tangent plane to γ^i (see (3.10)).

Given $\mathcal{T} \in \mathbb{T}$, we next formulate an approximation to the Laplace-Beltrami operator on the piecewise polynomial interpolant $\Gamma = \Gamma_{\mathcal{T}}$ of γ as follows. If $F_\Gamma \in L_{2,\#}(\Gamma)$ is a suitable approximation of f, then the finite element solution $U : \Gamma \to \mathbb{R}$ solves

$$U \in \mathbb{V}(\mathcal{T}): \quad \int_\Gamma \nabla_\Gamma U \nabla_\Gamma^T V = \int_\Gamma F_\Gamma V \qquad \forall V \in \mathbb{V}(\mathcal{T}), \tag{3.16}$$

where again $\int_\Gamma g = \sum_{i=1}^M \int_{\Gamma^i} g^i$. To this end we choose F_Γ to be

$$F_\Gamma := f\frac{q}{q_\Gamma}, \tag{3.17}$$

because this specific choice of F_Γ satisfies the compatibility property

$$\int_\Gamma F_\Gamma = \int_\gamma f = 0, \tag{3.18}$$

whenever γ is closed, and allows us to handle separately the approximation of surface γ and forcing f. In particular, (3.16) admits a unique solution U as a consequence of the Lax-Milgram theorem.

4. A Posteriori Error Analysis. In order to study the discrepancy between u and U we need to agree on comparing them in a common domain, say γ. Our goal is thus to obtain a posteriori error estimates for the energy error $\|\nabla_\gamma(u-U)\|_{L_2(\gamma)}$. This entails developing an a priori error analysis for the interpolation error committed in replacing γ by Γ in (3.16), which is a sort of consistency error, and its impact on the PDE error. We are concerned with these issues in this section and refer to [16, 25] where they are addressed for different surface representations as well as [17, 6] that discusses the case $n = 1$. We again drop the superscript i that identifies the surface patch.

4.1. Geometric Error and Estimator. We now quantify the error arising from interpolating γ, the so-called *geometric error*. To this end we resort to the matrix formulation of § 3.1 to relate the geometric error with the geometric estimator $\lambda_{\mathcal{T}}(\gamma)$ of (1.2).

Given $T \in \mathcal{T}$, we will deal with the regions $\widehat{T} \in \widehat{\mathcal{T}}$ and $\widetilde{T} \subset \gamma$ given by

$$\widehat{T} := X_{\mathcal{T}}^{-1}(T), \qquad \widetilde{T} := \chi(\widehat{T}). \tag{4.1}$$

On mapping back and forth to $\widehat{T}$, and using (3.5), we easily see that

$$\int_T v = \int_{\widetilde{T}} v \frac{q_\Gamma}{q}. \tag{4.2}$$

The consistency error stems from the different bilinear forms of the continuous and discrete equations [6, Lemma 5.1]. From (3.5), (3.7), and (3.2), we realize that

$$\int_\Gamma \nabla_\Gamma v \nabla_\Gamma^T w = \int_\gamma \nabla_\Gamma v \mathbf{T}\mathbf{D}_\Gamma \mathbf{D}_\Gamma^T \mathbf{T}^T \nabla_\Gamma^T w \frac{q_\Gamma}{q}. \tag{4.3}$$

Using that $\mathbf{T}\mathbf{D} = \mathbf{I} - \boldsymbol{\nu}\boldsymbol{\nu}^T$ is the projection onto the tangent plane to γ, we obtain

$$\int_\gamma \nabla_\gamma v \nabla_\gamma^T w = \int_\gamma \nabla_\Gamma v \mathbf{T}\mathbf{D}\mathbf{D}^T \mathbf{T}^T \nabla_\Gamma^T w.$$

15

These two expressions, in conjunction with $\mathbf{G}^{-1} = \mathbf{D}\mathbf{D}^T$ and $\mathbf{G}_\Gamma^{-1} = \mathbf{D}_\Gamma \mathbf{D}_\Gamma^T$, yield

$$\int_\Gamma \nabla_\Gamma v \nabla_\Gamma^T w - \int_\gamma \nabla_\gamma v \nabla_\gamma^T w = \int_\gamma \nabla_\gamma v \mathbf{E}_\Gamma \nabla_\gamma^T w \qquad \forall\, v, w \in H^1_\#(\gamma), \qquad (4.4)$$

where $\mathbf{E}_\Gamma \in \mathbb{R}^{(d+1)\times(d+1)}$ stands for the following error matrix

$$\mathbf{E}_\Gamma := \frac{1}{q}\mathbf{T}(q_\Gamma \mathbf{G}_\Gamma^{-1} - q\mathbf{G}^{-1})\mathbf{T}^T. \qquad (4.5)$$

Corollary 5.1 in [6] provides the following conditional estimate on the consistency error: If $\lambda_{\mathcal{T}_0}(\gamma)$ satisfies

$$\lambda_{\mathcal{T}_0}(\gamma) \leq \frac{1}{6\Lambda_0 L^3}, \qquad (4.6)$$

then we have, for $\mathcal{T} \in \mathbb{T}$,

$$\|\mathbf{E}_\Gamma\|_{L_\infty(\widehat{T})} \preccurlyeq \lambda_{\mathcal{T}}(\gamma, T) \qquad \forall\, T \in \mathcal{T}, \qquad (4.7)$$

where the hidden constant depends on $\mathcal{T}_0$ and the Lipschitz constant L of γ appearing in (2.1). The consistency error estimate (4.7) relies on the following properties for q_Γ, r_Γ, $\mathbf{G}_\Gamma$, $\mathbf{D}_\Gamma$ and $\boldsymbol{\nu}_\Gamma$ which will be used again later. Their proofs can be found in [6, Lemmas 5.2 and 5.4] except that of r_Γ, which is analogous and thus omitted.

LEMMA 4.1 (Properties of q_Γ, r_Γ, $\boldsymbol{\nu}_\Gamma$, $\mathbf{G}_\Gamma$ and $\mathbf{D}_\Gamma$). *If $\lambda_{\mathcal{T}_0}(\gamma)$ satisfies (4.6), then the matrices $\mathbf{G}$ and $\mathbf{G}_\Gamma$ have eigenvalues in the interval $[L^{-2}, L^2]$ and $[\frac{1}{2}L^{-2}, \frac{3}{2}L^2]$, respectively. Moreover, the forest $\mathbb{T}$ is shape regular, $L^{-d} \preccurlyeq q, q_\Gamma \preccurlyeq L^d$, and for $\mathcal{T} \in \mathbb{T}$*

$$\begin{aligned} \|q - q_\Gamma\|_{L_\infty(\widehat{T})} + \|r - r_\Gamma\|_{L_\infty(\partial\widehat{T})} + \|\boldsymbol{\nu} - \boldsymbol{\nu}_\Gamma\|_{L_\infty(\widehat{T})} \\ + \|\mathbf{G} - \mathbf{G}_\Gamma\|_{L_\infty(\widehat{T})} + \|\mathbf{D} - \mathbf{D}_\Gamma\|_{L_\infty(\widehat{T})} \preccurlyeq \lambda_{\mathcal{T}}(\gamma, T) \quad \forall\, T \in \mathcal{T} \end{aligned} \qquad (4.8)$$

where we recall that $\Gamma = \Gamma_{\mathcal{T}}$.

We stress that if $\mathcal{T}_0$ does not satisfy (4.6), then the algorithm AFEM of §5 will first refine $\mathcal{T}_0$ to make it comply with (4.6) without ever solving the discretized PDE. In this sense, (4.6) is not a serious restriction for AFEM, although necessary for the subsequent theory. We also note that (4.6) implies (2.15) because $L \geq 1$ in (2.1).

We finally point out the equivalence of norms on γ and Γ provided (4.6) is valid.

LEMMA 4.2 (Equivalence of norms). *If $\lambda_{\mathcal{T}_0}(\gamma)$ satisfies (4.6), then the following equivalence of norms holds for all $\mathcal{T} \in \mathbb{T}$ with constants depending on $\mathcal{T}_0$ and L*

$$\|v\|_{L_2(\widetilde{T})} \approx \|v\|_{L_2(T)} \approx \|v\|_{L_2(\widehat{T})}, \quad |v|_{H^1(\widetilde{T})} \approx |v|_{H^1(T)} \approx |v|_{H^1(\widehat{T})} \quad \forall\, T \in \mathcal{T}, \quad (4.9)$$

where $\widehat{T} = X_{\mathcal{T}}^{-1}(T)$ and $\widetilde{T} = \chi(\widehat{T})$.

Proof. The first assertion follows directly from (4.2) and Lemma 4.1, which implies $L^{-2d} \preccurlyeq \frac{q_\Gamma}{q} \preccurlyeq L^{2d}$. We next rewrite the integrals in (4.3) over $T \in \mathcal{T}$ and $\widetilde{T}$. This, combined with the spectral estimate given in Lemma 4.1 for $\mathbf{G}_\Gamma^{-1} = \mathbf{D}_\Gamma \mathbf{D}_\Gamma^T$ and (2.2) for $\mathbf{T} = \widehat{\nabla}\chi$, yields the second equivalence. Similar reasoning applies to $\widehat{T}$. $\square$

4.2. Inverse Estimates for Discrete Geometric Quantities. We now establish some inverse estimates for the discrete quantities q_Γ and $\mathbf{G}_\Gamma$ that are instrumental to derive Lemma 4.9 (reduction of residual estimator) and Lemma 7.8 (local decay of oscillation). These estimates are only required when the polynomial degree n is strictly greater than 1, which is a key distinction between this work and [6].

In the following, for $T \in \mathcal{T}$, we set $h_T := |\widehat{T}|^{\frac{1}{d}}$ where $\widehat{T}$ is defined in (4.1). This choice is motivated by the resulting reduction property after $b \geq 1$ bisections of $\widehat{T}$.

$$h_{T'} \leq 2^{-b/d} h_T, \tag{4.10}$$

where T' is the curvilinear element corresponding to any element $\widehat{T}' \subset \widehat{T}$.

LEMMA 4.3 (Inverse inequalities in L_p). *If $\lambda_{\mathcal{T}_0}(\gamma)$ satisfies (4.6), then the following estimates hold for all $1 \leq p \leq \infty$, $\mathcal{T}, \mathcal{T}_* \in \mathbb{T}$ and $\mathcal{T} \leq \mathcal{T}_*$, with constants depending on $\mathcal{T}_0$ and L*

$$\|Dq_\Gamma\|_{L_p(\widehat{T})} \lesssim h_T^{\frac{d}{p}-1}, \qquad \|D(q_{\Gamma_*} - q_\Gamma)\|_{L_p(\widehat{T}_*)} \lesssim h_T^{\frac{d}{p}} h_{T_*}^{-1} \lambda_{\mathcal{T}}(\gamma, T), \tag{4.11}$$

$$\|D\mathbf{G}_\Gamma^{-1}\|_{L_p(\widehat{T})} \lesssim h_T^{\frac{d}{p}-1}, \quad \|D(\mathbf{G}_{\Gamma_*}^{-1} - \mathbf{G}_\Gamma^{-1})\|_{L_p(\widehat{T}_*)} \lesssim h_T^{\frac{d}{p}} h_{T_*}^{-1} \lambda_{\mathcal{T}}(\gamma, T), \tag{4.12}$$

whenever $T \in \mathcal{T}$, $T_* \in \mathcal{T}_*$ satisfy $\widehat{T}_* \subset \widehat{T}$.

Proof. We start with $q_\Gamma = \sqrt{\det \mathbf{G}_\Gamma}$ and observe that $\partial_j q_\Gamma = \frac{1}{2q_\Gamma} \partial_j \det \mathbf{G}_\Gamma$ and $\det \mathbf{G}_\Gamma$ is polynomial. Using an inverse inequality for $\det \mathbf{G}_\Gamma$, along with the fact that q_Γ is bounded from above and below (see Lemma 4.1), we obtain

$$\|\partial_j q_\Gamma\|_{L_p(\widehat{T})} \leq \frac{1}{2}\|q_\Gamma^{-1}\|_{L_\infty(\widehat{T})}\|\partial_j \det \mathbf{G}_\Gamma\|_{L_p(\widehat{T})}$$
$$\lesssim \frac{1}{h_T}\|\det \mathbf{G}_\Gamma\|_{L_p(\widehat{T})} \lesssim \frac{1}{h_T}\|q_\Gamma^2\|_{L_p(\widehat{T})} \lesssim h_T^{\frac{d}{p}-1}.$$

We now deal with $q_{\Gamma_*} - q_\Gamma$ as follows. We first write

$$\partial_j(q_{\Gamma_*} - q_\Gamma) = \frac{1}{2}\left(\frac{1}{q_{\Gamma_*}} - \frac{1}{q_\Gamma}\right)\partial_j \det \mathbf{G}_{\Gamma_*} + \frac{1}{2q_\Gamma}\partial_j\left(\det \mathbf{G}_{\Gamma_*} - \det \mathbf{G}_\Gamma\right),$$

whence, for $T \in \mathcal{T}$, $T_* \in \mathcal{T}_*$ with $\widehat{T}_* \subset \widehat{T}$,

$$\|\partial_j(q_{\Gamma_*} - q_\Gamma)\|_{L_p(\widehat{T}_*)} \lesssim \|q_\Gamma - q_{\Gamma_*}\|_{L_p(\widehat{T})}\|\partial_j \det \mathbf{G}_{\Gamma_*}\|_{L_\infty(\widehat{T}_*)}$$
$$+ \|\partial_j\left(\det \mathbf{G}_{\Gamma_*} - \det \mathbf{G}_\Gamma\right)\|_{L_p(\widehat{T}_*)}.$$

Using an inverse inequality for $\det \mathbf{G}_{\Gamma_*} - \det \mathbf{G}_\Gamma = q_{\Gamma_*}^2 - q_\Gamma^2$, the bounds (4.8) on q_Γ and $q_{\Gamma*}$ in terms of $\lambda_{\mathcal{T}}(\gamma, T)$ and $\lambda_{\mathcal{T}_*}(\gamma, T)$, and the quasi-monotonicity (2.14) of $\lambda_{\mathcal{T}}(\gamma, T)$, we get

$$\|\partial_j(q_{\Gamma_*} - q_\Gamma)\|_{L_p(\widehat{T}_*)} \leq h_{T_*}^{-1}\|q_{\Gamma_*} - q_\Gamma\|_{L_p(\widehat{T})} \lesssim h_T^{d/p}h_{T_*}^{-1}\lambda_{\mathcal{T}}(\gamma, T).$$

To estimate $D\mathbf{G}_\Gamma^{-1}$ we see that $\partial_j(\mathbf{G}_\Gamma^{-1}\mathbf{G}_\Gamma) = \partial_j\mathbf{G}_\Gamma^{-1}\mathbf{G}_\Gamma + \mathbf{G}_\Gamma^{-1}\partial_j\mathbf{G}_\Gamma = 0$, whence $\partial_j\mathbf{G}_\Gamma^{-1} = -\mathbf{G}_\Gamma^{-1}\partial_j\mathbf{G}_\Gamma\mathbf{G}_\Gamma^{-1}$. This, an inverse inequality for $\mathbf{G}$ and the lower bound of the eigenvalues of $\mathbf{G}_\Gamma$ in Lemma 4.1 imply

$$\|\partial_j\mathbf{G}_\Gamma^{-1}\|_{L_p(\widehat{T})} \lesssim \|\mathbf{G}_\Gamma^{-1}\|_{L_\infty(\widehat{T})}^2\|\partial_j\mathbf{G}_\Gamma\|_{L_p(\widehat{T})} \lesssim h_T^{-1}\|\mathbf{G}_\Gamma\|_{L_p(\widehat{T})} \lesssim h_T^{\frac{d}{p}-1}.$$

17

Finally, $\mathbf{G}_{\Gamma_*}^{-1} - \mathbf{G}_{\Gamma}^{-1} = \mathbf{G}_{\Gamma_*}^{-1}(\mathbf{G}_{\Gamma} - \mathbf{G}_{\Gamma_*})\mathbf{G}_{\Gamma}^{-1}$, so that the partial derivatives can be computed with the product rule and always keeping the L_p norm in the middle term and the L_∞ norm in the other two. Then, making use of some inverse inequalities together with (4.8) and (2.14), we arrive at

$$\|\partial_j(\mathbf{G}_{\Gamma_*}^{-1} - \mathbf{G}_{\Gamma}^{-1})\|_{L_p(\widehat{T}_*)} \preccurlyeq h_{T_*}^{-1}\|\mathbf{G}_{\Gamma_*} - \mathbf{G}_{\Gamma}\|_{L_p(\widehat{T})} \preccurlyeq h_T^{d/p}\,h_{T_*}^{-1}\,\lambda_{\mathcal{T}}(\gamma, T),$$

as asserted. $\square$

We now establish an inverse estimate in Besov spaces. We refer to § 9 for the definition (9.2) of the Besov seminorm $|V|_{B_\infty^s(L_p(\widehat{T}))}$ in terms of the modulus of smoothness of order $k = \lfloor s \rfloor + 1$

$$\omega_k(V, t)_p = \sup_{|h| \leq t} \|\Delta_h^k V\|_{L_p(\widehat{T})},$$

where Δ_h^k are the k-th order differences defined in (9.1).

LEMMA 4.4 (Inverse estimate in Besov space). *Let $\mathcal{T} \in \mathbb{T}$ and $s > 0$, $0 < p \leq \infty$. Then, the following inequality holds*

$$|\partial_i V|_{B_\infty^s(L_p(\widehat{T}))} \preccurlyeq \frac{1}{h_T}|V|_{B_\infty^s(L_p(\widehat{T}))},$$

for any $\widehat{T} \in \widehat{\mathcal{T}}$, and function $V \in \mathbb{P}_n(\widehat{T})$ or $V = q_\Gamma \mathbf{G}_\Gamma^{-T}$ (with $\Gamma = \Gamma_{\mathcal{T}}$).

Proof. We prove the estimate for $V \in \mathbb{P}_n(\widehat{T})$ because dealing with $q_\Gamma \mathbf{G}_\Gamma^{-T}$ reduces to repeating the steps in the proof of Lemma 4.3 and applying the inverse inequality for polynomials. Since the k-th order differences satisfy

$$\Delta_h^k(\partial_i V)(\widehat{\mathbf{x}}) = \partial_i(\Delta_h^k V)(\widehat{\mathbf{x}}), \qquad \forall \widehat{\mathbf{x}} \in \widehat{T}_{kh},$$

and $\Delta_h^k V \in \mathbb{P}_n(\widehat{T}_{kh})$, in view of (9.1) the usual inverse inequality gives

$$\|\Delta_h^k \partial_i V\|_{L_q(\widehat{T})} = \|\Delta_h^k \partial_i V\|_{L_q(\widehat{T}_{kh})} = \|\partial_i \Delta_h^k V\|_{L_q(\widehat{T}_{kh})} \preccurlyeq \frac{1}{h_T}\|\Delta_h^k V\|_{L_q(\widehat{T})}.$$

Invoking the definition (9.2) yields the desired estimate. $\square$

4.3. Upper and Lower Bounds for the Energy Error. We now derive an error representation formula leading to lower and upper a posteriori bounds for the energy error. Given $\mathcal{T} \in \mathbb{T}$, we recall the notation $\Gamma = \Gamma_{\mathcal{T}}$ and introduce the usual interior and jump residuals suggested by (3.14) and (3.15) for arbitrary $V \in \mathbb{V}(\mathcal{T})$

$$\mathcal{R}_T(V, F_\Gamma) := F_\Gamma|_T + \Delta_\Gamma V|_T, \qquad \mathcal{J}_{\partial T}(V) := \{\mathcal{J}_S(V)\}_{S \subset \partial T} \qquad \forall T \in \mathcal{T},$$
$$\mathcal{J}_S(V) := \nabla_\Gamma V^+|_S \, \mathbf{n}_S^+ + \nabla_\Gamma V^-|_S \, \mathbf{n}_S^- \qquad \forall S \in \mathcal{S}_{\mathcal{T}},$$

where, for each $\mathbf{x} \in S$, $\mathbf{n}_S^\pm(\mathbf{x})$ denotes the outward unit normal to S and tangent to $T^\pm$ at $\mathbf{x}$, and T^+, T^- are curvilinear elements in $\mathcal{T}$ that share the side $S \in \mathcal{S}_{\mathcal{T}}$; recall that $\mathcal{S}_{\mathcal{T}}$ denotes the set of interior faces of $T \in \mathcal{T}$. We emphasize that, in contrast to flat domains, $\mathbf{n}_S^+ \neq -\mathbf{n}_S^-$. Similarly, if $\mathbf{D}_\Gamma^\pm$ denote the matrices associated to $T^\pm$, $\nabla_\Gamma V^\pm|_S = \widehat{\nabla} V^\pm \mathbf{D}_\Gamma^\pm|_{\widehat{S}}$ are tangential gradients of V on $T^\pm$ restricted to S. Moreover, according to (3.9), see that $\Delta_\Gamma V|_T = q_\Gamma^{-1}\widehat{\mathrm{div}}(q_\Gamma \widehat{\nabla} V \mathbf{G}_\Gamma^{-1})|_{\widehat{T}} \neq 0$ in general for $T \in \mathcal{T}$ when the polynomial degree $n > 1$. This is a major difference relative to [6], which deals with $n = 1$ and $V|_{\widehat{T}} \in \mathbb{P}_1(\widehat{T})$, $q_\Gamma \in \mathbb{P}_0(\widehat{T})$, $\mathbf{G}_\Gamma \in \mathbb{P}_0(\widehat{T})^{d \times d}$ imply $\Delta_\Gamma V|_T = 0$.

18

Subtracting the weak formulations (3.13) and (3.16), and employing (3.6) to integrate by parts elementwise, we obtain for all $v \in H_\#^1(\gamma)$:

$$\int_\gamma \nabla_\gamma(u - U) \, \nabla_\gamma^T v = I_1 + I_2 + I_3, \tag{4.13}$$

with

$$I_1 := \sum_{T \in \mathcal{T}} \int_T \mathcal{R}_T(U, F_\Gamma)(v - V) - \sum_{S \in \mathcal{S}_\mathcal{T}} \int_S \mathcal{J}_S(U)(v - V),$$

$$I_2 := \int_\Gamma \nabla_\Gamma U \, \nabla_\Gamma^T v - \int_\gamma \nabla_\gamma U \, \nabla_\gamma^T v = \int_\gamma \nabla_\gamma U \mathbf{E}_\Gamma \nabla_\gamma^T v,$$

$$I_3 := \int_\gamma f v - \int_\Gamma F_\Gamma v.$$

The choice $F_\Gamma = \frac{q}{q_\Gamma} f$ of (3.17) implies $I_3 = 0$ so that only I_1 and I_2 need to be estimated. Observe that I_1 is the usual residual term, whereas I_2 is the geometry consistency term (4.4) and accounts for the discrepancy between γ and Γ. An estimate for the error matrix $\mathbf{E}_\Gamma$ is given in (4.7).

The PDE error indicator stems from I_1 and is defined as follows for any $V \in \mathbb{V}(\mathcal{T})$

$$\eta_\mathcal{T}(V, F_\Gamma, T)^2 := h_T^2 \|\mathcal{R}_T(V, F_\Gamma)\|_{L_2(T)}^2 + h_T \|\mathcal{J}_{\partial T}(V)\|_{L_2(\partial T)}^2 \qquad \forall T \in \mathcal{T}.$$

We recall that the definition $h_T = |\widehat{T}|^{\frac{1}{d}}$ with $\widehat{T} = X_\mathcal{T}^{-1}(T)$ guarantees the strict reduction property (4.10).

We also introduce the *oscillation* for any $V \in \mathbb{V}(\mathcal{T})$ and $T \in \mathcal{T}$

$$\operatorname{osc}_\mathcal{T}(V, f, T)^2 := h_T^2 \left\| (\mathrm{id} - \Pi_{2n-2}^2) \left(fq + \widehat{\operatorname{div}}\big(q_\Gamma \widehat{\nabla} V \mathbf{G}_\Gamma^{-1}\big) \right) \right\|_{L_2(\widehat{T})}^2$$
$$+ h_T \left\| (\mathrm{id} - \Pi_{2n-1}^2) \left(q_\Gamma^+ \widehat{\nabla} V^+ (\mathbf{G}_\Gamma^+)^{-1} \widehat{\mathbf{n}}^+ + q_\Gamma^- \widehat{\nabla} V^- (\mathbf{G}_\Gamma^-)^{-1} \widehat{\mathbf{n}}^- \right) \right\|_{L_2(\partial \widehat{T})}^2, \tag{4.14}$$

where $\widehat{\mathbf{n}}^\pm$ is defined according to (3.10), $\mathbf{G}_\Gamma^\pm$ and $q_\Gamma^\pm = \sqrt{\det \mathbf{G}_\Gamma^\pm}$ are the first fundamental form and area element associated to $T^\pm$, and Π_m^p denotes the best L_p-approximation operator onto the space $\mathbb{P}_m$ of polynomials of degree $\le m$; the domain is implicit from the context. Notice that we used scaled local versions of the residual $q(f + \Delta_\Gamma V)$ (see (3.9)) and co-normal derivatives $r_\Gamma \nabla_\Gamma V \mathbf{n}$ (see (3.11)) to define the oscillation. We refer to Remark 4.7 for an alternative definition of oscillation.

Finally, for any subset $\tau \subset \mathcal{T}$ we set

$$\eta_\mathcal{T}(V, F_\Gamma, \tau)^2 := \sum_{T \in \tau} \eta_\mathcal{T}(V, F_\Gamma, T)^2, \qquad \operatorname{osc}_\mathcal{T}(V, f, \tau)^2 := \sum_{T \in \tau} \operatorname{osc}_\mathcal{T}(V, f, T)^2,$$

and simply write $\eta_\mathcal{T}(V, F_\Gamma)$ and $\operatorname{osc}_\mathcal{T}(V, f)$ whenever $\tau = \mathcal{T}$.

Standard arguments [2, 34] to derive upper and lower bounds for the energy error on flat domains can be extended to this case; see [17, 25, 6].

LEMMA 4.5 (A posteriori upper and lower bounds). *Let* $\lambda_{\mathcal{T}_0}(\gamma)$ *satisfy* (4.6). *Let* $u \in H_\#^1(\gamma)$ *be the solution of* (3.13), $(\mathcal{T}, \Gamma)$ *be a pair of mesh-surface approximations*

and $U \in \mathbb{V}(\mathcal{T})$ be the Galerkin solution of (3.16). Then there exist constants C_1, C_2 and Λ_1 depending only on $\mathcal{T}_0$, the Lipschitz constant of γ, and $\|f\|_{L_2(\gamma)}$, such that

$$\|\nabla_\gamma(u - U)\|^2_{L_2(\gamma)} \le C_1 \eta_{\mathcal{T}}(U, F_\Gamma)^2 + \Lambda_1 \lambda_{\mathcal{T}}(\gamma)^2, \tag{4.15}$$

$$C_2 \eta_{\mathcal{T}}(U, F_\Gamma)^2 \le \|\nabla_\gamma(u - U)\|^2_{L_2(\gamma)} + \operatorname{osc}_{\mathcal{T}}(U, f)^2 + \Lambda_1 \lambda_{\mathcal{T}}(\gamma)^2. \tag{4.16}$$

Proof. Our departing point is (4.13) with $v \in H^1_\#(\gamma)$ arbitrary and $V \in \mathbb{V}(\mathcal{T})$ being its Scott-Zhang interpolant built over the partition $\overline{\mathcal{T}}$ of $\overline{\Gamma}$ and lifted to Γ using $X_{\mathcal{T}} \circ X_0^{-1}$. Using interpolation estimates and (4.9) yields

$$|I_1| \preccurlyeq \eta_{\mathcal{T}}(U, F_\Gamma) \|\nabla_\gamma v\|_{L_2(\gamma)}.$$

Since $\|\nabla_\Gamma U\|_{L_2(\gamma)} \preccurlyeq \|f\|_{L_2(\gamma)}$, the estimate (4.7) on the error matrix $\mathbf{E}_\Gamma$ gives

$$|I_2| \preccurlyeq \lambda_{\mathcal{T}}(\gamma) \|\nabla_\gamma v\|_{L_2(\gamma)}.$$

The upper bound (4.15) follows from $I_3 = 0$. The lower bound (4.16) can be proved locally over an element $\widehat{T} \in \widehat{\mathcal{T}}$ in Ω using standard arguments for flat domains. $\square$

To prove optimality of AFEM we need a localized upper bound for the distance between two discrete solutions U and U_*. This bound measures $\|\nabla_\gamma(U_* - U)\|_{L_2(\gamma)}$ in terms of the PDE estimator restricted to the refined set and geometric estimator; we refer to [6, Lemma 4.13] for a similar estimate for $n = 1$.

LEMMA 4.6 (Localized upper bound). *Let $\lambda_{\mathcal{T}_0}(\gamma)$ satisfy (4.6). For $(\mathcal{T}, \Gamma)$, $(\mathcal{T}_*, \Gamma_*)$ pairs of mesh-surface approximations with $\mathcal{T} \le \mathcal{T}_*$, let $\mathcal{R} := \mathcal{R}_{\mathcal{T} \to \mathcal{T}_*} \subset \mathcal{T}$ be the set of elements refined in $\mathcal{T}$ to obtain $\mathcal{T}_*$ i.e., $\mathcal{R} = \mathcal{T} \setminus \mathcal{T}_*$. Let $U \in \mathbb{V}(\mathcal{T})$ and $U_* \in \mathbb{V}(\mathcal{T}_*)$ be the corresponding discrete solutions of (3.16) on Γ and Γ_*, respectively. Then the following localized upper bound is valid*

$$\|\nabla_\gamma(U_* - U)\|^2_{L_2(\gamma)} \le C_1 \eta_{\mathcal{T}}(U, F_\Gamma, \mathcal{R})^2 + \Lambda_1 \lambda_{\mathcal{T}}(\gamma)^2, \tag{4.17}$$

with constants C_1, Λ_1 as in Lemma 4.5.

Proof. We start from the error representation formula (4.13) by replacing γ by Γ_*, u by U_*, and taking as a test function $v = E_* := U_* - U \in \mathbb{V}(\mathcal{T}_*)$

$$\int_{\Gamma_*} \nabla_{\Gamma_*}(U_* - U) \nabla^T_{\Gamma_*} E_* = I_1 + I_2 + I_3.$$

To estimate I_1, we proceed as in the flat case [12, 28, 30]. We first construct an approximation $V \in \mathbb{V}(\mathcal{T})$ of $E_* \in \mathbb{V}(\mathcal{T}_*)$. Let ω be the union of elements of $\mathcal{R} = \mathcal{T} \setminus \mathcal{T}_*$ and let $\overline{\omega}$ be the corresponding union in $\overline{\mathcal{T}}$. Let ω_j (resp. $\overline{\omega_j}$), $1 \le j \le J$, denote the connected components of the interior of ω (resp. $\overline{\omega}$). We stress that ω_j may intersect several patches Γ^i and likewise $\overline{\omega_j}$ may intersect several copies of Ω. Let $\overline{\mathcal{T}}_j$ be the subset of elements in $\overline{\mathcal{T}}$ contained in $\overline{\omega_j}$ and let $\mathbb{V}(\overline{\mathcal{T}}_j)$ be the restriction of $\mathbb{V}(\overline{\mathcal{T}})$ to $\overline{\omega_j}$. We now construct the Scott-Zhang operator [29] on $\overline{\omega_j}$ and use the map $X_{\mathcal{T}} \circ X_0^{-1}$ to lift it to Γ. We denote this lift by $\pi_j : H^1(\overline{\omega_j}) \to \mathbb{V}(\mathcal{T}_j)$, with

$$\mathcal{T}_j := \left\{ T = X_{\mathcal{T}} \circ X_0^{-1}(\overline{T}) \ : \ \overline{T} \in \overline{\mathcal{T}}_j \right\} \subset \mathcal{T}.$$

Let $V \in \mathbb{V}(\mathcal{T})$ be the following approximation of the error $E_* \in \mathbb{V}(\mathcal{T}_*)$:

$$V := \pi_j E_* \quad \text{in } \omega_j, \qquad V := E_* \quad \text{elsewhere.}$$

By construction, V has conforming boundary values on $\partial\omega_j$, $V \in \mathbb{V}(\mathcal{T})$, and is an H^1-stable approximation to E_*. Since $V = E_*$ in $\Gamma\backslash\omega$, by the same standard argument for flat domains, we obtain

$$|I_1| \leq C_1 \eta_{\mathcal{T}}(U, F_\Gamma, \mathcal{R}) \|\nabla_\Gamma E_*\|_{L_2(\Gamma)}.$$

To estimate I_2, we note that Γ and Γ_* coincide in the unrefined region $\Gamma\backslash\omega$, so that

$$I_2 = \sum_{j=1}^{J} \int_{\widetilde{\omega}_j} \nabla_\gamma U \mathbf{E}_\Gamma \nabla_\gamma^T E_* - \nabla_\gamma U \mathbf{E}_{\Gamma_*} \nabla_\gamma^T E_*$$

with $\widetilde{\omega}_j := \chi \circ X_{\mathcal{T}}^{-1}(\omega_j)$. Combining the estimate (4.7) on the error matrices $\mathbf{E}_\Gamma$ and $\mathbf{E}_{\Gamma_*}$ with (4.9) and (2.14), in its elementwise form, we obtain

$$|I_2| \preccurlyeq (\lambda_{\mathcal{T}}(\gamma) + \lambda_{\mathcal{T}_*}(\gamma)) \|\nabla_\Gamma E_*\|_{L_2(\gamma)} \preccurlyeq (1 + \Lambda_0) \|f\|_{L_2(\gamma)} \lambda_{\mathcal{T}}(\gamma).$$

Since $I_3 = 0$ in view of the choice (3.17) of F_{Γ_*} and F_Γ, collecting the preceding estimates we finally conclude (4.17). $\square$

4.4. Properties of the PDE Estimator and Oscillation. As indicated in (4.15)–(4.16), we have access to the energy error $\|\nabla_\gamma(u - U)\|_{L_2(\gamma)}$ only through the PDE estimator $\eta_{\mathcal{T}}(U, F_\Gamma)$, the geometric estimator $\lambda_{\mathcal{T}}(\gamma)$, and the oscillation quantity $\mathrm{osc}_{\mathcal{T}}(U, f)$. As is customary for flat domains, the definition (4.14) of oscillation guarantees that $\mathrm{osc}_{\mathcal{T}}(U, f)$ is dominated by $\eta_{\mathcal{T}}(U, F_\Gamma)$, namely

$$\mathrm{osc}_{\mathcal{T}}(U, f, T)^2 \leq C_3 \eta_{\mathcal{T}}(U, F_\Gamma, T)^2 \qquad \forall T \in \mathcal{T}, \tag{4.18}$$

where the constant C_3 depends on the surface γ.

REMARK 4.7 (Definition of oscillation). The alternative definition to (4.14):

$$\mathrm{osc}_{\mathcal{T}}(V, f, T)^2 = h_T^2 \left\| (\mathrm{id} - \Pi_{2n-2}^2) \left(f q q_\Gamma^{-1/2} - q_\Gamma^{-1/2} \widehat{\mathrm{div}} \left(q_\Gamma \widehat{\nabla} V \mathbf{G}_\Gamma^{-1} \right) \right) \right\|_{L_2(\widehat{T})}^2$$

$$+ h_T^2 \left\| (\mathrm{id} - \Pi_{2n-1}^2) \left(r_\Gamma^{-1/2} \left(q_\Gamma^+ \widehat{\nabla} V^+ (\mathbf{G}_\Gamma^+)^{-1} \widehat{\mathbf{n}}^+ + q_\Gamma^- \widehat{\nabla} V^- (\mathbf{G}_\Gamma^-)^{-1} \widehat{\mathbf{n}}^- \right) \right) \right\|_{L_2(\widehat{\partial T})}^2$$

would imply (4.18) with an optimal constant $C_3 = 1$. However, this would be at the expense of a more intricate proof of Proposition 7.8 (local decay of oscillation). We opted to use definition (4.14) to simplify the presentation.

The main novelty in (4.15)–(4.17) relative to flat domains, which is also the chief challenge of the present analysis, is the presence of $\lambda_{\mathcal{T}}(\gamma)$. In this respect, we show now the equivalence of $\eta_{\mathcal{T}}(U, F_\Gamma)$ and the *PDE error*

$$\mathcal{E}_{\mathcal{T}}(U, f) := \left(\|\nabla_\gamma(u - U)\|_{L_2(\gamma)}^2 + \mathrm{osc}_{\mathcal{T}}(U, f)^2 \right)^{\frac{1}{2}} \tag{4.19}$$

provided $\lambda_{\mathcal{T}}(\gamma)$ is small relative to $\eta_{\mathcal{T}}(U, F_\Gamma)$. We refer to [12, 28] for a similar result for flat domains, and to [6] for parametric surfaces and $n = 1$.

LEMMA 4.8 (Equivalence of error and estimator). *Let C_1, C_2, Λ_1 be given in Lemma 4.5 and C_3 be as in (4.18). If*

$$\lambda_{\mathcal{T}}(\gamma)^2 \leq \frac{C_2}{2\Lambda_1} \eta_{\mathcal{T}}(U, F_\Gamma)^2, \tag{4.20}$$

21

then there exist explicit constants $C_4 \geq C_5 > 0$, depending on C_1, C_2 and C_3, such that

$$C_5 \eta_{\mathcal{T}}(U, F_\Gamma) \leq \mathcal{E}_{\mathcal{T}}(U, f) \leq C_4 \eta_{\mathcal{T}}(U, F_\Gamma). \tag{4.21}$$

Proof. Combining (4.15) with (4.20), we infer that

$$\|\nabla_\gamma(u - U)\|^2_{L_2(\gamma)} \leq \left(C_1 + \frac{C_2}{2}\right) \eta_{\mathcal{T}}(U, F_\Gamma)^2. \tag{4.22}$$

This, together with (4.18), gives the upper bound in (4.21). We next resort to (4.16) and (4.20) to obtain

$$C_2 \eta_{\mathcal{T}}(U, F_\Gamma)^2 \leq \|\nabla_\gamma(u - U)\|^2_{L_2(\gamma)} + \mathrm{osc}_{\mathcal{T}}(U, f)^2 + \frac{C_2}{2} \eta_{\mathcal{T}}(U, F_\Gamma)^2,$$

which implies the lower bound in (4.21) and concludes the proof. $\square$

It turns out that the usual reduction property of $\eta_{\mathcal{T}}(U, F_\Gamma)$ [12, Corollary 3.4], [28], which is instrumental to prove a contraction property of AFEM, is also compromised by the presence of $\lambda_{\mathcal{T}}(\gamma)$ as stated below.

LEMMA 4.9 (Reduction of residual error estimator). *Let $\lambda_{\mathcal{T}_0}(\gamma)$ satisfy (4.6). Given a mesh-surface pair $(\mathcal{T}, \Gamma)$, let $\mathcal{M} \subset \mathcal{T}$ be the subset of elements bisected at least $b \geq 1$ times in refining $\mathcal{T}$ to obtain $\mathcal{T}_* \geq \mathcal{T}$. If $\xi := 1 - 2^{-\frac{b}{d}}$, then there exist constants Λ_2 and Λ_3, solely depending on the shape regularity of $\mathbb{T}$, the Lipschitz constant L of γ, and $\|f\|_{L_2(\gamma)}$, such that for any $\delta > 0$*

$$\begin{aligned}
\eta_{\mathcal{T}_*}(U_*, F_{\Gamma_*})^2 &\leq (1 + \delta)\left(\eta_{\mathcal{T}}(U, F_\Gamma)^2 - \xi \eta_{\mathcal{T}}(U, F_\Gamma, \mathcal{M})^2\right) \\
&\quad + (1 + \delta^{-1})\left(\Lambda_3 \|\nabla_\gamma(U_* - U)\|^2_{L_2(\gamma)} + \Lambda_2 \lambda_{\mathcal{T}}(\gamma)^2\right).
\end{aligned} \tag{4.23}$$

Proof. We first examine the residual $\mathcal{R}_T(U, F_\Gamma)$. If $T_* \in \mathcal{T}_*$ and $T \in \mathcal{T}$ satisfy $\widehat{T}_* \subset \widehat{T}$, and $T' := X_{\mathcal{T}} \circ X_{\mathcal{T}_*}^{-1}(T_*) \subset T$, then the bound on q_{Γ_*} given in Lemma 4.1 yields

$$\begin{aligned}
\|\mathcal{R}_{T_*}(U_*, F_{\Gamma_*})\|_{L_2(T_*)} &= \|q_{\Gamma_*}^{\frac{1}{2}} \mathcal{R}_{T_*}(U_*, F_{\Gamma_*})\|_{L_2(\widehat{T}_*)} \\
&\precsim \left(\|F_{\Gamma_*} - F_\Gamma\|_{L_2(\widehat{T}_*)} + \|\Delta_{\Gamma_*}(U_* - U)\|_{L_2(\widehat{T}_*)} + \|(\Delta_{\Gamma_*} - \Delta_\Gamma)U\|_{L_2(\widehat{T}_*)}\right) \\
&\quad + \|(q_{\Gamma_*}^{1/2} - q_\Gamma^{1/2})\mathcal{R}_T(U, F_\Gamma)\|_{L_2(\widehat{T}_*)} + \|q_\Gamma^{1/2} \mathcal{R}_T(U, F_\Gamma)\|_{L_2(\widehat{T}_*)}.
\end{aligned}$$

Now, from (4.8) and the local form of (2.14) we bound the first term

$$\|F_{\Gamma_*} - F_\Gamma\|_{L_2(\widehat{T}_*)} \leq \|\left(q_{\Gamma_*}^{-1} - q_\Gamma^{-1}\right) q f\|_{L_2(\widehat{T}_*)} \precsim \lambda_{\mathcal{T}}(\gamma, T')\|f\|_{L_2(T')}.$$

Recalling the expression (3.9) for the Laplace-Beltrami operator and taking $V = U_* - U$, we can write

$$\begin{aligned}
\Delta_{\Gamma_*} V &= q_{\Gamma_*}^{-1} \widehat{\mathrm{div}}\,(q_{\Gamma_*} \widehat{\nabla} V \mathbf{G}_{\Gamma_*}^{-1}) \\
&= q_{\Gamma_*}^{-1}\left(\widehat{\nabla} q_{\Gamma_*} \cdot \widehat{\nabla} V \mathbf{G}_{\Gamma_*}^{-1} + q_{\Gamma_*} \widehat{D}^2 V : \mathbf{G}_{\Gamma_*}^{-1} + q_{\Gamma_*} \widehat{\nabla} V \cdot \widehat{\mathrm{div}}\,\mathbf{G}_{\Gamma_*}^{-1}\right),
\end{aligned}$$

and using bounds for $\|q_{\Gamma_*}\|_{L_\infty(\widehat{T}_*)}$ and $\|\mathbf{G}_{\Gamma_*}^{-1}\|_{L_\infty(\widehat{T}_*)}$ from Lemma 4.1, the inverse inequalities (4.11), (4.12) and a third one for $\widehat{D}^2 V$, we get

$$\|\Delta_{\Gamma_*}(U_* - U)\|_{L_2(\widehat{T}_*)} \precsim \frac{1}{h_{T_*}}\|\nabla_\gamma(U_* - U)\|_{L_2(T')}.$$

22

Again by virtue of (3.9) we rewrite the third term above

$$\|(\Delta_{\Gamma_*} - \Delta_\Gamma)U\|_{L_2(\widehat{T}_*)} \leq \|(q_{\Gamma_*}^{-1} - q_\Gamma^{-1})\widehat{\mathrm{div}}\,(q_{\Gamma_*}\widehat{\nabla}U\mathbf{G}_{\Gamma_*}^{-1})\|_{L_2(\widehat{T}_*)}$$
$$+ \|q_\Gamma^{-1}\widehat{\mathrm{div}}\big((q_{\Gamma_*} - q_\Gamma)\widehat{\nabla}U\mathbf{G}_{\Gamma_*}^{-1}\big)\|_{L_2(\widehat{T}_*)}$$
$$+ \|q_\Gamma^{-1}\widehat{\mathrm{div}}\,(q_\Gamma\widehat{\nabla}U(\mathbf{G}_{\Gamma_*}^{-1} - \mathbf{G}_\Gamma^{-1}))\|_{L_2(\widehat{T}_*)} \preccurlyeq \frac{1}{h_{T_*}}\lambda_{\mathcal{T}}(\gamma, T')\|\nabla U\|_{L_2(T')}$$

due to an inverse inequality for $\widehat{D}^2 U$, (4.11), (4.12) and Lemma 4.1. Finally, using the same arguments for the fourth term we obtain

$$\|(q_{\Gamma_*}^{1/2} - q_\Gamma^{1/2})\mathcal{R}_T(U, F_\Gamma)\|_{L_2(\widehat{T}_*)} = \|(q_{\Gamma_*} - q_\Gamma)(q_{\Gamma_*}^{1/2} + q_\Gamma^{1/2})^{-1}\mathcal{R}_T(U, F_\Gamma)\|_{L_2(\widehat{T}_*)}$$
$$\preccurlyeq \lambda_{\mathcal{T}}(\gamma, T')\|\mathcal{R}_T(U, F_\Gamma)\|_{L_2(\widehat{T}')}$$
$$\preccurlyeq \lambda_{\mathcal{T}}(\gamma, T')\left(\frac{1}{h_{T_*}}\|\nabla U\|_{L_2(T')} + \|f\|_{L_2(T')}\right).$$

As a consequence, the interior residuals on Γ_* and Γ are related through the estimate

$$h_{T_*}\|\mathcal{R}_T(U_*, F_{\Gamma_*})\|_{L_2(T')} \leq h_{T_*}\|\mathcal{R}_T(U, F_\Gamma)\|_{L_2(T')}$$
$$+ C\|\nabla_\gamma(U_* - U)\|_{L_2(T')} + C\lambda_{\mathcal{T}}(\gamma, T)\left(\|\nabla U\|_{L_2(T')} + h_{T_*}\|f\|_{L_2(T')}\right), \tag{4.24}$$

for some constant C only depending on the shape regularity of $\mathbb{T}$ and the Lipschitz constant L of γ.

We now examine the jump residual $\mathcal{J}_{\partial T}(U)$. Let $S_* \in \mathcal{S}_{\Gamma_*}$ and $S' := X_{\mathcal{T}} \circ X_{\mathcal{T}_*}^{-1}(S_*) \subset \Gamma$. We denote by $T_*^\pm$ the two elements of $\mathcal{T}_*$ sharing S_* (resp. $[T^\pm]' := X_{\mathcal{T}} \circ X_{\mathcal{T}_*}^{-1}(T_*^\pm)$) and recall that the corresponding outward pointing co-normals $\mathbf{n}_{S_*}^\pm$ are not necessarily co-linear; moreover, $T_*^\pm$ may belong to different surface patches, i.e. $T_*^+ \in \mathcal{T}_*^i$ and $T_*^- \in \mathcal{T}_*^j$ for some $1 \leq i, j \leq M$. Still, observe that the jump $\mathcal{J}_{S_*}(U_*)$ can be rewritten as follows

$$\mathcal{J}_{S_*}(U_*) = \mathcal{J}_S(U)|_{S_*} + \big(\nabla_{\Gamma_*}U_*^+|_{S_*}\mathbf{n}_{S_*}^+ - \nabla_\Gamma U^+|_{S_*}\mathbf{n}_S^+|_{S_*}\big)$$
$$+ \big(\nabla_{\Gamma_*}U_*^-|_{S_*}\mathbf{n}_{S_*}^- - \nabla_\Gamma U^-|_{S_*}\mathbf{n}_S^-|_{S_*}\big),$$

regardless of Γ^i and Γ^j. Therefore, the last two terms in the right hand side can now be estimated using the geometric error estimates (4.8). Note that on S_*

$$\nabla_{\Gamma_*}U_*^\pm\,\mathbf{n}_{S_*}^\pm - \nabla_\Gamma U^\pm\,\mathbf{n}_S^\pm = \nabla_{\Gamma_*}(U_*^\pm - U^\pm)\,\mathbf{n}_{S_*}^\pm$$
$$+ (\nabla_{\Gamma_*} - \nabla_\Gamma)\,U^\pm\,\mathbf{n}_{S_*}^\pm + \nabla_\Gamma U^\pm\,\big(\mathbf{n}_{S_*}^\pm - \mathbf{n}_S^\pm\big) = I + II + III.$$

We bound each term using their parametric representation on $\widehat{S}_* := X_{\mathcal{T}_*}^{-1}(S_*)$. For the first term, we use the expression (3.11) of the tangential derivative in the co-normal direction, the spectral bounds on $\mathbf{G}_{\Gamma_*}$ and q_{Γ_*} given in Lemma 4.1, and a scaled trace estimate to deduce

$$\|\nabla_{\Gamma_*}(U_*^\pm - U^\pm)\mathbf{n}_{S_*}^\pm\|_{L_2(S_*)} \preccurlyeq |\widehat{T}_*^\pm|^{-\frac{1}{2d}}\|\widehat{\nabla}(\widehat{U}_*^\pm - \widehat{U}^\pm)\|_{L_2(\widehat{T}_*)}.$$

Recalling that $h_{T_*^\pm}^d = |\widehat{T}_*^\pm|$, we see that

$$\|I\|_{L_2(S_*)} \preccurlyeq h_{T_*^\pm}^{-1/2}\|\nabla_{\Gamma_*}(U_*^\pm - U^\pm)\|_{L_2(T_*^\pm)}.$$

23

Similarly, in view of (3.7) and (3.11), we obtain

$$\|II\|_{L_2(S_*)} + \|III\|_{L_2(S_*)} \lesssim h_{T_*^\pm}^{-1/2} \|\nabla_\Gamma U\|_{L_2(\widehat{T}_*^\pm)} \Big(\|\mathbf{D}_{\Gamma_*} - \mathbf{D}_\Gamma)\|_{L_\infty(\widehat{T}_*^\pm)}$$
$$+ \|q_{\Gamma_*} - q_\Gamma\|_{L_\infty(\widehat{T}_*^\pm)} + \|r_{\Gamma_*} - r_\Gamma\|_{L_\infty(\partial\widehat{T}_*^\pm)} \Big),$$

where r_{Γ_*} and r_Γ denote the area elements associated with S_* and $S' := X_\mathcal{T} \circ X_{\mathcal{T}_*}^{-1}(S_*)$ respectively. Utilizing the geometry error estimate (4.8), we further get

$$\|II\|_{L_2(S_*)} + \|III\|_{L_2(S_*)} \lesssim h_{T_*^\pm}^{-1/2} \lambda_\mathcal{T}(\gamma, [T^\pm]') \|\nabla_\Gamma U\|_{L_2(\widehat{T}_*^\pm)}.$$

Hence, combining the previous two estimates, we get

$$\|\nabla_{\Gamma_*} U_*^\pm \mathbf{n}_{S_*}^\pm - \nabla_\Gamma U^\pm \mathbf{n}_S^\pm\|_{L_2(S_*)}$$
$$\lesssim h_{T_*^\pm}^{-1/2} \left(\|\nabla_\gamma(U_*^\pm - U_*^\pm)\|_{L_2(T_*^\pm)} + \lambda_\mathcal{T}(\gamma, [T^\pm]') \|\nabla_\Gamma U\|_{L_2(\widehat{T}_*^\pm)} \right),$$

whence

$$\|\mathcal{J}_S(U)\|_{L_2(S_*)} = \|r_{\Gamma_*}^{1/2} \mathcal{J}_S(U)\|_{L_2(\widehat{S}_*)} \leq \|(r_{\Gamma_*}^{1/2} - r_\Gamma^{1/2})\mathcal{J}_S(U)\|_{L_2(\widehat{S}_*)} + \|\mathcal{J}_S(U)\|_{L_2(S')}.$$

Invoking again (4.8) we realize that $\|r_{\Gamma_*} - r_\Gamma\|_{L_\infty(\widehat{S})} \leq \lambda_\mathcal{T}(\gamma, [T^\pm]')$. Combining this with a scaled trace theorem, we deduce that

$$\|(r_{\Gamma_*}^{1/2} - r_\Gamma^{1/2})\mathcal{J}_S(U)\|_{L_2(\widehat{S}_*)} \lesssim h_{T_*^\pm}^{-1/2} \lambda_\mathcal{T}(\gamma, [T^\pm]') \|\nabla_\Gamma U\|_{L_2(T_*^\pm)}$$

whence

$$\|\mathcal{J}_S(U)\|_{L_2(S_*)} \leq \|\mathcal{J}_S(U)\|_{L_2(S')} + C h_{T_*^\pm}^{-1/2} \lambda_\mathcal{T}(\gamma, [T^\pm]') \|\nabla_\Gamma U\|_{L_2(T_*^\pm)}.$$

The above three estimates guarantee the existence of a constant C only depending on the shape regularity of $\mathbb{T}$ and the Lipschitz constant L of γ such that

$$h_{T_*^\pm}^{1/2} \|\mathcal{J}_{S_*}(U_*)\|_{L_2(S_*)} \leq h_{T_*^\pm}^{1/2} \|\mathcal{J}_{S'}(U)\|_{L_2(S')}$$
$$+ C \left(\|\nabla_\gamma(U_* - U)\|_{L_2(T^\pm)} + \lambda_\mathcal{T}(\gamma, [T^\pm]') \|\nabla_\Gamma U\|_{L_2(T_*^\pm)} \right). \tag{4.25}$$

To conclude the proof we proceed as for graphs [25, Lemma 4.2], basically squaring (4.24) and (4.25) via Young's inequality, adding over all elements $T_* \in \mathcal{T}_*$ and sides $S_* \in \mathcal{S}_*$, and using the strict reduction (4.10) of meshsize h_T for all refined elements. In addition, we employ the global bound $\|\nabla_\Gamma U\|_{L_2(\Gamma)} \lesssim \|f\|_{L_2(\gamma)}$. $\square$

Another difference with the theory of adaptivity for flat domains is the behavior of data oscillation under refinement. The usual situation is that $\mathrm{osc}_\mathcal{T}(U, f)$ does not increase upon refinement from $\mathcal{T}$ to $\mathcal{T}_*$ [27]. This is no longer true because $\mathrm{osc}_\mathcal{T}(U, f)$ and $\mathrm{osc}_{\mathcal{T}_*}(U_*, f)$ correspond to different domains Γ and Γ_*. We state a quasi-monotonicity property in Lemma 4.10 but omit its proof because it is similar and somewhat simpler than that of Lemma 4.9.

LEMMA 4.10 (Quasi-monotonicity of data oscillation). *Let* $\lambda_{\mathcal{T}_0}(\gamma)$ *satisfy* (4.6). *Let* $(\mathcal{T}, \Gamma)$, $(\mathcal{T}_*, \Gamma_*)$ *be mesh-surface pairs with* $\mathcal{T} \leq \mathcal{T}_*$. *Then, there exist constant* C_6, Λ_2 *and* Λ_3 *depending only on* $\mathcal{T}_0$, *the Lipschitz constant* L *of* γ, *and* $\|f\|_{L_2(\gamma)}$, *such that*

$$\mathrm{osc}_{\mathcal{T}_*}(V_*, f)^2 \leq C_6 \, \mathrm{osc}_\mathcal{T}(V, f)^2 + \Lambda_3 \|\nabla_\gamma(V_* - V)\|_{L_2(\gamma)}^2 + \Lambda_2 \lambda_\mathcal{T}(\gamma)^2. \tag{4.26}$$

24

REMARK 4.11 (Local perturbation of data oscillation). The previous result is also valid locally, that is for any subset $\tau \subset \mathcal{T}_*$. In fact, if $\tau = \mathcal{T} \cap \mathcal{T}_*$ the same proof gives (4.26) with $C_6 = 2$,

$$\text{osc}_{\mathcal{T}_*}(V, f, \tau)^2 \le 2 \, \text{osc}_{\mathcal{T}}(W, f, \tau)^2 + \Lambda_3 \|\nabla_\gamma (V - W)\|_{L_2(\gamma)}^2 + \Lambda_2 \lambda_{\mathcal{T}}(\gamma)^2, \qquad (4.27)$$

for any piecewise polynomials V, W subordinate to τ. Although the elements in τ describe (part of) the common surface $\Gamma \cap \Gamma_*$, whence there is no geometric discrepancy, the presence of the geometric estimator $\lambda_{\mathcal{T}}(\gamma)$ in (4.27) is due to the boundary of this common region. Note that the contribution to the oscillation associated to a side on the boundary of τ involves the terms $q_\Gamma^\pm$ according to (4.14).

5. AFEM: Design and Properties. Since $\lambda_{\mathcal{T}}(\gamma)$ and $\eta_{\mathcal{T}}(U, F_\Gamma)$ account for quite different effects, following [7], the algorithm AFEM is designed to handle them separately via the modules ADAPT_SURFACE and ADAPT_PDE.

> **AFEM:** Given $\mathcal{T}_0$ and parameters $\varepsilon_0 > 0$, $0 < \rho < 1$, and $\omega > 0$, set $k = 0$.
> 1. $\mathcal{T}_k^+ = \text{ADAPT\_SURFACE}(\mathcal{T}_k, \omega \varepsilon_k)$
> 2. $[U_{k+1}, \mathcal{T}_{k+1}] = \text{ADAPT\_PDE}(\mathcal{T}_k^+, \varepsilon_k)$
> 3. $\varepsilon_{k+1} = \rho \varepsilon_k$; $k = k + 1$
> 4. go to 1.

We notice the presence of the factor ω , which is employed to make the geometric error small relative to the current tolerance ε_k, thereby controlling the interactions between the geometry and the PDE. This turns out to be essential for both contraction and optimality of AFEM, even for polynomial degree $n = 1$ as discussed in [6].

5.1. Module ADAPT_SURFACE. Given a tolerance $\varepsilon > 0$ and an admissible subdivision $\mathcal{T}$, $\mathcal{T}_* = \text{ADAPT\_SURFACE}(\mathcal{T}, \varepsilon)$ improves the surface resolution until the new subdivision $\mathcal{T}_* \ge \mathcal{T}$ satisfies

$$\lambda_{\mathcal{T}_*}(\gamma) \le \varepsilon, \qquad (5.1)$$

where $\lambda_{\mathcal{T}}(\gamma)$ is the *geometric estimator* introduced in (1.2). This module is based on a *greedy* algorithm and acts on a generic mesh $\mathcal{T} = \cup_{i=1}^M \mathcal{T}^i \in \mathbb{T}$:

> $\mathcal{T}_* = \text{ADAPT\_SURFACE}(\mathcal{T}, \varepsilon)$
> 1. if $\mathcal{M} := \{T \in \mathcal{T} : \lambda_{\mathcal{T}}(\gamma, T) > \varepsilon\} = \emptyset$
>   return$(\mathcal{T})$ and exit
> 2. $\mathcal{T} = \text{REFINE}(\mathcal{T}, \mathcal{M})$
> 3. go to 1.

where $\text{REFINE}(\mathcal{T}, \mathcal{M})$ refines all elements in the marked set $\mathcal{M}$ and keeps conformity; more details are given in §5.2. To derive convergence rates for AFEM, we require that ADAPT_SURFACE is *t-optimal*, i.e. there exists a constant $C(\gamma)$ such that the set $\mathcal{M}$ of all the elements marked for refinement in a call to $\text{ADAPT\_SURFACE}(\mathcal{T}, \varepsilon)$ satisfies

$$\#\mathcal{M} \le C(\gamma) \, \varepsilon^{-1/t}. \qquad (5.2)$$

In § 7.3 we show that this assumption is satisfied by a greedy algorithm provided that $\chi^i \in B_q^{1+td}(L_q(\Omega))$ with $tq > 1$, $0 < q \le \infty$ and $td \le n$ for all $1 \le i \le M$.

25

5.2. Module ADAPT_PDE. Given a tolerance $\varepsilon > 0$ and an admissible subdivision $\mathcal{T} \in \mathbb{T}$, $[U_*, \mathcal{T}_*] = \mathsf{ADAPT\_PDE}(\mathcal{T}, \varepsilon)$ outputs a refinement $\mathcal{T}_* \geq \mathcal{T}$ and the associated finite element solution $U_* \in \mathbb{V}(\mathcal{T}_*)$ such that

$$\eta_{\mathcal{T}_*}(U_*, F_{\Gamma_*}) \leq \varepsilon. \tag{5.3}$$

This module is based on the standard adaptive sequence:

$[U_*, \mathcal{T}_*] = \mathsf{ADAPT\_PDE}(\mathcal{T}, \varepsilon)$
    1. $U = \mathsf{SOLVE}(\mathcal{T})$
    2. $\{\eta_{\mathcal{T}}(U, F_{\Gamma}, T)\}_{T \in \mathcal{T}} = \mathsf{ESTIMATE}(\mathcal{T}, U)$
    3. if $\eta_{\mathcal{T}}(U, F_{\Gamma}) < \varepsilon$
        return$(\mathcal{T}, U)$ and exit
    4. $\mathcal{M} = \mathsf{MARK}(\mathcal{T}, \{\eta_{\mathcal{T}}(U, F_{\Gamma}, T)\}_{T \in \mathcal{T}})$
    5. $\mathcal{T} = \mathsf{REFINE}(\mathcal{T}, \mathcal{M})$
    6. go to 1

We describe below the modules SOLVE, ESTIMATE, MARK and REFINE separately.

Procedure SOLVE. This procedure solves the SPD linear system resulting for (3.16) where we recall that $\Gamma = \Gamma_{\mathcal{T}}$. For simplicity we assume that (3.16) is solved exactly with linear complexity. We refer to [24] for a hierachical basis multigrid preconditioner and to [9] for standard variational and non-variational multigrid algorithms.

Procedure ESTIMATE. Given the Galerkin solution $U \in \mathbb{V}(\mathcal{T})$ of (3.16) the procedure ESTIMATE computes the PDE error indicators $\{\eta_{\mathcal{T}}(U, F_{\Gamma}, T)\}_{T \in \mathcal{T}}$. We emphasize that this procedure does not compute the oscillation terms, which are only needed to carry out the analysis.

Lemma 4.8 (equivalence of error and estimator) is critical to deduce that the ADAPT_PDE module, which reduces the error indicators $\eta_{\mathcal{T}}(U, F_{\Gamma})$, is successful in reducing the PDE error $\mathcal{E}_{\mathcal{T}}(U, f)$ of (4.19) provided the parameter ω satisfies

$$\omega \leq \omega_1 := \sqrt{\frac{C_2}{2\Lambda_0^2 \Lambda_1}}. \tag{5.4}$$

In fact, given a tolerance $\varepsilon > 0$ to be reached by ADAPT_PDE starting from the input subdivision $\mathcal{T}$ satisfying $\lambda_{\mathcal{T}}(\gamma) \leq \omega\varepsilon$, we observe that (2.14) guarantees that $\mathcal{T}$ as well as all subdivisions $\mathcal{T}_* \geq \mathcal{T}$ constructed within the inner iterates of ADAPT_PDE satisfy

$$\lambda_{\mathcal{T}_*}(\gamma)^2 \leq \Lambda_0^2 \lambda_{\mathcal{T}}(\gamma)^2 \leq \frac{C_2}{2\Lambda_1}\varepsilon^2.$$

Within the while loop of ADAPT_PDE we have $\eta_{\mathcal{T}}(U, F_{\Gamma}) > \varepsilon$, so we deduce the validity of (4.20) whence that of (4.21) within that loop.

Procedure MARK. We rely on an optimal *Dörfler's* marking strategy for the selection of elements. Given the set of indicators $\{\eta_{\mathcal{T}}(U, F_{\Gamma}, T)\}_{T \in \mathcal{T}}$ and a marking parameter $\theta \in (0, 1]$, MARK outputs a subset of marked elements $\mathcal{M} \subset \mathcal{T}$ such that

$$\eta_{\mathcal{T}}(U, F_{\Gamma}, \mathcal{M}) \geq \theta\eta_{\mathcal{T}}(U, F_{\Gamma}). \tag{5.5}$$

In contrast to [25], MARK only employs the error indicators and does not use either the oscillation or surface indicators. We will see that quasi-optimal cardinality requires that $\mathcal{M}$ is *minimal* and quasi-optimal workload that the sorting scales linearly.

Procedure REFINE. Given a subdivision $\mathcal{T}$ and a set $\mathcal{M} \subset \mathcal{T}$ of marked elements, the call REFINE$(\mathcal{T}, \mathcal{M})$ bisects all elements in $\mathcal{M}$ at least $b \geq 1$ times and performs additional refinements necessary to maintain conformity. The resulting subdivision is denoted by $\mathcal{T}_*$. Recall that the bisection procedure is first executed on faces of the corresponding flat subdivision $\overline{\mathcal{T}}$ and its effect is transferred to the actual subdivision via interpolation maps $X^i_{\mathcal{T}^i} \circ (X^i_0)^{-1}$ for $i = 1, ..., M$.

Since the refinement procedure is performed on $\overline{\mathcal{T}}$ or similarly on $\widehat{\mathcal{T}}$, the complexity results of the overall refinement algorithm proved by Binev, Dahmen, and DeVore for $d = 2$ [3] and Stevenson [31] for $d > 2$ hold in our setting. In order to state them precisely, following [3, 31, 28], we need the concept of admissible labeling [3, 31].

REMARK 5.1 (Admissible labeling). For $d = 1$, any subdivision is said to have an admissible labeling. For $d = 2$, we say that $\mathcal{T}_0$ has an admissible labeling if each edge of $\mathcal{T}_0$ is labeled either 0 or 1 such that each element of $\mathcal{T}_0$ has exactly two edges with label 1 and one with label 0 [3]; refining an element entails connecting the middle of the edge labeled 0 with the opposite angle. For $d > 2$, the corresponding condition (b) of §4 in [31] is much more technical and is omitted here. In short, an admissible initial labeling guarantees that the bisection procedure terminates in finite steps with a conforming mesh, and that any uniform refinement of $\mathcal{T}_0$ is conforming.

LEMMA 5.2 (Complexity of REFINE). *Assume that the initial triangulation $\mathcal{T}_0$ has an admissible labeling. Let $\{\mathcal{T}_k\}_{k \geq 0}$ be a sequence of triangulations produced by successive calls to $\mathcal{T}_{k+1} = $ REFINE$(\mathcal{T}_k, \mathcal{M}_k)$, where $\mathcal{M}_k$ is any subset of $\mathcal{T}_k$, $k \geq 0$. Then, there exists a constant C_7 solely depending on $\mathcal{T}_0$, its labeling, and the refinement depth b such that*

$$\# \mathcal{T}_k - \# \mathcal{T}_0 \leq C_7 \sum_{j=0}^{k-1} \# \mathcal{M}_j, \qquad \forall k \geq 1. \tag{5.6}$$

It is worth noticing that the user parameter $b \geq 1$ can be chosen equal to one, which only implies a minimal refinement, and does not force an interior node property [27, 26] or an extra refinement to improve the surface approximation [25].

REMARK 5.3 (Alternative subdivision strategies). For simplicity we only discuss the refinement strategy based on simplex bisection. However, all the results obtained here can be extended to quadrilaterals with fixed number of hanging nodes or red refinements. We refer to [8, Section 6] for details.

6. Conditional Contraction Property. The procedure ADAPT_PDE is known to yield a contraction property in the *flat* case. In the present context, however, the surface approximation is responsible for lack of consistency in that the sequence of finite element spaces is no longer nested. This in turn leads to failure of a key orthogonality property between discrete solutions, the Pythagoras property. We have, instead, a perturbation result referred to as *quasi-orthogonality* below. Its proof follows the steps of that for graphs [25, Lemma 4.4]. In this section, we use the notation

$$e^j := \|\nabla_\gamma (u - U^j)\|_{L_2(\gamma)}, \quad E^j := \|\nabla_\gamma (U^{j+1} - U^j)\|_{L_2(\gamma)},$$
$$\eta^j := \eta_{\mathcal{T}^j}(U^j, F^j), \quad \eta^j(\mathcal{M}^j) := \eta_{\mathcal{T}^j}(U^j, F^j, \mathcal{M}^j), \quad \lambda^j := \lambda_{\mathcal{T}^j}(\gamma),$$

where $\mathcal{T}^j$ are meshes obtained after each inner iteration of ADAPT_PDE, starting with $\mathcal{T}^0 = \mathcal{T}$, $\mathcal{M}^j \subset \mathcal{T}^j$ are the subsets of elements selected by the marking procedure, F^j are the scaled right hand sides defined in (3.17) with Γ replaced by Γ^j, the surface

associated to $\mathcal{T}^j$, and $U^j \in \mathbb{V}(\mathcal{T}^j)$ are the corresponding Galerkin solutions.

LEMMA 6.1 (Quasi-orthogonality). *There exists a constant $\Lambda_2 > 0$ solely depending on the Lipschitz constant L of γ and $\|f\|_{L_2(\gamma)}$ such that for $i = j, j+1$ with $j \geq 0$, we have*

$$(e^j)^2 - \frac{3}{2}(E^j)^2 - \Lambda_2(\lambda^i)^2 \leq (e^{j+1})^2 \leq (e^j)^2 - \frac{1}{2}(E^j)^2 + \Lambda_2(\lambda^i)^2. \qquad (6.1)$$

Before proceeding with the proof of the above lemma, we point out that the constant Λ_2 was already defined in Lemma 4.9. This is to simplify the notations below and is without loss of generality (upon redefining Λ_2 as the maximum of the two constants).

Proof. Since the symmetry of the Dirichlet form implies

$$(e^j)^2 = (e^{j+1})^2 + (E^j)^2 + 2\int_\gamma \nabla_\gamma(u - U^{j+1})\nabla_\gamma^T(U^{j+1} - U^j),$$

we just have to examine the last term. Utilizing the error representation (4.13) with $v = U^{j+1} - U^j$ and realizing that $I_1 = I_3 = 0$, we readily obtain

$$\int_\gamma \nabla_\gamma(u - U^{j+1})\nabla_\gamma^T(U^{j+1} - U^j) = \int_\gamma \nabla_\gamma U^{j+1}\mathbf{E}_{\Gamma^{j+1}}\nabla_\gamma^T(U^{j+1} - U^j).$$

Invoking (4.7) yields

$$\left|\int_\gamma \nabla_\gamma(u - U^{j+1})\nabla_\gamma^T(U^{j+1} - U^j)\right| \preccurlyeq \|f\|_{L_2(\gamma)}\lambda^{j+1}E^j.$$

This leads to (6.1) after applying Young's inequality and using (2.14). $\square$

Notice that relation (6.1) also holds when (i) $\mathcal{T}^j$, $\mathcal{T}^{j+1}$ are replaced with $\mathcal{T}$, $\mathcal{T}^*$ satisfying $\mathcal{T}^* \geq \mathcal{T}$; (ii) U^{j+1} is replaced by $U^* \in \mathbb{V}(\mathcal{T}^*)$ and (iii) U^j is replaced by any $V \in \mathbb{V}(\mathcal{T})$ because V need not be the Galerkin solution over $\mathcal{T}$. The parameter

$$\omega_2 := \frac{\xi\theta^2}{\Lambda_0\sqrt{32\Lambda_2(2\Lambda_3 + 1)}}, \qquad (6.2)$$

where $\xi := 1 - 2^{-b/d}$ is defined in Lemma 4.9, is used subsequently as a threshold for the AFEM parameter ω.

THEOREM 6.2 (Conditional contraction property). *Let $\theta \in (0, 1]$ be the marking parameter of MARK and let $\{\mathcal{T}^j, U^j\}_{j=0}^J$ be a sequence of meshes and discrete solutions produced by one call to procedure ADAPT_PDE$(\mathcal{T}^0, \varepsilon)$ inside AFEM, i.e., $\lambda^0 := \lambda_{\mathcal{T}^0}(\gamma) \leq \omega\varepsilon$. Assume that the AFEM parameter ω satisfies*

$$\omega \leq \min\{\omega_1, \omega_2\},$$

where ω_1 and ω_2 are given in (5.4) and (6.2), respectively. Then there exist constants $0 < \alpha < 1$ and $\beta > 0$ such that

$$(e^{j+1})^2 + \beta(\eta^{j+1})^2 \leq \alpha^2\left((e^j)^2 + \beta(\eta^j)^2\right) \qquad \forall\, 0 \leq j < J. \qquad (6.3)$$

Moreover, the number of inner iterates J of ADAPT_PDE is uniformly bounded.

Proof. We proceed in four steps. Note that $\eta^j \geq \varepsilon$ for $0 \leq j < J$.

[1] Let $\beta > 0$ be a scaling parameter to be found later. We combine (6.1) and (4.23) to write

$$(e^{j+1})^2 + \beta(\eta^{j+1})^2 \leq (e^j)^2 + \left(-\frac{1}{2} + \beta(1+\delta^{-1})\Lambda_3 \right)(E^j)^2$$
$$+ \Lambda_2\left(1 + \beta(1+\delta^{-1}) \right)(\lambda^j)^2 + \beta(1+\delta)\left((\eta^j)^2 - \xi\eta^j(\mathcal{M}^j)^2 \right),$$

where $\mathcal{M}^j$ is the set of elements marked for refinement at the j-th subiteration. To remove the factor of E^j we now choose β dependent on δ, to be

$$\beta(1+\delta^{-1})\Lambda_3 = \frac{1}{2} \quad \Rightarrow \quad \beta(1+\delta) = \frac{\delta}{2\Lambda_3}, \tag{6.4}$$

and thereby obtain

$$(e^{j+1})^2 + \beta(\eta^{j+1})^2 \leq (e^j)^2 + \Lambda_2\left(1 + \frac{1}{2\Lambda_3} \right)(\lambda^j)^2 + \beta(1+\delta)\left((\eta^j)^2 - \xi\eta^j(\mathcal{M}^j)^2 \right).$$

[2] Invoking Dörfler marking (5.5), we deduce

$$(\eta^j)^2 - \xi\eta^j(\mathcal{M}^j)^2 \leq (1 - \xi\theta^2)(\eta^j)^2.$$

Since the initial mesh $\mathcal{T}^0$ comes from ADAPT_SURFACE we know that $\lambda^0 \leq \omega\varepsilon \leq \omega\eta^j$ for $1 \leq j < J$. Using (2.14) yields $\lambda^j \leq \Lambda_0\omega\eta^j$, whence

$$(e^{j+1})^2 + \beta(\eta^{j+1})^2 \leq (e^j)^2 - \beta(1+\delta)\frac{\xi\theta^2}{2}(\eta^j)^2$$
$$+ \beta\left((1+\delta)\left(1 - \frac{\xi\theta^2}{2} \right) + \Lambda_2\left(1 + \frac{1}{2\Lambda_3} \right)\frac{\Lambda_0^2\omega^2}{\beta} \right)(\eta^j)^2.$$

Moreover, $\omega \leq \omega_1$ implies $(\lambda^j)^2 \leq \frac{C_2}{2\Lambda_1}(\eta^j)^2$ which turns out to be (4.20). Therefore, applying the bound (4.22) and replacing β according to (6.4), we obtain

$$(e^{j+1})^2 + \beta(\eta^{j+1})^2 \leq \alpha_1(\delta)(e^j)^2 + \alpha_2(\delta)\beta(\eta^j)^2$$

with

$$\alpha_1(\delta)^2 := 1 - \delta\frac{\xi\theta^2}{4\Lambda_3(C_1 + \frac{C_2}{2})}, \quad \alpha_2(\delta)^2 := (1+\delta)\left(1 - \frac{\xi\theta^2}{2} \right) + \Lambda_2\left(1 + \frac{1}{2\Lambda_3} \right)\frac{\Lambda_0^2\omega^2}{\beta}.$$

[3] It remains to prove that δ can be chosen so that $\alpha_2(\delta)^2 < 1$. We then fix the parameter δ so that

$$(1+\delta)\left(1 - \frac{\xi\theta^2}{2} \right) = 1 - \frac{\xi\theta^2}{4} \quad \Rightarrow \quad \delta = \frac{\xi\theta^2}{4 - 2\xi\theta^2},$$

We now realize that (6.4) gives $\beta = \frac{\xi\theta^2}{2\Lambda_3(4-\xi\theta^2)} \geq \frac{\xi\theta^2}{8\Lambda_3}$ and, since $\omega \leq \omega_2$, we infer that

$$\Lambda_2\left(1 + \frac{1}{2\Lambda_3} \right)\frac{\Lambda_0^2\omega^2}{\beta} \leq \frac{4\Lambda_2(2\Lambda_3 + 1)}{\xi\theta^2}\Lambda_0^2\omega^2 \leq \frac{\xi\theta^2}{8}.$$

Hence $\alpha_2^2 \leq 1 - \frac{\xi\theta^2}{8} < 1$, and choosing $\alpha := \max\{\alpha_1, \alpha_2\} < 1$ yields (6.3).

4 The contraction property (6.3) guarantees that ADAPT_PDE stops in a finite number of iterations J. To show that J is independent of the outer iteration counter k, take $k \geq 1$ and note that before the call $\mathsf{ADAPT\_PDE}(\mathcal{T}_k^+, \varepsilon_k)$ in AFEM of § 5, we have

$$\eta_k := \eta_{\mathcal{T}_k}(U_k, F_k) \leq \varepsilon_{k-1} = \frac{\varepsilon_k}{\rho}, \qquad \lambda_k := \lambda_{\mathcal{T}_k}(\gamma) \leq \Lambda_0 \lambda_{\mathcal{T}_{k-1}^+}(\gamma) \leq \frac{\Lambda_0 \omega}{\rho} \varepsilon_k.$$

We next combine (4.23), with $\delta = 1$, and the estimate $\|\nabla_\gamma(U_k^+ - U_k)\|_{L_2(\gamma)}^2 \precsim \|\nabla_\gamma(u - U_k)\|_{L_2(\gamma)}^2 + \lambda_k^2$ arising from (6.1), to get

$$\eta_{\mathcal{T}_k^+}(U_k^+, F_k^+)^2 \precsim \eta_k^2 + \lambda_k^2 + \|\nabla_\gamma(U_k^+ - U_k)\|_{L_2(\gamma)}^2 \precsim \eta_k^2 + \lambda_k^2 + \|\nabla_\gamma(u - U_k)\|_{L_2(\gamma)}^2,$$

where F_k^+ is the right hand side associated to $\Gamma_{\mathcal{T}_k^+}$ defined in (3.17) and the hidden constants depend on Λ_2, Λ_3. The bounds on η_k, λ_k, together with (4.15), yield

$$(\eta^0)^2 = \eta_{\mathcal{T}_k^+}(U_k^+, F_k^+)^2 \precsim \eta_k^2 + \lambda_k^2 \precsim \varepsilon_k^2.$$

Since the stopping condition of ADAPT_PDE is $\eta^J \leq \varepsilon_k$, (6.3) implies that J is bounded independently of k, as asserted. □

The fact that J is uniformly bounded controls the complexity of ADAPT_PDE because the most expensive module SOLVE is run just J times. However, this property is not required for the study of cardinality of §8.

7. Approximation Class. In this section we discuss the approximation classes $\mathbb{A}_s$ and their connection with Besov regularity. We start with the notion of total error in §7.1 leading to the definition of $\mathbb{A}_s$. We then introduce and discuss a *greedy* algorithm in §7.2, that we use repeatedly in the rest of the section. We study the best approximation error achievable with piecewise polynomials of degree $n \geq 1$ for the surface γ in §7.3 and for the solution u in §7.4. We analyze the decay rate of oscillation in §7.5. Finally in §7.6 we conclude with our second main result: the membership $(u, f, \gamma) \in \mathbb{A}_s$ in terms of Besov regularity of u, f and γ.

In the discussion below we also remove the hat on functions, when no confusion is possible.

7.1. The Total Error. Let $\mathbb{T}_N \subset \mathbb{T} := \mathbb{T}(\mathcal{T}_0)$ be the set of all possible conforming triangulations, generated on γ with at most N elements more than $\mathcal{T}_0$ by successive bisection of $\mathcal{T}_0$:

$$\mathbb{T}_N := \big\{\mathcal{T} \in \mathbb{T} \mid \#\mathcal{T} - \#\mathcal{T}_0 \leq N\big\}.$$

Given $v \in H_\#^1(\gamma), f \in L_2(\gamma)$ and $V \in \mathbb{V}(\mathcal{T})$, we recall the notion of *total error*

$$E_{\mathcal{T}}(V; v, f, \gamma)^2 = \|\nabla_\gamma(v - V)\|_{L_2(\gamma)}^2 + \mathrm{osc}_{\mathcal{T}}(V, f)^2 + \omega^{-1}\lambda_{\mathcal{T}}(\gamma)^2,$$

or $E_{\mathcal{T}}(V; v, f, \gamma)^2 = \mathcal{E}_{\mathcal{T}}(V, f)^2 + \omega^{-1}\lambda_{\mathcal{T}}(\gamma)^2$. Owing to the equivalence of norms (4.9) we rewrite the first term in the parametric domain Ω, and omit the factor ω, to obtain the following equivalent notion of total error provided $\lambda_{\mathcal{T}_0}(\gamma)$ satisfies (4.6):

$$\widehat{E}_{\mathcal{T}}(V; v, f, \gamma)^2 := \sum_{T \in \mathcal{T}} \|\widehat{\nabla}(v - V)\|_{L_2(\widehat{T})}^2 + \mathrm{osc}_{\mathcal{T}}(V, f)^2 + \lambda_{\mathcal{T}}(\gamma)^2, \qquad (7.1)$$

Note that the last two terms are already evaluated in Ω according to definitions (2.13), (4.14). Yet, there is a nonlinear interaction between the approximations of γ and of u, f defined on γ. At this point, we recall the convention of dropping the patch index when no confusion arises, for example $v|_{\widehat{T}}$ for $\widehat{T} \in \widehat{\mathcal{T}}^i$ in (7.1) stands for $v^i|_{\widehat{T}}$.

Assuming that the parameter ω satisfies $\omega \leq \omega_1$ in (5.4) to guarantee the validity of Lemma 4.8 (equivalence of error and estimator), along with the fact that AFEM is driven by $\eta_{\mathcal{T}}(U, F_\Gamma)$ and $\lambda_{\mathcal{T}}(\gamma)$, we assess the quality of the best approximation of (v, f, γ) with N degrees of freedom in terms of the following modulus of smoothness:

$$\sigma(N; v, f, \gamma) := \inf_{\mathcal{T} \in \mathbb{T}_N} \inf_{V \in \mathbb{V}(\mathcal{T})} \widehat{E}_{\mathcal{T}}(V; v, f, \gamma).$$

This is thus consistent with the approach taken for flat domains in [12, 28]. For $s > 0$, we define the nonlinear (algebraic) approximation class $\mathbb{A}_s$ to be

$$\mathbb{A}_s := \Big\{ (v, f, \gamma) : |v, f, \gamma|_{\mathbb{A}_s} := \sup_{N \geq 1} \big(N^s\, \sigma(N; v, f, \gamma) \big) < \infty \Big\}. \tag{7.2}$$

The generic range of s is dictated by the polynomial degree, namely $0 < s \leq n/d$.

A useful and alternative definition to $(u, f, \gamma) \in \mathbb{A}_s$ follows: given $\varepsilon > 0$, there exists a subdivision $\mathcal{T}_\varepsilon$ with $\mathcal{T}_\varepsilon \geq \mathcal{T}_0$ and a discrete function $V_\varepsilon \in \mathbb{V}(\mathcal{T}_\varepsilon)$ such that

$$\widehat{E}_{\mathcal{T}_\epsilon}(V_\epsilon; u, f, \gamma) \leq \varepsilon, \qquad \text{and} \qquad \#\mathcal{T}_\varepsilon - \#\mathcal{T}_0 \leq |u, f, \gamma|_{\mathbb{A}_s}^{\frac{1}{s}} \varepsilon^{-\frac{1}{s}}. \tag{7.3}$$

The characterization of $\mathbb{A}_s$ in terms of Besov regularity is an open issue but we give sufficient conditions for membership in $\mathbb{A}_s$ in § 7.3 and § 7.4. Before doing so, we discuss in § 7.2 *greedy* algorithms suited to our particular framework, where the subdivisions consist of a collection of compatible subdivisions.

7.2. Greedy Algorithm. In this section we present and discuss a *greedy* algorithm to construct a *near best* piecewise polynomial approximation of a vector-valued function $\mathbf{g} := \{g^i\}_{i=1}^M : \Omega \to \mathbb{R}^M$ in a suitable semi-norm. Given a mesh $\mathcal{T} := \cup_{i=1}^M \mathcal{T}^i \in \mathbb{T}$, and a corresponding parametric mesh $\widehat{\mathcal{T}} := \cup_{i=1}^M \widehat{\mathcal{T}}^i \in \widehat{\mathbb{T}}$, the algorithm requires a *local* error estimator $\zeta_{\mathcal{T}^i}(g^i, T)$ for $T \in \mathcal{T}^i$, $1 \leq i \leq M$. To simplify the notations, we set

$$\zeta_{\mathcal{T}}(\mathbf{g}, T) := \zeta_{\mathcal{T}^i}(g^i, T), \quad T \in \mathcal{T}^i, \ 1 \leq i \leq M.$$

We emphasize at this point that approximating the functions g^i requires mesh compatibility conditions on $\partial\Omega$ to account for adjacent components $\mathcal{T}^i$ of the entire conforming mesh $\mathcal{T}$. This explains why we employ this somewhat cumbersome notation, which however we will simplify as much as possible below. Given a conforming refinement $\mathcal{T}$ of an initial triangulation $\mathcal{T}_0$ and a prescribed tolerance δ, the algorithm reads:

$\mathcal{T}_* = \mathsf{GREEDY}(\mathbf{g}, \mathcal{T}, \delta)$
    1. if $\mathcal{M} := \{T \in \mathcal{T} : \zeta_{\mathcal{T}}(\mathbf{g}, T) > \delta\} = \emptyset$
        return $(\mathcal{T})$ and exit
    2. $\mathcal{T} = \mathsf{REFINE}(\mathcal{T}, \mathcal{M})$
    3. go to 1

where the module REFINE bisects all elements in the marked set $\mathcal{M}$ and keeps conformity as described in §5.2. Note that $\mathsf{ADAPT\_SURFACE}$ is a particular instance of

31

GREEDY that uses $\zeta_{\mathcal{T}^i}(\chi^i, T) := \lambda_{\mathcal{T}^i}(\gamma, T)$ as local error estimator to approximate the patch of the surface γ^i parameterized by $\chi^i : \Omega \to \mathbb{R}^{d+1}$. We now discuss some properties of the greedy algorithm following [28]. Results of this type started with Birman and Solomyak for Sobolev spaces [5], and continued with [4] for Besov spaces and [15] for wavelet tree approximation. We do not refer to any specific norm below.

PROPOSITION 7.1 (Performance of GREEDY). *Let* $\mathcal{T} := \cup_{i=1}^M \mathcal{T}^i$ *be created by successive bisections of* $\mathcal{T}_0$, *which has an admissible labeling. Let* $0 < \mathtt{p} \le \infty$ *and let* $\mathbf{g} := \{g^i\}_{i=1}^M : \Omega \to \mathbb{R}^M$ *be a vector-valued function and* $\{\zeta_{\mathcal{T}}(\mathbf{g}, T)\}_{T \in \mathcal{T}}$ *be corresponding local error estimators that satisfy*

$$\zeta_{\mathcal{T}}(\mathbf{g}, T) \preccurlyeq h_T^r |g^i|_T, \quad r > 0, \quad T \in \mathcal{T}^i, \quad 1 \le i \le M, \tag{7.4}$$

where $h_T := |\widehat{T}|^{1/d}$ *and* $\left(\sum_{i=1}^M \sum_{T \in \mathcal{T}^i} |g^i|_T^{\mathtt{p}} \right)^{1/\mathtt{p}} \le |\mathbf{g}|_\Omega$ *is a given semi-(quasi) norm (with the convention that* $\max_{i=1,\dots,M} \max_{T \in \mathcal{T}^i} |g^i|_T \le |\mathbf{g}|_\Omega$ *if* $\mathtt{p} = \infty$*).*

If $|\mathbf{g}|_\Omega < \infty$, *then the module* GREEDY$(\mathbf{g}, \mathcal{T}, \delta)$ *terminates in a finite number of steps and the number of elements marked* $\mathcal{M}$ *within* GREEDY *satisfies*

$$\#\mathcal{M} \preccurlyeq |\mathbf{g}|_\Omega^{\frac{d\mathtt{p}}{d+r\mathtt{p}}} \delta^{-\frac{d\mathtt{p}}{d+r\mathtt{p}}}. \tag{7.5}$$

Proof. The algorithm stops in a finite number of steps because the local estimator $\zeta_{\mathcal{T}}(\mathbf{g}, T)$ is bounded by a positive power of h_T according to (7.4) and the admissibility of the labeling of $\mathcal{T}_0$ ensures that a finite number of refinements is required to guarantee the conformity of $\mathcal{T}$. To prove (7.5) we organize the elements in $\mathcal{M}$ by size so that it allows for a counting argument. Let $\mathcal{P}_j$ be the set of elements T of $\mathcal{M}$ with size (in the parametric domain) satisfying $2^{-(j+1)} \le |\widehat{T}| < 2^{-j}$, so that

$$T \in \mathcal{P}_j \iff 2^{-(j+1)} \le |\widehat{T}| < 2^{-j} \iff 2^{-(j+1)/d} \le h_T < 2^{-j/d}.$$

We assume that $\mathcal{T} = \mathcal{T}_0$ and proceed in several steps.

$\boxed{1}$ We first observe that all T's in $\mathcal{P}_j$ are *disjoint*. In fact, if $T_1, T_2 \in \mathcal{P}_j$ and they overlap (their interiors have a nonempty intersection), then one of them is contained in the other, say $T_1 \subset T_2$, due to the bisection procedure, thus $|\widehat{T}_1| \le \frac{1}{2} |\widehat{T}_2|$, contradicting the definition of $\mathcal{P}_j$. Then, recalling that we have M copies of Ω, we deduce

$$2^{-(j+1)} \#\mathcal{P}_j \le M|\Omega| \implies \#\mathcal{P}_j \le M|\Omega| 2^{j+1}. \tag{7.6}$$

$\boxed{2}$ We note that $\mathcal{P}_j = \cup_{i=1}^M \mathcal{P}_j^i$ where $\mathcal{P}_j^i$ contains the elements of $\mathcal{P}_j$ which are refinements of elements in $\mathcal{T}_0^i$. Each element $T \in \mathcal{P}_j$ belongs to a subdivision $\mathcal{T}$ created by GREEDY so that, in light of (7.4) and the fact that $\mathcal{P}_j \subset \mathcal{M}$, we have

$$T \in \mathcal{P}_j^i \implies \delta \le \zeta_{\mathcal{T}}(\mathbf{g}, T) \preccurlyeq 2^{-(j/d)r} |g^i|_T.$$

Therefore we have $\delta^{\mathtt{p}} \#\mathcal{P}_j \preccurlyeq 2^{-(j/d)r\mathtt{p}} \sum_{i=1}^M \sum_{T \in \mathcal{P}_j^i} |g^i|_{\widehat{T}}^{\mathtt{p}}$ and

$$\#\mathcal{P}_j \preccurlyeq \delta^{-\mathtt{p}} 2^{-(j/d)r\mathtt{p}} |\mathbf{g}|_\Omega^{\mathtt{p}}. \tag{7.7}$$

$\boxed{3}$ The two bounds for $\#\mathcal{P}_j$ in (7.6) and (7.7) are complementary. The first one is good for j small whereas the second one is suitable for j large (think of $\delta \ll 1$). The crossover takes place for j_0 such that

$$2^{j_0+1} M|\Omega| \approx \delta^{-\mathbf{P}} 2^{-j_0 r \mathbf{P}/d} |\mathbf{g}|_\Omega^{\mathbf{P}} \quad \Longleftrightarrow \quad 2^{j_0} \approx \left(M^{-1}|\Omega|^{-1}\delta^{-\mathbf{P}}|\mathbf{g}|_\Omega^{\mathbf{P}}\right)^{\frac{d}{d+r\mathbf{P}}}.$$

$\boxed{4}$ We now compute

$$\#\mathcal{M} = \sum_j \#\mathcal{P}_j \preccurlyeq \sum_{j \leq j_0} 2^j + \delta^{-\mathbf{P}} |\mathbf{g}|_\Omega^{\mathbf{P}} \sum_{j > j_0} (2^{-r\mathbf{P}/d})^j.$$

Since $\sum_{j \leq j_0} 2^j \approx 2^{j_0}$ and $\sum_{j > j_0} (2^{-r\mathbf{P}/d})^j \preccurlyeq 2^{-(r\mathbf{P}/d)j_0}$, we can write

$$\#\mathcal{M} \preccurlyeq \left(|\mathbf{g}|_\Omega \delta^{-1}\right)^{\frac{d\mathbf{P}}{d+r\mathbf{P}}},$$

which is the desired estimate.

$\boxed{5}$ It remains to remove the simplifying assumption $\mathcal{T} = \mathcal{T}_0$. Since $\mathcal{T}$ is a conforming refinement of $\mathcal{T}_0$, [7, Proposition 2] shows that the number of elements marked by $\mathsf{GREEDY}(\mathbf{g}, \mathcal{T}, \delta)$ does not exceed those marked by $\mathsf{GREEDY}(\mathbf{g}, \mathcal{T}_0, \delta)$ and estimated in step 4. This concludes the proof. $\square$

We consider the estimator

$$\zeta_\mathcal{T}(\mathbf{g}) := \{\zeta_\mathcal{T}(\mathbf{g}, T)\}_{T \in \mathcal{T}}$$

and its accumulation in $\ell_\mathbf{q}$ with $0 < \mathbf{p} < \mathbf{q} \leq \infty$. Its decay rate is assessed next.

COROLLARY 7.2 (Estimate in ℓ_q). *Let $\zeta_\mathcal{T}(\mathbf{g})$ satisfy (7.4) with $r := d(s - 1/\mathbf{p} + 1/\mathbf{q}) > 0$. Let the initial subdivision $\mathcal{T}_0$ have an admissible labeling. Given $\delta > 0$ there exists a conforming mesh refinement $\mathcal{T} \in \mathbb{T}$ such that*

$$\|\zeta_\mathcal{T}(\mathbf{g})\|_{\ell_\mathbf{q}} \preccurlyeq \delta, \qquad \#\mathcal{T} - \#\mathcal{T}_0 \preccurlyeq \#\mathcal{M} \preccurlyeq |\mathbf{g}|_\Omega^{1/s} \delta^{-1/s}. \tag{7.8}$$

Proof. Since $\frac{d\mathbf{p}}{d+r\mathbf{p}} = \frac{\mathbf{q}}{1+\mathbf{q}s}$, the output of the call $\mathcal{T} = \mathsf{GREEDY}(\mathbf{g}, \mathcal{T}_0, \epsilon)$ satisfies

$$\zeta_\mathcal{T}(\mathbf{g}, T) \preccurlyeq \epsilon, \qquad \#\mathcal{M} \preccurlyeq |\mathbf{g}|_\Omega^{\frac{\mathbf{q}}{1+\mathbf{q}s}} \epsilon^{-\frac{\mathbf{q}}{1+\mathbf{q}s}}, \qquad \forall T \in \mathcal{T},$$

for any $\epsilon > 0$ according to Proposition 7.1. Combining this with the complexity estimate (5.6) readily implies

$$\|\zeta_\mathcal{T}(\mathbf{g})\|_{\ell_\mathbf{q}} \preccurlyeq \#\mathcal{T}^{1/\mathbf{q}} \epsilon \preccurlyeq \#\mathcal{M}^{1/\mathbf{q}} \epsilon \preccurlyeq \epsilon^{\frac{\mathbf{q}s}{1+\mathbf{q}s}} |\mathbf{g}|_\Omega^{\frac{1}{1+\mathbf{q}s}}.$$

If ϵ satisfies $\delta = \epsilon^{\frac{\mathbf{q}s}{1+\mathbf{q}s}} |\mathbf{g}|_\Omega^{\frac{1}{1+\mathbf{q}s}}$, then it is easy to see that (7.8) is valid. $\square$

REMARK 7.3 (Scale invariant error estimates). The interpolation estimate (7.4) will be derived and utilized below for functions g which are products or composition of functions to account for geometry. To illustrate the importance of scale invariance, we consider now the product $g = vw$ with $v, w \in W_p^2(\widehat{T})$ and $2 > d/p$ ($1 \leq p \leq \infty$); this implies $L_\infty(\widehat{T}), W_{2p}^1(\widehat{T}) \subset W_p^2(\widehat{T})$ and $g \in W_p^2(\widehat{T})$. Therefore $|g|_T := |g|_{W_p^2(\widehat{T})}$ satisfies

$$|g|_{W_p^2(\widehat{T})} \preccurlyeq |v|_{W_p^2(\widehat{T})} \|w\|_{L_\infty(\widehat{T})} + |v|_{W_{2p}^1(\widehat{T})} |w|_{W_{2p}^1(\widehat{T})} + \|v\|_{L_\infty(\widehat{T})} |w|_{W_p^2(\widehat{T})}.$$

33

This scale invariant expression accumulates correctly in ℓ_p,

$$\sum_{T \in \mathcal{T}} |g|^p_{W^2_p(\widehat{T})} \precsim |v|^p_{W^2_p(\Omega)} \|w\|^p_{L_\infty(\Omega)} + |v|^p_{W^1_{2p}(\Omega)} |w|^p_{W^1_{2p}(\Omega)} + \|v\|^p_{L_\infty(\Omega)} |w|^p_{W^2_p(\Omega)}$$

as a consequence of applying the Cauchy-Schwarz inequality to the middle term. In §9 we extend estimates for the product and composition of functions to Besov spaces $B^s_q(L_p(\widehat{T}))$ with any order of differentiability s and integrability p, provided $s > d/p$.

7.3. Constructive Approximation of γ. We now analyze a constructive approximation of γ by piecewise polynomials based on the GREEDY algorithm. We also show that this algorithm, and hence ADAPT_SURFACE, is t-optimal, i.e. the set of marked elements satisfies (1.4), provided that γ belongs to a suitable Besov space. The case of polynomial degree $n = 1$ with regularity of γ in terms of Sobolev scales is discussed in [6]. We establish here a result for higher order degree $n \geq 1$ for which the regularity of γ must be measured in Besov scales.

We recall the following compact notation (2.4):

$$\boldsymbol{\chi} := \{\chi^i\}^M_{i=1}, \quad |\boldsymbol{\chi}|_{B^{1+td}_q(L_q(\Omega))} := \max_{i=1,\dots,M} |\chi^i|_{B^{1+td}_q(L_q(\Omega))}.$$

We also note that the superscript i indicating the patch label is dropped when no confusion arises.

COROLLARY 7.4 (Constructive approximation of γ). *Let γ be globally of class W^1_∞ and be parameterized by $\boldsymbol{\chi} \in [B^{1+td}_q(L_q(\Omega))]^M$ with $tq > 1, 0 < q \leq \infty, td \leq n$. Let $\mathcal{T}_0$ have an admissible labeling. Then $\mathcal{T}_* = \text{ADAPT\_SURFACE}(\mathcal{T}, \delta)$ is t-optimal, i.e.*

$$\lambda_{\mathcal{T}_*}(\gamma) \leq \delta, \qquad \#\mathcal{M} \precsim C_1(\gamma)^{1/t} \delta^{-1/t},$$

where $\mathcal{M}$ denotes the number of elements marked during the execution of the procedure $\text{ADAPT\_SURFACE}(\mathcal{T}, \delta)$ and $C_1(\gamma) \leq |\boldsymbol{\chi}|_{B^{1+td}_q(L_q(\Omega))}$ is the constant in (1.4).

Proof. Observe that $B^{1+td}_q(L_q(\widehat{T}))$, with $tq > 1$ and $0 < q \leq \infty$ is just above the nonlinear Sobolev scale of W^1_∞ in dimension d [32, p. 482], [23, Lemma 4.12], so that

$$B^{1+td}_q(L_q(\widehat{T})) \subset B^1_\infty(L_\infty(\widehat{T})) \subset W^1_\infty(\widehat{T}).$$

Therefore, a scaling argument and local interpolation estimates give the following bound with $r = dt - d/q > 0$ and $T \in \mathcal{T}$

$$\lambda_{\mathcal{T}}(\gamma, T) = \|\widehat{\nabla}(\chi - X_{\mathcal{T}})\|_{L_\infty(\widehat{T})} \precsim h^r_T |\chi|_{B^{1+td}_q(L_q(\widehat{T}))}. \tag{7.9}$$

We then can apply Proposition 7.1 with $\mathrm{p} = q$, $\mathbf{g} = \boldsymbol{\chi}$, $|\boldsymbol{\chi}|_\Omega = |\boldsymbol{\chi}|_{B^{1+td}_q(L_q(\Omega))}$, $r = dt - d/q$ and $\zeta_{\mathcal{T}}(\mathbf{g}, T) = \lambda_{\mathcal{T}}(\gamma, T)$. Upon termination of $\mathcal{T}_* = \text{ADAPT\_SURFACE}(\mathcal{T}, \delta)$, we obtain $\lambda_{\mathcal{T}_*}(\gamma) \leq \delta$ along with the asserted estimate on $\#\mathcal{M}$ because $\frac{dq}{d+rq} = \frac{1}{t}$. $\square$

7.4. Constructive Approximation of u. We use the vector notation (2.4)

$$\mathbf{u} := \{u^i\}^M_{i=1}, \quad |\mathbf{u}|_{B^{1+sd}_p(L_p(\Omega))} := \max_{i=1,\dots,M} |u^i|_{B^{1+sd}_p(L_p(\Omega))},$$

where $u^i := u|_{\gamma^i} \circ \chi^i$ and γ^i is the i-th surface patch, along with $\mathbf{V} = \{V^i\}^M_{i=1}$ and

$$\|\widehat{\nabla}(\mathbf{u} - \mathbf{V})\|^2_{L^2(\Omega)} := \sum_{i=1}^M \|\widehat{\nabla}(u^i - V^i)\|^2_{L^2(\Omega)}.$$

34

COROLLARY 7.5 (Constructive approximation of u). *Let $u \in H^1_\#(\gamma)$ be piecewise of class $B^{1+sd}_p(L_p(\Omega))$, namely $\mathbf{u} \in [B^{1+sd}_p(L_p(\Omega))]^M$, with $s - 1/p + 1/2 > 0$, $0 < p \leq \infty$ and $0 < sd \leq n$. Let $\mathcal{T}_0$ have an admissible labeling. Then, given $\delta > 0$ there exists a triangulation $\mathcal{T} \in \mathbb{T}$ such that*

$$\inf_{\mathbf{V} \in \mathbb{V}(\mathcal{T})} \|\widehat{\nabla}(\mathbf{u} - \mathbf{V})\|_{L^2(\Omega)} \lesssim \delta, \qquad \#\mathcal{M} \lesssim C(\mathbf{u})^{1/s} \delta^{-1/s}, \tag{7.10}$$

where $\mathcal{M}$ is the set of marked elements to create $\mathcal{T}$ and $C(\mathbf{u}) = |\mathbf{u}|_{B^{1+sd}_p(L_p(\Omega))}$.

Proof. Taking $\mathbf{g} := \widehat{\nabla}\mathbf{u} \in B^{sd}_p(L_p(\Omega))$ and applying Corollary 7.2 with $\mathsf{q} := 2$ we obtain the desired estimate provided we employ discontinuous piecewise polynomials of degree $\leq n$ over $\widehat{\mathcal{T}}$. We finally resort to [33], which shows that the error decay is in fact the same regardless of continuity for approximation of globally $H^1(\Omega)$-functions. This takes care also of continuity of traces across patches. $\square$

7.5. Decay Rate of Oscillation. In order to study the decay rate of the oscillation $\mathrm{osc}_{\mathcal{T}}(U, f)^2$, we split it into two terms that we analyze separately, namely

$$\mathrm{osc}_{\mathcal{T}}(U, f)^2 \leq \mathrm{osc}_{\mathcal{T}}(U)^2 + \mathrm{osc}_{\mathcal{T}}(f)^2, \tag{7.11}$$

where

$$\mathrm{osc}_{\mathcal{T}}(U)^2 := \sum_{T \in \mathcal{T}} \mathrm{osc}_{\mathcal{T}}(U, T)^2, \qquad \mathrm{osc}_{\mathcal{T}}(f)^2 := \sum_{T \in \mathcal{T}} \mathrm{osc}_{\mathcal{T}}(f, T)^2, \tag{7.12}$$

and for $T \in \mathcal{T}$ and $V \in \mathbb{V}(\mathcal{T})$

$$\mathrm{osc}_{\mathcal{T}}(V, T)^2 := h_T^2 \left\| (\mathrm{id} - \Pi^2_{2n-2}) \widehat{\mathrm{div}}\, (q_\Gamma \widehat{\nabla} V \mathbf{G}_\Gamma^{-1}) \right\|^2_{L_2(\widehat{T})} \tag{7.13}$$

$$+ h_T \left\| (\mathrm{id} - \Pi^2_{2n-1}) \big(q_\Gamma^+ \widehat{\nabla} V^+ (\mathbf{G}_\Gamma^+)^{-1} \widehat{\mathbf{n}}^+ - q_\Gamma^- \widehat{\nabla} V^- (\mathbf{G}_\Gamma^-)^{-1} \widehat{\mathbf{n}}^- \big) \right\|^2_{L_2(\partial\widehat{T})}$$

$$\mathrm{osc}_{\mathcal{T}}(f, T)^2 := h_T^2 \left\| (\mathrm{id} - \Pi^2_{2n-2})(fq) \right\|^2_{L_2(\widehat{T})}. \tag{7.14}$$

To assess their decay rate, we resort to the following bound [12, Lemma 3.2]:

$$\|(\mathrm{id} - \Pi^2_m)(vV)\|_{L_2(\omega)} \leq \|(\mathrm{id} - \Pi^\infty_{m-n})v\|_{L_\infty(\omega)} \|V\|_{L_2(\omega)}, \tag{7.15}$$

which is valid for $0 \leq n \leq m$, any domain ω of $\mathbb{R}^d$ or $\mathbb{R}^{d-1}$, $V \in \mathbb{P}_n(\omega)$ and $v \in L_\infty(\omega)$. In fact, since Π^2_m is invariant over $\mathbb{P}_m$, we see that $(\mathrm{id} - \Pi^2_m)(\Pi^\infty_{m-n} vV) = 0$ whence

$$(\mathrm{id} - \Pi^2_m)(vV) = (\mathrm{id} - \Pi^2_m)[(\mathrm{id} - \Pi^\infty_{m-n})vV].$$

This yields (7.15) for any interpolant $\Pi^\infty_{m-n} v$ via the L^2-stability of Π^2_m.

We now embark on the study of the decay rate of oscillation: we investigate $\mathrm{osc}_{\mathcal{T}}(f)$ in § 7.5.1 and $\mathrm{osc}_{\mathcal{T}}(U)$ in § 7.5.2.

7.5.1. Decay Rate of $\mathrm{osc}_{\mathcal{T}}(f)$. We employ the vector notation (2.4) with $\mathbf{f} := \{f^i\}_{i=1}^M$ and $\mathbf{q} = \{q^i\}_{i=1}^M$. We also recall that the superscript i indicating the patch label is omitted when no confusion arises.

COROLLARY 7.6 (Decay rate of $\mathrm{osc}_{\mathcal{T}}(f)$). *Let γ be globally of class W^1_∞ and be parameterized by $\boldsymbol{\chi} \in [B^{1+td}_q(L_q(\Omega))]^M$ with $tq > 1$, $0 < q \leq \infty$, $td \leq n$, and let $k =$*

$\lfloor td \rfloor + 1$. Let $f \in L_2(\gamma)$ be piecewise of class $B_p^{sd}(L_p(\Omega))$, namely $\mathbf{f} \in [B_p^{sd}(L_p(\Omega))]^M$, with $s - 1/p + 1/2 > 0, 0 < p \leq \infty$ and $sd \leq n$. Let $\mathcal{T}_0$ have an admissible labeling. Then, given $\delta > 0$ there exists a triangulation $\mathcal{T} \in \mathbb{T}$ such that

$$\mathrm{osc}_{\mathcal{T}}(f) \preccurlyeq \delta, \qquad \#\mathcal{T} - \#\mathcal{T}_0 \preccurlyeq C(\mathbf{f}, \gamma)^{\frac{1}{s \wedge t + 1/d}} \delta^{-\frac{1}{s \wedge t + 1/d}}, \tag{7.16}$$

where $s \wedge t := \min\{s, t\}$ and

$$C(\mathbf{f}, \gamma) := \|\mathbf{f}\|_{B_p^{sd}(L_p(\Omega))} \left(\|\boldsymbol{\chi}\|_{B_q^{1+td}(L_q(\Omega))} + \|\boldsymbol{\chi}\|^k_{B_q^{1+td}(L_q(\Omega))} \right).$$

Proof. Since $\Pi^2_{2n-2}(fq) \in \mathbb{P}_{2n-2}$ is the best L_2-approximation of fq, we see that

$$\|fq - \Pi^2_{2n-2}(fq)\|_{L_2(\widehat{T})} \leq \|fq - \Pi^2_{2n-2}(Vq)\|_{L_2(\widehat{T})}$$
$$\leq \|(f - V)q\|_{L_2(\widehat{T})} + \|Vq - \Pi^2_{2n-2}(Vq)\|_{L_2(\widehat{T})}$$
$$\leq \|f - V\|_{L_2(\widehat{T})}\|q\|_{L_\infty(\widehat{T})} + \|q - \Pi^\infty_{n-1}q\|_{L_\infty(\widehat{T})}\|V\|_{L_2(\widehat{T})},$$

for all $V \in \mathbb{P}_{n-1}$ due to (7.15). Taking $V = \Pi^2_{n-1}f$ we have $\|V\|_{L_2(\widehat{T})} \leq \|f\|_{L_2(\widehat{T})}$ and

$$\mathrm{osc}_{\mathcal{T}}(f, T) \leq h_T \|f - \Pi^2_{n-1}f\|_{L_2(\widehat{T})}\|q\|_{L_\infty(\widehat{T})} + h_T \|q - \Pi^\infty_{n-1}q\|_{L_\infty(\widehat{T})}\|f\|_{L_2(\widehat{T})}.$$

We now introduce for each $T \in \mathcal{T}$

$$E_{2,\mathcal{T}}(\mathbf{f}, T) := h_T \|f^i - \Pi^2_{n-1}f^i\|_{L_2(\widehat{T})}, \quad E_{\infty,\mathcal{T}}(\mathbf{q}, T) := h_T \|q^i - \Pi^\infty_{n-1}q^i\|_{L_\infty(\widehat{T})} \tag{7.17}$$

and notice that an immediate generalization of [23, Lemma 4.15] implies the local error estimates

$$E_{2,\mathcal{T}}(\mathbf{f}, T) \preccurlyeq h_T^{r_2}|f|_{B_p^{sd}(L_p(\widehat{T}))}, \qquad E_{\infty,\mathcal{T}}(\mathbf{q}, T) \preccurlyeq h_T^{r_\infty}|q|_{B_q^{td}(L_q(\widehat{T}))}$$

with $r_2 = d[(s + 1/d) - 1/p + 1/2] > 1$ and $r_\infty = d[(t + 1/d) - 1/q] > 1$. Moreover, Corollary 9.5 yields $\mathbf{q} \in [B_q^{td}(L_q(\Omega))]^M$ because $\widehat{\nabla}\boldsymbol{\chi} \in [B_q^{td}(L_q(\Omega))]^M$ and

$$\|\mathbf{q}\|_{L_\infty(\Omega)} \preccurlyeq \|\mathbf{q}\|_{B_q^{td}(L_q(\Omega))} \preccurlyeq \max\left\{ \|\boldsymbol{\chi}\|_{B_q^{1+td}(L_q(\Omega))}, \|\boldsymbol{\chi}\|^k_{B_q^{1+td}(L_q(\Omega))} \right\}, \tag{7.18}$$

with $k = \lfloor td \rfloor + 1$. Given a constant $c_2 > 0$ to be determined later, we resort to Corollary 7.2 with tolerance $c_2\delta$, the local indicator $E_{2,\mathcal{T}}(\mathbf{f}, T)$, $\mathsf{q} = 2$ and $r = r_2$, to obtain a mesh $\mathcal{T}_2 \in \mathbb{T}$ that satisfies

$$E_{2,\mathcal{T}_2}(\mathbf{f}) \preccurlyeq c_2\delta, \qquad \#\mathcal{T}_2 - \#\mathcal{T}_0 \preccurlyeq |\mathbf{f}|^{\frac{1}{s+1/d}}_{B_p^{sd}(L_p(\Omega))}(c_2\delta)^{-\frac{1}{s+1/d}}. \tag{7.19}$$

Invoking Corollary 7.2 once again, this time with tolerance $c_\infty\delta$, constant c_∞ to be chosen later, local indicator $E_{\infty,\mathcal{T}}(\mathbf{q}, T)$, $\mathsf{q} = \infty$ and $r = r_\infty$, we find a mesh $\mathcal{T}_\infty \in \mathbb{T}$ such that

$$E_{\infty,\mathcal{T}_\infty}(\mathbf{q}) \preccurlyeq c_\infty\delta, \qquad \#\mathcal{T}_\infty - \#\mathcal{T}_0 \preccurlyeq |\mathbf{q}|^{\frac{1}{t+1/d}}_{B_q^{td}(L_q(\Omega))}(c_\infty\delta)^{-\frac{1}{t+1/d}}. \tag{7.20}$$

If $\mathcal{T} = \mathcal{T}_2 \oplus \mathcal{T}_\infty$ is the overlay of the meshes $\mathcal{T}_2$ and $\mathcal{T}_\infty$, then it remains to show that $\mathcal{T}$ satisfies (7.16). Since the local indicators (7.17) are monotone, i.e they do not increase with refinement, we deduce from (7.19) and (7.20)

$$\mathrm{osc}_{\mathcal{T}}(f)^2 \preccurlyeq E^2_{2,\mathcal{T}}(\mathbf{f}) \|\mathbf{q}\|^2_{L_\infty(\Omega)} + E^2_{\infty,\mathcal{T}}(\mathbf{q}) \|\mathbf{f}\|^2_{L_2(\Omega)}$$
$$\preccurlyeq \delta^2 \left(c_2^2 \|\mathbf{q}\|^2_{L_\infty(\Omega)} + c_\infty^2 \|\mathbf{f}\|^2_{L_2(\Omega)} \right).$$

We now choose the constants c_2 and c_∞ as follows:

$$c_2 = \|\mathbf{q}\|_{B_q^{td}(L_q(\Omega))}^{-1}, \qquad c_\infty = \|\mathbf{f}\|_{B_p^{sd}(L_p(\Omega))}^{-1}.$$

This implies $\mathrm{osc}_{\mathcal{T}}(f) \preccurlyeq \delta$ in view of (7.18) and $\|\mathbf{f}\|_{L_2(\Omega)} \preccurlyeq \|\mathbf{f}\|_{B_p^{sd}(L_p(\Omega))}$. Finally, since $\#\mathcal{T} \leq \#\mathcal{T}_2 + \#\mathcal{T}_\infty - \#\mathcal{T}_0$ according to [12, Lemma 3.7], we obtain

$$\#\mathcal{T} - \#\mathcal{T}_0 \leq (\#\mathcal{T}_2 - \#\mathcal{T}_0) + (\#\mathcal{T}_\infty - \#\mathcal{T}_0) \preccurlyeq C(\mathbf{f}, \gamma)\, \delta^{-\frac{1}{s \wedge t + 1/d}}.$$

This follows from (7.19) and (7.20) upon replacing the exponents s and t by $s \wedge t$ because their left-hand sides are always larger than or equal to 1, and eventually using Corollary 9.5 to estimate $\|\mathbf{q}\|_{B_q^{td}(L_q(\Omega))}$. This concludes the proof. $\square$

REMARK 7.7 (Besov regularity of f). If $u^i \in B_p^{1+sd}(L_p(\Omega))$ and γ is smooth, then $f^i = -q^{-1}\widehat{\mathrm{div}}(q\widehat{\nabla}u^i\mathbf{G}^{-1}) \in B_p^{sd-1}(L_p(\Omega))$ is the natural Besov regularity for f^i. However, we require that $f^i \in B_p^{sd}(L_p(\Omega))$ because the data oscillation is evaluated in $L_2(\Omega)$ rather than $H^{-1}(\Omega)$. This additional degree of regularity of f is responsible for the faster decay of $\mathrm{osc}_{\mathcal{T}}(f)$ reported in Corollary 7.6.

7.5.2. Decay Rate of $\mathrm{osc}_{\mathcal{T}}(U)$**.** In this section we study the decay rates of $\mathrm{osc}_{\mathcal{T}}(U)$ defined in (7.13). We again use the GREEDY algorithm where now the local indicator will be $\mathrm{osc}_{\mathcal{T}}(V, T)$ for $V \in \mathbb{V}(\mathcal{T})$.

We start with an estimate for $\mathrm{osc}_{\mathcal{T}}(V, T)$ in terms of a positive power of h_T. In view of expression (7.13) for $\mathrm{osc}_{\mathcal{T}}(V, T)$, the major non-standard obstruction is the presence of the surface dependent and non-polynomial term $q_\Gamma \mathbf{G}_\Gamma^{-1}$. This requires two auxiliary results about Besov spaces, namely Corollary 9.2 (scale-invariant Besov semi-norm of products of functions) and Lemma 9.4 (scale-invariant Besov norm of composition), which we prove later in § 9 not to interrupt the flow.

We are now in position to show that $\mathrm{osc}_{\mathcal{T}}(V, T)$ is bounded by a positive power of h_T. The proof of Proposition 7.8 is a consequence of the subsequent three lemmas.

PROPOSITION 7.8 (Local decay of oscillation). *Let the surface γ be globally of class W_∞^1 and be parameterized by $\boldsymbol{\chi} \in [B_q^{1+td}(L_q(\Omega))]^M$ with $tq > 1$, $0 < q \leq \infty$, $td \leq n$. For all $\mathcal{T} \in \mathbb{T}$ and all $V \in \mathbb{V}(\mathcal{T})$*

$$\mathrm{osc}_{\mathcal{T}}(V, T) \preccurlyeq h_T^r \left(\sum_{j=1, \ell_j = j/k^k}^{k^k} |\boldsymbol{\chi}|_{B_q^{1+td\ell_j}(L_{q/\ell_j}(N_{\mathcal{T}}(\widehat{T}));\mathcal{T})}^{1/\ell_j} \right) \|\widehat{\nabla}V\|_{L_2(N_{\mathcal{T}}(\widehat{T}))}, \quad (7.21)$$

with $r = td - d/q > 0$ and $k = \lfloor td \rfloor + 1$, $N_{\mathcal{T}}(\widehat{T})$ is the set containing $\widehat{T}$ and its adjacent elements, and $B_q^{1+td\ell_j}(L_{q/\ell_j}(N_{\mathcal{T}}(\widehat{T}));\mathcal{T})$ indicates the broken Besov space.

LEMMA 7.9 (Element oscillation). *Under the assumptions of Proposition 7.8, for all $V \in \mathbb{V}(\mathcal{T})$ and all $T \in \mathcal{T}$, we have*

$$h_T \left\| (id - \Pi_{2n-2}^2)\widehat{\mathrm{div}}(q_\Gamma \widehat{\nabla}V\mathbf{G}_\Gamma^{-1}) \right\|_{L_2(\widehat{T})} \preccurlyeq h_T^r |q_\Gamma \mathbf{G}_\Gamma^{-1}|_{B_q^{td}(L_q(\widehat{T}))} \|\widehat{\nabla}V\|_{L_2(\widehat{T})},$$

with $r = td - d/q > 0$.

Proof. We first observe that

$$\widehat{\mathrm{div}}(q_\Gamma \widehat{\nabla}V\mathbf{G}_\Gamma^{-1}) = \widehat{\mathrm{div}}\left(q_\Gamma \mathbf{G}_\Gamma^{-1}\right) \cdot \widehat{\nabla}V + q_\Gamma \mathbf{G}_\Gamma^{-1} : \widehat{D}^2 V,$$

and by (7.15)

$$\left\| (\mathrm{id} - \Pi^2_{2n-2})\widehat{\mathrm{div}}(q_\Gamma \widehat{\nabla} V \mathbf{G}_\Gamma^{-1}) \right\|_{L_2(\widehat{T})} \preccurlyeq \left\| (\mathrm{id} - \Pi^\infty_{n-1})\widehat{\mathrm{div}}\left(q_\Gamma \mathbf{G}_\Gamma^{-1}\right) \right\|_{L_\infty(\widehat{T})} \left\| \widehat{\nabla} V \right\|_{L_2(\widehat{T})}$$
$$+ \left\| (\mathrm{id} - \Pi^\infty_{n-1})(q_\Gamma \mathbf{G}_\Gamma^{-1}) \right\|_{L_\infty(\widehat{T})} \left\| \widehat{D}^2 V \right\|_{L_2(\widehat{T})}.$$

Using interpolation estimates in Besov norms of an immediate generalization of [23, Lemma 4.15] we have

$$\left\| (\mathrm{id} - \Pi^\infty_{n-1})\widehat{\mathrm{div}}\left(q_\Gamma \mathbf{G}_\Gamma^{-1}\right) \right\|_{L_\infty(\widehat{T})} \lesssim h_T^r \left| \widehat{\mathrm{div}}\left(q_\Gamma \mathbf{G}_\Gamma^{-1}\right) \right|_{B_q^{td}(L_q(\widehat{T}))},$$

with $0 < r \leq n$. By the inverse inequality of Lemma 4.4, we readily get

$$\left| \widehat{\mathrm{div}}\left(q_\Gamma \mathbf{G}_\Gamma^{-1}\right) \right|_{B_q^{td}(L_q(\widehat{T}))} \lesssim \frac{1}{h_T} \left| q_\Gamma \mathbf{G}_\Gamma^{-1} \right|_{B_q^{td}(L_q(\widehat{T}))}.$$

A similar argument applied to the second term gives

$$\left\| (\mathrm{id} - \Pi^\infty_{n-1})(q_\Gamma \mathbf{G}_\Gamma^{-1}) \right\|_{L_\infty(\widehat{T})} \left\| \widehat{D}^2 V \right\|_{L_2(\widehat{T})} \lesssim h_T^r \left| q_\Gamma \mathbf{G}_\Gamma^{-1} \right|_{B_q^{td}(L_q(\widehat{T}))} \frac{\left\| \widehat{\nabla} V \right\|_{L_2(\widehat{T})}}{h_T}.$$

This proves the asserted estimate. $\square$

LEMMA 7.10 (Jump oscillation). *Let the assumptions of Proposition 7.8 be valid. For all $T \in \mathcal{T}$ and all $V \in \mathbb{V}(\mathcal{T})$, there holds*

$$h_T^{1/2} \left\| (\mathrm{id} - \Pi^2_{2n-1})\left(q_\Gamma^+ \widehat{\nabla} V^+ (\mathbf{G}_\Gamma^+)^{-1} \widehat{\mathbf{n}}^+ - q_\Gamma^- \widehat{\nabla} V^- (\mathbf{G}_\Gamma^-)^{-1} \widehat{\mathbf{n}}^- \right) \right\|_{L_2(\partial \widehat{T})}$$
$$\preccurlyeq h_T^r | q_\Gamma \mathbf{G}_\Gamma^{-1} |_{B_q^{td}(L_q(N_T(\widehat{T}));\mathcal{T})} \| \widehat{\nabla} V \|_{L_2(N_T(\widehat{T}))},$$

with $r = td - d/q > 0$.

Proof. Let $\widehat{S} = \widehat{T}^+ \cap \widehat{T}^-$ be any side of $\widehat{T}^+ := \widehat{T}$. Since

$$h_T^{1/2} \left\| (\mathrm{id} - \Pi^2_{2n-1})\left(q_\Gamma^+ \widehat{\nabla} V^+ (\mathbf{G}_\Gamma^+)^{-1} \widehat{\mathbf{n}}^+ - q_\Gamma^- \widehat{\nabla} V^- (\mathbf{G}_\Gamma^-)^{-1} \widehat{\mathbf{n}}^- \right) \right\|_{L_2(\widehat{S})}$$
$$\preccurlyeq h_T^{1/2} \left\| (\mathrm{id} - \Pi^2_{2n-1})\left(q_\Gamma^+ \widehat{\nabla} V^+ (\mathbf{G}_\Gamma^+)^{-1} \widehat{\mathbf{n}}^+ \right) \right\|_{L_2(\widehat{S})}$$
$$+ h_T^{1/2} \left\| (\mathrm{id} - \Pi^2_{2n-1})\left(q_\Gamma^- \widehat{\nabla} V^- (\mathbf{G}_\Gamma^-)^{-1} \widehat{\mathbf{n}}^- \right) \right\|_{L_2(\widehat{S})},$$

we estimate each term separately, dropping the $\pm$ superscript. We invoke (7.15) to deduce that

$$\left\| (\mathrm{id} - \Pi^2_{2n-1})\left(q_\Gamma \widehat{\nabla} V \mathbf{G}_\Gamma^{-1} \widehat{\mathbf{n}} \right) \right\|_{L_2(\widehat{S})} \leq \left\| (\mathrm{id} - \Pi^\infty_{n-1})\left(q_\Gamma \mathbf{G}_\Gamma^{-1} \widehat{\mathbf{n}} \right) \right\|_{L_\infty(\widehat{S})} \| \widehat{\nabla} V \|_{L_2(\widehat{S})},$$

where Π^∞_{n-1} is the Lagrange interpolation operator onto $\mathbb{P}_{n-1}(\widehat{S})$. Since the unit normal $\widehat{\mathbf{n}}$ is constant on $\widehat{S}$, the interpolation estimate from an immediate generalization of [23, Lemma 4.15] reveals that

$$\| (\mathrm{id} - \Pi^\infty_{n-1})\left(q_\Gamma \mathbf{G}_\Gamma^{-1} \widehat{\mathbf{n}} \right) \|_{L_\infty(\widehat{S})} \preccurlyeq h_T^r | q_\Gamma \mathbf{G}_\Gamma^{-1} |_{B_q^{td}(L_q(\widehat{T}^\pm))},$$

where we have used the assumption $r \leq n$. This together with a scaled trace estimate $\| \widehat{\nabla} V \|_{L_2(\widehat{S})} \preccurlyeq h_T^{-1/2} \| \widehat{\nabla} V \|_{L_2(\widehat{T}^\pm)}$ yields the desired estimate. $\square$

We see from Lemmas 7.9 and 7.10 that the discrete surface Γ enters the estimates via $|q_\Gamma \mathbf{G}_\Gamma^{-1}|_{B_q^{td}(L_q(\widehat{T}))}$. The next lemma provides control of this term.

LEMMA 7.11 (Besov semi-norm of $q_\Gamma \mathbf{G}_\Gamma^{-1}$). *Let the assumptions of Proposition 7.8 hold and $k = \lfloor td \rfloor + 1$. We then have for $T \in \mathcal{T}^i$, $1 \le i \le M$,*

$$|q_\Gamma \mathbf{G}_\Gamma^{-1}|_{B_q^{td}(L_q(\widehat{T}))} \precsim \sum_{j=1, \ell_j = j/k^k}^{k^k} |\chi^i|_{B_q^{1+td\ell_j}(L_{q/\ell_j}(\widehat{T}))}^{1/\ell_j}.$$

Proof. We fix i and drop it from the notation throughout this proof. We invoke Lemma 9.1 (scale-invariant Besov semi-norm of the product of two functions) along with the Hölder inequality to write

$$|q_\Gamma \mathbf{G}_\Gamma^{-1}|_{B_q^{td}(L_q(\widehat{T}))} \precsim \|q_\Gamma\|_{L_\infty(\widehat{T})} |\mathbf{G}_\Gamma^{-1}|_{B_q^{td}(L_q(\widehat{T}))}$$

$$+ \sum_{\substack{m=1 \\ \ell_m := m/k}}^{k-1} |q_\Gamma|_{B_q^{td\ell_m}(L_{q/\ell_m}(\widehat{T}))}^{1/\ell_m} + |\mathbf{G}_\Gamma^{-1}|_{B_q^{td(1-\ell_m)}(L_{q/(1-\ell_m)}(\widehat{T}))}^{1/(1-\ell_m)}$$

$$+ |q_\Gamma|_{B_q^{td}(L_q(\widehat{T}))} \|\mathbf{G}_\Gamma^{-1}\|_{L_\infty(\widehat{T})}$$

$$\precsim \sum_{\substack{m=1 \\ \ell_m := m/k}}^{k} |q_\Gamma|_{B_q^{td\ell_m}(L_{q/\ell_m}(\widehat{T}))}^{1/\ell_m} + |\mathbf{G}_\Gamma^{-1}|_{B_q^{td\ell_m}(L_{q/\ell_m}(\widehat{T}))}^{1/\ell_m},$$

because $\|q_\Gamma\|_{L_\infty(\widehat{T})}, \|\mathbf{G}_\Gamma^{-1}\|_{L_\infty(\widehat{T})} \precsim 1$ for γ being globally Lipschitz. We denote

$$s^* = td\ell_m, \quad q^* = q/\ell_m, \quad 1 \le m \le k,$$

and bound $|q_\Gamma|_{B_q^{s^*}(L_{q^*}(\widehat{T}))}$ using Lemma 9.4 for $q_\Gamma = \left(\det(\widehat{\nabla}X_\mathcal{T}^T \widehat{\nabla} X_\mathcal{T})\right)^{1/2}$:

$$|q_\Gamma|_{B_q^{s^*}(L_{q^*}(\widehat{T}))} \precsim \sum_{\ell=1}^{k} \sum_{i=1}^{\ell} \|\widehat{\nabla}X_\mathcal{T}\|_{L_\infty(\widehat{T})}^{\ell-i} \sum_{\sum_{j=1}^{i}\ell_j^*=1} \prod_{j=1}^{i} |\widehat{\nabla}X_\mathcal{T}|_{B_q^{s^*\ell_j^*}(L_{q^*/\ell_j^*}(\widehat{T}))};$$

a similar bound is valid for $\mathbf{G}_\Gamma^{-1}$. To get a simpler expression, we observe that $\sum_{j=1}^{i} \ell_j^* = 1$ with $0 \le \ell_j^* = i_j/k^{i-1} \le 1$ and $i_j \in \mathbb{N}_0$, whence there cannot be more than $i-1$ vanishing ℓ_j^*'s in each product $\prod_{j=1}^{i} |\widehat{\nabla}X_\mathcal{T}|_{B_q^{s^*\ell_j^*}(L_{q^*/\ell_j^*}(\widehat{T}))}$. Therefore

$$\prod_{j=1}^{i} |\widehat{\nabla}X_\mathcal{T}|_{B_q^{s^*\ell_j^*}(L_{q^*/\ell_j^*}(\widehat{T}))} \le \max\left(1, \|\widehat{\nabla}X_\mathcal{T}\|_{L_\infty(\widehat{T})}^{i-1}\right) \prod_{\substack{j=1 \\ \ell_j^* \ne 0}}^{i} |\widehat{\nabla}X_\mathcal{T}|_{B_q^{s^*\ell_j^*}(L_{q^*/\ell_j^*}(\widehat{T}))},$$

because $|\widehat{\nabla}X_\mathcal{T}|_{B_\infty^0(L_\infty(\widehat{T}))} = \|\widehat{\nabla}X_\mathcal{T}\|_{L_\infty(\widehat{T})}$. Now, using $i \le \ell \le k$, we obtain

$$|q_\Gamma|_{B_q^{s^*}(L_{q^*}(\widehat{T}))} \precsim \max\left(1, \|\widehat{\nabla}X_\mathcal{T}\|_{L_\infty(\widehat{T})}^{k-1}\right) \sum_{\ell=1}^{k} \sum_{i=1}^{\ell} \sum_{\sum_{j=1}^{i}\ell_j^*=1} \prod_{\substack{j=1 \\ \ell_j^* \ne 0}}^{i} |\widehat{\nabla}X_\mathcal{T}|_{B_q^{s^*\ell_j^*}(L_{q^*/\ell_j^*}(\widehat{T}))},$$

and remove the first factor in light of (2.9) and χ being globally Lipschitz. Since $\sum_{j=1}^{i} \ell_j^* = 1$ and $\ell_j^* > 0$, we employ Hölder's inequality to estimate each product as

$$\prod_{\substack{j=1 \\ \ell_j^* \neq 0}}^{i} |\widehat{\nabla} X_\mathcal{T}|_{B_q^{s^* \ell_j^*}(L_{q^*/\ell_j^*}(\widehat{T}))} \precsim \sum_{\substack{j=1 \\ \ell_j^* \neq 0}}^{i} |\widehat{\nabla} X_\mathcal{T}|_{B_q^{s^* \ell_j^*}(L_{q^*/\ell_j^*}(\widehat{T}))}^{1/\ell_j^*}.$$

Combining this with the preceding expression, and taking into account that the numbers of appearances of each $|\widehat{\nabla} X_\mathcal{T}|_{B_q^{s^* \ell_j^*}(L_{q^*/\ell_j^*}(\widehat{T}))}$ only depends of k, we have

$$|q_\Gamma|_{B_q^{s^*}(L_{q^*}(\widehat{T}))}^{1/\ell_m} \precsim \sum_{\substack{j=1 \\ \ell_j^*=j/k^{k-1}}}^{k^{k-1}} |\widehat{\nabla} X_\mathcal{T}|_{B_q^{s^* \ell_j^*}(L_{q^*/\ell_j^*}(\widehat{T}))}^{1/\ell_j^* \ell_m},$$

where $\ell_j^* = j/k^{k-1}$ has been redefined to fit all possible cases. It suffices now to realize that $\ell_j^* \ell_m$ can be written as $\ell_n = \ell_j^* \ell_m = n/k^k$ for $1 \leq n \leq k^k$ and

$$s^* \ell_j^* = td\ell_n, \quad q^*/\ell_j^* = q/\ell_n.$$

Finally, applying Lemma 2.1 (local stability of Lagrange interpolation) with $s = 1 + td\ell_n$, so that $d\ell_n/q < s \leq n+1$, we may replace $X_\mathcal{T}$ by χ in $\widehat{T}$ and thereby obtain the asserted estimate. $\square$

PROPOSITION 7.12 (Uniform decay rate of $\mathrm{osc}_\mathcal{T}(V)$). *Let γ be globally of class W_∞^1 and be parameterized by $\chi \in [B_q^{1+td}(L_q(\Omega))]^M$ with $tq > 1$, $td \leq n$. Let $\mathcal{T}_0$ have an admissible labeling and let $\mathcal{T} \geq \mathcal{T}_0$ be a refinement of $\mathcal{T}_0$. Then, for any tolerance $\delta > 0$ there exists a subdivision $\mathcal{T}_\delta \in \mathbb{T}$ such that $\mathcal{T}_\delta \geq \mathcal{T}$ and*

$$\max_{V \in \mathbb{V}(\mathcal{T}_\delta)} \frac{\mathrm{osc}_{\mathcal{T}_\delta}(V)}{\|\widehat{\nabla}\mathbf{V}\|_{L_2(\Omega)}} \precsim \delta, \qquad \#\mathcal{M} \precsim C_2(\gamma)^{1/t} \delta^{-1/t},$$

where $\mathcal{M}$ is the set of elements marked to create $\mathcal{T}_\delta$ from $\mathcal{T}$ and the constant $C_2(\gamma)$ depends on γ and is given explicitly by

$$C_2(\gamma) := \max\left(\|\chi\|_{B_q^{1+td}(L_q(\Omega))}, \|\chi\|_{B_q^{1+td}(L_q(\Omega))}^{k^k}\right),$$

with $k = \lfloor td \rfloor + 1$.

Proof. We make use of the GREEDY algorithm upon taking $\mathsf{p} = q$, $\mathsf{g} = \chi$, and

$$\zeta_\mathcal{T}(\chi, T) = h_T^r |\chi|_{\widehat{T}}, \qquad |\chi|_{\widehat{T}} = \sum_{j=1,\ell_j=j/k^k}^{k^k} |\chi|_{B_q^{1+td\ell_j}(L_{q/\ell_j}(\widehat{T}))}^{1/\ell_j}, \qquad (7.22)$$

with $r = d(t - \frac{1}{q}) > 0$. Since

$$|\chi|_\Omega^q = \sum_{\widehat{T} \in \widehat{\mathcal{T}}} |\chi|_{\widehat{T}}^q = \sum_{\widehat{T} \in \widehat{\mathcal{T}}} \left(\sum_{\substack{j=1 \\ \ell_j=j/k^k}}^{k^k} |\chi|_{B_q^{1+td\ell_j}(L_{q/\ell_j}(\widehat{T}))}^{1/\ell_j}\right)^q$$

$$\precsim \sum_{\widehat{T} \in \widehat{\mathcal{T}}} \sum_{\substack{j=1 \\ \ell_j=j/k^k}}^{k^k} |\chi|_{B_q^{1+td\ell_j}(L_{q/\ell_j}(\widehat{T}))}^{q/\ell_j} \precsim \sum_{\substack{j=1 \\ \ell_j=j/k^k}}^{k^k} |\chi|_{B_q^{1+td\ell_j}(L_{q/\ell_j}(\Omega))}^{q/\ell_j},$$

40

and

$$B_q^{1+td}(L_q(\Omega)) \subset B_q^{1+td\ell_j}(L_{q/\ell_j}(\Omega)) \quad \text{for all } 0 < \ell_j \leq 1,$$

we deduce $|\boldsymbol{\chi}|_{B_q^{1+td\ell_j}(L_{q/\ell_j}(\Omega))} \precsim \|\boldsymbol{\chi}\|_{B_q^{1+td}(L_q(\Omega))}$ along with

$$|\boldsymbol{\chi}|_\Omega \precsim C_2(\gamma).$$

Then, Proposition 7.1 guarantees that the call $\mathcal{T}_\delta = \mathsf{GREEDY}(\boldsymbol{\chi}, \mathcal{T}, \delta)$ stops in a finite number of steps and the resulting subdivision $\mathcal{T}_\delta$ satisfies

$$\zeta_{\mathcal{T}_\delta}(\boldsymbol{\chi}, T) \leq \delta \qquad \forall\, T \in \mathcal{T}_\delta.$$

This, in conjunction with Proposition 7.8 and the finite overlapping property of the sets $N_{\mathcal{T}}(T)$, implies that $\mathcal{T}_\delta$ satisfies

$$\mathrm{osc}_{\mathcal{T}_\delta}(V)^2 \precsim \sum_{T \in \mathcal{T}_\delta} \zeta_{\mathcal{T}}(\boldsymbol{\chi}, N_{\mathcal{T}_\delta}(T))^2 \|\widehat{\nabla} V\|_{L_2(N_{\mathcal{T}_\delta}(\widehat{T}))}^2 \leq \delta^2 \|\widehat{\nabla}\mathbf{V}\|_{L_2(\Omega)}^2, \quad \forall\, V \in \mathbb{V}(\mathcal{T}_\delta).$$

This proves the first assertion. In order to bound the cardinality of $\mathcal{M}$ we rely on the estimate (7.5) on the elements marked by GREEDY

$$\#\mathcal{M} \precsim |\boldsymbol{\chi}|_\Omega^{\frac{dq}{d+rq}} \delta^{-\frac{dq}{d+rq}}.$$

The proof concludes upon realizing that $\frac{dq}{d+rq} = \frac{1}{t}$. $\square$

REMARK 7.13 (Approximation of γ). Since $\lambda_{\mathcal{T}}(\gamma, T)$ and $\zeta_{\mathcal{T}}(\boldsymbol{\chi}, T)$ satisfy

$$\lambda_{\mathcal{T}}(\gamma, T) = \|\nabla(\chi - X_{\mathcal{T}})\|_{L_\infty(\widehat{T})} \precsim h_T^r |\chi|_{B_q^{1+td}(L_q(\widehat{T}))} \leq h_T^r |\chi|_{\widehat{T}} = \zeta_{\mathcal{T}}(\boldsymbol{\chi}, T),$$

we deduce $\lambda_{\mathcal{T}_\delta}(\gamma) \precsim \delta$ for the mesh $\mathcal{T}_\delta$ of Proposition (7.12).

7.6. Membership in $\mathbb{A}_s$. We now collect the estimates derived earlier in this section and prove our second main result.

THEOREM 7.14 (Membership in $\mathbb{A}_s$). *Let γ be globally of class W_∞^1 and be parameterized by $\boldsymbol{\chi} \in [B_q^{1+td}(L_q(\Omega))]^M$ with $tq > 1, 0 < q \leq \infty$ and $td \leq n$, and let $k := \lfloor td \rfloor + 1$. Let $u \in H_\#^1(\gamma)$ and $f \in L_2(\gamma)$ be piecewise of class $B_p^{1+sd}(L_p(\Omega))$ and $B_p^{sd}(L_p(\Omega))$, respectively, namely $\mathbf{u} \in [B_p^{1+sd}(L_p(\Omega))]^M$ and $\mathbf{f} \in [B_p^{sd}(L_p(\Omega))]^M$, with $s - 1/p + 1/2 > 0, 0 < p \leq \infty$ and $0 < sd \leq n$. Let $\mathcal{T}_0$ have an admissible labeling and $\lambda_{\mathcal{T}_0}(\gamma)$ satisfy (4.6). Then,*

$$(u, f, \gamma) \in \mathbb{A}_{s \wedge t},$$

i.e, given $\delta > 0$ there exists a conforming refinement $\mathcal{T}$ such that

$$\inf_{V \in \mathbb{V}(\mathcal{T})} E_{\mathcal{T}}(V; u, f, \gamma) \precsim \delta, \qquad \#\mathcal{T} - \#\mathcal{T}_0 \precsim |u, f, \gamma|_{\mathbb{A}_{s \wedge t}}^{\frac{1}{s \wedge t}} \delta^{-\frac{1}{s \wedge t}}, \tag{7.23}$$

and

$$\begin{aligned}
|u, f, \gamma|_{\mathbb{A}_{s \wedge t}} \leq &|\mathbf{u}|_{B_p^{1+sd}(L_p(\Omega))} \\
&+ \left(\|\boldsymbol{\chi}\|_{B_q^{1+td}(L_q(\Omega))} + \|\boldsymbol{\chi}\|_{B_q^{1+td}(L_q(\Omega))}^{k^k} \right) \left(1 + \|\mathbf{f}\|_{B_p^{sd}(L_p(\Omega))} \right).
\end{aligned} \tag{7.24}$$

41

Proof. Since $\lambda_{\mathcal{T}_0}(\gamma)$ satisfies (4.6), instead of dealing with $E_{\mathcal{T}}(V; u, f, \gamma)$, we argue with the equivalent quantity $\widehat{E}_{\mathcal{T}}(V; u, f, \gamma)$ from (7.1), which is evaluated in Ω. Given $\delta > 0$, we invoke Corollaries 7.5 and 7.6 to obtain meshes $\mathcal{T}_u, \mathcal{T}_f \in \mathbb{T}$ satisfying

$$\inf_{V \in \mathbb{V}(\mathcal{T}_u)} \|\widehat{\nabla}(\mathbf{u} - \mathbf{V})\|_{L^2(\Omega)} \preccurlyeq \delta, \qquad \#\mathcal{M}_u \preccurlyeq C(\mathbf{u})^{1/s} \delta^{-1/s},$$

$$\mathrm{osc}_{\mathcal{T}_f}(f) \preccurlyeq \delta, \qquad \#\mathcal{T}_f - \#\mathcal{T}_0 \preccurlyeq C(\mathbf{f}, \gamma)^{1/(s \wedge t + 1/d)} \delta^{-1/(s \wedge t + 1/d)}.$$

We next apply Proposition 7.12 and Remark 7.13, starting from $\mathcal{T}_u$, to obtain a refinement $\mathcal{T}_\gamma \in \mathbb{T}$ such that $\mathcal{T}_\gamma \geq \mathcal{T}_u$,

$$\lambda_{\mathcal{T}_\gamma}(\gamma) \preccurlyeq \delta, \qquad \max_{V \in \mathbb{V}(\mathcal{T}_\gamma)} \frac{\mathrm{osc}_{\mathcal{T}_\gamma}(V)}{\|\widehat{\nabla}\mathbf{V}\|_{L_2(\Omega)}} \preccurlyeq \delta, \qquad \#\mathcal{M}_\gamma \preccurlyeq C_2(\gamma)^{1/t} \delta^{-1/t}.$$

The cardinality of $\mathcal{T}_\gamma$ can be estimated via Lemma 5.2 (complexity of REFINE)

$$\#\mathcal{T}_\gamma - \#\mathcal{T}_0 \preccurlyeq \#\mathcal{M}_u + \#\mathcal{M}_\gamma \preccurlyeq C(\mathbf{u})^{1/s}\delta^{-1/s} + C_2(\gamma)^{1/t}\delta^{-1/t}.$$

Since the cardinalities can be assumed to be at least 1, we can replace the exponents $1/s$ and $1/t$ of $C(\mathbf{u})\delta^{-1}$ and $C_2(\gamma)\delta^{-1}$ by $1/s \wedge t$. Let $\mathcal{T} = \mathcal{T}_\gamma \oplus \mathcal{T}_f$ be the overlay of the two meshes $\mathcal{T}_\gamma$ and $\mathcal{T}_f$. According to [12, Lemma 3.7] the cardinality of $\#\mathcal{T} - \#\mathcal{T}_0$ is bounded by $\#\mathcal{T}_\gamma + \#\mathcal{T}_f - 2\#\mathcal{T}_0$, whence

$$\#\mathcal{T} - \#\mathcal{T}_0 \preccurlyeq |u, \gamma, f|_{\mathbb{A}_{s \wedge t}}^{1/s \wedge t} \delta^{-1/s \wedge t},$$

with the nonlinear quantity $|u, \gamma, f|_{\mathbb{A}_{s \wedge t}}$ satisfying (7.24).

It remains to show the first estimate in (7.23). We first observe that

$$\inf_{V \in \mathbb{V}(\mathcal{T})} \|\widehat{\nabla}(\mathbf{u} - \mathbf{V})\|_{L^2(\Omega)} \leq \inf_{V \in \mathbb{V}(\mathcal{T}_u)} \|\widehat{\nabla}(\mathbf{u} - \mathbf{V})\|_{L^2(\Omega)}$$

because $\mathcal{T} \geq \mathcal{T}_u$. We choose $V_u \in \mathbb{V}(\mathcal{T}_u)$ to be the function that realizes the minimum, whence $\|\widehat{\nabla}(\mathbf{u} - \mathbf{V}_u)\|_{L_2(\Omega)} \preccurlyeq \delta$. Since the definition (7.13) of $\mathrm{osc}_{\mathcal{T}}(V)$ involves the best L_2-approximation, we can argue as in Lemma 4.9 to deduce for $\mathcal{T} \geq \mathcal{T}_\gamma$ (see Lemma 4.10)

$$\mathrm{osc}_{\mathcal{T}}(V_u) \preccurlyeq \mathrm{osc}_{\mathcal{T}_\gamma}(V_u) + \lambda_{\mathcal{T}_\gamma}(\gamma) \leq \delta \|\widehat{\nabla}\mathbf{V}_u\|_{L_2(\Omega)} + \delta,$$

because $V_u \in \mathbb{V}(\widehat{\mathcal{T}}_u) \subset \mathbb{V}(\widehat{\mathcal{T}}_\gamma)$. Upon adding and subtracting u, we readily see that

$$\mathrm{osc}_{\mathcal{T}}(V_u) \preccurlyeq \delta \|\widehat{\nabla}\mathbf{u}\|_{L_2(\Omega)} + \delta \|\widehat{\nabla}(\mathbf{u} - \mathbf{V}_u)\|_{L_2(\Omega)} + \delta \preccurlyeq \delta.$$

Since the definition (7.14) utilizes the L_2-projection and $\mathcal{T} \geq \mathcal{T}_f$ we infer that

$$\mathrm{osc}_{\mathcal{T}}(f) \leq \mathrm{osc}_{\mathcal{T}_f}(f) \preccurlyeq \delta.$$

Collecting the preceding estimates and using the definition (7.1) of total error gives the desired estimate $\widehat{E}_{\mathcal{T}}(V_u; u, f, \gamma) \preccurlyeq \delta$, and finishes the proof. $\square$

42

8. Convergence rates. In this section we study the cardinality of AFEM, which is dictated by the regularity of u, f and γ. We now prove that AFEM achieves the asymptotic decay rate s dictated by the class $\mathbb{A}_s$. We establish the link between the performance of AFEM and the best possible error by adapting a clever idea of Stevenson [30] for the Laplace operator, further extended by Cascón et al [12] to general elliptic PDE, in flat domains. We refer to the survey [28] for a thorough discussion and to [11]. The insight of [30] is the following

> *Any marking strategy that reduces the total error to a fraction of its current value must contain a substantial portion of the error estimator, and so it can* (8.1) *be related to Dörfler Marking.*

Exploiting next the minimality of Dörfler marking we can compare meshes generated by AFEM with the best meshes within $\mathbb{T}$. The approach of [12, 28, 30] does not apply directly to the present context because of the consistency error due to surface interpolation. We account for this discrepancy below upon making the parameter ω of ADAPT_SURFACE sufficiently small. Let

$$\omega_3 := \frac{C_5}{\Lambda_0\sqrt{3\Lambda_1 + 4\Lambda_2 + 2\Lambda_1\Lambda_3}}, \tag{8.2}$$

be a threshold for ω to be used next and let θ_* be a threshold for Dörfler parameter θ

$$\theta_* := \frac{C_5}{\sqrt{2C_3 + C_1(3 + 2\Lambda_3)}}; \tag{8.3}$$

since $C_5 = \sqrt{C_2/2}$ and $C_2 \leq C_1$, we see that $\theta_* < 1$.

LEMMA 8.1 (Dörfler marking). *Let $\lambda_{\mathcal{T}_0}(\gamma)$ satisfy (4.6), and the parameters θ and ω satisfy*

$$0 < \theta < \theta_*, \qquad 0 < \omega \leq \min\{\omega_1, \omega_3\}, \tag{8.4}$$

where θ_*, ω_3 are defined in (8.2), (8.3), and ω_1 in (5.4). Let $\mu := \frac{1}{2}\sqrt{1 - \frac{\theta^2}{\theta_*^2}}$ and $(\Gamma, \mathcal{T}, U)$ be the approximate surface, mesh and discrete solution produced by an inner iterate of ADAPT_PDE. If $(\Gamma_*, \mathcal{T}_*, U_*)$ is a surface-mesh-solution triple with $\mathcal{T}_* \geq \mathcal{T}$, such that the PDE error satisfies

$$\mathcal{E}_{\mathcal{T}_*}(U_*, f) \leq \mu\,\mathcal{E}_{\mathcal{T}}(U, f), \tag{8.5}$$

then the refined set $\mathcal{R} := \mathcal{T} \setminus \mathcal{T}_*$ satisfies Dörfler property with parameter θ, namely

$$\eta_{\mathcal{T}}(U, F_\Gamma, \mathcal{R}) \geq \theta\eta_{\mathcal{T}}(U, F_\Gamma). \tag{8.6}$$

Proof. We proceed as in [12, Lemma 5.9] using the notation $e(U) := \|\nabla_\gamma(u - U)\|_{L_2(\gamma)}$. Since $\omega \leq \omega_1$, we combine the lower bound of (4.21) with (8.5) to write

$$(1 - 2\mu^2)C_5^2\eta_{\mathcal{T}}(U, F_\Gamma)^2 \leq (1 - 2\mu^2)\big(e(U)^2 + \mathrm{osc}_{\mathcal{T}}(U, f)^2\big)$$
$$\leq e(U)^2 - e(U_*)^2 + \mathrm{osc}_{\mathcal{T}}(U, f)^2 - 2\,\mathrm{osc}_{\mathcal{T}_*}(U_*, f)^2.$$

We now estimate separately error and oscillation terms. According to (6.1) and (4.17), we obtain

$$e(U)^2 - e(U_*)^2 \leq \frac{3}{2}\|\nabla_\gamma(U_* - U)\|_{L_2(\gamma)}^2 + \Lambda_2\lambda_{\mathcal{T}}(\gamma)^2$$
$$\leq \frac{3}{2}C_1\eta_{\mathcal{T}}(U, F_\Gamma, \mathcal{R})^2 + \Big(\frac{3}{2}\Lambda_1 + \Lambda_2\Big)\lambda_{\mathcal{T}}(\gamma)^2.$$

For the oscillation terms we argue according to whether an element $T \in \mathcal{T}$ belongs to the set of refined elements $\mathcal{R}$ or not. We use the dominance bound (4.18) to arrive at

$$\mathrm{osc}_{\mathcal{T}}(U, f, \mathcal{R})^2 \leq C_3 \eta_{\mathcal{T}}(U, F_{\Gamma}, \mathcal{R})^2.$$

On the other hand, using (4.27) for $\mathcal{T}_* \cap \mathcal{T}$ with $V = U$ and $W = U_*$ yields

$$\mathrm{osc}_{\mathcal{T}}(U, f, \mathcal{T}_* \cap \mathcal{T})^2 - 2\,\mathrm{osc}_{\mathcal{T}_*}(U_*, f)^2 \leq \Lambda_3 \|\nabla_{\gamma}(U_* - U)\|^2_{L_2(\gamma)} + \Lambda_2 \lambda_{\mathcal{T}}(\gamma)^2.$$

By combining these two estimates with (4.17) we infer that

$$\mathrm{osc}_{\mathcal{T}}(U, f)^2 - 2\,\mathrm{osc}_{\mathcal{T}_*}(U_*, f)^2 \leq (C_3 + C_1\Lambda_3)\eta_{\mathcal{T}}(U, F_{\Gamma}, \mathcal{R})^2 + (\Lambda_1\Lambda_3 + \Lambda_2)\lambda_{\mathcal{T}}(\gamma)^2.$$

Since $\mathcal{T}$ is produced within ADAPT_PDE, which is initialized with $\mathcal{T}^+$ in AFEM of § 5, we have $\eta_{\mathcal{T}}(U, F_{\Gamma}) \geq \varepsilon$, $\lambda_{\mathcal{T}^+}(\gamma) \leq \omega\varepsilon$, and $\mathcal{T} \geq \mathcal{T}^+$, whence

$$\lambda_{\mathcal{T}}(\gamma) \leq \Lambda_0 \lambda_{\mathcal{T}^+}(\gamma) \leq \Lambda_0 \omega\varepsilon \leq \Lambda_0 \omega \eta_{\mathcal{T}}(U, F_{\Gamma}).$$

Collecting these three estimates, we deduce

$$(1 - 2\mu^2)C_5^2 \eta_{\mathcal{T}}(U, F_{\Gamma})^2 \leq \frac{1}{2}\Big(2C_3 + C_1(3 + 2\Lambda_3)\Big)\eta_{\mathcal{T}}(U, F_{\Gamma}, \mathcal{R})^2$$
$$+ \frac{1}{2}(3\Lambda_1 + 4\Lambda_2 + 2\Lambda_1\Lambda_3)\Lambda_0^2 \omega^2 \eta_{\mathcal{T}}(U, F_{\Gamma})^2.$$

Finally, using that $\omega \leq \omega_3$, along with (8.2) and (8.3), we infer that

$$(1 - 4\mu^2)\theta_*^2 \eta_{\mathcal{T}}(U, F_{\Gamma})^2 \leq \eta_{\mathcal{T}}(U, F_{\Gamma}, \mathcal{R})^2.$$

The choice of μ implies the asserted estimate (8.6). $\square$

LEMMA 8.2 (Cardinality of $\mathcal{M}$). *Let* $\lambda_{\mathcal{T}_0}(\gamma)$ *satisfy* (4.6) *and the procedure MARK select a set $\mathcal{M}$ with minimal cardinality. Let the parameters θ and ω satisfy*

$$0 < \theta < \theta_*, \qquad 0 < \omega \leq \min\{\omega_1, \omega_3\} \tag{8.7}$$

with θ_*, ω_1 and ω_3 given in (8.3), (5.4), *and* (8.2), *respectively. Let u be the solution of* (3.13), *and let* $(\Gamma, \mathcal{T}, U)$ *be produced within ADAPT_PDE. If* $(u, f, \gamma) \in \mathbb{A}_s$, *then*

$$\#\mathcal{M} \preccurlyeq |u, f, \gamma|_{\mathbb{A}_s}^{\frac{1}{s}} \mathcal{E}_{\mathcal{T}}(U, f)^{-\frac{1}{s}}.$$

Proof. We set

$$\delta^2 = \hat{\mu}^2 \mathcal{E}_{\mathcal{T}}(U, f)^2 = \hat{\mu}^2\Big(e(U)^2 + \mathrm{osc}_{\mathcal{T}}(U, f)^2\Big),$$

for $0 < \hat{\mu} < \mu = \frac{1}{2}\sqrt{1 - \frac{\theta^2}{\theta_*^2}} < 1$ sufficiently small to be determined later. Since $(u, f, \gamma) \in \mathbb{A}_s$, there exists a subdivision $\mathcal{T}_\delta \in \mathbb{T}$ and $V_\delta \in \mathbb{V}(\mathcal{T}_\delta)$ such that

$$\#\mathcal{T}_\delta - \#\mathcal{T}_0 \preccurlyeq |u, f, \gamma|_{\mathbb{A}_s}^{\frac{1}{s}} \delta^{-\frac{1}{s}}, \qquad e(V_\delta)^2 + \mathrm{osc}_{\mathcal{T}_\delta}(V_\delta, f)^2 + \lambda_{\mathcal{T}_\delta}(\gamma)^2 \leq \delta^2. \tag{8.8}$$

Let $\mathcal{T}_* = \mathcal{T} \oplus \mathcal{T}_\delta$ be the overlay of $\mathcal{T}$ and $\mathcal{T}_\delta$, which satisfies [12, Lemma 3.7], [28],

$$\#\mathcal{T}_* \leq \#\mathcal{T} + \#\mathcal{T}_\delta - \#\mathcal{T}_0. \tag{8.9}$$

44

Let $U_* \in \mathbb{V}(\mathcal{T}_*)$ be the corresponding Galerkin solution. We observe that $\mathcal{T}_* \geq \mathcal{T}_\delta, \mathcal{T}$, and invoke the upper bound of (6.1) in conjunction with (4.26) to write

$$e(U_*)^2 + \mathrm{osc}_{\mathcal{T}_*}(U_*, f)^2 \leq e(V_\delta)^2 + \Lambda_2 \lambda_{\mathcal{T}_\delta}(\gamma)^2$$
$$+ C_6 \, \mathrm{osc}_{\mathcal{T}_\delta}(V_\delta, f)^2 + \Lambda_3 \|\nabla_\gamma(U_* - V_\delta)\|^2_{L_2(\gamma)} + \Lambda_2 \lambda_{\mathcal{T}_\delta}(\gamma)^2.$$

Applying (6.1) again gives $\|\nabla_\gamma(U_* - V_\delta)\|^2_{L_2(\gamma)} \leq 2e(V_\delta)^2 + 2\Lambda_2\lambda_{\mathcal{T}_\delta}(\gamma)^2$, whence

$$e(U_*)^2 + \mathrm{osc}_{\mathcal{T}_*}(U_*, f)^2 \leq (1 + 2\Lambda_3)e(V_\delta)^2 + C_6 \, \mathrm{osc}_{\mathcal{T}_\delta}(V_\delta, f)^2 + 2\Lambda_2(1 + \Lambda_3)\lambda_{\mathcal{T}_\delta}(\gamma)^2.$$

We now choose $\hat{\mu} = \frac{\mu}{\sqrt{\max\{C_6, 1+2\Lambda_3, 2\Lambda_2(1+\Lambda_3)\}}}$ to end up with

$$e(U_*)^2 + \mathrm{osc}_{\mathcal{T}_*}(U_*, f)^2 \leq \max\{C_6, 1 + 2\Lambda_3, 2\Lambda_2(1+\Lambda_3)\}\delta^2 = \mu^2\Big(e(U)^2 + \mathrm{osc}_{\mathcal{T}}(U, f)^2\Big).$$

We thus deduce from Lemma 8.1 that the subset $\mathcal{R} = \mathcal{R}_{\mathcal{T} \to \mathcal{T}_*} \subset \mathcal{T}$ satisfies Dörfler property (8.6). Since the set $\mathcal{M} \subset \mathcal{T}$ also satisfies this property, but with minimal cardinality, we infer from (8.8)–(8.9)

$$\#\mathcal{M} \leq \#\mathcal{R} \leq \#\mathcal{T}_* - \#\mathcal{T} \leq \#\mathcal{T}_\delta - \#\mathcal{T}_0 \preccurlyeq |u, f, \gamma|^{\frac{1}{s}}_{\mathbb{A}_s} \delta^{-\frac{1}{s}},$$

The asserted estimate finally follows upon using the definition of δ. $\square$

The quasi-optimal cardinality of AFEM is a direct consequence of Lemma 8.2 and Theorem 6.2. This is our third main result and we prove it next.

THEOREM 8.3 (Convergence rate of AFEM). *Let $\varepsilon_0 \leq (6\omega\Lambda_0 L^3)^{-1}$ be the initial tolerance, and the parameters θ, ω, ρ of AFEM satisfy*

$$0 < \theta \leq \theta_*, \qquad 0 < \omega \leq \omega_* := \min\{\omega_1, \omega_2, \omega_3\}, \qquad 0 < \rho < 1, \qquad (8.10)$$

where $\theta_*, \omega_1, \omega_2, \omega_3$ are given in (8.3), (5.4), (6.2), and (8.2), respectively. Let $\mathcal{T}_0$ have an admissible labeling, and let the procedure MARK select sets with minimal cardinality. Let u be the solution of (3.13) and $\{\Gamma_k, \mathcal{T}_k, U_k\}_{k \geq 0}$ be the sequence of approximate surfaces, meshes and discrete solutions generated by AFEM.

If $(u, f, \gamma) \in \mathbb{A}_s$ for some $0 < s \leq n/d$, then there exists a constant C, depending on the Lipschitz constant L of γ, $\|f\|_{L_2(\gamma)}$, the refinement depth b, the initial triangulation $\mathcal{T}_0$, and AFEM parameters θ, ω, ρ such that

$$e(U_k) + \mathrm{osc}_{\mathcal{T}_k}(U_k, f) + \omega^{-1}\lambda_{\mathcal{T}_k}(\gamma) \leq C|u, f, \gamma|_{\mathbb{A}_s}\big(\#\mathcal{T}_k - \#\mathcal{T}_0\big)^{-s}, \qquad (8.11)$$

where $|u, f, \gamma|_{\mathbb{A}_s}$ is defined in (7.2).

Proof. We start by noting that since $\omega\varepsilon_0 \leq \frac{1}{6\Lambda_0 L^3}$ the first output of the procedure ADAPT_SURFACE fulfills $\lambda_{\mathcal{T}_0^+}(\gamma) \leq \frac{1}{6\Lambda_0 L^3}$ which is (4.6) and implies that $\mathbb{T}(\mathcal{T}_0^+)$ is shape regular.

There are two instances where elements are added, inside ADAPT_SURFACE and ADAPT_PDE. In light of (1.4) and (7.3) we observe that ADAPT_SURFACE is s-optimal with $C(\gamma) \preccurlyeq |u, f, \gamma|^{1/s}_{\mathbb{A}_s}$, whence the set of all the elements marked for refinement in the k-th call to ADAPT_SURFACE satisfies

$$\#\mathcal{M}_k \preccurlyeq C(\gamma) \, \omega^{-\frac{1}{s}} \, \varepsilon_k^{-\frac{1}{s}} \preccurlyeq |u, f, \gamma|^{\frac{1}{s}}_{\mathbb{A}_s} \, \varepsilon_k^{-\frac{1}{s}}.$$

For ADAPT_PDE, Lemma 8.2 (cardinality of $\mathcal{M}$) yields

$$\#\mathcal{M}_k^j \preccurlyeq |u, f, \gamma|^{\frac{1}{s}}_{\mathbb{A}_s}\Big(e(U_k^j) + \mathrm{osc}_{\mathcal{T}_k^j}(U_k^j, f)\Big)^{-\frac{1}{s}} \qquad 0 \leq j < J.$$

45

where $\mathcal{M}_k^j$ denotes the subset of elements selected by the marking procedure at the j-th subiteration of the k-th step of ADAPT_PDE. Since the inner iterates of ADAPT_PDE satisfy Theorem 6.2 (conditional contraction property) and

$$e(U_k^j) + \operatorname{osc}_{\mathcal{T}_k^j}(U_k^j, f) \approx e(U_k^j) + \eta_{\mathcal{T}_k^j}(U_k^j, F_k^j),$$

according to (4.21), we deduce that

$$\left(e(U_k^j) + \operatorname{osc}_{\mathcal{T}_k^j}(U_k^j, f)\right)^{-\frac{1}{s}} \preccurlyeq \alpha^{\frac{J-j-1}{s}} \left(e(U_k^{J-1}) + \eta_{\mathcal{T}_k^{J-1}}(U_k^{J-1}, F_k^{J-1})\right)^{-\frac{1}{s}}$$
$$\leq \alpha^{\frac{J-j-1}{s}} \varepsilon_k^{-\frac{1}{s}}$$

by virtue of $\eta_{\mathcal{T}_k^{J-1}}(U_k^{J-1}, F_k^{J-1}) > \varepsilon_k$. This implies

$$\sum_{j=0}^{J-1} \#\mathcal{M}_k^j \preccurlyeq |u, f, \gamma|_{\mathbb{A}_s}^{\frac{1}{s}} \varepsilon_k^{-\frac{1}{s}} \sum_{j=0}^{J-1} \alpha^{\frac{J-j-1}{s}} \preccurlyeq |u, f, \gamma|_{\mathbb{A}_s}^{\frac{1}{s}} \varepsilon_k^{-\frac{1}{s}}.$$

To do a full counting argument, we resort to the crucial estimate (5.6) which, combined with the estimates above, gives

$$\#\mathcal{T}_k - \#\mathcal{T}_0 \leq C_7 \sum_{i=0}^{k-1} \left(\#\mathcal{M}_i + \sum_{j=0}^{J-1} \#\mathcal{M}_i^j\right) \preccurlyeq |u, f, \gamma|_{\mathbb{A}_s}^{\frac{1}{s}} \sum_{i=0}^{k-1} \varepsilon_i^{-\frac{1}{s}}.$$

We now use the relation $\varepsilon_{k+1} = \rho \varepsilon_k$ of step 3 of AFEM, together with $\rho < 1$, to obtain $\sum_{i=0}^{k-1} \varepsilon_i^{-\frac{1}{s}} = \varepsilon_{k-1}^{-\frac{1}{s}} \sum_{i=0}^{k-1} \rho^{\frac{i}{s}} \preccurlyeq \varepsilon_k^{-\frac{1}{s}}$, whence

$$\#\mathcal{T}_k - \#\mathcal{T}_0 \preccurlyeq |u, f, \gamma|_{\mathbb{A}_s}^{\frac{1}{s}} \varepsilon_k^{-\frac{1}{s}}. \tag{8.12}$$

Moreover, the stopping criteria (5.1) and (5.3) guarantee that

$$e(U_k) + \operatorname{osc}_{\mathcal{T}_k}(U_k, f) + \omega^{-1} \lambda_{\mathcal{T}_k}(\gamma) \preccurlyeq \varepsilon_k, \tag{8.13}$$

which implies the desired estimate (8.11). $\square$

The precise constant on the right-hand side of (8.11) is $\omega^{-1} C(\gamma)^s + |u, f, \gamma|_{\mathbb{A}_s}$. This and the condition $\omega \leq \omega_*$ in (8.10) suggest that ω should not be too small to optimize (8.11). An optimal choice of ω, which unfortunately is not computable, appears to be

$$\omega = \min\left\{\omega_*, |u, f, \gamma|_{\mathbb{A}_s}^{-1} C(\gamma)^s\right\}.$$

We also provide an estimate on the workload in the following corollary. We assume that the adaptive loop (1.6) on a subdivision $\mathcal{T} \in \mathbb{T}$ requires $O(\#\mathcal{T})$ computations and in particular (i) the linear algebra solver scales like $\#\mathcal{T}$ [9, 24] and (ii) an approximate sort requiring $O(\#\mathcal{T})$ arithmetic operations is used to select the local estimators $\eta_{\mathcal{T}}(U, F_\Gamma, T)$ for all $T \in \mathcal{T}$ (see e.g. [8, Remark 5.3]).

COROLLARY 8.4 (Workload estimate). *In addition to the assumptions of Theorem 8.3, suppose that each inner loop of ADAPT_PDE on a subdivision $\mathcal{T} \in \mathbb{T}$ requires $O(\#\mathcal{T})$ arithmetic operations. If $\varepsilon \leq \varepsilon_0$, then the number of arithmetic operations W for AFEM to construct a triple $(\Gamma, \mathcal{T}, U)$ such that*

$$e(U) + \operatorname{osc}_{\mathcal{T}}(U, f) + \omega^{-1} \lambda_{\mathcal{T}}(\gamma) \leq \varepsilon \tag{8.14}$$

46

satisfies

$$W \preccurlyeq \varepsilon^{-1/s}.$$

Proof. Let $C \geq 1$ be the hidden constant in (8.13) and set K to be the integer such that $\varepsilon_{K+1} := \rho^{K+1}\varepsilon_0 \leq \varepsilon/C \leq \varepsilon_K$. Moreover, we define W_j^+ to be the number of arithmetic operations performed within the call $\mathcal{T}_j^+ = \mathsf{ADAPT\_SURFACE}(\mathcal{T}_j, \omega\varepsilon_j)$ and W_{j+1} those within the call $[U_{j+1}, \mathcal{T}_{j+1}] = \mathsf{ADAPT\_PDE}(\mathcal{T}_j^+, \varepsilon_j)$. With these notations, the total number of operations to achieve (8.14) satisfies

$$W \preccurlyeq \sum_{j=0}^{K} (W_j^+ + W_{j+1}).$$

We now bound each term separately starting with W_j^+. The computation of each local geometric estimator requires $O(1)$ arithmetic operations and is performed

$$\#\mathcal{T}_j + \#\mathcal{M}_j \leq \#\mathcal{T}_j^+ \leq \#\mathcal{T}_{j+1}$$

times. Since $\mathsf{ADAPT\_SURFACE}$ does not involve sorting the local geometric estimators, we readily deduce that

$$W_j^+ \preccurlyeq \#\mathcal{T}_{j+1}.$$

Regarding W_{j+1}, we recall that Theorem 6.2 guarantees that the number of inner iterations J within $\mathsf{ADAPT\_PDE}$ is uniformly bounded. This, together with the complexity assumption on the inner loops of $\mathsf{ADAPT\_PDE}$, yields

$$W_{j+1} \preccurlyeq \#\mathcal{T}_{j+1}.$$

Now combining the above two estimates and invoking (8.12), we deduce that

$$W_j^+ + W_{j+1} \preccurlyeq \varepsilon_{j+1}^{-1/s}.$$

Going back to the total number of operations W, we find

$$W \preccurlyeq \rho^{-1/s} \varepsilon_K^{-1/s} \sum_{j=0}^{K} \rho^{j/s} \preccurlyeq \varepsilon_K^{-1/s},$$

where we used the relations $\varepsilon_{j+1} = \rho^{-(K-j-1)}\varepsilon_K$ for $j = 0, ..., K$. The desired estimate follows from the definition of ε_K. $\square$

9. Products and Compositions in Besov Spaces. In this section we derive scale-invariant estimates for Besov (quasi-)semi-norms for products and compositions of functions. This extends estimates for Sobolev norms of products [1], as alluded to in Remark 7.3, and for Besov norms of compositions [10]. Our estimates, however, possess a structure that does not seem to be available in the literature.

We recall the definition of Besov spaces via modulus of smoothness; a thorough discussion can be found in [23]. Let Ω be a Lipschitz domain in $\mathbb{R}^d$, $0 < p \leq \infty$ and $u \in L_p(\Omega)$, we define the differences as follow, for $h \in \mathbb{R}^d$, $k \in \mathbb{N}$: $\Delta_h u : \Omega \to \mathbb{R}$, with

$$\Delta_h u(x) = \begin{cases} u(x+h) - u(x), & \text{if } x \in \Omega_h = \{x \in \Omega : x + th \in \Omega, \forall t \in [0,1]\}, \\ 0, & \text{otherwise,} \end{cases}$$

47

and $\Delta_h^{k+1} u : \Omega \to \mathbb{R}$ as $\Delta_h^{k+1} u(x) = \Delta_h \Delta_h^k u(x)$ for $k \in \mathbb{N}$ and $x \in \Omega_{(k+1)h}$ and 0 otherwise. Therefore,

$$\Delta_h^k u(x) = \begin{cases} \sum_{j=0}^k (-1)^{k+j} \binom{k}{j} u(x + jh), & \text{if } x \in \Omega_{kh}, \\ 0, & \text{otherwise.} \end{cases} \tag{9.1}$$

Using these difference operators we define the *modulus of smoothness* of order k in $L_p(\Omega)$ as:

$$\omega_k(u,t)_p = \sup_{|h| \le t} \|\Delta_h^k u\|_{L_p(\Omega)}, \qquad t > 0,$$

Given $s > 0$ and $0 < q, p \le \infty$, the Besov space $B_q^s(L_p(\Omega))$, is the set of all functions $f \in L_p(\Omega)$ such that the semi-(quasi)norm $|f|_{B_q^s(L_p(\Omega))}$ is finite, with

$$|f|_{B_q^s(L_p(\Omega))} := \begin{cases} \left(\int_0^\infty [t^{-s} \omega_k(u,t)_p]^q \dfrac{dt}{t} \right)^{\frac{1}{q}}, & \text{if } 0 < q < \infty \\ \sup_{t>0} t^{-s} \omega_k(f,t)_p, & \text{if } q = \infty, \end{cases} \tag{9.2}$$

where $k \in \mathbb{N}$ is such that $s < k$. Although this definition only requires $k > s$, it turns out that the space $B_q^s(L_p(\Omega))$ is independent of k [18, Theorem 10.1]. From now on, we fix the value of $k := \lfloor s \rfloor + 1$ in most of our results for simplicity. The (quasi) norm of $B_q^s(L_p(\Omega))$ is defined by:

$$\|f\|_{B_q^s(L_p(\Omega))} = \|f\|_{L_p(\Omega)} + |f|_{B_q^s(L_p(\Omega))}. \tag{9.3}$$

Two important properties that we will exploit in what follows are the embeddings of $B_q^s(L_p(\Omega))$ in $L_\infty(\Omega)$, whenever $sp > d$, and of $B_{q_1}^{s_1}(L_{p_1}(\Omega))$ in $B_{q_2}^{s_2}(L_{p_2}(\Omega))$, whenever $s_1 - d/p_1 > s_2 - d/p_2$ [23].

9.1. Product of Functions. The following result, essential for our discussion, is analogous to [1, Theorem 4.39] for Besov spaces and is scale-invariant.

LEMMA 9.1 (Scale-invariant Besov semi-norm of the product of two functions). *Let $u, v \in B_q^s(L_p(\Omega))$ with $s > 0$ and $0 < p, q \le \infty$ satisfying $s > d/p$ (i.e. $B_q^s(L_p(\Omega)) \subset L_\infty(\Omega)$) and $k = \lfloor s \rfloor + 1$. Then $uv \in B_q^s(L_p(\Omega))$ and*

$$|uv|_{B_q^s(L_p(\Omega))} \precsim \sum_{j=0}^k |u|_{B_q^{sj/k}(L_{pk/j}(\Omega))} |v|_{B_q^{s(k-j)/k}(L_{pk/(k-j)}(\Omega))}, \tag{9.4}$$

with the convention that $B_q^0(L_{pk/0}(\Omega)) = L_\infty(\Omega)$ and $| \cdot |_{B_q^0(L_{pk/0}(\Omega))} = \| \cdot \|_{L_\infty(\Omega)}$.

Proof. Recall that $\omega_k(uv,t)_p = \sup_{|h| \le t} \|\Delta_h^k(uv)\|_{L_p(\Omega)}$ and that the k-th differences obey a rule similar to Leibniz rule. This translates into the expression

$$\|\Delta_h^k(uv)\|_{L_p(\Omega)} \precsim \sum_{j=0}^k \|\Delta_h^j u \Delta_h^{k-j} v\|_{L_p(\Omega)} \le \sum_{j=0}^k \|\Delta_h^j u\|_{L_{pk/j}(\Omega)} \|\Delta_h^{k-j} v\|_{L_{pk/(k-j)}(\Omega)},$$

where we have used Hölder inequality in the last step and the conventions $\Delta_h^0 = Id$ and $L_{pk/0}(\Omega) = L_\infty(\Omega)$. Therefore,

$$\omega_k(uv,t)_p \precsim \sum_{j=0}^k \omega_j(u,t)_{pk/j} \, \omega_{k-j}(v,t)_{pk/(k-j)},$$

48

where we use the convention $\omega_0(v,t)_{pk/0} = \|v\|_{L_\infty(\Omega)}$.

We now consider the cases $q = \infty$ and $q < \infty$ separately. Observe that

$$|uv|_{B^s_\infty(L_p(\Omega))} = \sup_{t>0} t^{-s}\omega_k(uv,t)_p$$

$$\preccurlyeq \sum_{j=0}^{k} \sup_{t>0} t^{-js/k}\omega_j(u,t)_{pk/j} \sup_{t>0} t^{-(k-j)s/k}\omega_{k-j}(v,t)_{pk/(k-j)}.$$

Utilizing that $|u|_{B^{js/k}_\infty(L_{pk/j}(\Omega))} \simeq \sup_{t>0} t^{-js/k}\omega_j(u,t)_{pk/j}$, because $k > s$ so that $js/k < j \le j - 1 + \max\{1, j/(pk)\}$ for $0 \le j \le k$, we have

$$|uv|_{B^s_\infty(L_p(\Omega))} \preccurlyeq \sum_{j=0}^{k} |u|_{B^{js/k}_\infty(L_{pk/j}(\Omega))} |v|_{B^{(k-j)s/k}_\infty(L_{pk/(k-j)}(\Omega))}.$$

If $0 < q < \infty$, we define $q^* = \max\{1, q\}$ and notice that by the triangle and Hölder inequalities

$$|uv|^{q/q^*}_{B^s_q(L_p(\Omega))} = \left(\int_0^\infty t^{-sq}\omega_k(uv,t)_p^q \frac{dt}{t} \right)^{1/q^*}$$

$$\le \sum_{j=0}^{k} \left(\int_0^\infty t^{-sq}\omega_j(u,t)_{pk/j}^q \omega_{k-j}(v,t)_{pk/(k-j)}^q \frac{dt}{t} \right)^{1/q^*}$$

$$\le \sum_{j=0}^{k} \left(\int_0^\infty t^{-sq}\omega_j(v,t)_{pk/j}^{qk/j} \frac{dt}{t} \right)^{\frac{j}{kq^*}} \left(\int_0^\infty t^{-sq}\omega_{k-j}(v,t)_{pk/(k-j)}^{qk/(k-j)} \frac{dt}{t} \right)^{\frac{k-j}{kq^*}}$$

$$\preccurlyeq \sum_{j=0}^{k} |u|^{q/q^*}_{B^{sj/k}_{qk/j}(L_{pk/j}(\Omega))} |v|^{q/q^*}_{B^{s(k-j)/k}_{qk/(k-j)}(L_{pk/(k-j)}(\Omega))}.$$

Upon raising both sides to the power $q^*/q \ge 1$, this inequality implies (9.4). $\square$

We make the important observation that (9.4) is scale-invariant: simply scale Ω by a constant h and realize that both sides of (9.4) scale by the same factor $h^{s-d/p}$. This implies that (9.4) can be used at the element level. Upon iterating (9.4) we obtain the following simple, but technical, generalization of (9.4).

COROLLARY 9.2 (Scale-invariant Besov semi-norm of products of functions). *Let $s > 0$ and $0 < p, q \le \infty$ satisfy $s > d/p$ (i.e. $B^s_q(L_p(\Omega)) \subset L_\infty(\Omega)$), and let $k = \lfloor s \rfloor + 1$. If $u_i \in B^s_q(L_p(\Omega))$ for $1 \le i \le m$, then $\prod_{i=1}^m u_i \in B^s_q(L_p(\Omega))$ and*

$$\left| \prod_{i=1}^m u_i \right|_{B^s_q(L_p(\Omega))} \preccurlyeq \sum_{\sum_{i=1}^m \ell_i = 1} \prod_{i=1}^m |u_i|_{B^{s\ell_i}_q(L_{p/\ell_i}(\Omega))}, \tag{9.5}$$

where $0 \le \ell_i = m_i/k^{m-1} \le 1$ and the sum ranges over all choices of $m_i \in \mathbb{N}_0$ such that $\sum_{i=1}^m m_i = k^{m-1}$.

Using embedding theorems for Besov spaces, (9.5) can be further simplified by replacing the semi-norms by norms. However, this is at the expense of having a constant depending on $|\Omega|$.

COROLLARY 9.3 (Besov norm of products of functions). *Let $s > 0$ and $0 < p, q \le \infty$ satisfy $s > d/p$ (i.e. $B^s_q(L_p(\Omega)) \subset L_\infty(\Omega)$). If $u_i \in B^s_q(L_p(\Omega))$ for $1 \le i \le m$, then*

49

$\prod_{i=1}^{m} u_i \in B_q^s(L_p(\Omega))$ and then there exists a constant $C(\Omega)$ such that

$$\Big\| \prod_{i=1}^{m} u_i \Big\|_{B_q^s(L_p(\Omega))} \leq C(\Omega) \prod_{i=1}^{m} \|u_i\|_{B_q^s(L_p(\Omega))}$$

Proof. Since $s\ell_i - d/(p/\ell_i) = \ell_i(s - d/p) < s - d/p$ for $0 < \ell_i < 1$, we deduce $B_q^s(L_p(\Omega)) \subset B_{q'}^{s\ell_i}(L_{p/\ell_i}(\Omega))$ for any $0 < q, q' \leq \infty$. This, together with $B_q^s(L_p(\Omega)) \subset L_\infty(\Omega)$, enables us to replace the semi-norms in (9.4) by the full Besov norms absorbing the scaling into the constant $C(\Omega)$. $\square$

9.2. Composition of Functions. The following result is a scale-invariant generalization of a result in [10] related to the Besov regularity of the composition of functions.

LEMMA 9.4 (Scale-invariant Besov semi-norm of composition). *Let* $u : \Omega \to \mathbb{R}$ *be of class* $B_q^s(L_p(\Omega))$ *with* $0 < p, q \leq \infty$ *and* $s > d/p$, *and let* R *be a closed interval in* $\mathbb{R}$ *that contains the range of* u. *If* $f \in C^k(R)$, *with* $k > s$, *then the composite function* $f \circ u \in B_q^s(L_p(\Omega))$ *and there exists a constant* $C(f)$ *depending on* $\max_{1 \leq j \leq k} \|f^{(j)}\|_{L_\infty(R)}$ *such that*

$$|f \circ u|_{B_q^s(L_p(\Omega))} \preccurlyeq C(f) \sum_{\ell=1}^{k} \sum_{i=1}^{\ell} \|u\|_{L_\infty(\Omega)}^{\ell-i} \sum_{\sum_{j=1}^{i} \ell_j = 1} \prod_{j=1}^{i} |u|_{B_q^{s\ell_j}(L_{p/\ell_j}(\Omega))}, \qquad (9.6)$$

where the inner sum ranges over all choices of fractions of the form $0 \leq \ell_j = m_j/k^{i-1} \leq 1$ *with* $m_j \in \mathbb{N}_0$ *such that* $\sum_{j=1}^{i} m_j = k^{i-1}$.

Proof. Recall formula (9.1) and notice that $\Delta_h^k 1 = 0$. Then for $x \in \Omega_{kh}$,

$$\Delta_h^k (f \circ u)(x) = \sum_{j=1}^{k} \binom{k}{j} (-1)^{k+j} \big[f(u(x + jh)) - f(u(x)) \big].$$

Using Taylor's formula

$$f(u(x + jh)) - f(u(x)) = \sum_{\ell=1}^{k-1} \frac{f^{(\ell)}(u(x))}{\ell!} \big(\Delta_{jh}^1 u(x) \big)^\ell$$

$$+ \int_0^1 \frac{f^{(k)} \big(u(x) + t\Delta_{jh}^1 u(x) \big)}{(k-1)!} (1-t)^{k-1} dt \, \big(\Delta_{jh}^1 u(x) \big)^k.$$

Therefore, $\|\Delta_h^k (f \circ u)\|_{L_p(\Omega)} \leq \Big\| \sum_{\ell=1}^{k} I_\ell \Big\|_{L_p(\Omega)}$ where

$$I_\ell(x) := \sum_{j=1}^{k} \binom{k}{j} (-1)^{k+j} \frac{f^{(\ell)}(u(x))}{\ell!} \big(\Delta_{jh}^1 u(x) \big)^\ell \quad 1 \leq \ell < k,$$

$$I_k(x) := \sum_{j=1}^{k} \binom{k}{j} (-1)^{k+j} \int_0^1 \frac{f^{(k)} \big(u(x) + t\Delta_{jh}^1 u(x) \big)}{(k-1)!} (1-t)^{k-1} dt \, \big(\Delta_{jh}^1 u(x) \big)^k.$$

In order to bound the terms corresponding to $\ell < k$ we use Newton's binomial formula:

$$
\begin{aligned}
I_\ell &= \sum_{j=1}^{k} \binom{k}{j} (-1)^{k+j} \frac{f^{(\ell)}(u(x))}{\ell!} \sum_{i=0}^{\ell} \binom{\ell}{i} (-1)^{\ell-i} u(x+jh)^i u(x)^{\ell-i} \\
&= \sum_{i=0}^{\ell} \binom{\ell}{i} (-1)^{\ell-i} \frac{f^{(\ell)}(u(x))}{\ell!} \sum_{j=1}^{k} \binom{k}{j} (-1)^{k+j} u(x+jh)^i u(x)^{\ell-i} \\
&= \sum_{i=1}^{\ell} \binom{\ell}{i} (-1)^{\ell-i} \frac{f^{(\ell)}(u(x))}{\ell!} \Delta_h^k u^i(x) u(x)^{\ell-i},
\end{aligned}
$$

because $\Delta_h^k u^0 = 0$. Consequently,

$$
\left\| \sum_{\ell=1}^{k-1} I_\ell \right\|_{L_p(\Omega)} \le C(f) \sum_{\ell=1}^{k-1} \sum_{i=1}^{\ell} \| \Delta_h^k u^i \|_{L_p(\Omega)} \| u \|_{L_\infty(\Omega)}^{\ell-i}.
$$

A similar formula is valid for I_k, whence

$$
\| \Delta_h^k (f \circ u) \|_{L_p(\Omega)} \le C(f) \sum_{\ell=1}^{k} \sum_{i=1}^{\ell} \| \Delta_h^k u^i \|_{L_p(\Omega)} \| u \|_{L_\infty(\Omega)}^{\ell-i}.
$$

The modulus of smoothness $\omega_k(f \circ u, t)_p$ in turn satisfies

$$
\omega_k(f \circ u, t)_p = \sup_{|h| \le t} \| \Delta_h^k f \circ u \|_{L_p(\Omega)} \le C(f) \sum_{\ell=1}^{k} \sum_{i=1}^{\ell} \omega_k(u^i, t)_p \| u \|_{L_\infty(\Omega)}^{\ell-i}.
$$

Consequently, the Besov seminorm satisfies

$$
\begin{aligned}
|f \circ u|_{B_q^s(L_p(\Omega))} &:= \left(\int_0^\infty t^{-sq} \omega_k(f \circ u, t)_p^q \frac{dt}{t} \right)^{1/q} \\
&\precsim C(f) \sum_{\ell=1}^{k} \sum_{i=1}^{\ell} \left(\int_0^\infty t^{-sp} \omega_k(u^i, t)_p^q \frac{dt}{t} \right)^{1/q} \| u \|_{L_\infty(\Omega)}^{\ell-i} \\
&\precsim C(f) \sum_{\ell=1}^{k} \sum_{i=1}^{\ell} |u^i|_{B_q^s(L_p(\Omega))} \| u \|_{L_\infty(\Omega)}^{\ell-i}.
\end{aligned}
$$

Employing Corollary 9.2, we eventually infer the desired bound (9.6). $\square$

The inequality (9.6) is scale-invariant and, as such, can be used at the element level. However, it can be further simplified at the expense of having a constant depending on $|\Omega|$.

COROLLARY 9.5 (Besov norm of composition). *Under the assumptions of Lemma 9.4, there exists a constant $C(f, \Omega)$ such that*

$$
\begin{aligned}
|f \circ u|_{B_q^s(L_p(\Omega))} &\precsim C(f, \Omega) \sum_{\ell=1}^{k} \| u \|_{B_q^s(L_p(\Omega))}^\ell \\
&\le C(f, \Omega) \, k \, \max \left\{ \| u \|_{B_q^s(L_p(\Omega))}, \| u \|_{B_q^s(L_p(\Omega))}^k \right\}.
\end{aligned}
\tag{9.7}
$$

Proof. It suffices to use the embeddings $B_q^s(L_p(\Omega)) \subset B_q^{s\ell_j}(L_{p/\ell_j}(\Omega))$ for $0 < \ell_j < 1$ as well as $B_q^{s\ell_j}(L_{p/\ell_j}(\Omega)) \subset L_\infty(\Omega)$, which are valid because $s > d/p$, to convert (9.6) into (9.7). $\square$

Glossary. No glossary available in the arxiv version.

REFERENCES

[1] R. A. Adams and J. J. F. Fournier. *Sobolev spaces*, volume 140 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, second edition, 2003.

[2] M. Ainsworth and J. T. Oden. *A posteriori error estimation in finite element analysis*. Pure and Applied Mathematics (New York). Wiley-Interscience [John Wiley & Sons], New York, 2000.

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

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

[5] M. Sh. Birman and M.Z. Solomyak. Piecewise-polynomial approximations of functions of the classes W_p^α. *Mat. Sb. (N.S.)*, 73(115)(3):331–355, 1967.

[6] A. Bonito, J. M. Cascón, P. Morin, and R. H. Nochetto. AFEM for geometric PDE: the Laplace-Beltrami operator. In *Analysis and numerics of partial differential equations*, volume 4 of *Springer INdAM Ser.*, pages 257–306. Springer, Milan, 2013.

[7] A. Bonito, R. A. DeVore, and R. H. Nochetto. Adaptive finite element methods for elliptic problems with discontinuous coefficients. *SIAM J. Numer. Anal.*, 51(6):3106–3134, 2013.

[8] A. Bonito and R. H. Nochetto. Quasi-optimal convergence rate of an adaptive discontinuous Galerkin method. *SIAM J. Numer. Anal.*, 48(2):734–771, 2010.

[9] A. Bonito and J.E. Pasciak. Convergence analysis of variational and non-variational multigrid algorithm for the Laplace-Beltrami operator. *Math. Comp.*, 81:1263–1288, 2012.

[10] G. Bourdaud and W. Sickel. Composition operators on function spaces with fractional order of smoothness. In *Harmonic analysis and nonlinear partial differential equations*, RIMS Kôkyûroku Bessatsu, B26, pages 93–132. Res. Inst. Math. Sci. (RIMS), Kyoto, 2011.

[11] C. Carstensen, M. Feischl, M. Page, and D. Praetorius. Axioms of adaptivity. *Comput. Math. Appl.*, 67(6):1195–1253, 2014.

[12] 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(5):2524–2550, 2008.

[13] J. M. Cascón and R. H. Nochetto. Quasioptimal cardinality of AFEM driven by nonresidual estimators. *IMA J. Numer. Anal.*, 32(1):1–29, 2012.

[14] P. G. Ciarlet. *The finite element method for elliptic problems*. North-Holland Publishing Co., Amsterdam, 1978. Studies in Mathematics and its Applications, Vol. 4.

[15] A. Cohen, W. Dahmen, I. Daubechies, and R. DeVore. Tree approximation and optimal encoding. *Appl. Comput. Harmon. Anal.*, 11(2):192–226, 2001.

[16] A. Demlow. Higher-order finite element methods and pointwise error estimates for elliptic problems on surfaces. *SIAM J. Numer. Anal.*, 47(2):805–827, 2009.

[17] A. Demlow and G. Dziuk. An adaptive finite element method for the Laplace-Beltrami operator on implicitly defined surfaces. *SIAM J. Numer. Anal.*, 45(1):421–442 (electronic), 2007.

[18] R. A. DeVore and G. G. Lorentz. *Constructive approximation*, volume 303 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1993.

[19] W. Dörfler. A convergent adaptive algorithm for Poisson's equation. *SIAM J. Numer. Anal.*, 33(3):1106–1124, 1996.

[20] G. Dziuk. Finite elements for the Beltrami operator on arbitrary surfaces. In *Partial differential equations and calculus of variations*, volume 1357 of *Lecture Notes in Math.*, pages 142–155. Springer, Berlin, 1988.

[21] T. Gantumur. Convergence rates of adaptive methods, besov spaces, and multilevel approximation. arXiv:1408.3889, 2015.

[22] E. M. Garau and P. Morin. Convergence and quasi-optimality of adaptive FEM for Steklov eigenvalue problems. *IMA J. Numer. Anal.*, 31(3):914–946, 2011.

[23] F. D. Gaspoz and P. Morin. Approximation classes for adaptive higher order finite element approximation. *Math. Comp.*, 83(289):2127–2160, 2014.

[24] R. Kornhuber and H. Yserentant. Multigrid methods for discrete elliptic problems on triangular surfaces. *Comput. Vis. Sci.*, 11(4-6):251–257, 2008.

[25] K. Mekchay, P. Morin, and R. H. Nochetto. AFEM for the Laplace-Beltrami operator on graphs: design and conditional contraction property. *Math. Comp.*, 80(274):625–648, 2011.

[26] P. Morin, R. H. Nochetto, and K. G. Siebert. Convergence of adaptive finite element methods.

SIAM Rev., 44(4):631–658 (electronic) (2003), 2002. Revised reprint of "Data oscillation and convergence of adaptive FEM" [SIAM J. Numer. Anal. **3**8 (2000), no. 2, 466–488 (electronic); MR1770058 (2001g:65157)].

[27] P. Morin, R.H. Nochetto, and K. G. Siebert. Data oscillation and convergence of adaptive FEM. *SIAM J. Numer. Anal.*, 38(2):466–488 (electronic), 2000.

[28] R. H. Nochetto, K. G. Siebert, and A. Veeser. Theory of adaptive finite element methods: an introduction. In *Multiscale, nonlinear and adaptive approximation*, pages 409–542. Springer, Berlin, 2009.

[29] L. R. Scott and S. Zhang. Finite element interpolation of nonsmooth functions satisfying boundary conditions. *Math. Comp.*, 54(190):483–493, 1990.

[30] R. Stevenson. Optimality of a standard adaptive finite element method. *Found. Comput. Math.*, 7(2):245–269, 2007.

[31] R. Stevenson. The completion of locally refined simplicial partitions created by bisection. *Math. Comp.*, 77(261):227–241 (electronic), 2008.

[32] H. Triebel. Function spaces in Lipschitz domains and on Lipschitz manifolds. Characteristic functions as pointwise multipliers. *Rev. Mat. Complut.*, 15(2):475–524, 2002.

[33] A. Veeser. Approximating gradients with continuous piecewise polynomial functions. *Found. Comput. Math.*, pages 1–28, 2015.

[34] R. Verfürth. *A Review of A Posteriori Error Estimation and Adaptive Mesh-Refinement Technique*. Wiley-Teubner, Chichester, 1996.