SIAM J. NUMER. ANAL.
Vol. 57, No. 5, pp. 2359–2382

CONVERGENCE OF ADAPTIVE STOCHASTIC GALERKIN FEM*

ALEX BESPALOV†, DIRK PRAETORIUS‡, LEONARDO ROCCHI†, AND MICHELE RUGGERI§

Abstract. We propose and analyze novel adaptive algorithms for the numerical solution of elliptic partial differential equations with parametric uncertainty. Four different marking strategies are employed for refinement of stochastic Galerkin finite element approximations. The algorithms are driven by the energy error reduction estimates derived from two-level a posteriori error indicators for spatial approximations and hierarchical a posteriori error indicators for parametric approximations. The focus of this work is on the mathematical foundation of the adaptive algorithms in the sense of rigorous convergence analysis. In particular, we prove that the proposed algorithms drive the underlying energy error estimates to zero.

Key words. adaptive methods, a posteriori error analysis, two-level error estimate, stochastic Galerkin methods, finite element methods, parametric PDEs

AMS subject classifications. 35R60, 65C20, 65N12, 65N15, 65N30

DOI. 10.1137/18M1229560

1. Introduction. The design and analysis of adaptive algorithms for the numerical solution of partial differential equations (PDEs) with parametric or uncertain inputs have been active research themes in the last decade. Adaptive algorithms are indispensable when solving a particularly challenging class of parametric problems represented by PDEs whose inputs depend on infinitely many uncertain parameters. For this class of problems, adaptive algorithms have been shown, on the one hand, to yield approximations that are immune to the curse of dimensionality and, on the other hand, to outperform standard sampling methods (see [11, 12]).

It is well known in the finite element community that adaptive strategies based on rigorous a posteriori error analysis of computed solutions provide an effective mechanism for building approximation spaces and accelerating convergence. Several adaptive strategies of this type have been proposed in the context of stochastic Galerkin finite element method (sGFEM) for PDE problems with parametric or uncertain inputs. Typically, they are developed by extending the a posteriori error estimation techniques commonly used for deterministic problems to parametric settings. For example, dual-based a posteriori error estimates are employed in [27]; implicit error estimators (in the spirit of [1]) are used in [32] for the sGFEM based on multi-element generalized polynomial chaos expansions; explicit residual-based a posteriori error estimators provide spatial and stochastic error indicators for adaptive refinement in [24, 15, 16]; local equilibration error estimators are utilized in [17]; and hierarchi-

*Received by the editors December 4, 2018; accepted for publication (in revised form) July 25, 2019; published electronically October 3, 2019.
https://doi.org/10.1137/18M1229560
Funding: The work of the first author was supported by the EPSRC under grant EP/P013791/1 and by The Alan Turing Institute under the EPSRC grant EP/N510129/1. The work of the second and fourth authors was supported by the Austrian Science Fund (FWF) under grants W1245 and F65.
†School of Mathematics, University of Birmingham, Birmingham, UK (a.bespalov@bham.ac.uk, leonardo.rocchi@yahoo.it).
‡Institute for Analysis and Scientific Computing, TU Wien, Vienna, Austria (dirk.praetorius@asc.tuwien.ac.at).
§Faculty of Mathematics, University of Vienna, Vienna, Austria (michele.ruggeri@univie.ac.at).

cal error estimators and the associated estimates of error reduction drive adaptive algorithms proposed in [8, 6, 5, 13, 26].

In contrast to the design of algorithms, convergence analysis of adaptive sGFEM is much less developed. In [16], convergence of the adaptive algorithm driven by residual-based error estimators is proved in the spirit of the convergence analysis for deterministic FEM in [10]; moreover, the quasi-optimality of the generated sequence of meshes, in a suitable sense, is established. The analysis in [16], however, requires that the adaptive algorithm enforces additional spatial refinements during the iterations where parametric enrichment is performed (see [16, section 6]). This is caused by a purely theoretical artifact associated with using inverse estimates for the residual-based error estimators (see [16, section 6.1]).

In this paper, we study convergence of adaptive algorithms which are driven by the energy error reduction estimates derived from two-level a posteriori error indicators for spatial approximations and hierarchical a posteriori error indicators for parametric approximations. The underlying a posteriori error estimate that combines these two types of indicators has been recently introduced and analyzed in [5]. We employ four practical marking criteria which are combinations of Dörfler [14] and maximum [2] marking strategies. At each step, the algorithm performs either solely mesh refinement or solely polynomial enrichment. Our central result in Theorem 5 shows that each proposed adaptive algorithm generates a sequence of Galerkin approximations such that the corresponding sequence of energy error estimates converges to zero. Therefore, this result provides a theoretical guarantee that, for any given positive tolerance, the algorithms stop after a finite number of iterations. We note in Remark 6 that the proof of Theorem 5 is given for more general marking strategies, which are inspired by [29, section 2.2]. As an immediate consequence of Theorem 5, we show that, under the saturation assumption, the Galerkin approximations generated by the algorithms converge to the true parametric solution (Corollary 7). Additionally, in the case of Dörfler marking, we prove linear convergence of the energy error in Theorem 8.

We note that, although the results in this paper are presented for a simple model problem—steady-state diffusion equation whose coefficient has affine dependence on infinitely many parameters—our analysis will apply to more general elliptic linear problems with affine-parametric inputs (e.g., to linear elasticity models; see [26]). Furthermore, in practical applications, the goal of the simulation can be a specific quantity of interest, represented by a (non)linear functional of the solution to a PDE problem. In these cases, it is important to design adaptive strategies that directly reduce errors in the goal functional—the so-called goal-oriented adaptivity (see, e.g., [20, 21, 19] for a sampling-based approach to adaptive computation of empirical distribution functions and [27, 9, 5] for adaptive goal-oriented sGFEMs). While the present work focuses on approximating solutions in the energy norm, we expect that the main ideas will apply in the context of goal-oriented adaptivity (see, e.g., [23, 22] for the deterministic case).

The paper is organized as follows. Section 2 introduces the parametric model problem and its weak formulation. In section 3, we introduce the approximation spaces, define sGFEM formulations, and recall the a posteriori error estimates derived in [5]. In section 4, we present adaptive algorithms with four different marking criteria and formulate the main results of this work. The results of numerical experiments are reported in section 5, where, in particular, we compare the computational cost associated with employing different marking criteria. Technical details and the proofs of theorems are given in sections 6–8.

2. Parametric model problem. Let $D \subset \mathbb{R}^d$ ($d = 2, 3$) be a bounded Lipschitz domain with polytopal boundary ∂D and let $\Gamma := \prod_{m=1}^{\infty}[-1, 1]$ denote the infinitely dimensional hypercube. We consider the elliptic boundary value problem

$$\begin{aligned} -\nabla \cdot (a\nabla u) &= f \qquad \text{in } D \times \Gamma, \\ u &= 0 \qquad \text{on } \partial D \times \Gamma, \end{aligned} \tag{2.1}$$

where the scalar coefficient a and the right-hand side function f (and, hence, the solution u) depend on a countably infinite number of scalar parameters, i.e., $a = a(x, \mathbf{y})$, $f = f(x, \mathbf{y})$, and $u = u(x, \mathbf{y})$ with $x \in D$ and $\mathbf{y} \in \Gamma$. For the coefficient a, we assume linear dependence on the parameters, i.e.,

$$a(x, \mathbf{y}) = a_0(x) + \sum_{m=1}^{\infty} y_m a_m(x) \quad \text{for } x \in D \text{ and } \mathbf{y} = (y_m)_{m\in\mathbb{N}} \in \Gamma, \tag{2.2}$$

whereas for the right-hand side of (2.1) we assume that $f \in L^2_\pi(\Gamma; H^{-1}(D))$. Here, $\pi = \pi(\mathbf{y})$ is a probability measure on $(\Gamma, \mathcal{B}(\Gamma))$ with $\mathcal{B}(\Gamma)$ being the Borel σ-algebra on Γ, and we assume that $\pi(\mathbf{y})$ is the product of symmetric Borel probability measures π_m on $[-1, 1]$, i.e., $\pi(\mathbf{y}) = \prod_{m=1}^{\infty} \pi_m(y_m)$. The scalar functions $a_m \in W^{1,\infty}(D)$ ($m \in \mathbb{N}_0$) in (2.2) are required to satisfy the following inequalities

$$0 < a_0^{\min} \le a_0(x) \le a_0^{\max} < \infty \quad \text{for almost all } x \in D \tag{2.3}$$

$$\text{and} \quad \tau := \frac{1}{a_0^{\min}} \sum_{m=1}^{\infty} \|a_m\|_{L^\infty(D)} < 1. \tag{2.4}$$

With the Sobolev space $\mathbb{X} := H^1_0(D)$, consider the Bochner space $\mathbb{V} := L^2_\pi(\Gamma; \mathbb{X})$. On $\mathbb{V}$, define the bilinear forms

$$\begin{aligned} B_0(u, v) &:= \int_\Gamma \int_D a_0(x) \nabla u(x, \mathbf{y}) \cdot \nabla v(x, \mathbf{y}) \,\mathrm{d}x \,\mathrm{d}\pi(\mathbf{y}), \\ B(u, v) &:= B_0(u, v) + \sum_{m=1}^{\infty} \int_\Gamma \int_D y_m a_m(x) \nabla u(x, \mathbf{y}) \cdot \nabla v(x, \mathbf{y}) \,\mathrm{d}x \,\mathrm{d}\pi(\mathbf{y}). \end{aligned}$$

An elementary computation shows that assumptions (2.2)–(2.4) ensure that the bilinear forms $B_0(\cdot, \cdot)$ and $B(\cdot, \cdot)$ are symmetric, continuous, and elliptic on $\mathbb{V}$. Let $|\!|\!| \cdot |\!|\!|$ (resp., $|\!|\!| \cdot |\!|\!|_0$) denote the norm induced by $B(\cdot, \cdot)$ (resp., $B_0(\cdot, \cdot)$). Then, there holds

$$\lambda \, |\!|\!| v |\!|\!|^2 \le |\!|\!| v |\!|\!|_0^2 \le \Lambda \, |\!|\!| v |\!|\!|^2 \quad \text{for all } v \in \mathbb{V}, \tag{2.5}$$

where $0 < \lambda := \frac{a_0^{\min}}{a_0^{\max}(1+\tau)} < 1 < \Lambda := \frac{a_0^{\max}}{a_0^{\min}(1-\tau)} < \infty$.

The parametric problem (2.1) is understood in the weak sense: Given $f \in L^2_\pi(\Gamma; H^{-1}(D))$, find $u \in \mathbb{V}$ such that

$$B(u, v) = F(v) := \int_\Gamma \int_D f(x, \mathbf{y}) v(x, \mathbf{y}) \,\mathrm{d}x \,\mathrm{d}\pi(\mathbf{y}) \quad \text{for all } v \in \mathbb{V}. \tag{2.6}$$

The well-posedness of (2.6) follows by the Riesz theorem.

3. Finite element discretization and a posteriori error analysis.

3.1. Approximation spaces. Let $\mathcal{T}_\bullet$ be a mesh, i.e., a conforming triangulation of D into compact nondegenerate simplices $T \in \mathcal{T}_\bullet$. Let $\mathcal{E}_\bullet$ be the corresponding set of facets. Let $\mathcal{E}_\bullet^{\text{int}} \subset \mathcal{E}_\bullet$ be the set of interior facets, i.e., for each $E \in \mathcal{E}_\bullet^{\text{int}}$, there exist unique $T, T' \in \mathcal{T}_\bullet$ such that $E = T \cap T'$. Let $\mathcal{N}_\bullet$ be the set of vertices of $\mathcal{T}_\bullet$. For $z \in \mathcal{N}_\bullet$, let $\varphi_{\bullet,z}$ be the associated hat function, i.e., $\varphi_{\bullet,z}$ is piecewise affine, globally continuous, and satisfies the Kronecker property $\varphi_{\bullet,z}(z') = \delta_{zz'}$ for all $z' \in \mathcal{N}_\bullet$. We consider the space of continuous piecewise linear finite elements

$$\mathbb{X}_\bullet := \mathcal{S}_0^1(\mathcal{T}_\bullet) := \{v_\bullet \in \mathbb{X} : v_\bullet|_T \text{ is affine for all } T \in \mathcal{T}_\bullet\} \subset \mathbb{X} = H_0^1(D).$$

Recall that $\{\varphi_{\bullet,z} : z \in \mathcal{N}_\bullet \setminus \partial D\}$ is the standard basis of $\mathbb{X}_\bullet$.

Let us now introduce the polynomial spaces on Γ. For each $m \in \mathbb{N}$, let $(P_n^m)_{n\in\mathbb{N}_0}$ denote the sequence of univariate polynomials which are orthogonal with respect to π_m such that P_n^m is a polynomial of degree $n \in \mathbb{N}_0$ with $\|P_n^m\|_{L^2_{\pi_m}(-1,1)} = 1$ and $P_0^m \equiv 1$. It is well known that $\{P_n^m : n \in \mathbb{N}_0\}$ is an orthonormal basis of $L^2_{\pi_m}(-1,1)$. With $\mathbb{N}_0^{\mathbb{N}} := \{\nu = (\nu_m)_{m\in\mathbb{N}} : \nu_m \in \mathbb{N}_0 \text{ for all } m \in \mathbb{N}\}$ and $\operatorname{supp}(\nu) := \{m \in \mathbb{N} : \nu_m \neq 0\}$, let $\mathfrak{I} := \{\nu \in \mathbb{N}_0^{\mathbb{N}} : \#\operatorname{supp}(\nu) < \infty\}$ be the set of finitely supported multi-indices. Note that $\mathfrak{I}$ is countable. With

$$P_\nu(\mathbf{y}) := \prod_{m\in\mathbb{N}} P_{\nu_m}^m(y_m) = \prod_{m\in\operatorname{supp}(\nu)} P_{\nu_m}^m(y_m) \quad \text{for all } \nu \in \mathfrak{I} \text{ and all } \mathbf{y} \in \Gamma,$$

the set $\{P_\nu : \nu \in \mathfrak{I}\}$ is an orthonormal basis of $\mathbb{P} := L^2_\pi(\Gamma)$; see [30, Theorem 2.12].

The Bochner space $\mathbb{V} = L^2_\pi(\Gamma; \mathbb{X})$ is isometrically isomorphic to $\mathbb{X} \otimes \mathbb{P}$ and each function $v \in \mathbb{V}$ can be represented in the form

$$v(x, \mathbf{y}) = \sum_{\nu\in\mathfrak{I}} v_\nu(x) P_\nu(\mathbf{y}) \quad \text{with unique coefficients } v_\nu \in \mathbb{X}. \tag{3.1}$$

Moreover, there holds (see [5, Lemma 2.1])

$$B_0(v, w) = \sum_{\nu\in\mathfrak{I}} \int_D a_0(x)\, \nabla v_\nu(x) \cdot \nabla w_\nu(x)\, \mathrm{d}x \quad \text{for all } v, w \in \mathbb{V} \tag{3.2}$$

and, in particular,

$$\|v\|_0^2 = \sum_{\nu\in\mathfrak{I}} \|a_0^{1/2} \nabla v_\nu\|_{L^2(D)}^2 = \sum_{\nu\in\mathfrak{I}} \|v_\nu P_\nu\|_0^2 \quad \text{for all } v \in \mathbb{V}. \tag{3.3}$$

Let $\mathbf{0} = (0, 0, \dots)$ and consider a finite index set $\mathfrak{P}_\bullet \subset \mathfrak{I}$ with $\mathbf{0} \in \mathfrak{P}_\bullet$. We denote by $\operatorname{supp}(\mathfrak{P}_\bullet) := \bigcup_{\nu\in\mathfrak{P}_\bullet} \operatorname{supp}(\nu)$ the set of active parameters in $\mathfrak{P}_\bullet$.

Our discretization of (2.6) is based on the finite-dimensional tensor-product space

$$\mathbb{V}_\bullet := \mathbb{X}_\bullet \otimes \mathbb{P}_\bullet \subset \mathbb{X} \otimes \mathbb{P} = \mathbb{V} \quad \text{with} \quad \mathbb{P}_\bullet := \operatorname{span}\{P_\nu : \nu \in \mathfrak{P}_\bullet\} \subset \mathbb{P} = L^2_\pi(\Gamma).$$

The Galerkin discretization of (2.6) reads as follows: Find $u_\bullet \in \mathbb{V}_\bullet$ such that

$$B(u_\bullet, v_\bullet) = F(v_\bullet) \quad \text{for all } v_\bullet \in \mathbb{V}_\bullet. \tag{3.4}$$

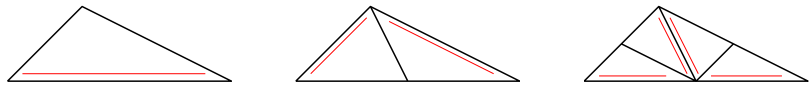

FIG. 1. *For NVB in two dimensions (2D), each triangle $T \in \mathcal{T}_\bullet$ has one* reference edge, *indicated by the double line (left). Bisection of T is achieved by halving the reference edge. The reference edges of the sons are always opposite to the new vertex (middle). Recursive application of this rule leads to conforming meshes. After three bisections per element all edges of a triangle are halved (right). If all elements $T \in \mathcal{T}_\bullet$ are refined by three bisections, the resulting uniform refinement is conforming.*

3.2. Mesh refinement and parametric enrichment. For mesh refinement, we employ newest vertex bisection (NVB); see Figure 1 for $d = 2$ and, e.g., [18, Figure 2] for $d = 3$ as well as [31, 25]. We assume that any mesh $\mathcal{T}_\bullet$ employed for the spatial discretization can be obtained by applying NVB refinement(s) to a given initial mesh $\mathcal{T}_0$. For a given mesh $\mathcal{T}_\bullet$, let $\widehat{\mathcal{T}}_\bullet$ be the coarsest mesh obtained from $\mathcal{T}_\bullet$ such that: (i) for $d = 2$, all edges of $\mathcal{T}_\bullet$ have been bisected once (i.e., uniform refinement of all elements by three bisections; see Figure 1); (ii) for $d = 3$, all faces contain an interior vertex (see [18, Figure 3] and the associated discussion therein). Then $\widehat{\mathcal{N}}_\bullet$ denotes the set of vertices of $\widehat{\mathcal{T}}_\bullet$ and $\{\widehat{\varphi}_{\bullet,z} : z \in \widehat{\mathcal{N}}_\bullet\}$ is the corresponding set of hat functions. The finite element space associated with $\widehat{\mathcal{T}}_\bullet$ is denoted by $\widehat{\mathbb{X}}_\bullet := \mathcal{S}_0^1(\widehat{\mathcal{T}}_\bullet)$. With $\mathcal{N}_\bullet^+ := (\widehat{\mathcal{N}}_\bullet \setminus \mathcal{N}_\bullet) \setminus \partial D$ being the set of new interior vertices created by uniform refinement of $\mathcal{T}_\bullet$, one has $\widehat{\mathbb{X}}_\bullet = \mathbb{X}_\bullet \oplus \operatorname{span}\{\widehat{\varphi}_{\bullet,z} : z \in \mathcal{N}_\bullet^+\}$. For a later use, we note that there exists a constant $K \geq 1$ depending only on the initial mesh $\mathcal{T}_0$ such that

$$\#\{z \in \mathcal{N}_\bullet^+ : |T \cap \operatorname{supp}(\widehat{\varphi}_{\bullet,z})| > 0\} \leq K < \infty \quad \text{for all } T \in \mathcal{T}_\bullet. \tag{3.5}$$

For a set of marked vertices $\mathcal{M}_\bullet \subseteq \mathcal{N}_\bullet^+$, let $\mathcal{T}_\circ := \operatorname{refine}(\mathcal{T}_\bullet, \mathcal{M}_\bullet)$ be the coarsest mesh such that $\mathcal{M}_\bullet \subseteq \mathcal{N}_\circ$, i.e., all marked vertices are vertices of $\mathcal{T}_\circ$. Since NVB is a binary refinement rule, this implies that $\mathcal{N}_\circ \subseteq \widehat{\mathcal{N}}_\bullet$ and $(\mathcal{N}_\circ \setminus \mathcal{N}_\bullet) \setminus \partial D = \mathcal{N}_\bullet^+ \cap \mathcal{N}_\circ$. In particular, the choices $\mathcal{M}_\bullet = \emptyset$ and $\mathcal{M}_\bullet = \mathcal{N}_\bullet^+$ lead to the meshes $\mathcal{T}_\bullet = \operatorname{refine}(\mathcal{T}_\bullet, \emptyset)$ and $\widehat{\mathcal{T}}_\bullet = \operatorname{refine}(\mathcal{T}_\bullet, \mathcal{N}_\bullet^+)$, respectively.

For parametric enrichment, we follow [8, 6, 5] and consider the *detail index set*

$$\mathfrak{Q}_\bullet := \{\mu \in \mathfrak{I} \setminus \mathfrak{P}_\bullet : \mu = \nu \pm \varepsilon_m \text{ for all } \nu \in \mathfrak{P}_\bullet \text{ and all } m = 1, \dots, M_{\mathfrak{P}_\bullet} + 1\}, \tag{3.6}$$

where $\varepsilon_m \in \mathfrak{I}$ denotes the mth unit sequence, i.e., $(\varepsilon_m)_i = \delta_{mi}$ for all $i \in \mathbb{N}$, and $M_{\mathfrak{P}_\bullet} \in \mathbb{N}$ is given by

$$M_{\mathfrak{P}_\bullet} := \begin{cases} 0 & \text{if } \mathfrak{P}_\bullet = \{\mathbf{0}\}, \\ \max\{\max(\operatorname{supp}(\nu)) : \nu \in \mathfrak{P}_\bullet \setminus \{\mathbf{0}\}\} & \text{otherwise.} \end{cases}$$

Then an enriched polynomial space $\mathbb{P}_\circ$ with $\mathbb{P}_\bullet \subset \mathbb{P}_\circ \subset \mathbb{P}$ can be obtained by adding some marked indices $\mathfrak{M}_\bullet \subseteq \mathfrak{Q}_\bullet$ to the current index set $\mathfrak{P}_\bullet$, i.e., $\mathbb{P}_\circ := \operatorname{span}\{P_\nu : \nu \in \mathfrak{P}_\circ\}$ with $\mathfrak{P}_\circ := \mathfrak{P}_\bullet \cup \mathfrak{M}_\bullet$. We denote by $\widehat{\mathbb{P}}_\bullet \subset \mathbb{P}$ the polynomial space obtained by adding to $\mathfrak{P}_\bullet$ all indices of $\mathfrak{Q}_\bullet$, i.e., $\widehat{\mathbb{P}}_\bullet := \operatorname{span}\{P_\nu : \nu \in \widehat{\mathfrak{P}}_\bullet\}$ with $\widehat{\mathfrak{P}}_\bullet := \mathfrak{P}_\bullet \cup \mathfrak{Q}_\bullet$.

The analysis of the forthcoming adaptive algorithm will also rely on the spaces

$$\widehat{\mathbb{V}}_\bullet := (\widehat{\mathbb{X}}_\bullet \otimes \mathbb{P}_\bullet) + (\mathbb{X}_\bullet \otimes \widehat{\mathbb{P}}_\bullet) \quad \text{and} \quad \widehat{\mathbb{V}}'_\bullet := \mathbb{X}_\bullet \otimes \widehat{\mathbb{P}}_\bullet. \tag{3.7}$$

3.3. A posteriori error estimation. In order to estimate the error due to spatial discretization, we employ the two-level error estimation strategy from [5]. Specifically, our spatial error estimate is given by

$$\eta_\bullet(\mathcal{N}_\bullet^+)^2 := \sum_{z\in\mathcal{N}_\bullet^+} \eta_\bullet(z)^2 \quad \text{with} \quad \eta_\bullet(z)^2 := \sum_{\nu\in\mathfrak{P}_\bullet} \frac{|F(\widehat{\varphi}_{\bullet,z}P_\nu) - B(u_\bullet, \widehat{\varphi}_{\bullet,z}P_\nu)|^2}{\|a_0^{1/2}\nabla\widehat{\varphi}_{\bullet,z}\|^2_{L^2(D)}}. \tag{3.8}$$

Remark 1. For $d = 2$, we have $\#\mathcal{N}_\bullet^+ = \#\mathcal{E}_\bullet^{\mathrm{int}}$, and the new degrees of freedom correspond to the midpoints of interior edges. Then, the spatial error estimate can be indexed by $E \in \mathcal{E}_\bullet^{\mathrm{int}}$ rather than by $z \in \mathcal{N}_\bullet^+$; see [5]. Furthermore, in this case, one has $K = 3$ in (3.5).

In order to estimate the error due to polynomial approximation on the parameter domain Γ, we employ the hierarchical error estimator from [3, 8]. First, for each $\nu \in \mathfrak{Q}_\bullet$, we define the estimator $e_\bullet^\nu \in \mathbb{X}_\bullet$ satisfying

$$B_0(e_\bullet^\nu P_\nu, v_\bullet P_\nu) = F(v_\bullet P_\nu) - B(u_\bullet, v_\bullet P_\nu) \quad \text{for all } v_\bullet \in \mathbb{X}_\bullet. \tag{3.9}$$

Then, the parametric error estimate is defined as follows:

$$\eta_\bullet(\mathfrak{Q}_\bullet)^2 := \sum_{\nu\in\mathfrak{Q}_\bullet} \eta_\bullet(\nu)^2 \quad \text{with} \quad \eta_\bullet(\nu) := \|a_0^{1/2}\nabla e_\bullet^\nu\|_{L^2(D)}. \tag{3.10}$$

From now on, for any $\mathcal{M}_\bullet \subseteq \mathcal{N}_\bullet^+$ and $\mathfrak{M}_\bullet \subseteq \mathfrak{Q}_\bullet$, we use the following notation:

$$\eta_\bullet(\mathcal{M}_\bullet)^2 := \sum_{z\in\mathcal{M}_\bullet} \eta_\bullet(z)^2, \quad \eta_\bullet(\mathfrak{M}_\bullet)^2 := \sum_{\nu\in\mathfrak{M}_\bullet} \eta_\bullet(\nu)^2,$$
$$\eta_\bullet(\mathcal{M}_\bullet, \mathfrak{M}_\bullet)^2 := \eta_\bullet(\mathcal{M}_\bullet)^2 + \eta_\bullet(\mathfrak{M}_\bullet)^2.$$

We define the overall error estimate as follows:

$$\eta_\bullet^2 := \eta_\bullet(\mathcal{N}_\bullet^+, \mathfrak{Q}_\bullet)^2 = \eta_\bullet(\mathcal{N}_\bullet^+)^2 + \eta_\bullet(\mathfrak{Q}_\bullet)^2. \tag{3.11}$$

Let us now consider the enriched space $\widehat{\mathbb{V}}_\bullet$ defined in (3.7). According to the Riesz theorem, there exists a unique $\widehat{u}_\bullet \in \widehat{\mathbb{V}}_\bullet$ such that

$$B(\widehat{u}_\bullet, \widehat{v}_\bullet) = F(\widehat{v}_\bullet) \quad \text{for all } \widehat{v}_\bullet \in \widehat{\mathbb{V}}_\bullet. \tag{3.12}$$

Since $\mathbb{V}_\bullet \subset \widehat{\mathbb{V}}_\bullet$, the Galerkin orthogonality implies that

$$|\!|\!| u - \widehat{u}_\bullet |\!|\!|^2 + |\!|\!| \widehat{u}_\bullet - u_\bullet |\!|\!|^2 = |\!|\!| u - u_\bullet |\!|\!|^2. \tag{3.13}$$

In [5, Theorem 3.1], we prove the following theorem for the overall error estimate $\eta_\bullet$. The main result is the estimate (3.14), while efficiency (3.15) and reliability (3.17) then follow easily from (3.13).

THEOREM 2. *There exists a constant $C_{\mathrm{thm}} \geq 1$, which depends only on the initial mesh $\mathcal{T}_0$ and the mean field a_0, such that*

$$\frac{\lambda}{K}\,\eta_\bullet^2 \leq |\!|\!| \widehat{u}_\bullet - u_\bullet |\!|\!|^2 \leq \Lambda C_{\mathrm{thm}}\,\eta_\bullet^2, \tag{3.14}$$

where λ, Λ are the constants in (2.5) *and K is the constant in* (3.5)*. In particular, there holds efficiency*

$$\frac{\lambda}{K}\,\eta_\bullet^2 \leq |\!|\!| \widehat{u}_\bullet - u_\bullet |\!|\!|^2 \overset{(3.13)}{\leq} |\!|\!| u - u_\bullet |\!|\!|^2. \tag{3.15}$$

Moreover, under the saturation assumption

$$\|u - \widehat{u}_\bullet\| \le q_{\rm sat} \|u - u_\bullet\| \quad \text{with some constant } 0 < q_{\rm sat} < 1, \tag{3.16}$$

there holds reliability

$$\|u - u_\bullet\|^2 \overset{(3.13)}{\le} \frac{1}{1 - q_{\rm sat}^2} \|\widehat{u}_\bullet - u_\bullet\|^2 \le \frac{\Lambda C_{\rm thm}}{1 - q_{\rm sat}^2} \eta_\bullet^2. \tag{3.17}$$

The proof of Theorem 2 given in [5] relies on the stable subspace decompositions

$$\widehat{\mathbb{X}}_\bullet = \mathbb{X}_\bullet \oplus \bigoplus_{z \in \mathcal{N}_\bullet^+} \operatorname{span}\{\widehat{\varphi}_{\bullet,z}\} \quad \text{and} \quad \widehat{\mathbb{P}}_\bullet = \mathbb{P}_\bullet \oplus \bigoplus_{\nu \in \mathfrak{Q}_\bullet} \operatorname{span}\{P_\nu\}.$$

For $d = 2$, the analysis in [5], in fact, proves a more general result than estimate (3.14). Let $\mathcal{T}_\circ = \operatorname{refine}(\mathcal{T}_\bullet, \mathcal{M}_\bullet)$ and consider $z \in (\mathcal{N}_\circ \setminus \mathcal{N}_\bullet) \setminus \partial D = \mathcal{N}_\bullet^+ \cap \mathcal{N}_\circ \subseteq \mathcal{N}_\bullet^+$. Let $\varphi_{\circ,z} \in \mathbb{X}_\circ$ and $\widehat{\varphi}_{\bullet,z} \in \widehat{\mathbb{X}}_\bullet$ be the corresponding hat functions. Then, two-dimensional NVB refinement ensures that $\varphi_{\circ,z} = \widehat{\varphi}_{\bullet,z}$, which yields the stable decomposition

$$\mathbb{X}_\circ = \mathbb{X}_\bullet \oplus \bigoplus_{z \in \mathcal{N}_\bullet^+ \cap \mathcal{N}_\circ} \operatorname{span}\{\varphi_{\circ,z}\} = \mathbb{X}_\bullet \oplus \bigoplus_{z \in \mathcal{N}_\bullet^+ \cap \mathcal{N}_\circ} \operatorname{span}\{\widehat{\varphi}_{\bullet,z}\}.$$

As a consequence, the analysis from [5] also proves the following result that allows one to control the error reduction due to *adaptive* enrichment of both components of the approximation space $\mathbb{V}_\bullet = \mathbb{X}_\bullet \otimes \mathbb{P}_\bullet$.

COROLLARY 3. *Let $d = 2$. Let $C_{\rm thm} \ge 1$ be the constant from Theorem 2. Suppose that $\mathcal{T}_\circ = \operatorname{refine}(\mathcal{T}_\bullet, \mathcal{M}_\bullet)$ and $\widehat{\mathcal{T}}_\bullet = \operatorname{refine}(\mathcal{T}_\bullet, \mathcal{N}_\bullet^+)$ are obtained by two-dimensional NVB refinement and $\mathfrak{P}_\circ = \mathfrak{P}_\bullet \cup \mathfrak{M}_\bullet$ for an index set $\mathfrak{M}_\bullet \subseteq \mathfrak{Q}_\bullet$. If $u_\bullet \in \mathbb{V}_\bullet$ and $u_\circ \in \mathbb{V}_\circ$ are two Galerkin approximations, then there holds*

$$\frac{\lambda}{K} \eta_\bullet\big(\mathcal{N}_\bullet^+ \cap \mathcal{N}_\circ, \mathfrak{M}_\bullet\big)^2 \le \|u_\circ - u_\bullet\|^2 \le \Lambda C_{\rm thm} \eta_\bullet\big(\mathcal{N}_\bullet^+ \cap \mathcal{N}_\circ, \mathfrak{M}_\bullet\big)^2. \tag{3.18}$$

4. Main results.

4.1. Adaptive algorithms. Let $\mathcal{T}_0$ be the initial mesh and let the initial index set $\mathfrak{P}_0$ contain only the zero index, i.e., $\mathfrak{P}_0 := \{\mathbf{0}\}$. The adaptive algorithm below generates a sequence $(\mathcal{T}_\ell)_{\ell \in \mathbb{N}_0}$ of adaptively refined meshes and a sequence $(\mathfrak{P}_\ell)_{\ell \in \mathbb{N}_0}$ of adaptively enriched index sets such that, for all $\ell \in \mathbb{N}_0$, there holds

$$\mathcal{T}_{\ell+1} = \operatorname{refine}(\mathcal{T}_\ell, \mathcal{M}_\ell) \text{ for some } \mathcal{M}_\ell \subseteq \mathcal{N}_\ell^+ \quad \text{and} \quad \mathfrak{P}_\ell \subseteq \mathfrak{P}_{\ell+1} \subseteq \widehat{\mathfrak{P}}_\ell = \mathfrak{P}_\ell \cup \mathfrak{Q}_\ell.$$

In particular, by the definition of the detail index set (3.6), one has $\mathfrak{Q}_\ell \setminus \mathfrak{P}_{\ell+1} \subseteq \mathfrak{Q}_{\ell+1}$ and $\widehat{\mathfrak{P}}_\ell \subseteq \widehat{\mathfrak{P}}_{\ell+1}$. Thus, the following inclusions hold:

$$\mathbb{X}_\ell \subseteq \mathbb{X}_{\ell+1} \subseteq \widehat{\mathbb{X}}_\ell \subset \mathbb{X} \quad \text{and} \quad \mathbb{P}_\ell \subseteq \mathbb{P}_{\ell+1} \subseteq \widehat{\mathbb{P}}_\ell \subseteq \widehat{\mathbb{P}}_{\ell+1} \subset \mathbb{P}.$$

Furthermore, since the adaptive algorithm presented below performs either mesh refinement or parametric enrichment at each iteration $\ell \in \mathbb{N}_0$, one of the inclusions $\mathbb{X}_\ell \subseteq \mathbb{X}_{\ell+1}$ or $\mathbb{P}_\ell \subseteq \mathbb{P}_{\ell+1}$ is strict. Therefore, recalling the definition of the enriched spaces $\widehat{\mathbb{V}}_\ell$ and $\widehat{\mathbb{V}}'_\ell$ (see (3.7)), we conclude that

$$\mathbb{V}_\ell \subset \widehat{\mathbb{V}}'_\ell \subset \widehat{\mathbb{V}}_\ell \subset \mathbb{V}, \quad \mathbb{V}_\ell \subset \mathbb{V}_{\ell+1}, \quad \widehat{\mathbb{V}}'_\ell \subset \widehat{\mathbb{V}}'_{\ell+1}, \quad \text{and } \widehat{\mathbb{V}}_\ell \subset \widehat{\mathbb{V}}_{\ell+1} \quad \text{for all } \ell \in \mathbb{N}_0.$$

We consider the following basic loop of an adaptive algorithm, where the precise marking strategy is still left open, but will be specified subsequently.

ALGORITHM 4. Input: $\mathcal{T}_0$, $\mathfrak{P}_0 = \{\mathbf{0}\}$, *marking criterion. Set* $\ell = 0$.
(i) *Compute discrete solution* $u_\ell \in \mathbb{V}_\ell$.
(ii) *Compute error indicators* $\eta_\ell(z)$ *and* $\eta_\ell(\nu)$ *for all* $z \in \mathcal{N}_\ell^+$ *and all* $\nu \in \mathfrak{Q}_\ell$.
(iii) *Use marking criterion to obtain* $\mathcal{M}_\ell \subseteq \mathcal{N}_\ell^+$ *and* $\mathfrak{M}_\ell \subseteq \mathfrak{Q}_\ell$.
(iv) *Set* $\mathfrak{P}_{\ell+1} = \mathfrak{P}_\ell \cup \mathfrak{M}_\ell$ *and* $\mathcal{T}_{\ell+1} = \mathrm{refine}(\mathcal{T}_\ell, \mathcal{M}_\ell)$.
(v) *Increase the counter* $\ell \mapsto \ell + 1$ *and continue with* (i).
Output: $(\mathcal{T}_\ell, \mathfrak{P}_\ell, u_\ell, \eta_\ell)_{\ell \in \mathbb{N}_0}$.

The criteria below specify four different marking strategies for step (iii) of Algorithm 4 and, at the same time, determine the type of enrichment for the next iteration of the algorithm. Each strategy comes with three parameters: $\vartheta > 0$ is a weight modulating the choice between mesh refinement and parametric enrichment (with parametric enrichment being favored for $\vartheta > 1$), $0 < \theta_{\mathbb{X}} \le 1$ controls the marking of nodes in $\mathcal{N}_\ell^+$ (always based on the Dörfler criterion), whereas $0 < \theta_{\mathbb{P}} \le 1$ controls the marking of indices in $\mathfrak{Q}_\ell$ (based on either the Dörfler criterion or the maximum criterion).

The first criterion enforces spatial refinement if the spatial error estimate is comparably large; otherwise, parametric enrichment is chosen for the next iteration. The marked facets (resp., marked indices) are obtained via Dörfler marking.

CRITERION A (see [15, 6]).
Input: *error indicators* $\{\eta_\ell(z) : z \in \mathcal{N}_\ell^+\}, \{\eta_\ell(\nu) : \nu \in \mathfrak{Q}_\ell\}$*; marking parameters* $0 < \theta_{\mathbb{X}}, \theta_{\mathbb{P}} \le 1$*, and* $\vartheta > 0$.
Case (a)*:* $\vartheta\, \eta_\ell(\mathfrak{Q}_\ell) \le \eta_\ell(\mathcal{N}_\ell^+)$.
– *Set* $\mathfrak{M}_\ell = \emptyset$.
– *Find* $\mathcal{M}_\ell \subseteq \mathcal{N}_\ell^+$ *with minimal cardinality such that* $\theta_{\mathbb{X}}\, \eta_\ell(\mathcal{N}_\ell^+) \le \eta_\ell(\mathcal{M}_\ell)$.
Case (b)*:* $\vartheta\, \eta_\ell(\mathfrak{Q}_\ell) > \eta_\ell(\mathcal{N}_\ell^+)$.
– *Find* $\mathfrak{M}_\ell \subseteq \mathfrak{Q}_\ell$ *with minimal cardinality such that* $\theta_{\mathbb{P}}\, \eta_\ell(\mathfrak{Q}_\ell) \le \eta_\ell(\mathfrak{M}_\ell)$.
– *Set* $\mathcal{M}_\ell = \emptyset$.
Output: $\mathcal{M}_\ell \subseteq \mathcal{N}_\ell^+$ *and* $\mathfrak{M}_\ell \subseteq \mathfrak{Q}_\ell$*, where one of the subsets is empty.*

Criterion B is based on the idea that the error estimate η_ℓ on the refined elements (resp., added indices) provides information about the associated error reduction (see Corollary 3). This criterion enforces either spatial refinement (if the error reduction for spatial mesh refinement is comparably large) or parametric enrichment (otherwise).

CRITERION B (see [6]).
Input: *error indicators* $\{\eta_\ell(z) : z \in \mathcal{N}_\ell^+\}$, $\{\eta_\ell(\nu) : \nu \in \mathfrak{Q}_\ell\}$*; marking parameters* $0 < \theta_{\mathbb{X}}, \theta_{\mathbb{P}} \le 1$*, and* $\vartheta > 0$.
– *Find* $\widetilde{\mathfrak{M}}_\ell \subseteq \mathfrak{Q}_\ell$ *with minimal cardinality such that* $\theta_{\mathbb{P}}\, \eta_\ell(\mathfrak{Q}_\ell) \le \eta_\ell(\widetilde{\mathfrak{M}}_\ell)$.
– *Find* $\widetilde{\mathcal{M}}_\ell \subseteq \mathcal{N}_\ell^+$ *with minimal cardinality such that* $\theta_{\mathbb{X}}\, \eta_\ell(\mathcal{N}_\ell^+) \le \eta_\ell(\widetilde{\mathcal{M}}_\ell)$.
– *Define* $\widetilde{\mathcal{R}}_\ell := \mathcal{N}_\ell^+ \cap \widetilde{\mathcal{N}}_\ell$*, where* $\widetilde{\mathcal{N}}_\ell$ *is associated with* $\widetilde{\mathcal{T}}_\ell = \mathrm{refine}(\mathcal{T}_\ell, \widetilde{\mathcal{M}}_\ell)$.
Case (a)*:* $\vartheta\, \eta_\ell(\widetilde{\mathfrak{M}}_\ell) \le \eta_\ell(\widetilde{\mathcal{R}}_\ell)$*. Set* $\mathfrak{M}_\ell = \emptyset$ *and* $\mathcal{M}_\ell = \widetilde{\mathcal{M}}_\ell$.
Case (b)*:* $\vartheta\, \eta_\ell(\widetilde{\mathfrak{M}}_\ell) > \eta_\ell(\widetilde{\mathcal{R}}_\ell)$*. Set* $\mathfrak{M}_\ell = \widetilde{\mathfrak{M}}_\ell$ *and* $\mathcal{M}_\ell = \emptyset$.
Output: $\mathcal{M}_\ell \subseteq \mathcal{N}_\ell^+$ *and* $\mathfrak{M}_\ell \subseteq \mathfrak{Q}_\ell$*, where one of the subsets is empty.*

Criterion C is a modification of Criterion A. It employs a maximum criterion in the parameter domain, while using Dörfler marking in the physical domain. As in Criterion A, the enrichment type is determined by the dominant contributing error estimate.

Criterion C.

Input: *error indicators* $\{\eta_\ell(z) : z \in \mathcal{N}_\ell^+\}$, $\{\eta_\ell(\nu) : \nu \in \mathfrak{Q}_\ell\}$*; marking parameters* $0 < \theta_{\mathbb{X}} \le 1$, $0 \le \theta_{\mathbb{P}} \le 1$, *and* $\vartheta > 0$.

Case (a)*:* $\vartheta\, \eta_\ell(\mathfrak{Q}_\ell) \le \eta_\ell(\mathcal{N}_\ell^+)$.

– *Set* $\mathfrak{M}_\ell = \emptyset$.

– *Find* $\mathcal{M}_\ell \subseteq \mathcal{N}_\ell^+$ *with minimal cardinality such that* $\theta_{\mathbb{X}}\, \eta_\ell(\mathcal{N}_\ell^+) \le \eta_\ell(\mathcal{M}_\ell)$.

Case (b)*:* $\vartheta\, \eta_\ell(\mathfrak{Q}_\ell) > \eta_\ell(\mathcal{N}_\ell^+)$.

– *Define* $\mathfrak{M}_\ell := \{\mu \in \mathfrak{Q}_\ell : \eta_\ell(\mu) \ge (1-\theta_{\mathbb{P}}) \max_{\nu\in\mathfrak{Q}_\ell} \eta_\ell(\nu)\}$.

– *Set* $\mathcal{M}_\ell = \emptyset$.

Output: $\mathcal{M}_\ell \subseteq \mathcal{N}_\ell^+$ *and* $\mathfrak{M}_\ell \subseteq \mathfrak{Q}_\ell$*, where one of the subsets is empty.*

Finally, Criterion D is a modification of Criterion B in the same way as Criterion C is a modification of Criterion A. Namely, we employ Dörfler marking in the physical domain and use a maximum criterion in the parameter domain, while the refinement type for the next iteration is determined by the dominant error reduction.

Criterion D.

Input: *error indicators* $\{\eta_\ell(z) : z \in \mathcal{N}_\ell^+\}$, $\{\eta_\ell(\nu) : \nu \in \mathfrak{Q}_\ell\}$*; marking parameters* $0 < \theta_{\mathbb{X}} \le 1$, $0 \le \theta_{\mathbb{P}} \le 1$, *and* $\vartheta > 0$.

– *Define* $\widetilde{\mathfrak{M}}_\ell := \{\mu \in \mathfrak{Q}_\ell : \eta_\ell(\mu) \ge (1-\theta_{\mathbb{P}}) \max_{\nu\in\mathfrak{Q}_\ell} \eta_\ell(\nu)\}$.

– *Find* $\widetilde{\mathcal{M}}_\ell \subseteq \mathcal{N}_\ell^+$ *with minimal cardinality such that* $\theta_{\mathbb{X}}\, \eta_\ell(\mathcal{N}_\ell^+) \le \eta_\ell(\widetilde{\mathcal{M}}_\ell)$.

– *Define* $\widetilde{\mathcal{R}}_\ell := \mathcal{N}_\ell^+ \cap \widetilde{\mathcal{N}}_\ell$*, where* $\widetilde{\mathcal{N}}_\ell$ *is associated with* $\widetilde{\mathcal{T}}_\ell = \mathrm{refine}(\mathcal{T}_\ell, \widetilde{\mathcal{M}}_\ell)$.

Case (a)*:* $\vartheta\, \eta_\ell(\widetilde{\mathfrak{M}}_\ell) \le \eta_\ell(\widetilde{\mathcal{R}}_\ell)$. *Set* $\mathfrak{M}_\ell = \emptyset$ *and* $\mathcal{M}_\ell = \widetilde{\mathcal{M}}_\ell$.

Case (b)*:* $\vartheta\, \eta_\ell(\widetilde{\mathfrak{M}}_\ell) > \eta_\ell(\widetilde{\mathcal{R}}_\ell)$. *Set* $\mathfrak{M}_\ell = \widetilde{\mathfrak{M}}_\ell$ *and* $\mathcal{M}_\ell = \emptyset$.

Output: $\mathcal{M}_\ell \subseteq \mathcal{N}_\ell^+$ *and* $\mathfrak{M}_\ell \subseteq \mathfrak{Q}_\ell$*, where one of the subsets is empty.*

In what follows we will write, e.g., Algorithm 4.A to refer to the algorithm obtained by employing Criterion A in step (iii) of Algorithm 4. When we refer to Algorithm 4 without specifying the marking criterion, this will mean that the statement holds for any of the four proposed marking strategies.

4.2. Convergence results. The following theorem is the first main result of the present work. Its proof is postponed to section 6. It shows that Algorithm 4 ensures convergence of the underlying error estimates to zero. We emphasize that it is valid independently of the saturation assumption (3.16).

Theorem 5. *For any choice of the marking parameters* $\theta_{\mathbb{X}}$, $\theta_{\mathbb{P}}$, *and* ϑ*, Algorithm* 4 *yields a convergent sequence of error estimates, i.e.,* $\eta_\ell \to 0$ *as* $\ell \to \infty$.

Remark 6. The proof of Theorem 5 allows for more general marking strategies than those proposed in subsection 4.1 above (see Propositions 10 and 11 in section 6). However, we believe that the marking strategies proposed in Criteria A–D are natural candidates for the present setting.

The following result is an immediate consequence of Theorem 5 and the reliability (3.17) from Theorem 2.

Corollary 7. *Let* $(u_\ell)_{\ell\in\mathbb{N}_0}$ *be the sequence of Galerkin solutions generated by Algorithm* 4*. Denote by* $(\widehat{u}_\ell)_{\ell\in\mathbb{N}_0}$ *the associated sequence of Galerkin solutions satisfying* (3.12) *and suppose that the saturation assumption* (3.16) *holds for each pair* u_ℓ, $\widehat{u}_\ell$ *(*$\ell \in \mathbb{N}_0$*). Then, for any choice of marking parameters* $\theta_{\mathbb{X}}$, $\theta_{\mathbb{P}}$, *and* ϑ*, Algorithm* 4 *yields convergence, i.e.,* $|||u - u_\ell||| \to 0$ *as* $\ell \to \infty$.

In 2D and under the saturation assumption (3.16), Algorithm 4.A and Algorithm 4.B allow for a stronger convergence result than Corollary 7. The following theorem states linear convergence of the energy error. The proof is given in section 8.

THEOREM 8. *Let $d = 2$ and let $(u_\ell)_{\ell\in\mathbb{N}_0}$ be the sequence of Galerkin solutions generated by either Algorithm* 4.A *or Algorithm* 4.B *with arbitrary* $0 < \theta_{\mathbb{X}}, \theta_{\mathbb{P}} \leq 1$, *and* $\vartheta > 0$. *Denote by* $(\widehat{u}_\ell)_{\ell\in\mathbb{N}_0}$ *the associated sequence of Galerkin solutions satisfying* (3.12) *and suppose that the saturation assumption* (3.16) *holds for each pair* u_ℓ, $\widehat{u}_\ell$ $(\ell \in \mathbb{N}_0)$. *Then, there exists a constant* $0 < q_{\text{lin}} < 1$ *such that*

$$|\!|\!| u - u_{\ell+1} |\!|\!| \leq q_{\text{lin}} |\!|\!| u - u_\ell |\!|\!| \quad \textit{for all } \ell \in \mathbb{N}_0.$$

The constant q_{lin} *depends only on the mean field* a_0, *the constant* τ *in* (2.4), *the saturation constant* q_{sat} *in* (3.16), *the coarse mesh* $\mathcal{T}_0$, *and the marking parameters* $\theta_{\mathbb{X}}$, $\theta_{\mathbb{P}}$, ϑ.

5. Numerical results. In this section, we report the results of numerical experiments aiming to underpin our theoretical findings and compare the performance of Algorithms 4.A–4.D for a range of marking parameters. The experiments were performed using the open source MATLAB toolbox Stochastic T-IFISS [7].

We consider the parametric model problem (2.1) posed on the L-shaped domain $D = (-1, 1)^2 \setminus (-1, 0]^2 \subset \mathbb{R}^2$ and set $f \equiv 1$. Following [15, section 11.1], we choose the expansion coefficients a_m $(m \in \mathbb{N}_0)$ in (2.2) to represent planar Fourier modes of increasing total order, i.e.,

$$a_0(x) := 1, \quad a_m(x) := \alpha_m \cos(2\pi\beta_1(m)\, x_1) \cos(2\pi\beta_2(m)\, x_2), \quad x = (x_1, x_2) \in D.$$

Here, for all $m \in \mathbb{N}$, $\alpha_m := Am^{-\sigma}$ is the amplitude of the coefficient, where $\sigma > 1$ and $0 < A < 1/\zeta(\sigma)$, with ζ denoting the Riemann zeta function, while β_1 and β_2 are defined as $\beta_1(m) := m - k(m)(k(m)+1)/2$ and $\beta_2(m) := k(m) - \beta_1(m)$ with $k(m) := \lfloor -1/2 + \sqrt{1/4 + 2m} \rfloor$. Note that under these assumptions, both conditions (2.3) and (2.4) are satisfied with $a_0^{\min} = a_0^{\max} = 1$ and $\tau = A\zeta(\sigma)$, respectively. We consider the case of $\sigma = 2$, which corresponds to a slow decay of the coefficients; fixing $\tau = A\zeta(\sigma) = 0.9$, this results in $A \approx 0.547$. Furthermore, we assume that the parameters y_m $(m \in \mathbb{N})$ in (2.2) are the images of uniformly distributed independent mean-zero random variables on $[-1, 1]$. In this case, $\mathrm{d}\pi_m(y_m) = \mathrm{d}y_m/2$ and the orthonormal polynomial basis of $L^2_{\pi_m}(-1, 1)$ consists of scaled Legendre polynomials. Note that the same model problem was used in numerical experiments in, e.g., [15, 16, 17, 6, 5].

We compare the performance of Algorithms 4.A–4.D with respect to a measure of the total amount of work needed to reach a prescribed tolerance tol. Let $L = L(\mathsf{tol}) \in \mathbb{N}$ be the smallest integer such that $\eta_L \leq \mathsf{tol}$, and let $N_\ell := \dim(\mathbb{V}_\ell) = \dim(\mathbb{X}_\ell)\dim(\mathbb{P}_\ell)$ be the total number of degrees of freedom at the ℓth iteration. We define the computational *cost* of Algorithm 4 as the cumulative number of degrees of freedom for all iterations of the adaptive loop, i.e.,

$$\mathsf{cost} = \mathsf{cost}(L) := \sum_{\ell=0}^{L} N_\ell. \tag{5.1}$$

We set $\mathsf{tol} = $ 5e-03 and run Algorithms 4.A–4.D with marking parameters $\theta_{\mathbb{X}}$, $\theta_{\mathbb{P}} \in \Theta := \{0.1, 0.2, \ldots, 0.9\}$ (we set $\vartheta = 1$ in all Criteria A–D). The computational costs and the empirical convergence rates for each algorithm with 81 pairs $(\theta_{\mathbb{X}}, \theta_{\mathbb{P}}) \in \Theta \times \Theta$ of marking parameters are shown in [4, Tables 2–5]. A snapshot of these results

TABLE 1

Computational cost (5.1) *(shown here in units of* 10^6*) for Algorithms* 4.A–4.D. *For each algorithm, we choose the spatial marking parameter* $\theta_{\mathbb{X}} \in \Theta$ *for which the smallest cost is incurred (see* [4, Tables 2–5]*) and show the computational cost for all* $\theta_{\mathbb{P}} \in \Theta$. *The smallest cost for each algorithm is highlighted in boldface in the corresponding row. The boldface starred value shows the overall smallest cost, i.e., the smallest cost among all computations with* 81 *pairs* $(\theta_{\mathbb{X}}, \theta_{\mathbb{P}}) \in \Theta \times \Theta$ *for all four algorithms.*

	$\theta_{\mathbb{P}}$								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Algorithm 4.A ($\theta_{\mathbb{X}} = 0.8$)	2.454	2.454	2.454	2.454	2.454	2.403	1.628	**1.560**	1.731
Algorithm 4.B ($\theta_{\mathbb{X}} = 0.7$)	3.157	3.157	3.157	3.157	3.157	2.146	1.973	1.966	**1.488**
Algorithm 4.C ($\theta_{\mathbb{X}} = 0.7$)	2.094	1.891	1.970	2.014	**1.496**	1.710	1.793	2.185	1.837
Algorithm 4.D ($\theta_{\mathbb{X}} = 0.7$)	2.146	1.952	2.000	1.966	**1.460***	1.604	1.740	2.050	1.855

is presented in Table 1. We see that the overall smallest cost is achieved by Algorithm 4.D for the values $\theta_{\mathbb{X}} = 0.7$ and $\theta_{\mathbb{P}} = 0.5$. These values of marking parameters are the ones for which Algorithm 4.C also yields the smallest cost among all pairs $(\theta_{\mathbb{X}}, \theta_{\mathbb{P}}) \in \Theta \times \Theta$. This similarity does not hold for Algorithms 4.A–4.B, for which the smallest cost is achieved with $\theta_{\mathbb{X}} = \theta_{\mathbb{P}} = 0.8$ for Algorithm 4.A and with $\theta_{\mathbb{X}} = 0.7$ and $\theta_{\mathbb{P}} = 0.9$ for Algorithm 4.B. Thus, we conclude that, for the above values of marking parameters, the adaptive algorithms with refinements driven by dominant error reduction estimates (Algorithms 4.B and 4.D) incur less computational costs than their counterparts driven by dominant contributing error estimates (Algorithms 4.A and 4.C). On the other hand, the algorithms that employ the maximum criterion for parametric refinement (Algorithms 4.C and 4.D) incur less computational costs than their counterparts that use Dörfler marking (Algorithms 4.A and 4.B). Overall, the smallest computational cost is incurred by the algorithm that combines these two winning strategies—Algorithm 4.D.

Figure 2 shows the decay of the overall error estimate η_ℓ versus the number of degrees of freedom N_ℓ for different values of $\theta_{\mathbb{P}} \in \Theta$ with $\theta_{\mathbb{X}} = 0.8$ in Algorithm 4.A and $\theta_{\mathbb{X}} = 0.7$ in Algorithms 4.B–4.D. The aim of these plots is to show that the adaptive algorithm converges regardless of the marking criterion and the value of $\theta_{\mathbb{P}}$ used (similar decay rates are obtained for other values of $\theta_{\mathbb{X}}$, $\theta_{\mathbb{P}} \in \Theta$; see [4, Appendix A]). Observe that η_ℓ decays also in the case $\theta_{\mathbb{P}} = 1 \notin \Theta$ for all algorithms. However, in this case, significantly more degrees of freedom are needed to reach the prescribed tolerance, compared to the cases of $\theta_{\mathbb{P}} \in \Theta$. This is because, for $\theta_{\mathbb{P}} = 1$, each parametric enrichment is performed by augmenting the index set $\mathfrak{P}_\ell$ with the whole detail index set $\mathfrak{Q}_\ell$.

To conclude, we test the effectiveness of our error estimation strategy by computing a reference energy error as follows. We first compute an accurate solution $u_{\mathrm{ref}} \in \mathbb{V}_{\mathrm{ref}} := \mathbb{X}_{\mathrm{ref}} \otimes \mathbb{P}_{\mathrm{ref}}$ using quadratic (P2) finite element approximations over a fine mesh $\mathcal{T}_{\mathrm{ref}}$ and employing a large index set $\mathfrak{P}_{\mathrm{ref}}$. Then, we define the effectivity indices

$$\zeta_\ell := \frac{\eta_\ell}{\| u_{\mathrm{ref}} - u_\ell \|} = \frac{\eta_\ell}{(\| u_{\mathrm{ref}} \|^2 - \| u_\ell \|^2)^{1/2}} \quad \text{for all } \ell = 0, \ldots, L,$$

where the equality holds due to Galerkin orthogonality and the symmetry of the bilinear form $B(\cdot,\cdot)$. In this experiment, we choose $\mathcal{T}_{\mathrm{ref}}$ to be the uniform refinement

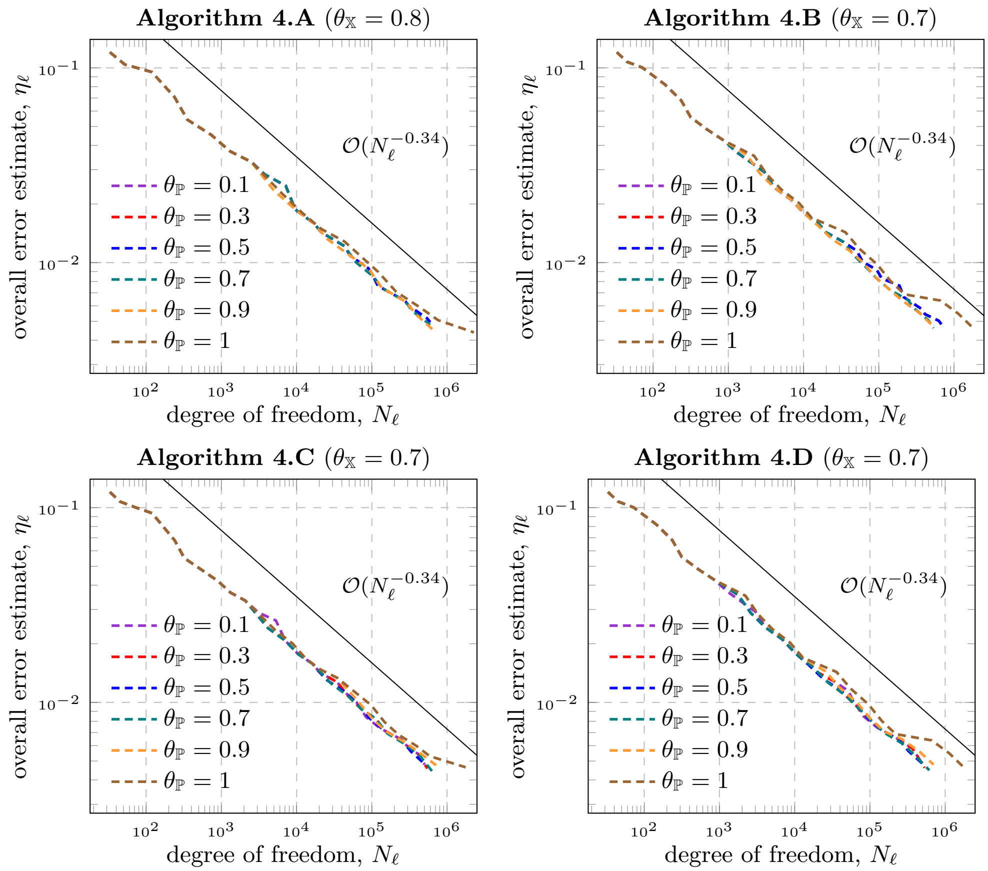

FIG. 2. *Decay of the overall error estimates η_ℓ computed at each iteration of Algorithm* 4.A *with $\theta_{\mathbb{X}} = 0.8$ and Algorithms* 4.B–4.D *with $\theta_{\mathbb{X}} = 0.7$ for $\theta_{\mathbb{P}} \in \{0.1, 0.3, 0.5, 0.7, 0.9, 1\}$.*

of the mesh $\mathcal{T}_L$ generated by Algorithm 4.B with $\theta_{\mathbb{P}} = 0.5$ (i.e., one of the final meshes with the largest number of elements) and $\mathfrak{P}_{\text{ref}}$ to be the final index set $\mathfrak{P}_L$ produced by Algorithm 4.D with $\theta_{\mathbb{P}} = 0.8$ (i.e., one of the largest index sets generated).

Figure 3 shows the effectivity indices ζ_ℓ obtained for Algorithms 4.A–4.B (left) and Algorithms 4.C–4.D (right) with the pairs of parameters $(\theta_{\mathbb{X}}, \theta_{\mathbb{P}})$ for which the smallest cost is attained. We observe that in all cases the error is slightly underestimated, as the effectivity indices vary in a range between 0.7 and 0.82 throughout all iterations.

6. Proof of Theorem 5 (plain convergence). We start with stating three propositions which address convergence of either the spatial component or the parametric component of the error estimate given by (3.11). To ease the readability, the proofs of propositions are postponed to section 7. The first proposition proves that each parametric error indicator converges to some limiting error indicator.

PROPOSITION 9. *For $\nu \in \mathfrak{Q}_\ell$, let $\eta_\ell(\nu) \geq 0$ be the parametric error indicator from* (3.10). *For $\nu \in \mathfrak{I} \setminus \mathfrak{Q}_\ell$, define $\eta_\ell(\nu) := 0$. Then, for each $\nu \in \mathfrak{I}$, there exists $\eta_\infty(\nu) \geq 0$ such that*

$$\sum_{\nu \in \mathfrak{I}} \eta_\infty(\nu)^2 < \infty \quad \text{and} \quad \sum_{\nu \in \mathfrak{I}} |\eta_\infty(\nu) - \eta_\ell(\nu)|^2 \to 0 \quad \text{as } \ell \to \infty. \tag{6.1}$$

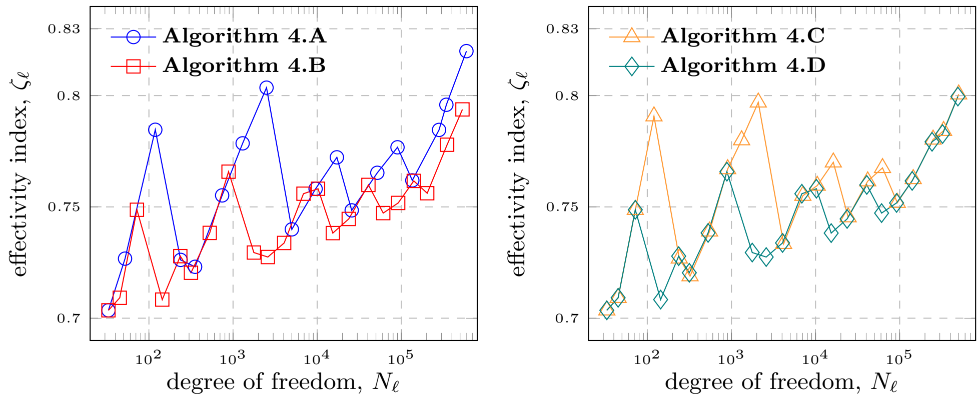

FIG. 3. *The effectivity indices ζ_ℓ for the sGFEM solutions at each iteration of Algorithms* 4.A–4.B *(left) and Algorithms* 4.C–4.D *(right) with the marking parameters $\theta_{\mathbb{X}}$, $\theta_{\mathbb{P}} \in \Theta$ that yield smallest cost (see Table* 1*).*

The second proposition states that the parametric enrichment satisfying a certain weak marking criterion along a subsequence guarantees convergence of the whole sequence of parametric error estimates.

PROPOSITION 10. *Let $g_{\mathbb{P}} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ be a continuous function with $g_{\mathbb{P}}(0) = 0$. Suppose that Algorithm* 4 *yields a subsequence $(\mathfrak{P}_{\ell_k})_{k\in\mathbb{N}_0} \subset (\mathfrak{P}_\ell)_{\ell\in\mathbb{N}_0}$ satisfying the following property:*

$$\eta_{\ell_k}(\mu) \leq g_{\mathbb{P}}\big(\eta_{\ell_k}(\mathfrak{M}_{\ell_k})\big) \quad \textit{for all } k \in \mathbb{N}_0 \textit{ and all } \mu \in \mathfrak{Q}_{\ell_k} \setminus \mathfrak{M}_{\ell_k}, \tag{6.2}$$

i.e., the nonmarked multi-indices are controlled by the marked ones. Then, the sequence of parametric error estimates converges to zero, i.e., $\eta_\ell(\mathfrak{Q}_\ell) \to 0$ as $\ell \to \infty$.

The third proposition addresses convergence of spatial error estimates. Unlike in Proposition 10 for parametric estimates, the convergence here is only shown along the subsequence for which spatial refinement takes place.

PROPOSITION 11. *Let $g_{\mathbb{X}} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ be a continuous function with $g_{\mathbb{X}}(0) = 0$. Suppose that Algorithm* 4 *yields a subsequence $(\mathcal{T}_{\ell_k})_{k\in\mathbb{N}_0} \subset (\mathcal{T}_\ell)_{\ell\in\mathbb{N}_0}$ satisfying the following property:*

$$\eta_{\ell_k}(z) \leq g_{\mathbb{X}}\big(\eta_{\ell_k}(\mathcal{M}_{\ell_k})\big) \quad \textit{for all } k \in \mathbb{N}_0 \textit{ and all } z \in \mathcal{N}^+_{\ell_k} \setminus \mathcal{M}_{\ell_k}, \tag{6.3}$$

i.e., the nonmarked vertices are controlled by the marked ones. Then, the corresponding subsequence of spatial error estimates converges to zero, i.e., $\eta_{\ell_k}(\mathcal{N}^+_{\ell_k}) \to 0$ as $k \to \infty$.

Remark 12. The marking strategies employed in Criteria A–D, i.e., the Dörfler marking strategy and the maximum criterion, satisfy the properties (6.2)–(6.3) assumed in Propositions 10 and 11. For example, let us show that (6.2) holds for parametric error indicators (the same arguments will apply to spatial error indicators). Suppose that the ℓ_kth step of the adaptive algorithm employs the maximum criterion, i.e., $\mathfrak{M}_{\ell_k} := \{\mu \in \mathfrak{Q}_{\ell_k} : \eta_{\ell_k}(\mu) \geq (1-\theta_{\mathbb{P}}) \max_{\nu\in\mathfrak{Q}_{\ell_k}} \eta_{\ell_k}(\nu)\}$. Then, for $\mu \in \mathfrak{Q}_{\ell_k} \setminus \mathfrak{M}_{\ell_k}$, there holds

$$\eta_{\ell_k}(\mu) < (1-\theta_{\mathbb{P}}) \max_{\nu\in\mathfrak{Q}_{\ell_k}} \eta_{\ell_k}(\nu) \leq (1-\theta_{\mathbb{P}})\, \eta_{\ell_k}(\mathfrak{M}_{\ell_k}),$$

which is (6.2) with $g_{\mathbb{P}}(s) := (1 - \theta_{\mathbb{P}})s$. Similarly, suppose that the ℓ_kth step of the algorithm employs Dörfler marking, i.e., $\mathfrak{M}_{\ell_k} \subseteq \mathfrak{Q}_{\ell_k}$ satisfies $\theta_{\mathbb{P}}\, \eta_{\ell_k}(\mathfrak{Q}_{\ell_k}) \le \eta_{\ell_k}(\mathfrak{M}_{\ell_k})$. Then, for $\mu \in \mathfrak{Q}_{\ell_k} \setminus \mathfrak{M}_{\ell_k}$, one has

$$\eta_{\ell_k}(\mu) \le \eta_{\ell_k}(\mathfrak{Q}_{\ell_k} \setminus \mathfrak{M}_{\ell_k}) = \big(\eta_{\ell_k}(\mathfrak{Q}_{\ell_k})^2 - \eta_{\ell_k}(\mathfrak{M}_{\ell_k})^2\big)^{1/2} \le (1-\theta_{\mathbb{P}}^2)^{1/2}\, \theta_{\mathbb{P}}^{-1}\, \eta_{\ell_k}(\mathfrak{M}_{\ell_k}),$$

which is (6.2) with $g_{\mathbb{P}}(s) := (1-\theta_{\mathbb{P}}^2)^{1/2}\, \theta_{\mathbb{P}}^{-1}\, s$.

With the aforegoing propositions, we can proceed to the proof of Theorem 5.

Proof of Theorem 5. We divide the proof into three steps.

Step 1. Consider Algorithms 4.A and 4.C. If case (a) in the corresponding marking strategies occurs only finitely many times, then there exists $\ell_0 \in \mathbb{N}$ such that case (b) (i.e., parametric enrichment) occurs for all $\ell \ge \ell_0$. Then, according to the criterion used to decide on the type of enrichment, one has $0 \le \eta_\ell(\mathcal{N}_\ell^+) < \vartheta\, \eta_\ell(\mathfrak{Q}_\ell)$ for all $\ell \ge \ell_0$. Since $\eta_\ell(\mathfrak{Q}_\ell) \to 0$ as $\ell \to \infty$ by Proposition 10, we conclude that $\eta_\ell \to 0$ as $\ell \to \infty$. If case (b) in Criteria A and C occurs finitely many times, then there exists $\ell_0 \in \mathbb{N}$ such that only case (a) (i.e., spatial refinement) occurs for all $\ell \ge \ell_0$. Hence, $0 \le \vartheta\, \eta_\ell(\mathfrak{Q}_\ell) \le \eta_\ell(\mathcal{N}_\ell^+)$ for all $\ell \ge \ell_0$. Since $\eta_\ell(\mathcal{N}_\ell^+) \to 0$ as $\ell \to \infty$ by Proposition 11, we conclude that $\eta_\ell \to 0$ as $\ell \to \infty$. Finally, if both cases (a) and (b) happen infinitely often, we split the sequence $(\eta_\ell)_{\ell \in \mathbb{N}_0}$ into two disjoint subsequences: $(\eta_{\ell_k^{(\mathrm{a})}})_{k \in \mathbb{N}}$, where only case (a) occurs, and $(\eta_{\ell_k^{(\mathrm{b})}})_{k\in\mathbb{N}}$, where only case (b) occurs. With the preceding argument, it follows that $\eta_{\ell_k^{(\mathrm{a})}}, \eta_{\ell_k^{(\mathrm{b})}} \to 0$ as $k \to \infty$. This implies the convergence of the sequence $\eta_\ell \to 0$ as $\ell \to \infty$.

Step 2. Let us now consider Algorithm 4.B. We argue as in Step 1. If case (a) in Criterion B occurs only finitely many times, then there exists $\ell_0 \in \mathbb{N}$ such that case (b) (i.e., parametric enrichment) occurs for all $\ell \ge \ell_0$. Then, according to the criterion used to decide on the type of enrichment, one has

$$0 \le \theta_{\mathbb{X}}\, \eta_\ell(\mathcal{N}_\ell^+) \le \eta_\ell(\widetilde{\mathcal{M}}_\ell) \le \eta_\ell(\widetilde{\mathcal{R}}_\ell) < \vartheta\, \eta_\ell(\mathfrak{M}_\ell) \le \vartheta\, \eta_\ell(\mathfrak{Q}_\ell) \quad \text{for all } \ell \ge \ell_0.$$

Since $\eta_\ell(\mathfrak{Q}_\ell) \to 0$ as $\ell \to \infty$ by Proposition 10, we conclude that $\eta_\ell \to 0$ as $\ell \to \infty$. If case (b) in Criterion B occurs finitely many times, then there exists $\ell_0 \in \mathbb{N}$ such that only case (a) (i.e., spatial refinement) occurs for all $\ell \ge \ell_0$ and hence

$$0 \le \theta_{\mathbb{P}}\, \eta_\ell(\mathfrak{Q}_\ell) \le \eta_\ell(\widetilde{\mathfrak{M}}_\ell) < \vartheta^{-1}\, \eta_\ell(\widetilde{\mathcal{R}}_\ell) \le \vartheta^{-1}\, \eta_\ell(\mathcal{N}_\ell^+) \quad \text{for all } \ell \ge \ell_0.$$

Since $\eta_\ell(\mathcal{N}_\ell^+) \to 0$ as $\ell \to \infty$ by Proposition 11, we conclude that $\eta_\ell \to 0$ as $\ell \to \infty$. If both cases (a) and (b) occur infinitely often, then we proceed as in Step 1 to show that $\eta_\ell \to 0$ as $\ell \to \infty$.

Step 3. Finally, consider Algorithm 4.D. Arguing as for Algorithm 4.B in Step 2, we prove that

$$\eta_\ell(\mathcal{N}_\ell^+) \to 0 \quad \text{as well as} \quad \eta_\ell(\widetilde{\mathfrak{M}}_\ell) \to 0 \quad \text{as } \ell \to \infty. \tag{6.4}$$

It remains to show that $\eta_\ell(\mathfrak{Q}_\ell) \to 0$ as $\ell \to \infty$. By Proposition 9, there exists a sequence $(\eta_\infty(\nu))_{\nu \in \mathfrak{I}}$ satisfying (6.1). In particular, $\sup_{\nu \in \mathfrak{I}} \eta_\infty(\nu) < \infty$. Let $\varepsilon > 0$ and choose $\mu \in \mathfrak{I}$ such that $\sup_{\nu\in\mathfrak{I}} \eta_\infty(\nu) \le \eta_\infty(\mu) + \varepsilon$. Together with (6.1) and (6.4), the triangle inequality yields that

$$0 \le \eta_\infty(\mu) + \varepsilon \le |\eta_\infty(\mu) - \eta_\ell(\mu)| + \eta_\ell(\mu) + \varepsilon \le |\eta_\infty(\mu) - \eta_\ell(\mu)| + \eta_\ell(\widetilde{\mathfrak{M}}_\ell) + \varepsilon \xrightarrow{\ell \to \infty} \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, we conclude that $\eta_\infty(\nu) = 0$ for all $\nu \in \mathfrak{I}$. With (6.1), this proves that $\eta_\ell(\mathfrak{Q}_\ell)^2 = \sum_{\nu \in \mathfrak{I}} \eta_\ell(\nu)^2 \to 0$ as $\ell \to \infty$. ∎

7. Proof of Propositions 9, 10, and 11. In this section, we collect some auxiliary results and prove Propositions 9, 10, and 11.

7.1. A priori convergence of adaptive algorithms. The following lemma is an early result from [2] which proves that adaptive algorithms (without coarsening) always lead to convergence of the discrete solutions.

LEMMA 13 (a priori convergence). *Let V be a Hilbert space. Let $a : V \times V \to \mathbb{R}$ be an elliptic and continuous bilinear form. Let $F \in V^*$ be a linear and continuous functional. For each $\ell \in \mathbb{N}_0$, let $V_\ell \subseteq V$ be a closed subspace such that $V_\ell \subseteq V_{\ell+1}$. Furthermore, define the limiting space $V_\infty := \overline{\bigcup_{\ell=0}^\infty V_\ell} \subseteq V$. Then, for all $\ell \in \mathbb{N}_0 \cup \{\infty\}$, there exists a unique Galerkin solution $u_\ell \in V_\ell$ satisfying*

$$a(u_\ell, v_\ell) = F(v_\ell) \quad \textit{for all } v_\ell \in V_\ell. \tag{7.1}$$

Moreover, there holds $\|u_\infty - u_\ell\|_V \xrightarrow{\ell\to\infty} 0$.

Proof. For each $\ell \in \mathbb{N}_0 \cup \{\infty\}$, the existence and uniqueness of the Galerkin solution $u_\ell \in V_\ell$ satisfying (7.1) follow from the Lax–Milgram theorem. Moreover, since $V_\ell \subseteq V_\infty$, u_ℓ is also a Galerkin approximation to u_∞. Therefore, by the definition of V_∞, the Céa lemma proves that $\|u_\infty - u_\ell\|_V \lesssim \min_{v_\ell \in V_\ell} \|u_\infty - v_\ell\|_V \xrightarrow{\ell\to\infty} 0$. ∎

7.2. Proof of Proposition 9. For $\nu \in \mathfrak{Q}_\ell$, recall the functions $e_\ell^\nu \in \mathbb{X}_\ell$ in (3.9). Define $\widehat{e}'_\ell := \sum_{\nu\in\mathfrak{I}} e_\ell^\nu P_\nu \in \mathbb{X}_\ell \otimes \widehat{\mathbb{P}}_\ell \overset{(3.7)}{=} \widehat{\mathbb{V}}'_\ell$, where $e_\ell^\nu := 0$ for all $\nu \in \mathfrak{I} \setminus \mathfrak{Q}_\ell$. Note that $\eta_\ell(\nu) = \|e_\ell^\nu P_\nu\|_0$ for all $\nu \in \mathfrak{Q}_\ell$ and define $\eta_\ell(\nu) := \|e_\ell^\nu P_\nu\|_0 = 0$ for all $\nu \in \mathfrak{I} \setminus \mathfrak{Q}_\ell$. The next lemma shows that the sequence $\widehat{e}'_\ell$ converges to some limit $\widehat{e}'_\infty$ in $\mathbb{V}$.

LEMMA 14. *There exists a sequence $(e_\infty^\nu)_{\nu\in\mathfrak{I}} \subset \mathbb{X}$ such that $\widehat{e}'_\infty := \sum_{\nu\in\mathfrak{I}} e_\infty^\nu P_\nu \in \mathbb{V}$ satisfies*

$$\|\widehat{e}'_\infty\|_0^2 = \sum_{\nu\in\mathfrak{I}} \|e_\infty^\nu P_\nu\|_0^2 < \infty \quad \textit{and} \quad \|\widehat{e}'_\infty - \widehat{e}'_\ell\|_0^2 = \sum_{\nu\in\mathfrak{I}} \|(e_\infty^\nu - e_\ell^\nu)\, P_\nu\|_0^2 \xrightarrow{\ell\to\infty} 0. \tag{7.2}$$

Proof. The tensor-product structure of $\widehat{\mathbb{V}}'_\ell = \mathbb{X}_\ell \otimes \widehat{\mathbb{P}}_\ell$ and pairwise orthogonality of subspaces $X_\ell \otimes \operatorname{span}\{P_\nu\}$ ($\nu \in \mathfrak{I}$) with respect to $B_0(\cdot,\cdot)$ imply that

$$B_0(\widehat{e}'_\ell, v_\ell P_\nu) \overset{(3.2)}{=} B_0(e_\ell^\nu P_\nu, v_\ell P_\nu) \overset{(3.9)}{=} F(v_\ell P_\nu) - B(u_\ell, v_\ell P_\nu)$$

for all $\nu \in \mathfrak{Q}_\ell$ and $v_\ell \in \mathbb{X}_\ell$. Moreover, there holds

$$B_0(\widehat{e}'_\ell, v_\ell P_\nu) \overset{(3.2)}{=} B_0(e_\ell^\nu P_\nu, v_\ell P_\nu) = 0 \overset{(3.4)}{=} F(v_\ell P_\nu) - B(u_\ell, v_\ell P_\nu)$$

for all $\nu \in \mathfrak{P}_\ell$ and $v_\ell \in \mathbb{X}_\ell$. Hence, $\widehat{e}'_\ell \in \widehat{\mathbb{V}}'_\ell$ is the unique solution of the problem

$$B_0(\widehat{e}'_\ell, \widehat{v}'_\ell) = F(\widehat{v}'_\ell) - B(u_\ell, \widehat{v}'_\ell) \quad \text{for all } \widehat{v}'_\ell \in \widehat{\mathbb{V}}'_\ell. \tag{7.3}$$

Lemma 13 proves that $\|u_\infty - u_\ell\| \to 0$ as $\ell \to \infty$ for some $u_\infty \in \mathbb{V}$. Consider the unique solution $\widehat{e}''_\ell \in \widehat{\mathbb{V}}'_\ell$ of the auxiliary problem

$$B_0(\widehat{e}''_\ell, \widehat{v}'_\ell) = F(\widehat{v}'_\ell) - B(u_\infty, \widehat{v}'_\ell) \quad \text{for all } \widehat{v}'_\ell \in \widehat{\mathbb{V}}'_\ell. \tag{7.4}$$

Since $\widehat{\mathbb{V}}'_\ell \subseteq \widehat{\mathbb{V}}'_{\ell+1}$, Lemma 13 also proves that $\|\widehat{e}'_\infty - \widehat{e}''_\ell\| \to 0$ as $\ell \to \infty$ for some $\widehat{e}'_\infty \in \mathbb{V}$. Exploiting (7.3) and (7.4) for $\widehat{v}'_\ell = \widehat{e}''_\ell - \widehat{e}'_\ell \in \widehat{\mathbb{V}}'_\ell$, we see that

$$\|\widehat{e}''_\ell - \widehat{e}'_\ell\|_0^2 = B_0(\widehat{e}''_\ell - \widehat{e}'_\ell, \widehat{e}''_\ell - \widehat{e}'_\ell) = -B(u_\infty - u_\ell, \widehat{e}''_\ell - \widehat{e}'_\ell) \le \|u_\infty - u_\ell\| \|\widehat{e}''_\ell - \widehat{e}'_\ell\|.$$

With the norm equivalence $\|\!|\cdot|\!\|_0 \simeq \|\!|\cdot|\!\|$, the triangle inequality thus proves that

$$\|\!|\widehat{e}'_\infty - \widehat{e}'_\ell|\!\|_0 \le \|\!|\widehat{e}'_\infty - \widehat{e}''_\ell|\!\|_0 + \|\!|\widehat{e}''_\ell - \widehat{e}'_\ell|\!\|_0 \lesssim \|\!|\widehat{e}'_\infty - \widehat{e}''_\ell|\!\|_0 + \|\!|u_\infty - u_\ell|\!\| \xrightarrow{\ell\to\infty} 0.$$

Hence, the proof is concluded by noticing that the existence of $(e^\nu_\infty)_{\nu\in\mathfrak{I}} \subset \mathbb{X}$ is a consequence of the representation in (3.1) and that the equalities in (7.2) then immediately follow from (3.3). ∎

With the above result, we can proceed to the proof of Proposition 9.

Proof of Proposition 9. Lemma 14 provides a sequence $(e^\nu_\infty)_{\nu\in\mathfrak{I}} \subset \mathbb{X}$ satisfying (7.2). For $\nu \in \mathfrak{I}$, we define $\eta_\infty(\nu) := \|\!|e^\nu_\infty P_\nu|\!\|_0$. From (7.2) it follows that

$$\sum_{\nu\in\mathfrak{I}} \eta_\infty(\nu)^2 = \sum_{\nu\in\mathfrak{I}} \|\!|e^\nu_\infty P_\nu|\!\|_0^2 < \infty,$$

and using (7.2) together with the definition of $\eta_\ell(\nu)$ in (3.10) we find that

$$\sum_{\nu\in\mathfrak{I}} |\eta_\infty(\nu) - \eta_\ell(\nu)|^2 = \sum_{\nu\in\mathfrak{I}} \big(\|\!|e^\nu_\infty P_\nu|\!\|_0 - \|\!|e^\nu_\ell P_\nu|\!\|_0\big)^2 \le \sum_{\nu\in\mathfrak{I}} \|\!|e^\nu_\infty P_\nu - e^\nu_\ell P_\nu|\!\|_0^2 \xrightarrow{\ell\to\infty} 0.$$

This yields (6.1) and concludes the proof. ∎

7.3. Proof of Proposition 10. We first state an auxiliary result for square summable sequences.

LEMMA 15. *Let $g : \mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}$ be a continuous function with $g(0) = 0$. Let $(x_n)_{n\in\mathbb{N}} \subset \mathbb{R}_{\ge 0}$ with $\sum_{n=1}^\infty x_n^2 < \infty$. For $k \in \mathbb{N}_0$, let $(x_n^{(k)})_{n\in\mathbb{N}} \subset \mathbb{R}_{\ge 0}$ with $\sum_{n=1}^\infty (x_n - x_n^{(k)})^2 \to 0$ as $k \to \infty$. In addition, let $(\mathcal{P}_k)_{k\in\mathbb{N}_0}$ be a sequence of nested subsets of $\mathbb{N}$ (i.e., $\mathcal{P}_k \subseteq \mathcal{P}_{k+1}$ for all $k \in \mathbb{N}_0$) satisfying the following property:*

$$x_m^{(k)} \le g\Bigg(\sum_{n\in\mathcal{P}_{k+1}\setminus\mathcal{P}_k} (x_n^{(k)})^2\Bigg) \quad \textit{for all } k \in \mathbb{N}_0 \textit{ and } m \in \mathbb{N}\setminus\mathcal{P}_{k+1}. \tag{7.5}$$

Then $\sum_{n\in\mathbb{N}\setminus\mathcal{P}_k} x_n^2 \to 0$ as $k \to \infty$.

Proof. We divide the proof into 3 steps.

Step 1. First, we show that $\min(\mathcal{P}_{k+1}\setminus\mathcal{P}_k) \to \infty$ as $k \to \infty$, where $\min(\emptyset) := \infty$. This statement is trivial if there exists $K \in \mathbb{N}$ such that $\mathcal{P}_k = \mathcal{P}_{k+1}$ for all $k \ge K$. Therefore, without loss of generality, we can consider a sequence of strictly nested sets, i.e., $\mathcal{P}_k \subset \mathcal{P}_{k+1}$ for all $k \in \mathbb{N}_0$. We argue by contradiction and assume the existence of $C > 0$ such that, for all $k_0 \in \mathbb{N}_0$, there exists $k \ge k_0$ such that $M_k := \min(\mathcal{P}_{k+1}\setminus\mathcal{P}_k) \le C$. In particular, we can construct a monotonic increasing sequence $(k_j)_{j\in\mathbb{N}_0} \subset \mathbb{N}_0$, i.e., $k_j \le k_{j+1}$ for all $j \in \mathbb{N}_0$, and consider the corresponding bounded sequence $(M_{k_j})_{j\in\mathbb{N}_0}$. Since this sequence is bounded, we can extract a convergent subsequence (not relabeled) and denote its limit by $m := \lim_{j\to\infty} M_{k_j}$. Since $(M_{k_j})_{j\in\mathbb{N}_0} \subset \mathbb{N}$, it follows that there exists $i \in \mathbb{N}_0$ such that $m = M_{k_j}$ for all $j \ge i$. In particular, $m = M_{k_i}$ and $m = M_{k_{i+1}}$, so that $m \in \mathcal{P}_{k_i+1} \cap \mathcal{P}_{k_{i+1}+1} = \mathcal{P}_{k_i+1}$. On the other hand, since the sets are nested and $k_i + 1 \le k_{i+1}$, we conclude that $\mathcal{P}_{k_i+1} \subseteq \mathcal{P}_{k_{i+1}}$. This leads to a contradiction:

$$m = M_{k_{i+1}} = \min(\mathcal{P}_{k_{i+1}+1}\setminus\mathcal{P}_{k_{i+1}}) \in \mathcal{P}_{k_{i+1}+1}\setminus\mathcal{P}_{k_{i+1}} \subseteq \mathcal{P}_{k_{i+1}+1}\setminus\mathcal{P}_{k_i+1} \not\ni m.$$

Step 2. Next, let us establish some auxiliary convergence statements. Using the summability assumption on $(x_n)_{n\in\mathbb{N}}$ and the convergence assumption on $(x_n^{(k)})_{n\in\mathbb{N}}$ $(k \in \mathbb{N}_0)$, it follows from Step 1 that

$$\Big(\sum_{n\in\mathcal{P}_{k+1}\setminus\mathcal{P}_k} (x_n^{(k)})^2\Big)^{1/2} \le \Big(\sum_{n=1}^{\infty}(x_n - x_n^{(k)})^2\Big)^{1/2} + \Big(\sum_{n=\min(\mathcal{P}_{k+1}\setminus\mathcal{P}_k)}^{\infty} x_n^2\Big)^{1/2} \xrightarrow{k\to\infty} 0.$$

Therefore, considering the set $\mathcal{P}_\infty^c := \{n \in \mathbb{N} : n \notin \mathcal{P}_k \text{ for all } k \in \mathbb{N}_0\}$, we deduce from (7.5) that

$$0 \le x_m^{(k)} \le g\Big(\sum_{n\in\mathcal{P}_{k+1}\setminus\mathcal{P}_k} (x_n^{(k)})^2\Big) \xrightarrow{k\to\infty} 0 \quad \text{for all } m \in \mathcal{P}_\infty^c. \tag{7.6}$$

To conclude this step, let us show that

$$\min((\mathbb{N}\setminus\mathcal{P}_{k+1})\setminus\mathcal{P}_\infty^c) \to \infty \quad \text{as} \quad k\to\infty, \quad \text{where } \min(\emptyset) := \infty. \tag{7.7}$$

Let $m_k := \min((\mathbb{N}\setminus\mathcal{P}_{k+1})\setminus\mathcal{P}_\infty^c)$ for all $k \in \mathbb{N}_0$. Note that the sequence $(m_k)_{k\in\mathbb{N}_0}$ is monotonic increasing, because the sets are nested. Since $m_k \in (\mathbb{N}\setminus\mathcal{P}_{k+1})\setminus\mathcal{P}_\infty^c$, there exists $j_0 \in \mathbb{N}_0$ with $j_0 > k+1$ such that $m_k \in \mathcal{P}_{j_0}$. Therefore, since the sets are nested, we conclude that $m_k \in \mathcal{P}_j$ for all $j \ge j_0$. In particular, $m_j \ge m_k + 1$ for all $j \ge j_0$. Together with the monotonicity of $(m_k)_{k\in\mathbb{N}_0}$, this implies that $\lim_{k\to\infty} m_k = \infty$, which yields (7.7).

Step 3. Finally, let us show that $\sum_{n\in\mathbb{N}\setminus\mathcal{P}_k} x_n^2 \to 0$ as $k\to\infty$. Let $N \in \mathbb{N}$ be an arbitrary free parameter and consider the following sets:

$$\begin{aligned}
\mathcal{A}_k^1[N] &:= (\mathbb{N}\setminus\mathcal{P}_k) \cap \{n \in \mathbb{N} : n \ge N\},\\
\mathcal{A}_k^2[N] &:= (\mathbb{N}\setminus\mathcal{P}_{k+1}) \cap \{n \in \mathbb{N} : n < N\} \cap \mathcal{P}_\infty^c,\\
\mathcal{A}_k^3[N] &:= (\mathbb{N}\setminus\mathcal{P}_{k+1}) \cap \{n \in \mathbb{N} : n < N\} \setminus \mathcal{P}_\infty^c,\\
\mathcal{A}_k^4[N] &:= (\mathcal{P}_{k+1}\setminus\mathcal{P}_k) \cap \{n \in \mathbb{N} : n < N\}.
\end{aligned}$$

Note that this defines a disjoint partition of $\mathbb{N}\setminus\mathcal{P}_k$, i.e.,

$$\mathbb{N}\setminus\mathcal{P}_k = \mathcal{A}_k^1[N] \cup \mathcal{A}_k^2[N] \cup \mathcal{A}_k^3[N] \cup \mathcal{A}_k^4[N] \quad \text{and} \quad \mathcal{A}_k^i[N] \cap \mathcal{A}_k^j[N] = \emptyset \ \text{ for } i \ne j.$$

For the sum over the set $\mathcal{A}_k^2[N]$, we have

$$\sum_{n\in\mathcal{A}_k^2[N]} x_n^2 \lesssim \sum_{n\in\mathcal{A}_k^2[N]} (x_n^{(k)})^2 + \sum_{n\in\mathcal{A}_k^2[N]} \big(x_n - x_n^{(k)}\big)^2 \le \sum_{n\in\mathcal{A}_k^2[N]} (x_n^{(k)})^2 + \sum_{n\in\mathbb{N}} \big(x_n - x_n^{(k)}\big)^2.$$

The second sum on the right-hand side of this estimate converges to 0 as $k\to\infty$ by assumption, whereas the first sum is finite, and therefore also converges to 0 as $k\to\infty$ because of (7.6).

For the sums over the sets $\mathcal{A}_k^3[N]$ and $\mathcal{A}_k^4[N]$, we use the convergence result in (7.7) and the result of Step 1, respectively. Along with the summability assumption on $(x_n)_{n\in\mathbb{N}}$, this proves that

$$\sum_{n\in\mathcal{A}_k^3[N]} x_n^2 \le \sum_{n\in(\mathbb{N}\setminus\mathcal{P}_{k+1})\setminus\mathcal{P}_\infty^c} x_n^2 \le \sum_{n=\min((\mathbb{N}\setminus\mathcal{P}_{k+1})\setminus\mathcal{P}_\infty^c)}^{\infty} x_n^2 \xrightarrow{k\to\infty} 0$$

and

$$\sum_{n\in\mathcal{A}_k^4[N]} x_n^2 \le \sum_{n\in\mathcal{P}_{k+1}\setminus\mathcal{P}_k} x_n^2 \le \sum_{n=\min(\mathcal{P}_{k+1}\setminus\mathcal{P}_k)}^{\infty} x_n^2 \xrightarrow{k\to\infty} 0.$$

We have thus shown that

$$\sum_{n\in\mathcal{A}_k^2[N]} x_n^2 + \sum_{n\in\mathcal{A}_k^3[N]} x_n^2 + \sum_{n\in\mathcal{A}_k^4[N]} x_n^2 \xrightarrow{k\to\infty} 0 \quad \text{for all } N\in\mathbb{N}.$$

In particular, for all $N\in\mathbb{N}$, one has

$$0 \le \liminf_{k\to\infty}\Bigg(\sum_{n\in\mathbb{N}\setminus\mathcal{P}_k} x_n^2\Bigg) \le \limsup_{k\to\infty}\Bigg(\sum_{n\in\mathbb{N}\setminus\mathcal{P}_k} x_n^2\Bigg) = \limsup_{k\to\infty}\Bigg(\sum_{n\in\mathcal{A}_k^1[N]} x_n^2\Bigg) \le \sum_{n=N}^{\infty} x_n^2.$$

Thus, the limit inferior and the limit superior of $\sum_{n\in\mathbb{N}\setminus\mathcal{P}_k} x_n^2$ are nonnegative and bounded from above by a tail of the convergent series. Since N is arbitrary, this leads to the desired convergence $\sum_{n\in\mathbb{N}\setminus\mathcal{P}_k} x_n^2 \to 0$ as $k\to\infty$. This concludes the proof. $\square$

With this lemma, we can proceed to the proof of Proposition 10.

Proof of Proposition 10. Proposition 9 yields a sequence $(\eta_\infty(\nu))_{\nu\in\mathfrak{I}}$ such that

$$\sum_{\nu\in\mathfrak{I}} \eta_\infty(\nu)^2 < \infty \quad \text{and} \quad \sum_{\nu\in\mathfrak{I}} \big(\eta_\infty(\nu) - \eta_{\ell_k}(\nu)\big)^2 \xrightarrow{k\to\infty} 0.$$

Let $g:\mathbb{R}_{\ge 0}\to\mathbb{R}_{\ge 0}$ be a continuous function defined by $g(s) := g_{\mathbb{P}}(\sqrt{s})$ for all $s\in\mathbb{R}_{\ge 0}$. Setting $\eta_{\ell_k}(\mu)=0$ for $\mu\in\mathfrak{I}\setminus\mathfrak{Q}_{\ell_k}$, we deduce from (6.2) that

$$\eta_{\ell_k}(\mu) \le g_{\mathbb{P}}\big(\eta_{\ell_k}(\mathfrak{M}_{\ell_k})\big) = g\big(\eta_{\ell_k}(\mathfrak{M}_{\ell_k})^2\big) \quad \text{for all } k\in\mathbb{N}_0 \text{ and } \mu\in\mathfrak{I}\setminus\mathfrak{M}_{\ell_k}.$$

Note that the index set $\mathfrak{I}$ is countable, since it can be understood as a countable union of countable sets, and that $\mathfrak{P}_{\ell_n}\subseteq\mathfrak{P}_{\ell_n+1}\subseteq\mathfrak{P}_{\ell_{n+1}}$, since $\ell_n+1\le\ell_{n+1}$. Therefore, we can establish a one-to-one map between $\mathfrak{I}$ and $\mathbb{N}$, which allows us to identify each index set $\mathfrak{P}_{\ell_k}\subset\mathfrak{I}$ ($k\in\mathbb{N}_0$) with a set $\mathcal{P}_k\subset\mathbb{N}$. Then $\mathcal{P}_k\subseteq\mathcal{P}_{k+1}$ and applying Lemma 15 to the sequences $(x_n)_{n\in\mathbb{N}} := (\eta_\infty(\nu))_{\nu\in\mathfrak{I}}$, $(x_n^{(k)})_{n\in\mathbb{N}} := (\eta_{\ell_k}(\nu))_{\nu\in\mathfrak{I}}$, we prove that

$$\sum_{\nu\in\mathfrak{I}\setminus\mathfrak{P}_{\ell_k}} \eta_\infty(\nu)^2 \xrightarrow{k\to\infty} 0.$$

Note that the sequence $(z_\ell)_{\ell\in\mathbb{N}_0} := \big(\sum_{\nu\in\mathfrak{I}\setminus\mathfrak{P}_\ell}\eta_\infty(\nu)^2\big)_{\ell\in\mathbb{N}_0}$ is monotonic decreasing and bounded from below. Hence, it is convergent. Moreover, it has a subsequence that converges to zero. We therefore conclude that

$$\sum_{\nu\in\mathfrak{Q}_\ell} \eta_\infty(\nu)^2 \le \sum_{\nu\in\mathfrak{I}\setminus\mathfrak{P}_\ell} \eta_\infty(\nu)^2 \xrightarrow{\ell\to\infty} 0.$$

Overall, we derive that

$$\eta_\ell(\mathfrak{Q}_\ell)^2 = \sum_{\nu\in\mathfrak{Q}_\ell}\eta_\ell(\nu)^2 \lesssim \sum_{\nu\in\mathfrak{Q}_\ell}\eta_\infty(\nu)^2 + \sum_{\nu\in\mathfrak{I}}\big(\eta_\infty(\nu)-\eta_\ell(\nu)\big)^2 \xrightarrow{\ell\to\infty} 0.$$

This concludes the proof. $\square$

7.4. Proof of Proposition 11. The proof of Proposition 11 essentially follows the same lines as that of [29, Theorem 2.1]. Therefore, here we only sketch the proof by demonstrating how the results of [29] for deterministic problems can be extended to the parametric setting in the present paper.

We start by observing that the variational problem (2.6), its discretization, and the proposed adaptive algorithm satisfy the general framework described in [29, section 2]: The variational formulation (2.6) clearly fits into the class of problems considered in [29, section 2.1]; Our Galerkin discretization (3.4) satisfies the assumptions in [29, eqs. (2.6)–(2.8)]; The spatial NVB refinement considered in the present paper satisfies the assumptions on the mesh refinement in [29, eqs. (2.5) and (2.14)]; The weak marking condition (6.3) in Proposition 11 is the same as the marking condition in [29, eq. (2.13)]. Finally, we prove in Lemma 16 below that the local discrete efficiency estimate holds in the parametric setting (cf. [29, eq. (2.9b)]). Note that the global reliability of the estimator (see (3.17) and [29, eq. (2.9a)]) is not exploited in this section (and, hence, not needed for the proof of Theorem 5). The reliability is only needed to establish convergence of the true error, i.e., $\|u - u_\ell\| \to 0$ as $\ell \to \infty$ (see Corollary 7).

We will use the following notation: For $\omega \subset D$, we define

$$B_\omega(v, w) := \int_\Gamma \int_\omega a_0 \nabla u \cdot \nabla v \,\mathrm{d}x \,\mathrm{d}\pi(\mathbf{y}) + \sum_{m=1}^{\infty} \int_\Gamma \int_\omega y_m a_m \nabla u \cdot \nabla v \,\mathrm{d}x \,\mathrm{d}\pi(\mathbf{y}) \text{ for } v, w \in \mathbb{V}.$$

Note that $B_\omega(\cdot,\cdot)$ is symmetric, bilinear, and positive semidefinite. We denote by $\|v\|_\omega := B_\omega(v, v)^{1/2}$ the corresponding induced seminorm. Furthermore, in addition to the limiting space $\mathbb{V}_\infty$ introduced in Lemma 13, we define the spatial limiting space $\mathbb{X}_\infty := \bigcup_{\ell=0}^{\infty} \mathbb{X}_\ell$.

LEMMA 16. *Let $z \in \mathcal{N}_\ell^+$ and denote by $\omega_\ell(z) := \bigcup\{T \in \mathcal{T}_\ell : z \in T\}$ the associated vertex patch. Then the following estimate holds:*

$$\eta_\ell(z) \le C \|u - u_\ell\|_{\omega_\ell(z)}. \tag{7.8a}$$

Furthermore, let $u_\infty \in \mathbb{V}$ be the limit of $(u_\ell)_{\ell \in \mathbb{N}_0}$ guaranteed by Lemma 13. If $\widehat{\varphi}_{\ell,z} \in \mathbb{X}_\infty$, then there holds

$$\eta_\ell(z) \le C \|u_\infty - u_\ell\|_{\omega_\ell(z)}. \tag{7.8b}$$

The constant $C > 0$ in (7.8a) *and* (7.8b) *depends only on a_0 and τ.*

Proof. We recall the definition of the spatial error indicators in (3.8):

$$\eta_\ell(z)^2 = \sum_{\nu \in \mathfrak{P}_\ell} \frac{|F(\widehat{\varphi}_{\ell,z} P_\nu) - B(u_\ell, \widehat{\varphi}_{\ell,z} P_\nu)|^2}{\|a_0^{1/2} \nabla \widehat{\varphi}_{\ell,z}\|_{L^2(D)}^2} = \sum_{\nu \in \mathfrak{P}_\ell} \|\mathcal{G}_{\ell,z,\nu} e_\ell\|_0^2 \quad \text{for all } z \in \mathcal{N}_\ell^+,$$

where $\mathcal{G}_{\ell,z,\nu} : \mathbb{V} \to \mathrm{span}\{\widehat{\varphi}_{\ell,z} P_\nu\}$ is the orthogonal projection onto $\mathrm{span}\{\widehat{\varphi}_{\ell,z} P_\nu\}$ with respect to $B_0(\cdot,\cdot)$, and $e_\ell \in \widehat{\mathbb{X}} \otimes \mathbb{P}_\ell$ solves

$$B_0(e_\ell, v_\ell) = F(v_\ell) - B(u_\ell, v_\ell) \quad \text{for all } v_\ell \in \widehat{\mathbb{X}} \otimes \mathbb{P}_\ell.$$

Note that the functions $\{\widehat{\varphi}_{\ell,z} P_\nu : \nu \in \mathfrak{P}_\ell\}$ are orthogonal with respect to $B_0(\cdot,\cdot)$. Hence, $\sum_{\nu \in \mathfrak{P}_\ell} \mathcal{G}_{\ell,z,\nu} : \mathbb{V} \to \mathrm{span}\{\widehat{\varphi}_{\ell,z} P_\nu : \nu \in \mathfrak{P}_\ell\} \subset \widehat{\mathbb{X}} \otimes \mathbb{P}_\ell$ is an orthogonal projection with respect to $B_0(\cdot,\cdot)$ as well. This yields that

$$\begin{aligned}\eta_\ell(z)^2 &= \sum_{\nu\in\mathfrak{P}_\ell} |||\mathcal{G}_{\ell,z,\nu}e_\ell|||_0^2 = \Big|\Big|\Big| \sum_{\nu\in\mathfrak{P}_\ell} \mathcal{G}_{\ell,z,\nu}e_\ell \Big|\Big|\Big|_0^2 = B_0\Big(e_\ell, \sum_{\nu\in\mathfrak{P}_\ell} \mathcal{G}_{\ell,z,\nu}e_\ell\Big) \\ &= F\Big(\sum_{\nu\in\mathfrak{P}_\ell} \mathcal{G}_{\ell,z,\nu}e_\ell\Big) - B\Big(u_\ell, \sum_{\nu\in\mathfrak{P}_\ell} \mathcal{G}_{\ell,z,\nu}e_\ell\Big) = B\Big(u-u_\ell, \sum_{\nu\in\mathfrak{P}_\ell} \mathcal{G}_{\ell,z,\nu}e_\ell\Big).\end{aligned}$$

Note that the spatial support of $\sum_{\nu\in\mathfrak{P}_\ell} \mathcal{G}_{\ell,z,\nu}e_\ell$ lies in $\omega := \mathrm{supp}(\widehat{\varphi}_{\ell,z})$. Then, the Cauchy–Schwarz inequality shows that

$$\begin{aligned}\Big|\Big|\Big| \sum_{\nu\in\mathfrak{P}_\ell} \mathcal{G}_{\ell,z,\nu}e_\ell \Big|\Big|\Big|_0^2 &= B\Big(u-u_\ell, \sum_{\nu\in\mathfrak{P}_\ell} \mathcal{G}_{\ell,z,\nu}e_\ell\Big) = B_\omega\Big(u-u_\ell, \sum_{\nu\in\mathfrak{P}_\ell} \mathcal{G}_{\ell,z,\nu}e_\ell\Big) \\ &\le |||u-u_\ell|||_\omega \Big|\Big|\Big| \sum_{\nu\in\mathfrak{P}_\ell} \mathcal{G}_{\ell,z,\nu}e_\ell \Big|\Big|\Big|_\omega \le |||u-u_\ell|||_\omega \Big|\Big|\Big| \sum_{\nu\in\mathfrak{P}_\ell} \mathcal{G}_{\ell,z,\nu}e_\ell \Big|\Big|\Big| \\ &\overset{(2.5)}{\simeq} |||u-u_\ell|||_\omega \Big|\Big|\Big| \sum_{\nu\in\mathfrak{P}_\ell} \mathcal{G}_{\ell,z,\nu}e_\ell \Big|\Big|\Big|_0.\end{aligned}$$

We have thus shown that

$$\eta_\ell(z)^2 = B\Big(u-u_\ell, \sum_{\nu\in\mathfrak{P}_\ell} \mathcal{G}_{\ell,z,\nu}e_\ell\Big) = \Big|\Big|\Big| \sum_{\nu\in\mathfrak{P}_\ell} \mathcal{G}_{\ell,z,\nu}e_\ell \Big|\Big|\Big|_0^2 \lesssim |||u-u_\ell|||_\omega^2.$$

Since $\omega = \mathrm{supp}(\widehat{\varphi}_{\ell,z}) \subseteq \omega_\ell(z)$, this proves (7.8a).

Finally, if $\widehat{\varphi}_{\ell,z} \in \mathbb{X}_\infty$, then $\sum_{\nu\in\mathfrak{P}_\ell} \mathcal{G}_{\ell,z,\nu}e_\ell \in \mathbb{V}_\infty$. Therefore, the same arguments as above yield (7.8b). ∎

Note that in the present setting, the estimates (7.8a) and (7.8b) from Lemma 16 replace [29, eq. (2.9b)] and [29, eq. (4.11)], respectively. Having these estimates, we can now proceed to the proof of Proposition 11.

Proof of Proposition 11. Let $\mathcal{T}_\infty := \bigcup_{k\ge0}\bigcap_{\ell\ge k}\mathcal{T}_\ell$ be the set of all elements which remain unrefined after finitely many steps of refinement. Following [29, eqs. (4.10)], for all $\ell\in\mathbb{N}_0$, we consider the decomposition $\mathcal{T}_\ell = \mathcal{T}_\ell^{\mathrm{good}} \cup \mathcal{T}_\ell^{\mathrm{bad}} \cup \mathcal{T}_\ell^{\mathrm{neither}}$, where

$$\begin{aligned}\mathcal{T}_\ell^{\mathrm{good}} &:= \{T\in\mathcal{T}_\ell : \widehat{\varphi}_{\ell,z}\in\mathbb{X}_\infty \text{ for all } z\in\mathcal{N}_\ell^+\cap T\}, \\ \mathcal{T}_\ell^{\mathrm{bad}} &:= \{T\in\mathcal{T}_\ell : T'\in\mathcal{T}_\infty \text{ for all } T'\in\mathcal{T}_\ell \text{ with } T\cap T'\neq\emptyset\}, \\ \mathcal{T}_\ell^{\mathrm{neither}} &:= \mathcal{T}_\ell \setminus (\mathcal{T}_\ell^{\mathrm{good}}\cup\mathcal{T}_\ell^{\mathrm{bad}}).\end{aligned}$$

The elements in $\mathcal{T}_\ell^{\mathrm{good}}$ are refined sufficiently many times in order to guarantee (7.8b). The set $\mathcal{T}_\ell^{\mathrm{bad}}$ consists of all elements such that the whole element patch remains unrefined. The remaining elements are collected in the set $\mathcal{T}_\ell^{\mathrm{neither}}$. We note that $\mathcal{T}_\ell^{\mathrm{good}}$ is slightly larger than the corresponding set $\mathcal{G}_\ell^0$ in [29, eq. (4.10a)], while $\mathcal{T}_\ell^{\mathrm{bad}}$ coincides with the corresponding set $\mathcal{G}_\ell^+$ in [29, eq. (4.10b)]. As a consequence, $\mathcal{T}_\ell^{\mathrm{neither}}$ is smaller than the corresponding set $\mathcal{G}_\ell^*$ in [29, eq. (4.10c)].

By arguing as in the proof of [29, Propostion 4.1], we exploit the uniform shape regularity of $\mathcal{T}_\ell$ guaranteed by NVB and use Lemmas 13 and 16 to prove that

$$\sum_{T\in\mathcal{T}_\ell^{\mathrm{good}}}\sum_{z\in\mathcal{N}_\ell^+\cap T}\eta_\ell(z)^2 \overset{(7.8b)}{\lesssim} \sum_{T\in\mathcal{T}_\ell^{\mathrm{good}}}\sum_{z\in\mathcal{N}_\ell^+\cap T} |||u_\infty-u_\ell|||_{\omega_\ell(z)}^2 \lesssim |||u_\infty-u_\ell|||^2 \xrightarrow{\ell\to\infty} 0. \tag{7.9}$$

Let $D_\ell^{\rm neither} := \bigcup\{T' \in \mathcal{T}_\ell : T \cap T' \neq \emptyset \text{ for some } T \in \mathcal{T}_\ell^{\rm neither}\}$. Since $\mathcal{T}_\ell^{\rm neither}$ is contained in the corresponding set $\mathcal{G}_\ell^*$ in [29, eq. (4.10c)], arguing as in Step 1 of the proof of [29, Proposition 4.2], we show that $|D_\ell^{\rm neither}| \to 0$ as $\ell \to \infty$. Hence, Lemma 16, uniform shape regularity, and the fact that the local energy seminorm is absolutely continuous with respect to the Lebesgue measure, i.e., $|\!|\!| v |\!|\!|_\omega \to 0$ as $|\omega| \to 0$ for all $v \in \mathbb{V}$, lead to

(7.10)
$$\sum_{T\in\mathcal{T}_\ell^{\rm neither}} \sum_{z\in\mathcal{N}_\ell^+\cap T} \eta_\ell(z)^2 \overset{(7.8a)}{\lesssim} \sum_{T\in\mathcal{T}_\ell^{\rm neither}} \sum_{z\in\mathcal{N}_\ell^+\cap T} |\!|\!| u - u_\ell |\!|\!|^2_{\omega_\ell(z)} \lesssim |\!|\!| u - u_\ell |\!|\!|^2_{D_\ell^{\rm neither}} \to 0$$

as $\ell \to \infty$. We note that (7.9) and (7.10) hold independently of the marking property (6.3), but rely only on the nestedness of the finite-dimensional subspaces $\mathbb{X}_\ell \subseteq \mathbb{X}_{\ell+1}$ and $\mathbb{V}_\ell \subseteq \mathbb{V}_{\ell+1}$ for all $\ell \in \mathbb{N}_0$.

To conclude the proof, it remains to consider the set $\mathcal{T}_\ell^{\rm bad}$. Let $(\mathcal{T}_{\ell_k})_{k\in\mathbb{N}_0}$ be the subsequence of $(\mathcal{T}_\ell)_{\ell\in\mathbb{N}_0}$ satisfying (6.3). If $z \in \mathcal{M}_{\ell_k}$ and $T \in \mathcal{T}_{\ell_k}$ with $z \in T$, then $T \in \mathcal{T}_{\ell_k} \setminus \mathcal{T}_{\ell_k}^{\rm bad} = \mathcal{T}_{\ell_k}^{\rm good} \cup \mathcal{T}_{\ell_k}^{\rm neither}$. Therefore, it follows from (7.9)–(7.10) that

$$\sum_{z\in\mathcal{M}_{\ell_k}} \eta_{\ell_k}(z)^2 \le \sum_{T\in\mathcal{T}_{\ell_k}^{\rm good}} \sum_{z\in\mathcal{N}_{\ell_k}^+\cap T} \eta_{\ell_k}(z)^2 + \sum_{T\in\mathcal{T}_{\ell_k}^{\rm neither}} \sum_{z\in\mathcal{N}_{\ell_k}^+\cap T} \eta_{\ell_k}(z)^2 \xrightarrow{k\to\infty} 0.$$

This implies that

$$0 \le \eta_{\ell_k}(z) \overset{(6.3)}{\le} g_{\mathbb{X}}\big(\eta_{\ell_k}(\mathcal{M}_{\ell_k})\big) \xrightarrow{k\to\infty} 0 \quad \text{for all } z \in \mathcal{N}_{\ell_k}^+ \setminus \mathcal{M}_{\ell_k}.$$

Hence, recalling the definition of $\mathcal{T}_\bullet^{\rm bad}$, we obtain (cf., [29, eq. (4.17)])

$$\sum_{z\in\mathcal{N}_{\ell_k}^+\cap T} \eta_{\ell_k}(z)^2 \xrightarrow{k\to\infty} 0 \quad \text{for all } T \in \mathcal{T}_{\ell_k}^{\rm bad}. \tag{7.11}$$

Finally, arguing as in Steps 2–5 of the proof of [29, Proposition 4.3], we use (7.11) and apply the Lebesgue dominated convergence theorem to derive that

$$\sum_{T\in\mathcal{T}_{\ell_k}^{\rm bad}} \sum_{z\in\mathcal{N}_{\ell_k}^+\cap T} \eta_{\ell_k}(z)^2 \xrightarrow{k\to\infty} 0. \tag{7.12}$$

Combining now (7.9)–(7.12), we find that

$$\begin{aligned} \eta_{\ell_k}(\mathcal{N}_{\ell_k}^+)^2 &= \sum_{z\in\mathcal{N}_{\ell_k}^+} \eta_{\ell_k}(z)^2 \\ &\le \sum_{T\in\mathcal{T}_{\ell_k}^{\rm good}} \sum_{z\in\mathcal{N}_{\ell_k}^+\cap T} \eta_{\ell_k}(z)^2 + \sum_{T\in\mathcal{T}_{\ell_k}^{\rm neither}} \sum_{z\in\mathcal{N}_{\ell_k}^+\cap T} \eta_{\ell_k}(z)^2 + \sum_{T\in\mathcal{T}_{\ell_k}^{\rm bad}} \sum_{z\in\mathcal{N}_{\ell_k}^+\cap T} \eta_{\ell_k}(z)^2 \to 0 \end{aligned}$$

as $k \to \infty$. This concludes the proof. ∎

8. Proof of Theorem 8 (linear convergence). In this section, we prove that in 2D the saturation assumption yields contraction of the energy error at each iteration of Algorithms 4.A and 4.B. In the proof, we adapt the arguments of [14, 28]. In particular, the following result holds for iterations where the spatial refinement is performed.

LEMMA 17. *Let $\ell \in \mathbb{N}_0$. Suppose that the saturation assumption* (3.16) *holds for two Galerkin solutions u_ℓ and $\widehat{u}_\ell$ satisfying* (3.4) *and* (3.12)*, respectively. Suppose that*

$$\eta_\ell(\mathfrak{Q}_\ell) \le C_\vartheta\, \eta_\ell(\mathcal{N}_\ell^+) \quad \text{with } C_\vartheta > 0 \tag{8.1}$$

and let $\mathcal{M}_\ell \subseteq \mathcal{N}_\ell^+ \cap \mathcal{N}_{\ell+1}$ be such that

$$\theta\, \eta_\ell(\mathcal{N}_\ell^+) \le \eta_\ell(\mathcal{M}_\ell) \quad \text{with } 0 < \theta \le 1. \tag{8.2}$$

Then, for the enhanced Galerkin solution $u_{\ell+1} \in \mathbb{X}_{\ell+1} \otimes \mathbb{P}_\ell$, there holds

$$|\!|\!| u - u_{\ell+1} |\!|\!|^2 \le (1-q)\, |\!|\!| u - u_\ell |\!|\!|^2,$$

where $0 < q < 1$ depends only on a_0, C_ϑ, $q_{\rm sat}$, $\mathcal{T}_0$, τ, and θ.

Proof. Using reliability (3.17), the refinement criterion (8.1), and the marking criterion (8.2), we obtain

$$\frac{1-q_{\rm sat}^2}{\Lambda C_{\rm thm}}\, |\!|\!| u - u_\ell |\!|\!|^2 \le \eta_\ell(\mathcal{N}_\ell^+)^2 + \eta_\ell(\mathfrak{Q}_\ell)^2 \le (1+C_\vartheta^2)\, \eta_\ell(\mathcal{N}_\ell^+)^2 \le (1+C_\vartheta^2)\theta^{-2}\, \eta_\ell(\mathcal{M}_\ell)^2.$$

Hence, using Corollary 3 and the fact that $\mathcal{M}_\ell \subseteq \mathcal{N}_\ell^+ \cap \mathcal{N}_{\ell+1}$, we derive that

$$\begin{aligned}
|\!|\!| u - u_{\ell+1} |\!|\!|^2 = |\!|\!| u - u_\ell |\!|\!|^2 - |\!|\!| u_{\ell+1} - u_\ell |\!|\!|^2 &\le |\!|\!| u - u_\ell |\!|\!|^2 - \frac{\lambda}{K}\, \eta_\ell(\mathcal{N}_\ell^+ \cap \mathcal{N}_{\ell+1})^2 \\
&\le |\!|\!| u - u_\ell |\!|\!|^2 - \frac{\lambda}{K}\, \eta_\ell(\mathcal{M}_\ell)^2 \le \left(1 - \frac{\lambda}{\Lambda} \cdot \frac{\theta^2(1-q_{\rm sat}^2)}{C_{\rm thm}(1+C_\vartheta^2)K}\right) |\!|\!| u - u_\ell |\!|\!|^2.
\end{aligned}$$

This concludes the proof. ∎

The next lemma concerns iterations where parametric enrichment is performed. The proof is similar to that of Lemma 17.

LEMMA 18. *Let $\ell \in \mathbb{N}_0$. Suppose that the saturation assumption* (3.16) *holds for two Galerkin solutions u_ℓ and $\widehat{u}_\ell$ satisfying* (3.4) *and* (3.12)*, respectively. Suppose that $\eta_\ell(\mathcal{N}_\ell^+) \le C_\vartheta\, \eta_\ell(\mathfrak{Q}_\ell)$ with $C_\vartheta > 0$ and let $\mathfrak{M}_\ell \subseteq \mathfrak{Q}_\ell \cap \mathfrak{P}_{\ell+1}$ be such that $\theta\, \eta_\ell(\mathfrak{Q}_\ell) \le \eta_\ell(\mathfrak{M}_\ell)$ with $0 < \theta \le 1$. Then, for the enhanced Galerkin solution $u_{\ell+1} \in \mathbb{X}_\ell \otimes \mathbb{P}_{\ell+1}$, there holds*

$$|\!|\!| u - u_{\ell+1} |\!|\!|^2 \le (1-q)\, |\!|\!| u - u_\ell |\!|\!|^2,$$

where $0 < q < 1$ depends only on a_0, C_ϑ, $q_{\rm sat}$, $\mathcal{T}_0$, τ, and θ.

With these results, we can prove Theorem 8.

Proof of Theorem 8. We divide the proof into two steps.

Step 1. Consider Algorithm 4.A. In case (a) of Criterion A, we apply Lemma 17 with $C_\vartheta = \vartheta^{-1}$ and $\theta = \theta_{\mathbb{X}}$, whereas in case (b) of this marking criterion, we use Lemma 18 with $C_\vartheta = \vartheta$ and $\theta = \theta_{\mathbb{P}}$. In both cases, this proves contraction of the energy error $|\!|\!| u - u_{\ell+1} |\!|\!| \le q_{\rm lin}\, |\!|\!| u - u_\ell |\!|\!|$ with $q_{\rm lin} \in (0,1)$.

Step 2. Consider now Algorithm 4.B. In case (a) of Criterion B one has

$$\theta_{\mathbb{P}}\, \eta_\ell(\mathfrak{Q}_\ell) \le \eta_\ell(\widetilde{\mathfrak{M}}_\ell) \le \vartheta^{-1} \eta_\ell(\widetilde{\mathcal{R}}_\ell) \le \vartheta^{-1} \eta_\ell(\mathcal{N}_\ell^+).$$

Hence, Lemma 17 applies to this case with $C_\vartheta = \theta_{\mathbb{P}}^{-1}\vartheta^{-1}$ and $\theta = \theta_{\mathbb{X}}$. Similarly, in case (b) of Criterion B, one has

$$\theta_{\mathbb{X}}\, \eta_\ell(\mathcal{N}_\ell^+) \le \eta_\ell(\widetilde{\mathcal{M}}_\ell) \le \eta_\ell(\widetilde{\mathcal{R}}_\ell) < \vartheta\, \eta_\ell(\widetilde{\mathfrak{M}}_\ell) \le \vartheta\, \eta_\ell(\mathfrak{Q}_\ell).$$

Hence, in this case, Lemma 18 applies with $C_\vartheta = \theta_{\mathbb{X}}^{-1}\vartheta$ and $\theta = \theta_{\mathbb{P}}$. Thus, in both cases, we obtain contraction of the energy error $\lVert u - u_{\ell+1} \rVert \le q_{\rm lin} \lVert u - u_\ell \rVert$ with $q_{\rm lin} \in (0,1)$. ◻

Acknowledgment. This work was initiated and part of it was undertaken when AB visited the Institute for Analysis and Scientific Computing at TU Wien in 2018. This author wishes to thank the colleagues in that institute for hospitality and stimulating research atmosphere.

REFERENCES

[1] M. Ainsworth and J. T. Oden, *A Posteriori Error Estimation in Finite Element Analysis*, Pure Appl. Math., Wiley, New York, 2000, https://doi.org/10.1002/9781118032824.

[2] I. Babuška and M. Vogelius, *Feedback and adaptive finite element solution of one-dimensional boundary value problems*, Numer. Math., 44 (1984), pp. 75–102, https://doi.org/10.1007/BF01389757.

[3] A. Bespalov, C. E. Powell, and D. Silvester, *Energy norm a posteriori error estimation for parametric operator equations*, SIAM J. Sci. Comput., 36 (2014), pp. A339–A363, https://doi.org/10.1137/130916849.

[4] A. Bespalov, D. Praetorius, L. Rocchi, and M. Ruggeri, *Convergence of Adaptive Stochastic Galerkin FEM*, 2018, https://arxiv.org/abs/1811.09462.

[5] A. Bespalov, D. Praetorius, L. Rocchi, and M. Ruggeri, *Goal-oriented error estimation and adaptivity for elliptic PDEs with parametric or uncertain inputs*, Comput. Methods Appl. Mech. Engrg., 345 (2019), pp. 951–982, https://doi.org/10.1016/j.cma.2018.10.041.

[6] A. Bespalov and L. Rocchi, *Efficient adaptive algorithms for elliptic PDEs with random data*, SIAM/ASA J. Uncertain. Quantif., 6 (2018), pp. 243–272, https://doi.org/10.1137/17M1139928.

[7] A. Bespalov and L. Rocchi, *Stochastic T-IFISS*, http://web.mat.bham.ac.uk/A.Bespalov/software/index.html#stoch_tifiss (2019).

[8] A. Bespalov and D. Silvester, *Efficient adaptive stochastic Galerkin methods for parametric operator equations*, SIAM J. Sci. Comput., 38 (2016), pp. A2118–A2140, https://doi.org/10.1137/15M1027048.

[9] T. Butler, P. Constantine, and T. Wildey, *A posteriori error analysis of parameterized linear systems using spectral methods*, SIAM J. Matrix Anal. Appl., 33 (2012), pp. 195–209, https://doi.org/10.1137/110840522.

[10] J. M. Cascon, C. Kreuzer, R. H. Nochetto, and K. G. Siebert, *Quasi-optimal convergence rate for an adaptive finite element method*, SIAM J. Numer. Anal., 46 (2008), pp. 2524–2550, https://doi.org/10.1137/07069047X.

[11] A. Cohen and R. DeVore, *Approximation of high-dimensional parametric PDEs*, Acta Numer., 24 (2015), pp. 1–159, https://doi.org/10.1017/S0962492915000033.

[12] A. Cohen, R. DeVore, and C. Schwab, *Convergence rates of best N-term Galerkin approximations for a class of elliptic sPDEs*, Found. Comput. Math., 10 (2010), pp. 615–646, https://doi.org/10.1007/s10208-010-9072-2.

[13] A. J. Crowder, C. E. Powell, and A. Bespalov, *Efficient adaptive multilevel stochastic Galerkin approximation using implicit a posteriori error estimation*, SIAM J. Sci. Comput., 41 (2019), pp. A1681–A1705, https://doi.org/10.1137/18M1194420.

[14] W. Dörfler, *A convergent adaptive algorithm for Poisson's equation*, SIAM J. Numer. Anal., 33 (1996), pp. 1106–1124, https://doi.org/10.1137/0733054.

[15] M. Eigel, C. J. Gittelson, C. Schwab, and E. Zander, *Adaptive stochastic Galerkin FEM*, Comput. Methods Appl. Mech. Engrg., 270 (2014), pp. 247–269, https://doi.org/10.1016/j.cma.2013.11.015.

[16] M. Eigel, C. J. Gittelson, C. Schwab, and E. Zander, *A convergent adaptive stochastic Galerkin finite element method with quasi-optimal spatial meshes*, ESAIM Math. Model. Numer. Anal., 49 (2015), pp. 1367–1398, https://doi.org/10.1051/m2an/2015017.

[17] M. Eigel and C. Merdon, *Local equilibration error estimators for guaranteed error control in adaptive stochastic higher-order Galerkin finite element methods*, SIAM/ASA J. Uncertain. Quantif., 4 (2016), pp. 1372–1397, https://doi.org/10.1137/15M102188X.

[18] C. Erath, G. Gantner, and D. Praetorius, *Optimal convergence behavior of adaptive FEM driven by simple* $(h-h/2)$*-type error estimators*, Comput. Math. Appl., (2019), https://doi.org/10.1016/j.camwa.2019.07.014.

[19] D. Estep, M. J. Holst, and A. Målqvist, *Nonparametric density estimation for randomly perturbed elliptic problems* III: *Convergence, computational cost, and generalizations*, J. Appl. Math. Comput., 38 (2012), pp. 367–387, https://doi.org/10.1007/s12190-011-0484-1.

[20] D. Estep, A. Målqvist, and S. Tavener, *Nonparametric density estimation for randomly perturbed elliptic problems* I: *Computational methods, a posteriori analysis, and adaptive error control*, SIAM J. Sci. Comput., 31 (2009), pp. 2935–2959, https://doi.org/10.1137/080731670.

[21] D. Estep, A. Målqvist, and S. Tavener, *Nonparametric density estimation for randomly perturbed elliptic problems* II: *Applications and adaptive modeling*, Internat. J. Numer. Methods Engrg., 80 (2009), pp. 846–867, https://doi.org/10.1002/nme.2547.

[22] M. Feischl, D. Praetorius, and K. G. van der Zee, *An abstract analysis of optimal goal-oriented adaptivity*, SIAM J. Numer. Anal., 54 (2016), pp. 1423–1448, https://doi.org/10.1137/15M1021982.

[23] M. B. Giles and E. Süli, *Adjoint methods for PDEs: A posteriori error analysis and post-processing by duality*, Acta Numer., 11 (2002), pp. 145–236, https://doi.org/10.1017/CBO9780511550140.003.

[24] C. J. Gittelson, *An adaptive stochastic Galerkin method for random elliptic operators*, Math. Comp., 82 (2013), pp. 1515–1541, https://doi.org/10.1090/S0025-5718-2013-02654-3.

[25] M. Karkulik, D. Pavlicek, and D. Praetorius, *On 2D newest vertex bisection: Optimality of mesh-closure and* H^1*-stability of* L_2*-projection*, Constr. Approx., 38 (2013), pp. 213–234, https://doi.org/10.1007/s00365-013-9192-4.

[26] A. Khan, A. Bespalov, C. E. Powell, and D. J. Silvester, *Robust A Posteriori Error Estimation for Stochastic Galerkin Formulations of Parameter-Dependent Linear Elasticity Equations*, 2018, https://arxiv.org/abs/1810.07440.

[27] L. Mathelin and O. Le Maître, *Dual-based a posteriori error estimate for stochastic finite element methods*, Commun. Appl. Math. Comput. Sci., 2 (2007), pp. 83–115, https://doi.org/10.2140/camcos.2007.2.83.

[28] P. Morin, R. H. Nochetto, and K. G. Siebert, *Data oscillation and convergence of adaptive FEM*, SIAM J. Numer. Anal., 38 (2000), pp. 466–488, https://doi.org/10.1137/S0036142999360044.

[29] P. Morin, K. G. Siebert, and A. Veeser, *A basic convergence result for conforming adaptive finite elements*, Math. Models Methods Appl. Sci., 18 (2008), pp. 707–737, https://doi.org/10.1142/S0218202508002838.

[30] C. Schwab and C. J. Gittelson, *Sparse tensor discretizations of high-dimensional parametric and stochastic PDEs*, Acta Numer., 20 (2011), pp. 291–467, https://doi.org/10.1017/S0962492911000055.

[31] R. Stevenson, *The completion of locally refined simplicial partitions created by bisection*, Math. Comp., 77 (2008), pp. 227–241, https://doi.org/10.1090/S0025-5718-07-01959-X.

[32] X. Wan and G. E. Karniadakis, *Error control in multi-element generalized polynomial chaos method for elliptic problems with random coefficients*, Commun. Comput. Phys., 5 (2009), pp. 793–820.