

*Energy norm a posteriori error estimation for  
parametric operator equations*

Bespalov, Alex and Powell, Catherine E. and Silvester,  
David

2014

MIMS EPrint: **2013.16**

Manchester Institute for Mathematical Sciences  
School of Mathematics

The University of Manchester

Reports available from: <http://eprints.maths.manchester.ac.uk/>

And by contacting: The MIMS Secretary  
School of Mathematics  
The University of Manchester  
Manchester, M13 9PL, UK

ISSN 1749-9097

## ENERGY NORM A POSTERIORI ERROR ESTIMATION FOR PARAMETRIC OPERATOR EQUATIONS\*

ALEX BESPALOV<sup>†</sup>, CATHERINE E. POWELL<sup>‡</sup>, AND DAVID SILVESTER<sup>‡</sup>

*Dedicated to Professor Norbert Heuer on the occasion of his 50th birthday*

**Abstract.** Stochastic Galerkin approximation is an increasingly popular approach for the solution of elliptic PDE problems with correlated random data. A typical strategy is to combine conventional ( $h$ -)finite element approximation on the spatial domain with spectral ( $p$ -)approximation on a finite-dimensional manifold in the (stochastic) parameter domain. The issues involved in a posteriori error analysis of computed solutions are outlined in this paper using an abstract setting of parametric operator equations. A novel energy error estimator that uses a parameter-free part of the underlying differential operator is introduced which effectively exploits the tensor product structure of the approximation space. We prove that our error estimator is reliable and efficient. We also discuss different strategies for enriching the approximation space and prove two-sided estimates of the error reduction for the corresponding enhanced approximations. These give computable estimates of the error reduction that depend only on the problem data and the original approximation.

**Key words.** stochastic Galerkin methods, stochastic finite elements, random data, Karhunen–Loève expansion, parametric operator equations, error estimation, a posteriori error analysis

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

**DOI.** 10.1137/130916849

**1. Introduction.** Stochastic Galerkin finite element methods (sGFEMs) for partial differential equations (PDEs) with random data have been well studied in the last decade. Typically, such discretizations generate solutions in tensor product spaces  $X \otimes P$ , where  $X$  is a finite element space associated with a physical domain  $D \subset \mathbb{R}^2$  or  $D \subset \mathbb{R}^3$ , and  $P$  (often called the stochastic approximation space) is a set of polynomials in  $\mathbf{y} = (y_1, y_2, \dots) \in \Gamma$ . The variables  $y_m$  are images of random variables and  $\Gamma$  is a (possibly) high-dimensional or even infinite-dimensional parameter domain containing the observations. There are several strategies for defining the stochastic Galerkin approximation space  $X \otimes P$ . For  $P$ , it is common to employ polynomials that are orthogonal with respect to an underlying probability measure, giving rise to the name polynomial chaos (PC). Although studies of sGFEM discretizations are relatively mature, a posteriori error estimation techniques and adaptive refinement schemes are in need of development. There are two distinct sources of discretization error arising from the choice of  $X$  and  $P$ . Even if an efficient and reliable estimator is available for the total error in the Galerkin approximation, the fact that there are two distinct discretizations complicates the design of adaptive refinement schemes. The optimal refinement strategy remains an open question.

One possibility for  $P$  is to use global multivariate polynomials on the parameter domain  $\Gamma$ . Alternatively,  $\Gamma$  may be partitioned into hypercubes or “elements” (just

---

\*Submitted to the journal’s Methods and Algorithms for Scientific Computing section April 12, 2013; accepted for publication (in revised form) November 13, 2013; published electronically March 13, 2014. This work was supported by the EPSRC under grant EP/H021205/1.

<http://www.siam.org/journals/sisc/36-2/91684.html>

<sup>†</sup>School of Mathematics, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK (a.bespalov@bham.ac.uk).

<sup>‡</sup>School of Mathematics, University of Manchester, Oxford Road, Manchester, M13 9PL, UK (c.powell@manchester.ac.uk, david.silvester@manchester.ac.uk).

as  $D$  is partitioned in a finite element method based on  $h$ -refinement) in combination with piecewise polynomials of a fixed low degree. For sGFEM discretizations based on the latter strategy,  $P$  can be enriched by increasing the local polynomial degrees in each element or by refining the partition of  $\Gamma$ , or, indeed, by a combination of the two. Adaptive approaches have been suggested for enriching such stochastic approximation spaces in [15], [16], [17], and [23]. A review is also given in [14, Chapter 9]. However, the schemes in these papers are not based on rigorous error estimates but on error indicators derived from heuristic arguments concerning the relative contributions of the higher order terms to the local solution expansion (in the  $y_m$  variables). A rigorous error estimator is derived in [18] that replaces the aforementioned heuristic error indicators. Various adaptive refinement strategies are discussed in [18] and then tested for a stochastic Burgers equation. However, the authors comment that a limitation of their method is the lack of information regarding the *structure* of the estimated error. That is, the individual contributions from the two different discretizations are not readily accessible.

In [24], an sGFEM based on so-called multielement generalized PC expansions is considered. Working with a moderate number of variables  $y_m$  and polynomials of low degree,  $\Gamma$  is once again partitioned into elements. The error control is based on implicit a posteriori error estimation (and is described in the spirit of Ainsworth and Oden [1]). A reduced space is also employed to keep the cost of the error estimation under control. The residual-based a posteriori error estimator developed in [8] considers two contributions to the total sGFEM error: the error incurred by truncating the PC expansion and the error associated with the finite element approximations of the PC coefficients. Computable a posteriori error estimates for linear functionals of solutions to nonlinear PDEs with random data can be found in [6]. Similarly, estimates for pointwise evaluation of linear functionals of solutions to parameterized linear systems can be found in [5]. An adaptive stochastic Galerkin method based on wavelet approximation is developed in [12]. This latter work extends the sGFEM approximation methods based on sparse tensor product spaces that are described in [3]. A problem-adapted algorithm is developed in the latter paper that selects the deterministic and stochastic approximation spaces based on a priori estimates of the norms of the PC coefficients of the unknown solution.

The error estimation strategy that is developed in this paper has a somewhat different flavor than the strategies described in the above-mentioned references. Our idea is to use a posteriori estimates of the error reduction in the energy norm (rather than the error itself) to guide the adaptivity process. Such a strategy is not new: for example, it is used in [20] in the context of  $hp$ -adaptivity for deterministic problems and allows the possibility of choosing between two distinct local refinement strategies so as to construct an updated combination of the mesh and polynomial degrees that achieves the best energy error reduction.

The outline of the paper is as follows. Our starting point in section 2 is an abstract parametric operator equation—this represents a general setting for the class of PDEs with random data that is our primary focus, and it mirrors the presentation in the excellent review article [21]. We will restrict our attention to self-adjoint PDE operators throughout. Extensions to non-self-adjoint and nonlinear PDE problems might be possible using duality techniques and adjoint operators, but this is not our primary focus here. A concrete model problem with random data is discussed in section 2.1. Galerkin approximation of the abstract operator equation is developed in section 3. A residual-based error estimator that uses a parameter-free part of the

operator is introduced in section 4 and is shown to be reliable and efficient. Possible strategies for enriching the discrete space are suggested in section 5, and two-sided estimates of the error reduction are established for the corresponding enhanced approximations. The effectiveness of our proposed enrichment strategy is demonstrated numerically in section 6. The concluding section looks towards further development of these ideas—specifically, the design of a practical adaptive algorithm based on our error estimation strategy. Testing a practical implementation of this strategy will be the focus of future publications.

**2. Parametric operator equations.** Let  $\Gamma$  be a topological space and let  $H$  be a separable Hilbert space over  $\mathbb{R}$  with natural norm  $\|\cdot\|_H$ . We denote the dual space of  $H$  by  $H'$  and the corresponding duality pairing by  $\langle \cdot, \cdot \rangle$ . Our starting point is the parametric operator equation

$$(2.1) \quad A(y)u(y) = f(y) \quad \forall y \in \Gamma,$$

where  $A : \Gamma \rightarrow \mathcal{L}(H, H')$  and  $f : \Gamma \rightarrow H'$  are given continuous maps defining a parametric bounded linear operator in  $\mathcal{L}(H, H')$  and a parametric linear functional in  $H'$ , respectively. We assume that  $A(y)$  has a bounded inverse for all  $y \in \Gamma$  so that (2.1) has a unique solution  $u : \Gamma \rightarrow H$  which is a continuous map (see [21, Proposition 2.17]).

Let  $\pi = \pi(y)$  be a measure on  $(\Gamma, \mathcal{B}(\Gamma))$ , where  $\mathcal{B}(\Gamma)$  is the Borel  $\sigma$ -algebra on  $\Gamma$ , and denote by  $V$  the Bochner space  $L^2_\pi(\Gamma; H)$  with the norm

$$\|\cdot\|_V := \left( \int_\Gamma \|\cdot\|_H^2 d\pi(y) \right)^{1/2}.$$

Then we have the following weak formulation of (2.1): find  $u \in V$  such that

$$(2.2) \quad B(u, v) = F(v) \quad \forall v \in V.$$

Here, the bilinear form and linear functional are given by

$$(2.3) \quad B(u, v) := \int_\Gamma \langle A(y)u(y), v(y) \rangle d\pi(y)$$

and

$$(2.4) \quad F(v) := \int_\Gamma \langle f(y), v(y) \rangle d\pi(y),$$

respectively, and depend on the chosen measure  $\pi$ .

To ensure that (2.2) is well posed, we assume that  $f \in L^2_\pi(\Gamma; H')$ , the operator  $A(y)$  is symmetric and positive definite for all  $y \in \Gamma$ , and there exist positive constants  $\alpha_{\min}$  and  $\alpha_{\max}$  such that

$$(2.5) \quad \|A(y)\|_{\mathcal{L}(H, H')} \leq \alpha_{\max}, \quad \|A(y)^{-1}\|_{\mathcal{L}(H', H)} \leq \alpha_{\min}^{-1} \quad \forall y \in \Gamma.$$

Then  $B(u, v)$  and  $F(v)$  in (2.2) are well defined for any  $u, v \in V$ . Moreover, the bilinear form  $B(\cdot, \cdot)$  is symmetric and the following hold:

$$(2.6) \quad |B(u, v)| \leq \alpha_{\max} \|u\|_V \|v\|_V \quad \forall u, v \in V,$$

$$(2.7) \quad B(v, v) \geq \alpha_{\min} \|v\|_V^2 \quad \forall v \in V.$$

Consequently, the solution  $u$  to (2.1) is the unique element of  $V$  satisfying (2.2) (see [21, Theorem 2.18]). In addition, note that  $B(\cdot, \cdot)$  defines an inner product in  $V$  which induces the norm  $\|v\|_B := (B(v, v))^{1/2}$  equivalent to  $\|v\|_V$ .

We now make some additional assumptions on the parameter domain  $\Gamma$ , the measure  $\pi$ , and the parametric operator  $A$ . These are motivated by our goal to use stochastic finite element techniques to solve PDE problems with random data. In particular, we are interested in using spectral expansions (e.g., Karhunen–Loève expansions) to represent the data (see [11], [13], [21], and references therein). First, let

$$(2.8) \quad \Gamma := \prod_{m=1}^{\infty} \Gamma_m,$$

with  $\Gamma_m$  being bounded intervals in  $\mathbb{R}$ , and let  $\pi$  be a product measure. Then the elements of  $\Gamma$  are vectors, denoted by  $\mathbf{y} = (y_1, y_2, \dots) \in \Gamma$  in the rest of the paper, and

$$(2.9) \quad \pi(\mathbf{y}) := \prod_{m=1}^{\infty} \pi_m(y_m),$$

where  $\pi_m$  is a measure on  $(\Gamma_m, \mathcal{B}(\Gamma_m))$ . We now assume  $A(\mathbf{y})$  has the decomposition

$$(2.10) \quad A(\mathbf{y}) = A_0 + \sum_{m=1}^{\infty} y_m A_m \quad \forall \mathbf{y} \in \Gamma,$$

with convergence of the series to be understood in  $\mathcal{L}(H, H')$  uniformly in  $\mathbf{y}$ . To ensure the unique solvability of (2.1) and (2.2) for parametric operators  $A(\mathbf{y})$  of the form (2.10), we must make the following assumptions on  $A_0$  and  $A_m$ ,  $m \in \mathbb{N}$  (see [12, section 1]). The operator  $A_0$  is symmetric and positive definite. Hence, the bilinear form  $\langle A_0 \cdot, \cdot \rangle$  defines an inner product in  $H$  and there hold

$$(2.11) \quad |\langle A_0 u, v \rangle| \leq \alpha_0^{\max} \|u\|_H \|v\|_H \quad \forall u, v \in H,$$

$$(2.12) \quad \langle A_0 v, v \rangle \geq \alpha_0^{\min} \|v\|_H^2 \quad \forall v \in H$$

with positive constants  $\alpha_0^{\min}, \alpha_0^{\max}$ . For  $m \in \mathbb{N}$ , the operators  $A_m \in \mathcal{L}(H, H')$  in (2.10) are symmetric and there exists a constant  $\tau \in [0, 1)$  such that for all  $\mathbf{y} \in \Gamma$ ,

$$(2.13) \quad \left| \left\langle \sum_{m=1}^{\infty} y_m A_m v, v \right\rangle \right| \leq \tau \langle A_0 v, v \rangle \quad \forall v \in H.$$

Now, substituting (2.10) into (2.3) allows us to split  $B(\cdot, \cdot)$  and rewrite (2.2) as

$$(2.14) \quad B_0(u, v) + \sum_{m=1}^{\infty} B_m(u, v) = F(v) \quad \forall v \in V.$$

Here, the component bilinear forms  $B_m(\cdot, \cdot)$  for  $m \in \mathbb{N}_0$  are defined as

$$(2.15) \quad \begin{aligned} B_0(u, v) &:= \int_{\Gamma} \langle A_0 u(\mathbf{y}), v(\mathbf{y}) \rangle d\pi(\mathbf{y}), \\ B_m(u, v) &:= \int_{\Gamma} \langle A_m u(\mathbf{y}), v(\mathbf{y}) \rangle y_m d\pi(\mathbf{y}) \quad \forall m \in \mathbb{N}. \end{aligned}$$

Due to the above assumptions on  $A_0$ , the bilinear form  $B_0(\cdot, \cdot)$  defines an inner product in  $V$  which induces the norm  $\|v\|_{B_0} := (B_0(v, v))^{1/2}$  equivalent to  $\|v\|_V$ . Furthermore, thanks to (2.6)–(2.7) and (2.11)–(2.12), there exist positive constants  $\lambda := \frac{\alpha_0^{\min}}{\alpha_{\max}}$  and  $\Lambda := \frac{\alpha_0^{\max}}{\alpha_{\min}}$  such that

$$(2.16) \quad \lambda B(v, v) \leq B_0(v, v) \leq \Lambda B(v, v) \quad \forall v \in V.$$

REMARK 2.1. *We will use all the assumptions made above throughout this paper. We note, however, that the main theoretical results, i.e., Theorems 4.1 and 5.1, will remain valid if, instead of (2.10),  $A(\mathbf{y})$  has the more general decomposition*

$$(2.17) \quad A(\mathbf{y}) = A_0 + R(\mathbf{y}) \quad \forall \mathbf{y} \in \Gamma,$$

where  $A_0$  satisfies the assumptions stated above,  $R(\mathbf{y}) \in \mathcal{L}(H, H')$  is symmetric for all  $\mathbf{y} \in \Gamma$ , and there exists a constant  $\tau \in [0, 1)$  such that for all  $\mathbf{y} \in \Gamma$  there holds  $|\langle R(\mathbf{y})v, v \rangle| \leq \tau \langle A_0 v, v \rangle$  for any  $v \in H$ .

**2.1. PDEs with random data.** An important class of problems that leads to abstract parametric operator equations of the form (2.1) is PDEs with random data. In this case, we have an operator equation of the form

$$(2.18) \quad A(\boldsymbol{\xi})u(\boldsymbol{\xi}) = f(\boldsymbol{\xi}),$$

where  $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots)$  is a vector of bounded real-valued random variables  $\xi_m : \Omega \rightarrow \Gamma_m \subset \mathbb{R}$ . Here, the data (and hence the solution) are functions of  $\boldsymbol{\xi}$ , a multivariate random variable taking values (observations) in a set  $\Gamma$  of the form (2.8). When the data are expressed in terms of random variables, there is an underlying probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , where  $\Omega$  is an abstract sample space,  $\mathcal{F}$  is a  $\sigma$ -algebra, and  $\mathbb{P}$  is a probability measure. The multivariate random variable  $\boldsymbol{\xi}$  generates a distribution  $\pi$  satisfying  $\pi(G) = \mathbb{P}(\boldsymbol{\xi}^{-1}(G))$  for  $G \in \mathcal{B}(\Gamma)$ , which is also a probability measure by definition. Thus, the observations  $\mathbf{y}$  of  $\boldsymbol{\xi}$  belong to  $\Gamma$  and  $(\Gamma, \mathcal{B}(\Gamma), \pi)$  is a probability space. By working on the observation set associated with  $\boldsymbol{\xi}$  we can rewrite (2.18) in the parametric form (2.1), and we can consider the weak formulation (2.2) associated with the probability measure  $\pi$ . Crucially, if  $\xi_1, \xi_2, \dots$  are independent, then  $\pi$  has the form (2.9), where  $\pi_m$  is the probability measure associated with  $\xi_m$  ( $m = 1, 2, \dots$ ).

We now illustrate the abstract problems (2.1) and (2.2) with a concrete example. Let  $D$  be a Lipschitz domain in  $\mathbb{R}^2$  and consider the homogeneous Dirichlet problem for the steady-state diffusion equation with a random, spatially varying diffusion coefficient. More precisely, it is assumed that the diffusion coefficient  $a = a(\mathbf{x}, \boldsymbol{\xi})$  is a second-order correlated random field that can be written as a function of a multivariate random variable  $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots)$  and that the right-hand side function  $f = f(\mathbf{x})$  is deterministic. It is known (see, e.g., [11], [13], [2], [21]) that we may rewrite this problem in the parametric form

$$(2.19) \quad \begin{aligned} -\nabla \cdot (a(\mathbf{x}, \mathbf{y})\nabla u(\mathbf{x}, \mathbf{y})) &= f(\mathbf{x}), & \mathbf{x} \in D, \mathbf{y} \in \Gamma, \\ u(\mathbf{x}, \mathbf{y}) &= 0, & \mathbf{x} \in \partial D, \mathbf{y} \in \Gamma, \end{aligned}$$

where  $\Gamma := \prod_{m=1}^{\infty} [-1, 1]$ , with the diffusion coefficient represented as

$$(2.20) \quad a(\mathbf{x}, \mathbf{y}) = a_0(\mathbf{x}) + \sum_{m=1}^{\infty} a_m(\mathbf{x}) y_m, \quad \mathbf{x} \in D, \mathbf{y} \in \Gamma,$$

and with the series converging uniformly in  $L^\infty(D)$ . We note that even if the underlying bounded random variables  $\xi_m$  do not take values in  $\Gamma_m = [-1, 1]$ , we can always rescale the coefficient functions  $a_m$  so that we may work with the stated  $\Gamma$ .

The parameter-free term  $a_0(\mathbf{x})$  in (2.20) may represent the mean: that is,

$$a_0(\mathbf{x}) = \int_{\Gamma} a(\mathbf{x}, \mathbf{y}) \, d\pi(\mathbf{y}) = \mathbb{E}[a](\mathbf{x}), \quad \mathbf{x} \in D.$$

This is certainly true for Karhunen–Loève expansions since in that case each  $y_m$  in (2.20) is the image of a mean-zero random variable and so

$$(2.21) \quad \int_{\Gamma_m} y_m \, d\pi_m(y_m) = 0.$$

Alternatively, if (2.20) represents another spectral expansion or just a synthetic parametric expansion, then  $a_0(\mathbf{x})$  is simply a specified function of  $\mathbf{x} \in D$ . The important point for our energy error estimation strategy is that  $a_0(\mathbf{x})$  does not depend on the parameter  $\mathbf{y}$ .

We now write (2.19) as an operator equation of the form (2.1). Let  $H := H_0^1(D)$ , let  $f(\mathbf{y}) := f \in H^{-1}(D)$  for all  $\mathbf{y} \in \Gamma$ , and define the operator  $A(\mathbf{y}) \in \mathcal{L}(H_0^1(D), H^{-1}(D))$  for all  $\mathbf{y} \in \Gamma$  by the identity

$$(2.22) \quad \langle A(\mathbf{y})v, w \rangle := \int_D a(\mathbf{x}, \mathbf{y}) \nabla v(\mathbf{x}) \cdot \nabla w(\mathbf{x}) \, d\mathbf{x} \quad \forall v, w \in H_0^1(D).$$

Thanks to (2.20), the operator  $A(\mathbf{y})$  admits decomposition (2.10) with  $A_0$  and  $A_m$ ,  $m \in \mathbb{N}$ , defined by

$$(2.23) \quad \langle A_0 v, w \rangle := \int_D a_0(\mathbf{x}) \nabla v(\mathbf{x}) \cdot \nabla w(\mathbf{x}) \, d\mathbf{x} \quad \forall v, w \in H_0^1(D),$$

$$(2.24) \quad \langle A_m v, w \rangle := \int_D a_m(\mathbf{x}) \nabla v(\mathbf{x}) \cdot \nabla w(\mathbf{x}) \, d\mathbf{x} \quad \forall v, w \in H_0^1(D).$$

In order to ensure (2.1) and (2.2) are well posed for problem (2.19), one has to make assumptions on the functions  $a_0(\mathbf{x})$  and  $a_m(\mathbf{x})$ ,  $m \in \mathbb{N}$ , in the representation of the diffusion coefficient. In (2.20), we assume that  $a_0(\mathbf{x}) \in L^\infty(D)$  is uniformly bounded away from zero, i.e.,

$$(2.25) \quad \exists \alpha_0^{\min}, \alpha_0^{\max} > 0 \text{ such that } \alpha_0^{\min} \leq a_0(\mathbf{x}) \leq \alpha_0^{\max} \text{ a.e. in } D.$$

Then the operator  $A_0$  defined by (2.23) is symmetric and positive definite, and inequalities (2.11)–(2.12) hold. We further assume that  $a_m(\mathbf{x}) \in L^\infty(D)$ ,  $m \in \mathbb{N}$ , and

$$(2.26) \quad \tau := \frac{1}{\alpha_0^{\min}} \sum_{m=1}^{\infty} \|a_m\|_{L^\infty(D)} < 1.$$

If (2.20) is a Karhunen–Loève expansion, then this is equivalent to assuming that the eigenvalues of the integral operator associated with the covariance of  $a(\mathbf{x}, \boldsymbol{\xi})$  decay sufficiently quickly. On one hand, (2.26) ensures convergence of the series in (2.10) uniformly in  $\mathbf{y}$ ; see [21, Lemma 2.21]. On the other hand, (2.26) together with (2.25)

imply bounded invertibility of  $A(\mathbf{y})$  for all  $\mathbf{y} \in \Gamma$  (and, hence, unique solvability of (2.1)) and inequalities (2.5) hold with

$$\alpha_{\max} := \alpha_0^{\max}(1 + \tau) \quad \text{and} \quad \alpha_{\min} := \alpha_0^{\min}(1 - \tau);$$

see [21, Proposition 2.22]. Note that this implies that

$$\alpha_{\min} < \alpha_0^{\min} \leq \alpha_0^{\max} < \alpha_{\max},$$

and hence the constants  $\lambda$  and  $\Lambda$  in (2.16) satisfy  $\lambda < 1 < \Lambda$ . This will be useful later.

**3. Discrete formulations.** We now return to the abstract parametric operator equation (2.1). The weak problem (2.2) can be discretized by choosing a finite-dimensional subspace  $V_N \subset V$  and using Galerkin projection onto  $V_N$ . This defines a unique element  $u_N \in V_N$  satisfying

$$(3.1) \quad B(u_N, v) = F(v) \quad \forall v \in V_N.$$

Our goal is twofold: first, to find reliable and efficient estimators for the approximation error  $u - u_N$  (measured in an appropriate norm), and second, to suggest an algorithm for adaptive selection of a sequence of finite-dimensional subspaces  $V_N \subset V$ .

To make progress we will need to specify the structure of  $V_N$  more precisely. We observe that due to the separability of  $H$ , the Bochner space  $V = L^2_\pi(\Gamma, H)$  is isometrically isomorphic to the tensor product Hilbert space  $L^2_\pi(\Gamma) \otimes H$  (see, e.g., [21, Theorem B.17, Remark C.24]). Hence,  $V_N$  can be defined by mimicking this tensor product structure. Bearing this in mind, let us first introduce finite-dimensional (polynomial) subspaces of  $L^2_\pi(\Gamma)$ . For any  $m \in \mathbb{N}$ , let  $\{P_n^m\}_{n \in \mathbb{N}_0}$  denote the set of univariate polynomials on  $\Gamma_m$  that are orthogonal with respect to the measure  $\pi_m$ . Note that for any polynomial  $P_n^m$ , the index  $n$  refers to the polynomial degree and we denote by  $c_n^m$  the leading coefficient of  $P_n^m$ . It is well known that these polynomials satisfy the following three-term recurrence (e.g., see [10], [22]):

$$(3.2) \quad P_0^m \equiv 1; \quad \beta_n^m P_{n+1}^m(t) = (t - \alpha_n^m)P_n^m(t) - \beta_{n-1}^m P_{n-1}^m(t), \quad n \in \mathbb{N}, \quad t \in \Gamma_m,$$

where

$$\alpha_n^m = \int_{\Gamma_m} t (P_n^m(t))^2 d\pi_m(t) \text{ for } n \in \mathbb{N} \quad \text{and} \quad \beta_n^m = \frac{c_n^m}{c_{n+1}^m} \text{ for } n \in \mathbb{N}_0.$$

Moreover,  $\{P_n^m\}_{n \in \mathbb{N}_0}$  is an orthonormal basis of  $L^2_{\pi_m}(\Gamma_m)$ .

We can now construct an orthonormal basis of multivariate polynomials for  $L^2_\pi(\Gamma)$ . To this end, we introduce the following set of finitely supported sequences:

$$\mathfrak{J} := \{ \nu = (\nu_1, \nu_2, \dots) \in \mathbb{N}_0^{\mathbb{N}}; \# \text{supp } \nu < \infty \},$$

where  $\text{supp } \nu := \{m \in \mathbb{N}; \nu_m \neq 0\}$  for any  $\nu \in \mathbb{N}_0^{\mathbb{N}}$ . We will call  $\mathfrak{J}$  and any of its subsets the *index sets*. Then the set of tensor product polynomials defined by

$$(3.3) \quad P_\nu(\mathbf{y}) = \prod_{m=1}^{\infty} P_{\nu_m}^m(y_m) \quad \forall \nu \in \mathfrak{J}$$

forms an orthonormal basis of  $L^2_\pi(\Gamma)$  (e.g., see [21, Theorem 2.12]). For any finite index set  $\mathfrak{P} \subset \mathfrak{J}$ ,

$$\mathcal{P}_{\mathfrak{P}} := \text{span}\{P_\nu; \nu \in \mathfrak{P}\}$$



is a finite-dimensional subspace of  $L^2_\pi(\Gamma)$ . Note that each polynomial  $P_\nu$  is a function of a finite number of variables (parameters)  $y_m$ ,  $m \in \mathbb{N}$ . Indeed, (3.3) can also be written as

$$P_\nu(\mathbf{y}) = \prod_{m \in \text{supp } \nu} P_{\nu_m}^m(y_m) \quad \forall \nu \in \mathcal{J},$$

since  $P_0^m(y_m) \equiv 1$  for any  $m \in \mathbb{N}$ . We can now define the finite-dimensional subspace  $V_N$  as

$$V_N = V_{X\mathfrak{P}} := X \otimes \mathcal{P}_{\mathfrak{P}},$$

where  $X$  is a finite-dimensional subspace of  $H$  and  $\mathfrak{P}$  is a finite subset of the index set  $\mathcal{J}$ . We note that the choice of the index set  $\mathfrak{P}$  for  $\mathcal{P}_{\mathfrak{P}}$  determines both the (finite) number of “active” parameters  $y_m$  in the polynomial approximation on  $\Gamma$  and the polynomial degrees in these “active” parameters. For the subsequent development it will be convenient to rewrite (3.1) as follows: find  $u_{X\mathfrak{P}} \in V_{X\mathfrak{P}}$  satisfying

$$(3.4) \quad B(u_{X\mathfrak{P}}, v) = F(v) \quad \forall v \in V_{X\mathfrak{P}}.$$

The approximation provided by (3.4) can be improved by enriching the subspace  $V_{X\mathfrak{P}}$ . This can be done by enriching the finite-dimensional subspace  $X \subset H$  and/or the polynomial space  $\mathcal{P}_{\mathfrak{P}} \subset L^2_\pi(\Gamma)$ .

Let  $X^*$  be a finite-dimensional subspace of  $H$  such that  $X^* \supset X$ . For example, in finite element methods,  $X^*$  could be obtained from  $X$  by adding new basis functions corresponding to nodes introduced by mesh refinement. Then  $X^*$  can be decomposed as

$$(3.5) \quad X^* = X \oplus Y,$$

where  $Y \subset H$  and  $X \cap Y = \{0\}$ . The subspace  $Y$  will be called the *detail space*. We observe that since  $\langle A_0 \cdot, \cdot \rangle$  defines an inner product in  $H$  and  $X \cap Y = \{0\}$ , there exists a constant  $\gamma \in [0, 1)$  such that the strengthened Cauchy–Schwarz inequality holds (see, for example, [9]). That is,

$$(3.6) \quad |\langle A_0 u_X, v_Y \rangle| \leq \gamma \langle A_0 u_X, u_X \rangle^{1/2} \langle A_0 v_Y, v_Y \rangle^{1/2} \quad \forall u_X \in X, \forall v_Y \in Y.$$

On the parameter domain  $\Gamma$ , we introduce an enriched polynomial space  $\mathcal{P}_{\mathfrak{P}^*}$  corresponding to an index set  $\mathfrak{P}^* \supset \mathfrak{P}$ . For example,  $\mathfrak{P}^*$  may be chosen such that more parameters  $y_m$  become “active” than in  $\mathfrak{P}$ , or, alternatively,  $\mathfrak{P}^*$  may be chosen such that it has the same “active” parameters as  $\mathfrak{P}$  but includes higher order polynomials. Moreover, the degrees can be increased uniformly for all parameters or selectively for only certain parameters. Thus,  $\mathfrak{P}^* = \mathfrak{P} \cup \mathfrak{Q}$  with  $\mathfrak{Q} \subset \mathcal{J}$  such that  $\mathfrak{P} \cap \mathfrak{Q} = \emptyset$ . We will call  $\mathfrak{Q}$  the *detail index set*. Then  $\mathcal{P}_{\mathfrak{P}^*}$  can be decomposed as

$$(3.7) \quad \mathcal{P}_{\mathfrak{P}^*} = \mathcal{P}_{\mathfrak{P}} \oplus \mathcal{P}_{\mathfrak{Q}}, \quad \mathcal{P}_{\mathfrak{P}} \cap \mathcal{P}_{\mathfrak{Q}} = \{0\}.$$

The decomposition in (3.7) is orthogonal with respect to the measure  $\pi$ . That is,

$$(3.8) \quad \int_\Gamma P_\nu(\mathbf{y}) P_\mu(\mathbf{y}) \, d\pi(\mathbf{y}) = 0 \quad \forall \nu \in \mathfrak{P}, \forall \mu \in \mathfrak{Q}.$$

Next, let us define the *enriched* finite-dimensional subspace of  $V$  as the space

$$(3.9) \quad V_{X^*\mathfrak{P}^*} := X^* \otimes \mathcal{P}_{\mathfrak{P}^*}.$$

Then, using (3.5) and (3.7), we have

$$(3.10) \quad V_{X\mathfrak{P}}^* := V_{X\mathfrak{P}} \oplus V_{X\Omega} \oplus V_{Y\mathfrak{P}} \oplus V_{Y\Omega},$$

where

$$(3.11) \quad V_{X\Omega} := X \otimes \mathcal{P}_\Omega, \quad V_{Y\mathfrak{P}} := Y \otimes \mathcal{P}_\mathfrak{P}, \quad V_{Y\Omega} := Y \otimes \mathcal{P}_\Omega.$$

We now use properties (3.6) and (3.8) to prove the following auxiliary result.

LEMMA 3.1. *For any finite index set  $\mathfrak{P} \subset \mathfrak{I}$ , the finite-dimensional subspaces  $V_{X\mathfrak{P}}, V_{Y\mathfrak{P}} \subset V$  are such that the following strengthened Cauchy–Schwarz inequality holds:*

$$(3.12) \quad |B_0(u, v)| \leq \gamma \|u\|_{B_0} \|v\|_{B_0} \quad \forall u \in V_{X\mathfrak{P}}, \forall v \in V_{Y\mathfrak{P}}.$$

Here,  $\gamma \in [0, 1)$  is the same constant as in the strengthened Cauchy–Schwarz inequality (3.6) for the spaces  $X, Y \subset H$ .

*Proof.* Let  $\mathfrak{P}$  be an arbitrary finite subset of the index set  $\mathfrak{I}$ . For any  $u \in V_{X\mathfrak{P}} = X \otimes \mathcal{P}_\mathfrak{P}$  we have that

$$u(\mathbf{y}) = \sum_{\nu \in \mathfrak{P}} \phi_\nu P_\nu(\mathbf{y}) \quad \text{with } \phi_\nu \in X.$$

Then, recalling the definition of  $B_0(\cdot, \cdot)$  in (2.15) and the orthonormality of the polynomials in  $\mathcal{P}_\mathfrak{P}$ , we find that

$$(3.13) \quad \|u\|_{B_0}^2 = \int_\Gamma \left\langle A_0 \sum_{\nu \in \mathfrak{P}} \phi_\nu P_\nu(\mathbf{y}), \sum_{\nu \in \mathfrak{P}} \phi_\nu P_\nu(\mathbf{y}) \right\rangle d\pi(\mathbf{y}) = \sum_{\nu \in \mathfrak{P}} \langle A_0 \phi_\nu, \phi_\nu \rangle.$$

Similarly, for any  $v \in V_{Y\mathfrak{P}} = Y \otimes \mathcal{P}_\mathfrak{P}$ ,

$$(3.14) \quad v = \sum_{\mu \in \mathfrak{P}} \psi_\mu P_\mu(\mathbf{y}) \quad \text{with } \psi_\mu \in Y, \quad \|v\|_{B_0}^2 = \sum_{\mu \in \mathfrak{P}} \langle A_0 \psi_\mu, \psi_\mu \rangle.$$

Furthermore,

$$\begin{aligned} B_0(u, v) &= \int_\Gamma \left\langle A_0 \sum_{\nu \in \mathfrak{P}} \phi_\nu P_\nu(\mathbf{y}), \sum_{\mu \in \mathfrak{P}} \psi_\mu P_\mu(\mathbf{y}) \right\rangle d\pi(\mathbf{y}) \\ &= \sum_{\nu \in \mathfrak{P}} \sum_{\mu \in \mathfrak{P}} \langle A_0 \phi_\nu, \psi_\mu \rangle \int_\Gamma P_\nu(\mathbf{y}) P_\mu(\mathbf{y}) d\pi(\mathbf{y}) = \sum_{\nu \in \mathfrak{P}} \langle A_0 \phi_\nu, \psi_\nu \rangle. \end{aligned}$$

Hence, applying the strengthened Cauchy–Schwarz inequality (3.6) and the standard Cauchy–Schwarz inequality for sums, we obtain

$$\begin{aligned} |B_0(u, v)| &\leq \sum_{\nu \in \mathfrak{P}} \gamma \langle A_0 \phi_\nu, \phi_\nu \rangle^{1/2} \langle A_0 \psi_\nu, \psi_\nu \rangle^{1/2} \\ &\leq \gamma \left( \sum_{\nu \in \mathfrak{P}} \langle A_0 \phi_\nu, \phi_\nu \rangle \right)^{1/2} \left( \sum_{\nu \in \mathfrak{P}} \langle A_0 \psi_\nu, \psi_\nu \rangle \right)^{1/2}. \end{aligned}$$

The result then follows thanks to (3.13) and (3.14).  $\square$

Now, let  $u_{X\mathfrak{P}}^* \in V_{X\mathfrak{P}}^*$  be the Galerkin projection onto the enriched subspace  $V_{X\mathfrak{P}}^*$ , so that

$$(3.15) \quad B(u_{X\mathfrak{P}}^*, v) = F(v) \quad \forall v \in V_{X\mathfrak{P}}^*.$$

The approximation  $u_{X\mathfrak{P}}^* \in V_{X\mathfrak{P}}^*$  generated by (3.15) is not worse than the approximation  $u_{X\mathfrak{P}} \in V_{X\mathfrak{P}}$ , in the following sense:

$$(3.16) \quad \|u - u_{X\mathfrak{P}}^*\|_B = \inf_{v \in V_{X\mathfrak{P}}^*} \|u - v\|_B \leq \|u - u_{X\mathfrak{P}}\|_B.$$

We will assume, as is commonly done in nonparametric a posteriori error analysis, that a stronger property holds. The following assumption is equivalent to the standard saturation assumption employed in adaptive finite element analysis, but here the approximation spaces are tensor product spaces of finite element functions on the computational domain  $D$  and polynomials on the parameter domain  $\Gamma$ .

ASSUMPTION 3.1 (saturation assumption). *Let  $u \in V$  solve (2.2), and let  $u_{X\mathfrak{P}} \in V_{X\mathfrak{P}}$  and  $u_{X\mathfrak{P}}^* \in V_{X\mathfrak{P}}^* \supset V_{X\mathfrak{P}}$  be two Galerkin approximations satisfying (3.4) and (3.15), respectively. We assume that there exists a constant  $\beta \in [0, 1)$  such that*

$$(3.17) \quad \|u - u_{X\mathfrak{P}}^*\|_B \leq \beta \|u - u_{X\mathfrak{P}}\|_B.$$

REMARK 3.1. *Thanks to (3.16), inequality (3.17) always holds with some  $\beta \leq 1$ . If  $V_{X\mathfrak{P}}^* = X^* \otimes \mathcal{P}_{\mathfrak{P}^*}$ , then following a similar argument gives*

$$(3.18) \quad \begin{aligned} \|u - u_{X\mathfrak{P}}^*\|_B &= \inf_{v \in V_{X\mathfrak{P}}^*} \|u - v\|_B \leq \beta_1 \inf_{v \in X^* \otimes \mathcal{P}_{\mathfrak{P}}} \|u - v\|_B \\ &\leq \beta_1 \beta_2 \inf_{v \in X \otimes \mathcal{P}_{\mathfrak{P}}} \|u - v\|_B = \beta_1 \beta_2 \|u - u_{X\mathfrak{P}}\|_B \end{aligned}$$

for some  $\beta_1, \beta_2 \leq 1$ . It is an open problem, however, to find assumptions on the data in (2.1) and to identify detail spaces  $Y$  and detail index sets  $\mathfrak{Q}$  that yield (3.17) with  $\beta < 1$  (or, equivalently, (3.18) with  $\beta_1 \beta_2 < 1$ ).

It is worth recalling here that if the operator  $A$  has affine dependence on the parameters  $y_m$  (see (2.10)), then the bilinear form  $B(\cdot, \cdot)$  is given by the left-hand side in (2.14). Hence, for the Galerkin approximation  $u_{X\mathfrak{P}}$  (and the enhanced approximation  $u_{X\mathfrak{P}}^*$ ) to be computable, the number of terms that are nonzero in the expansion of  $B(u_{X\mathfrak{P}}, v)$  in (3.4) (resp.,  $B(u_{X\mathfrak{P}}^*, v)$  in (3.15)) must be finite. To show that this is indeed the case when  $\mathfrak{P}$  (resp.,  $\mathfrak{P}^*$ ) is a finite set, we write

$$u_{X\mathfrak{P}}(\mathbf{y}) = \sum_{\nu \in \mathfrak{P}} \phi_\nu P_\nu(\mathbf{y}), \quad v(\mathbf{y}) = \sum_{\mu \in \mathfrak{P}} \psi_\mu P_\mu(\mathbf{y})$$

with coefficients  $\phi_\nu, \psi_\mu \in X$ . With this choice, the left-hand side of (3.4) is given by

$$B(u_{X\mathfrak{P}}, v) = \sum_{\nu \in \mathfrak{P}} \langle A_0 \phi_\nu, \psi_\nu \rangle + \sum_{m=1}^{\infty} \sum_{\nu \in \mathfrak{P}} \sum_{\mu \in \mathfrak{P}} \langle A_m \phi_\nu, \psi_\mu \rangle \int_{\Gamma} P_\nu(\mathbf{y}) P_\mu(\mathbf{y}) y_m \, d\pi(\mathbf{y}),$$

and expanding the integral over  $\Gamma$  using (2.8) and (2.9) gives the product

$$\int_{\Gamma_m} P_{\nu_m}^m(y_m) P_{\mu_m}^m(y_m) y_m \, d\pi_m(y_m) \left( \prod_{s \in \mathbb{N} \setminus \{m\}} \int_{\Gamma_s} P_{\nu_s}^s(y_s) P_{\mu_s}^s(y_s) \, d\pi_s(y_s) \right).$$

If a finite index set  $\mathfrak{P}$  is selected with, say,  $M$  active parameters, then, for all  $m > M$ ,  $\nu_m = \mu_m = 0$  and  $P_{\nu_m}^m(y_m) = 1 = P_{\mu_m}^m(y_m)$ . Thus, if the active parameters  $y_m$  have mean zero, (2.21) holds, in which case

$$\int_{\Gamma} P_{\nu}(\mathbf{y}) P_{\mu}(\mathbf{y}) y_m \, d\pi(\mathbf{y}) = 0 \quad \forall m > M, \forall \nu, \mu \in \mathfrak{P},$$

so the sum is effectively truncated at  $m = M$ . A similar argument applies to (3.15).

**4. A posteriori error estimation.** Our goal in this section is to develop an a posteriori estimator for the discretization error  $e := u - u_{X\mathfrak{P}} \in V$ . From (2.2),

$$(4.1) \quad B(e, v) = F(v) - B(u_{X\mathfrak{P}}, v) \quad \forall v \in V,$$

and Galerkin orthogonality immediately follows from (3.4); that is, the error  $e$  satisfies

$$(4.2) \quad B(e, v) = 0 \quad \forall v \in V_{X\mathfrak{P}}.$$

We can approximate the error  $e \in V$  by discretizing (4.1). A Galerkin projection onto the enriched subspace  $V_{X\mathfrak{P}}^*$  (see (3.9)) leads to our first estimator  $e^* \in V_{X\mathfrak{P}}^*$  satisfying

$$(4.3) \quad B(e^*, v) = F(v) - B(u_{X\mathfrak{P}}, v) \quad \forall v \in V_{X\mathfrak{P}}^*.$$

That is,  $e^*$  is sought in a space  $V_{X\mathfrak{P}}^*$  that is richer than the approximation space used for the Galerkin approximation  $u_{X\mathfrak{P}}$ . The following well-known result establishes the relation between the true error  $e$  in (4.1) and  $e^*$  satisfying (4.3).

**PROPOSITION 4.1** (see [1, Theorem 5.1]). *Suppose that Assumption 3.1 holds for the solution  $u$  to (2.2). Then the estimator  $e^*$  defined by (4.3) satisfies*

$$(4.4) \quad \|e^*\|_B \leq \|e\|_B \leq \frac{1}{\sqrt{1 - \beta^2}} \|e^*\|_B,$$

where  $\beta \in [0, 1)$  is the constant in (3.17).

There are several possibilities for  $V_{X\mathfrak{P}}^*$  in (4.3). Notice, however, that the constant  $\beta$  which appears in (4.4) depends on the specific choice of  $V_{X\mathfrak{P}}^*$ . For highly enriched spaces, we expect  $\beta$  to be small and the bound to be tight. However, the computational cost associated with solving (4.3) is comparable to the cost of finding the enhanced approximation  $u_{X\mathfrak{P}}^*$ , which makes the estimator  $e^*$  impractical. One way to avoid this is to consider an alternative bilinear form on the left-hand side of (4.3), which leads to a linear system that is cheaper to solve. In addition, it also makes sense to try and exploit the decomposition of the enriched subspace  $V_{X\mathfrak{P}}^*$  and perform computations on lower-dimensional spaces.

Using the bilinear form  $B_0(\cdot, \cdot)$  given by (2.15) instead of  $B(\cdot, \cdot)$  on the left-hand side of (4.3) results in our second error estimator  $e_0^* \in V_{X\mathfrak{P}}^*$  satisfying

$$(4.5) \quad B_0(e_0^*, v) = F(v) - B(u_{X\mathfrak{P}}, v) \quad \forall v \in V_{X\mathfrak{P}}^*.$$

The  $B_0$  inner product is particularly convenient from the point of view of linear algebra. Indeed, since it incorporates only the parameter-free part of the operator  $A(\mathbf{y})$ , it invariably leads to a block diagonal system matrix. Calculations can then be decomposed into multiple problems each having the dimension of the subspace  $X^*$ . The following result establishes the connection between the estimators  $e_0^*$  and  $e^*$ .

**PROPOSITION 4.2.** *Let  $e^* \in V_{X\mathfrak{P}}^*$  and  $e_0^* \in V_{X\mathfrak{P}}^*$  be the error estimators defined by (4.3) and (4.5), respectively. Then there holds*

$$(4.6) \quad \sqrt{\lambda} \|e_0^*\|_{B_0} \leq \|e^*\|_B \leq \sqrt{\Lambda} \|e_0^*\|_{B_0},$$

where  $\lambda$  and  $\Lambda$  are the constants in (2.16).

*Proof.* The proof is standard (see [1, Theorem 5.3]), and we include it for completeness. Equations (4.3) and (4.5) imply

$$(4.7) \quad B(e^*, v) = B_0(e_0^*, v) \quad \forall v \in V_{X\mathfrak{P}}^*.$$

Selecting  $v = e^*$  in (4.7), applying the Cauchy–Schwarz inequality, and using the right-hand inequality in (2.16), we obtain

$$\|e^*\|_B^2 = B(e^*, e^*) = B_0(e_0^*, e^*) \leq \|e_0^*\|_{B_0} \|e^*\|_{B_0} \leq \sqrt{\Lambda} \|e_0^*\|_{B_0} \|e^*\|_B.$$

This yields the right-hand inequality in (4.6).

Similarly, selecting  $v = e_0^*$  in (4.7), we get

$$\|e_0^*\|_{B_0}^2 = B_0(e_0^*, e_0^*) = B(e^*, e_0^*) \leq \|e^*\|_B \|e_0^*\|_B \leq \lambda^{-1/2} \|e^*\|_B \|e_0^*\|_{B_0},$$

and the left-hand inequality in (4.6) follows.  $\square$

The error estimators  $e^*$  and  $e_0^*$  provide error estimates  $\|e^*\|_B$  and  $\|e_0^*\|_{B_0}$ , respectively, for  $\|e\|_B$ . For the enriched space  $V_{X\mathfrak{P}}^*$ , we can make use of the decomposition (3.10) to obtain a third estimate. This is given by

$$(4.8) \quad \eta := \left( 2 \|\bar{e}_{X\Omega}\|_{B_0}^2 + \|\bar{e}_{Y\mathfrak{P}}\|_{B_0}^2 + 2 \|\bar{e}_{Y\Omega}\|_{B_0}^2 \right)^{1/2},$$

where the contributing estimators  $\bar{e}_{X\Omega} \in V_{X\Omega}$ ,  $\bar{e}_{Y\mathfrak{P}} \in V_{Y\mathfrak{P}}$ , and  $\bar{e}_{Y\Omega} \in V_{Y\Omega}$  satisfy

$$(4.9) \quad B_0(\bar{e}_{X\Omega}, v) = F(v) - B(u_{X\mathfrak{P}}, v) \quad \forall v \in V_{X\Omega},$$

$$(4.10) \quad B_0(\bar{e}_{Y\mathfrak{P}}, v) = F(v) - B(u_{X\mathfrak{P}}, v) \quad \forall v \in V_{Y\mathfrak{P}},$$

$$(4.11) \quad B_0(\bar{e}_{Y\Omega}, v) = F(v) - B(u_{X\mathfrak{P}}, v) \quad \forall v \in V_{Y\Omega},$$

respectively. Note that  $\mathfrak{P}$  and  $\Omega$  are finite index sets. Hence, if  $A$  has affine dependence on the parameters  $y_m$  and (2.21) holds, then  $\eta$  is computable.

The connection between  $\eta$  and  $\|e_0^*\|_{B_0}$  is established in the next lemma.

LEMMA 4.1. *Let  $e_0^* \in V_{X\mathfrak{P}}^*$  be defined by (4.5). Then the error estimate  $\eta$  defined by (4.8)–(4.11) satisfies*

$$(4.12) \quad \frac{1}{\sqrt{5}} \eta \leq \|e_0^*\|_{B_0} \leq \frac{1}{\sqrt{1-\gamma^2}} \eta,$$

where  $\gamma \in [0, 1)$  is the constant in the strengthened Cauchy–Schwarz inequality (3.6) for the subspaces  $X, Y \subset H$ .

*Proof.* Since  $V_{X\Omega}$ ,  $V_{Y\mathfrak{P}}$ , and  $V_{Y\Omega}$  are subspaces of  $V_{X\mathfrak{P}}^*$ , we deduce from (4.5) and (4.9)–(4.11) that

$$(4.13) \quad B_0(\bar{e}_{X\Omega}, v_{X\Omega}) = B_0(e_0^*, v_{X\Omega}) \quad \forall v_{X\Omega} \in V_{X\Omega},$$

$$(4.14) \quad B_0(\bar{e}_{Y\mathfrak{P}}, v_{Y\mathfrak{P}}) = B_0(e_0^*, v_{Y\mathfrak{P}}) \quad \forall v_{Y\mathfrak{P}} \in V_{Y\mathfrak{P}},$$

$$(4.15) \quad B_0(\bar{e}_{Y\Omega}, v_{Y\Omega}) = B_0(e_0^*, v_{Y\Omega}) \quad \forall v_{Y\Omega} \in V_{Y\Omega}.$$

Selecting  $v_{X\Omega} = \bar{e}_{X\Omega}$  in (4.13) and applying the Cauchy–Schwarz inequality, one has

$$\|\bar{e}_{X\Omega}\|_{B_0}^2 = B_0(\bar{e}_{X\Omega}, \bar{e}_{X\Omega}) = B_0(e_0^*, \bar{e}_{X\Omega}) \leq \|e_0^*\|_{B_0} \|\bar{e}_{X\Omega}\|_{B_0}.$$

This implies  $\|\bar{e}_{X\Omega}\|_{B_0} \leq \|e_0^*\|_{B_0}$ . Similarly,  $\|\bar{e}_{Y\mathfrak{P}}\|_{B_0} \leq \|e_0^*\|_{B_0}$ ,  $\|\bar{e}_{Y\Omega}\|_{B_0} \leq \|e_0^*\|_{B_0}$ , and the left-hand inequality in (4.12) follows from the definition of  $\eta$  in (4.8).

Let us now prove the right-hand inequality in (4.12). To that end, we represent  $e_0^* \in V_{X\mathfrak{F}}^*$  as

$$(4.16) \quad e_0^* = w_{X\mathfrak{F}} + w_{X\Omega} + w_{Y\mathfrak{F}} + w_{Y\Omega},$$

where  $w_{X\mathfrak{F}} \in V_{X\mathfrak{F}}$ ,  $w_{X\Omega} \in V_{X\Omega}$ ,  $w_{Y\mathfrak{F}} \in V_{Y\mathfrak{F}}$ , and  $w_{Y\Omega} \in V_{Y\Omega}$ . Note that it follows from (3.4) and (4.5) that

$$B_0(e_0^*, w_{X\mathfrak{F}}) = F(w_{X\mathfrak{F}}) - B(u_{X\mathfrak{F}}, w_{X\mathfrak{F}}) = 0.$$

Hence, we can write

$$\begin{aligned} \|e_0^*\|_{B_0}^2 &= B_0(e_0^*, w_{X\mathfrak{F}} + w_{X\Omega} + w_{Y\mathfrak{F}} + w_{Y\Omega}) \\ &= B_0(e_0^*, w_{X\Omega}) + B_0(e_0^*, w_{Y\mathfrak{F}}) + B_0(e_0^*, w_{Y\Omega}). \end{aligned}$$

Then selecting  $v_{X\Omega} = w_{X\Omega}$ ,  $v_{Y\mathfrak{F}} = w_{Y\mathfrak{F}}$ ,  $v_{Y\Omega} = w_{Y\Omega}$  in (4.13), (4.14), (4.15), respectively, and applying the Cauchy–Schwarz inequality gives

$$(4.17) \quad \begin{aligned} \|e_0^*\|_{B_0}^2 &= B_0(\bar{e}_{X\Omega}, w_{X\Omega}) + B_0(\bar{e}_{Y\mathfrak{F}}, w_{Y\mathfrak{F}}) + B_0(\bar{e}_{Y\Omega}, w_{Y\Omega}) \\ &\leq \|\bar{e}_{X\Omega}\|_{B_0} \|w_{X\Omega}\|_{B_0} + \|\bar{e}_{Y\mathfrak{F}}\|_{B_0} \|w_{Y\mathfrak{F}}\|_{B_0} + \|\bar{e}_{Y\Omega}\|_{B_0} \|w_{Y\Omega}\|_{B_0}. \end{aligned}$$

On the other hand, using representation (4.16) we have

$$(4.18) \quad \begin{aligned} \|e_0^*\|_{B_0}^2 &= \|w_{X\mathfrak{F}}\|_{B_0}^2 + \|w_{X\Omega}\|_{B_0}^2 + \|w_{Y\mathfrak{F}}\|_{B_0}^2 + \|w_{Y\Omega}\|_{B_0}^2 \\ &\quad + 2B_0(w_{X\mathfrak{F}}, w_{X\Omega}) + 2B_0(w_{X\mathfrak{F}}, w_{Y\mathfrak{F}}) + 2B_0(w_{X\mathfrak{F}}, w_{Y\Omega}) \\ &\quad + 2B_0(w_{X\Omega}, w_{Y\mathfrak{F}}) + 2B_0(w_{X\Omega}, w_{Y\Omega}) + 2B_0(w_{Y\mathfrak{F}}, w_{Y\Omega}). \end{aligned}$$

Recalling the definition of  $B_0(\cdot, \cdot)$  in (2.15) and the fact that the polynomial spaces  $\mathcal{P}_{\mathfrak{F}}$  and  $\mathcal{P}_{\Omega}$  are orthogonal with respect to the measure  $\pi$  (cf. (3.8)), we conclude

$$(4.19) \quad B_0(w_{X\mathfrak{F}}, w_{X\Omega}) = B_0(w_{X\mathfrak{F}}, w_{Y\Omega}) = B_0(w_{X\Omega}, w_{Y\mathfrak{F}}) = B_0(w_{Y\mathfrak{F}}, w_{Y\Omega}) = 0.$$

Furthermore, applying Lemma 3.1, we have the inequalities

$$(4.20) \quad \begin{aligned} B_0(w_{X\mathfrak{F}}, w_{Y\mathfrak{F}}) &\geq -\gamma \|w_{X\mathfrak{F}}\|_{B_0} \|w_{Y\mathfrak{F}}\|_{B_0}, \\ B_0(w_{X\Omega}, w_{Y\Omega}) &\geq -\gamma \|w_{X\Omega}\|_{B_0} \|w_{Y\Omega}\|_{B_0}. \end{aligned}$$

Using (4.19) and (4.20), we obtain from (4.18)

$$(4.21) \quad \begin{aligned} \|e_0^*\|_{B_0}^2 &\geq \|w_{X\mathfrak{F}}\|_{B_0}^2 + \|w_{X\Omega}\|_{B_0}^2 + \|w_{Y\mathfrak{F}}\|_{B_0}^2 + \|w_{Y\Omega}\|_{B_0}^2 \\ &\quad - 2\gamma \|w_{X\mathfrak{F}}\|_{B_0} \|w_{Y\mathfrak{F}}\|_{B_0} - 2\gamma \|w_{X\Omega}\|_{B_0} \|w_{Y\Omega}\|_{B_0}. \end{aligned}$$

We will now make use of the following inequalities:

$$(4.22) \quad 2\gamma \|w_{X\mathfrak{F}}\|_{B_0} \|w_{Y\mathfrak{F}}\|_{B_0} \leq \|w_{X\mathfrak{F}}\|_{B_0}^2 + \gamma^2 \|w_{Y\mathfrak{F}}\|_{B_0}^2,$$

$$(4.23) \quad 2\gamma \|w_{X\Omega}\|_{B_0} \|w_{Y\Omega}\|_{B_0} \leq \|w_{X\Omega}\|_{B_0}^2 + \gamma^2 \|w_{Y\Omega}\|_{B_0}^2,$$

$$(4.24) \quad 2\gamma \|w_{X\Omega}\|_{B_0} \|w_{Y\Omega}\|_{B_0} \leq \gamma^2 \|w_{X\Omega}\|_{B_0}^2 + \|w_{Y\Omega}\|_{B_0}^2.$$

Using (4.22) and (4.23) to bound the corresponding terms in (4.21) yields

$$(4.25) \quad \|e_0^*\|_{B_0}^2 \geq (1 - \gamma^2) (\|w_{Y\mathfrak{F}}\|_{B_0}^2 + \|w_{Y\Omega}\|_{B_0}^2).$$

On the other hand, applying (4.22) and (4.24) to the same terms gives

$$(4.26) \quad \|e_0^*\|_{B_0}^2 \geq (1 - \gamma^2) (\|w_{Y\mathfrak{P}}\|_{B_0}^2 + \|w_{X\Omega}\|_{B_0}^2).$$

Combining (4.25) and (4.26), we prove

$$\|e_0^*\|_{B_0}^2 \geq (1 - \gamma^2) \left( \frac{1}{2} \|w_{X\Omega}\|_{B_0}^2 + \|w_{Y\mathfrak{P}}\|_{B_0}^2 + \frac{1}{2} \|w_{Y\Omega}\|_{B_0}^2 \right).$$

Finally, thanks to the Cauchy–Schwarz inequality for sums, we obtain from (4.17)

$$\begin{aligned} \|e_0^*\|_{B_0}^2 &\leq \|\bar{e}_{X\Omega}\|_{B_0} \|w_{X\Omega}\|_{B_0} + \|\bar{e}_{Y\mathfrak{P}}\|_{B_0} \|w_{Y\mathfrak{P}}\|_{B_0} + \|\bar{e}_{Y\Omega}\|_{B_0} \|w_{Y\Omega}\|_{B_0} \\ &\leq (1 - \gamma^2)^{-1/2} \left( 2 \|\bar{e}_{X\Omega}\|_{B_0}^2 + \|\bar{e}_{Y\mathfrak{P}}\|_{B_0}^2 + 2 \|\bar{e}_{Y\Omega}\|_{B_0}^2 \right)^{1/2} \|e_0^*\|_{B_0}, \end{aligned}$$

and the right-hand inequality in (4.12) follows.  $\square$

Combining the results of Propositions 4.1 and 4.2 and Lemma 4.1 now gives two-sided bounds for the energy error  $\|e\|_B = \|u - u_{X\mathfrak{P}}\|_B$  in terms of the estimate  $\eta$ .

**THEOREM 4.1.** *Let  $u \in V$  be the solution to problem (2.2), and let  $u_{X\mathfrak{P}} \in V_{X\mathfrak{P}}$  be the Galerkin approximation satisfying (3.4). Suppose that Assumption 3.1 holds. Then the a posteriori error estimate  $\eta$  defined by (4.8)–(4.11) satisfies*

$$(4.27) \quad \frac{\sqrt{\lambda}}{\sqrt{5}} \eta \leq \|u - u_{X\mathfrak{P}}\|_B \leq \frac{\sqrt{\Lambda}}{\sqrt{1 - \beta^2} \sqrt{1 - \gamma^2}} \eta,$$

where  $\lambda, \Lambda$  are the constants in (2.16),  $\gamma \in [0, 1)$  is the constant in the strengthened Cauchy–Schwarz inequality (3.6), and  $\beta \in [0, 1)$  is the constant in (3.17).

**5. Estimates of the error reduction.** The upper bound in (4.27) can be used to control the error in the Galerkin approximation and, in particular, to decide whether or not a prescribed tolerance is met. If the tolerance is exceeded, one needs to decide how to enrich the subspace  $V_{X\mathfrak{P}}$  in order to compute more accurate approximations. For tensor product subspaces (recall that we choose  $V_{X\mathfrak{P}} := X \otimes \mathcal{P}_{\mathfrak{P}}$ ) there are several possibilities. One may, for example, enrich only the subspace  $X \subset H$  while keeping the same polynomial space  $\mathcal{P}_{\mathfrak{P}}$  on the parameter domain  $\Gamma$ . In this case, one computes the enhanced Galerkin approximation  $u_{X^*\mathfrak{P}} \in V_{X^*\mathfrak{P}} := X^* \otimes \mathcal{P}_{\mathfrak{P}}$  satisfying

$$(5.1) \quad B(u_{X^*\mathfrak{P}}, v) = F(v) \quad \forall v \in V_{X^*\mathfrak{P}}.$$

Alternatively, one may keep the same subspace  $X \subset H$  and enrich only the polynomial space on  $\Gamma$  (e.g., by adding new “active” parameters  $y_m$  and/or by using higher order polynomials in currently “active” parameters). In this case, the enhanced Galerkin approximation  $u_{X\mathfrak{P}^*} \in V_{X\mathfrak{P}^*} := X \otimes \mathcal{P}_{\mathfrak{P}^*}$  satisfies

$$(5.2) \quad B(u_{X\mathfrak{P}^*}, v) = F(v) \quad \forall v \in V_{X\mathfrak{P}^*}.$$

Obviously, the third possibility is to enrich both  $X$  and  $\mathcal{P}_{\mathfrak{P}}$  and to compute  $u_{X^*\mathfrak{P}^*}^* \in V_{X^*\mathfrak{P}^*}^* = X^* \otimes \mathcal{P}_{\mathfrak{P}^*}^*$  satisfying (3.15).

Let  $u_{X^*\mathfrak{P}}$  be the Galerkin approximation satisfying (5.1) and let  $e_{X^*\mathfrak{P}} = u - u_{X^*\mathfrak{P}}$  denote the corresponding error. Since the bilinear form  $B(\cdot, \cdot)$  is symmetric, the Galerkin orthogonality property (cf. (4.2)) yields the equality

$$(5.3) \quad \|e\|_B^2 = \|e_{X^*\mathfrak{P}}\|_B^2 + \|u_{X^*\mathfrak{P}} - u_{X\mathfrak{P}}\|_B^2,$$

where  $e = u - u_{X\mathfrak{P}}$ . Indeed, using the Galerkin orthogonality property, we have

$$B(e_{X^*\mathfrak{P}}, u_{X^*\mathfrak{P}} - u_{X\mathfrak{P}}) = 0,$$

and equality (5.3) then follows by the Pythagorean theorem, because

$$e_{X^*\mathfrak{P}} + (u_{X^*\mathfrak{P}} - u_{X\mathfrak{P}}) = u - u_{X\mathfrak{P}} = e.$$

From (5.3) we conclude that the quantity  $\|u_{X^*\mathfrak{P}} - u_{X\mathfrak{P}}\|_B$  characterizes the error reduction that is achieved by enriching only the subspace  $X \subset H$ . In the same way, the error reduction achieved by enriching only the polynomial subspace  $\mathcal{P}_{\mathfrak{P}}$  is characterized by the quantity  $\|u_{X\mathfrak{P}^*} - u_{X\mathfrak{P}}\|_B$ , where  $u_{X\mathfrak{P}^*} \in X \otimes \mathcal{P}_{\mathfrak{P}^*}$  solves (5.2).

We now prove two-sided bounds for the error reductions corresponding to the enhanced Galerkin approximations  $u_{X^*\mathfrak{P}}$  and  $u_{X\mathfrak{P}^*}$  satisfying (5.1) and (5.2).

**THEOREM 5.1.** *Let  $u_{X\mathfrak{P}} \in V_{X\mathfrak{P}}$  be the Galerkin approximation satisfying (3.4), and let  $u_{X^*\mathfrak{P}} \in V_{X^*\mathfrak{P}}$  and  $u_{X\mathfrak{P}^*} \in V_{X\mathfrak{P}^*}$  be the enhanced approximations satisfying (5.1) and (5.2), respectively. Then there hold the following estimates for the error reduction:*

$$(5.4) \quad \sqrt{\lambda} \|\bar{e}_{Y\mathfrak{P}}\|_{B_0} \leq \|u_{X^*\mathfrak{P}} - u_{X\mathfrak{P}}\|_B \leq \frac{\sqrt{\Lambda}}{\sqrt{1-\gamma^2}} \|\bar{e}_{Y\mathfrak{P}}\|_{B_0},$$

$$(5.5) \quad \sqrt{\lambda} \|\bar{e}_{X\Omega}\|_{B_0} \leq \|u_{X\mathfrak{P}^*} - u_{X\mathfrak{P}}\|_B \leq \sqrt{\Lambda} \|\bar{e}_{X\Omega}\|_{B_0}.$$

Here,  $\bar{e}_{Y\mathfrak{P}} \in Y \otimes \mathcal{P}_{\mathfrak{P}}$  and  $\bar{e}_{X\Omega} \in X \otimes \mathcal{P}_{\Omega}$  are defined by (4.10) and (4.9), respectively,  $\lambda$  and  $\Lambda$  are the constants in (2.16), and  $\gamma \in [0, 1)$  is the constant appearing in the strengthened Cauchy–Schwarz inequality (3.6).

*Proof.* Let us prove (5.4). For simplicity of presentation we denote  $\delta_{X^*\mathfrak{P}} := u_{X^*\mathfrak{P}} - u_{X\mathfrak{P}}$ . It follows from (4.10) and (5.1) that

$$(5.6) \quad B_0(\bar{e}_{Y\mathfrak{P}}, v_{Y\mathfrak{P}}) = B(\delta_{X^*\mathfrak{P}}, v_{Y\mathfrak{P}}) \quad \forall v_{Y\mathfrak{P}} \in V_{Y\mathfrak{P}}.$$

Then selecting  $v_{Y\mathfrak{P}} = \bar{e}_{Y\mathfrak{P}}$ , applying the Cauchy–Schwarz inequality, and using the left-hand inequality in (2.16) gives

$$(5.7) \quad \|\bar{e}_{Y\mathfrak{P}}\|_{B_0}^2 \leq \|\delta_{X^*\mathfrak{P}}\|_B \|\bar{e}_{Y\mathfrak{P}}\|_B \leq \lambda^{-1/2} \|\delta_{X^*\mathfrak{P}}\|_B \|\bar{e}_{Y\mathfrak{P}}\|_{B_0}.$$

Recalling our notation for  $\delta_{X^*\mathfrak{P}}$ , the last estimate immediately implies the left-hand inequality in (5.4).

In order to prove the upper bound for  $\|\delta_{X^*\mathfrak{P}}\|_B$  we write this norm as

$$(5.8) \quad \|\delta_{X^*\mathfrak{P}}\|_B = B\left(\delta_{X^*\mathfrak{P}}, \frac{\delta_{X^*\mathfrak{P}}}{\|\delta_{X^*\mathfrak{P}}\|_B}\right) \leq \sup_{\substack{v \in X^* \otimes \mathcal{P}_{\mathfrak{P}} \\ \|v\|_B=1}} B(\delta_{X^*\mathfrak{P}}, v).$$

Now let  $v$  be an arbitrary element in  $X^* \otimes \mathcal{P}_{\mathfrak{P}} = (X \oplus Y) \otimes \mathcal{P}_{\mathfrak{P}} = V_{X\mathfrak{P}} \oplus V_{Y\mathfrak{P}}$  (see (3.11)). Then  $v = v_{X\mathfrak{P}} + v_{Y\mathfrak{P}}$  with  $v_{X\mathfrak{P}} \in V_{X\mathfrak{P}}$ ,  $v_{Y\mathfrak{P}} \in V_{Y\mathfrak{P}}$ , and since  $B(\delta_{X^*\mathfrak{P}}, v_{X\mathfrak{P}}) = 0$ , we get

$$B(\delta_{X^*\mathfrak{P}}, v) = B(\delta_{X^*\mathfrak{P}}, v - v_{X\mathfrak{P}}) = B(\delta_{X^*\mathfrak{P}}, v_{Y\mathfrak{P}}).$$

Hence, using (5.6) and applying the Cauchy–Schwarz inequality gives

$$(5.9) \quad B(\delta_{X^*\mathfrak{P}}, v) = B_0(\bar{e}_{Y\mathfrak{P}}, v_{Y\mathfrak{P}}) \leq \|\bar{e}_{Y\mathfrak{P}}\|_{B_0} \|v_{Y\mathfrak{P}}\|_{B_0}.$$



It remains to estimate  $\|v_{Y\mathfrak{P}}\|_{B_0}$ . Making use of Lemma 3.1 and the algebraic inequality  $ab \leq (a^2 + b^2)/2$ , we find that

$$B_0(v_{X\mathfrak{P}}, v_{Y\mathfrak{P}}) \geq -\gamma \|v_{X\mathfrak{P}}\|_{B_0} \|v_{Y\mathfrak{P}}\|_{B_0} \geq -\frac{1}{2} \|v_{X\mathfrak{P}}\|_{B_0}^2 - \frac{\gamma^2}{2} \|v_{Y\mathfrak{P}}\|_{B_0}^2.$$

Then we have

$$\|v_{Y\mathfrak{P}}\|_{B_0}^2 = \|v\|_{B_0}^2 - 2B_0(v_{X\mathfrak{P}}, v_{Y\mathfrak{P}}) - \|v_{X\mathfrak{P}}\|_{B_0}^2 \leq \|v\|_{B_0}^2 + \gamma^2 \|v_{Y\mathfrak{P}}\|_{B_0}^2,$$

which yields

$$(5.10) \quad \|v_{Y\mathfrak{P}}\|_{B_0} \leq (1 - \gamma^2)^{-1/2} \|v\|_{B_0} \quad \forall v \in X^* \otimes \mathcal{P}_{\mathfrak{P}}.$$

Combining (5.8)–(5.10), using the right-hand inequality in (2.16), and recalling our notation for  $\delta_{X^*\mathfrak{P}}$  gives

$$\|u_{X^*\mathfrak{P}} - u_{X\mathfrak{P}}\|_B \leq \frac{1}{\sqrt{1 - \gamma^2}} \|\bar{e}_{Y\mathfrak{P}}\|_{B_0} \sup_{\substack{v \in X^* \otimes \mathcal{P}_{\mathfrak{P}} \\ \|v\|_B = 1}} \|v\|_{B_0} \leq \frac{\sqrt{\Lambda}}{\sqrt{1 - \gamma^2}} \|\bar{e}_{Y\mathfrak{P}}\|_{B_0}.$$

This completes the proof of (5.4).

The inequalities in (5.5) are proved in the same way. The constant in the strengthened Cauchy–Schwarz inequality is equal to zero in this case due to the orthogonality of the decomposition (3.7) (see (3.8)).  $\square$

The preceding results show that the terms  $\|\bar{e}_{Y\mathfrak{P}}\|_{B_0}$  and  $\|\bar{e}_{X\Omega}\|_{B_0}$  contributing to the error estimate  $\eta$  in (4.8) provide effective estimates for the error reductions  $\|u_{X^*\mathfrak{P}} - u_{X\mathfrak{P}}\|_B$  and  $\|u_{X\mathfrak{P}^*} - u_{X\mathfrak{P}}\|_B$ , respectively. Crucially, if  $A$  has affine dependence on the parameters  $y_m$  and (2.21) holds (or some other condition that ensures that only finitely many terms on the right-hand sides of (4.9) and (4.10) are nonzero), then  $\|\bar{e}_{Y\mathfrak{P}}\|_{B_0}$  and  $\|\bar{e}_{X\Omega}\|_{B_0}$  are computable in terms of the problem data and the original approximation  $u_{X\mathfrak{P}} \in V_{X\mathfrak{P}}$ . This means that the estimated error reductions can be evaluated before the enhanced approximations are actually computed. By considering the ratio of the estimated error reduction and the number of additional degrees of freedom associated with the enhanced approximation, the component estimators can be used to identify the enrichment strategy that reduces the energy error most efficiently.

Note that the true error reduction  $u_{X\mathfrak{P}^*} - u_{X\mathfrak{P}} \in V_{X\mathfrak{P}^*}$  (corresponding to enrichment of the polynomial space only) satisfies the identity

$$(5.11) \quad B(u_{X\mathfrak{P}^*} - u_{X\mathfrak{P}}, v) = F(v) - B(u_{X\mathfrak{P}}, v) \quad \forall v \in V_{X\mathfrak{P}^*}.$$

We have defined the associated error reduction estimator  $\bar{e}_{X\Omega} \in V_{X\Omega}$  via identity (4.9) which is a simplification of (5.11) in two ways. First, instead of the bilinear form  $B$  on the left-hand side of (5.11) we use the bilinear form  $B_0$ , which leads to simplified linear algebra; and second, the test and trial space  $V_{X\mathfrak{P}^*} = V_{X\mathfrak{P}} \oplus V_{X\Omega}$  is reduced to  $V_{X\Omega}$ . As a result of these simplifications,  $\bar{e}_{X\Omega}$  is the approximation of  $u_{X\mathfrak{P}^*} - u_{X\mathfrak{P}}$  satisfying the two-sided estimate (5.5) for any detail index set  $\Omega \subset \mathfrak{I}$  such that  $\mathfrak{P} \cap \Omega = \emptyset$ . However, for some specific detail index sets, one of the above simplifications, namely, the use of  $B_0$  instead of  $B$  in the left-hand side of (5.11), may become redundant. Evidently, this will be the case when

$$(5.12) \quad B(u, v) = B_0(u, v) \quad \forall u, v \in V_{X\Omega}.$$

Here,  $\|\bar{e}_{X\Omega}\|_B = \|\bar{e}_{X\Omega}\|_{B_0}$ , and similarly to (5.6)–(5.7) we have

$$\begin{aligned} \|\bar{e}_{X\Omega}\|_{B_0}^2 &= B_0(\bar{e}_{X\Omega}, \bar{e}_{X\Omega}) = B(u_{X\mathfrak{P}^*} - u_{X\mathfrak{P}}, \bar{e}_{X\Omega}) \\ &\leq \|u_{X\mathfrak{P}^*} - u_{X\mathfrak{P}}\|_B \|\bar{e}_{X\Omega}\|_B = \|u_{X\mathfrak{P}^*} - u_{X\mathfrak{P}}\|_B \|\bar{e}_{X\Omega}\|_{B_0}. \end{aligned}$$

This implies the following lower bound for the error reduction:

$$\|\bar{e}_{X\Omega}\|_{B_0} \leq \|u_{X\mathfrak{P}^*} - u_{X\mathfrak{P}}\|_B,$$

which is an improvement on the corresponding bound in (5.5) in the case when  $\lambda < 1$ . In particular, this is true for the model diffusion problem considered in section 2.1.

Let us further develop this line of thought. If

$$B(u, v) = B_0(u, v) + \sum_{m=1}^{\infty} \int_{\Gamma} \langle A_m u(\mathbf{y}), v(\mathbf{y}) \rangle y_m \, d\pi(\mathbf{y}),$$

we see that identity (5.12) holds if

$$(5.13) \quad \int_{\Gamma} y_m P_{\nu}(\mathbf{y}) P_{\mu}(\mathbf{y}) \, d\pi(\mathbf{y}) = 0 \quad \forall m \in \mathbb{N}, \quad \forall \mu, \nu \in \Omega.$$

This will be the case, for example, when the detail index set  $\Omega$  consists of only one index (this follows from the three-term recurrence (3.2)). The orthogonality property (5.13) (and hence the identity (5.12)) will also hold whenever  $\Omega$  defines a set of multivariate polynomials in  $y_m$  of only even or only odd total polynomial degree (this is again thanks to (3.2)). Two examples of such index sets are given by

$$\begin{aligned} \Omega &= \{(1, 1, 0, 0, \dots); (2, 0, 0, 0, \dots); (0, 2, 0, 0, \dots)\}, \\ \Omega &= \{(2, 1, 0, 0, \dots); (1, 2, 0, 0, \dots); (3, 0, 0, 0, \dots); (0, 3, 0, 0, \dots)\}. \end{aligned}$$

**6. Numerical experiments.** Some results of numerical experiments for a representative model problem are reported in this section. These results demonstrate the efficiency of the a posteriori error estimate  $\eta$  and illustrate the behavior of the component error estimators when the approximation space  $V_{X\mathfrak{P}}$  is selectively enriched.

**6.1. Model problem.** We adopt the notation of section 2.1 and consider the two-dimensional diffusion equation with random coefficient  $a$  and deterministic right-hand side function  $f$ , subject to homogeneous Dirichlet boundary conditions, as in (2.19). We assume that  $a$  is a second-order random field with prescribed mean  $\mathbb{E}[a]$  and covariance function  $C[a]$ . Specifically, we assume that  $C[a]$  is the separable exponential covariance function given by

$$(6.1) \quad C[a](\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\frac{|x_1 - x'_1|}{l_1} - \frac{|x_2 - x'_2|}{l_2}\right), \quad \mathbf{x}, \mathbf{x}' \in D \subset \mathbb{R}^2,$$

where  $\sigma$  denotes the standard deviation and  $l_1, l_2$  are correlation lengths. This is a convenient choice since there exist analytical expressions for the eigenpairs of the associated integral operator in the special case of a rectangular domain  $D$ . This leads in turn to an analytical representation of  $a$  in the parametric form (2.20) using the Karhunen–Loève expansion:

$$(6.2) \quad a(\mathbf{x}, \mathbf{y}) = \mathbb{E}[a](\mathbf{x}) + \sigma \sqrt{3} \sum_{m=1}^{\infty} \sqrt{\lambda_m} \varphi_m(\mathbf{x}) y_m,$$

where  $\mathbf{x} \in D$ ,  $\mathbf{y} \in \Gamma = \prod_{m=1}^\infty [-1, 1]$ , and  $\{(\lambda_m, \varphi_m)\}_{m=1}^\infty$  are the eigenpairs of the integral operator associated with  $C[a]$  in (6.1).<sup>1</sup> With this representation of the random coefficient  $a$ , our model problem can now be written in the parametric form (2.19). The variational formulation of (2.19) is given by (2.2)–(2.4), where the parametric operator  $A(\mathbf{y})$  is defined by (2.22) for all  $\mathbf{y} \in \Gamma$ ,  $f(\mathbf{y}) := f \in H^{-1}(D)$ , and  $V := L^2_\pi(\Gamma, H^1_0(D))$  with  $\pi$  defined by (2.9).

In the numerical experiments described below we chose  $D = (-1, 1) \times (-1, 1)$  and we set  $f(\mathbf{x}) = \frac{1}{8}(2 - x_1^2 - x_2^2)$ . The statistical parameters defining the random field  $a$  were selected as follows:

$$\mathbb{E}[a](\mathbf{x}) = 1 \ (\mathbf{x} \in D), \quad \sigma = 0.2, \quad l_1 = l_2 = 2.0.$$

We also assume that the parameters  $y_m$  in (6.2) are the images of uniformly distributed independent mean-zero random variables, and so  $\pi_m = \pi_m(y_m)$  is the associated probability measure on  $\Gamma_m = [-1, 1]$ . This assumption ensures (2.21). We note that the resulting model problem is essentially the same as that considered in [7], [19].

**6.2. Discretization details.** Turning now to approximation logistics, we construct a finite-dimensional subspace of  $V$  by tensorizing standard finite element functions of  $\mathbf{x} \in D$  and multivariate polynomials of  $\mathbf{y} \in \Gamma$ . For the finite element approximation we use a space  $X_h$  of continuous piecewise bilinear functions over a uniform partition  $\square_h$  of square elements (here,  $h > 0$  denotes the length of each element edge). Next, given integers  $M \geq 1$  and  $p \geq 0$ , we choose the finite-dimensional subspace  $\mathcal{P}_{M,p}$  of  $L^2_\pi(\Gamma)$  to be the space of polynomials of total degree  $\leq p$  in the first  $M$  parameters  $y_m$ ,  $m = 1, \dots, M$ . Note that  $\mathcal{P}_{M,p}$  can be equivalently defined as the span of the tensorized Legendre polynomials  $P_\nu(\mathbf{y})$  (cf. (3.3)) with

$$\nu \in \mathfrak{P}_{M,p} := \left\{ \nu = (\nu_1, \nu_2, \dots) \in \mathbb{N}_0^M; \text{supp } \nu \subset \{1, \dots, M\}, \sum_{m=1}^M \nu_m \leq p \right\}.$$

The resulting finite-dimensional subspace of  $V$  is  $V_N = V_{X\mathfrak{P}} := X_h \otimes \mathcal{P}_{M,p}$ . Here,  $N = N(h, M, p)$  denotes the total number of degrees of freedom and is simply the product of the number of spatial degrees of freedom and the dimension of the polynomial space  $\mathcal{P}_{M,p}$ . Since  $\dim(\mathcal{P}_{M,p}) = \frac{(p+M)!}{p! M!}$ , we have  $N = (2h^{-1} - 1)^2 \frac{(p+M)!}{p! M!}$ .

The sGFEM solution  $u_N = u_{X\mathfrak{P}} \in V_{X\mathfrak{P}}$  is uniquely defined by identity (3.4) and can be written as

$$u_{X\mathfrak{P}}(\mathbf{x}, \mathbf{y}) = \sum_{\nu \in \mathfrak{P}_{M,p}} \phi_\nu(\mathbf{x}) P_\nu(\mathbf{y}), \quad \phi_\nu(\mathbf{x}) \in X_h.$$

We fix the number of active parameters  $y_m$  in the approximation  $u_{X\mathfrak{P}}$  to be  $M = 3$ . To compute the error estimate  $\eta$  defined by (4.8)–(4.11), the detail spaces on  $D$  and  $\Gamma$  need to be specified. In our experiments,  $Y_h$  and  $\mathcal{Q}_{M,p}$  are chosen so that

$$X_{h/2} = X_h \oplus Y_h \quad \text{and} \quad \mathcal{P}_{M,p+1} = \mathcal{P}_{M,p} \oplus \mathcal{Q}_{M,p},$$

respectively. Thus,  $Y_h$  spans the set of bilinear bubble functions corresponding to the edge midpoints and element centroids of the original mesh  $\square_h$ , whereas  $\mathcal{Q}_{M,p}$

---

<sup>1</sup>See [11, pp. 28–29] for analytical expressions for  $\lambda_m$  and  $\varphi_m$  in one spatial dimension; the formulas for rectangular domains follow by tensorization.

spans  $M$ -variate polynomials of total degree exactly equal to  $p + 1$ . This gives a decomposition (3.11) with

$$V_{X\Omega} := X_h \otimes \mathcal{Q}_{M,p}, \quad V_{Y\mathfrak{F}} := Y_h \otimes \mathcal{P}_{M,p}, \quad V_{Y\Omega} := Y_h \otimes \mathcal{Q}_{M,p}.$$

The resulting three error estimators  $\bar{e}_{X\Omega}$ ,  $\bar{e}_{Y\mathfrak{F}}$ , and  $\bar{e}_{Y\Omega}$  are computed by solving linear systems associated with the discrete formulations (4.9)–(4.11) with bilinear form  $B_0(\cdot, \cdot)$  (cf. (2.15), (2.23)) given by

$$B_0(v, w) = \int_{\Gamma} \int_D \mathbb{E}[a](\mathbf{x}) \nabla v(\mathbf{x}, \mathbf{y}) \cdot \nabla w(\mathbf{x}, \mathbf{y}) \, d\mathbf{x} \, d\pi(\mathbf{y}).$$

Our smart choice of detail space  $Y_h$  is now evident: the error estimators  $\bar{e}_{Y\mathfrak{F}}$  and  $\bar{e}_{Y\Omega}$  can be computed locally on each element  $K \in \square_h$ . We use a standard element residual technique (see, e.g., [1]) to construct the following local residual problem corresponding to (4.10): find  $\bar{e}_{Y\mathfrak{F}}|_K \in Y_h|_K \otimes \mathcal{P}_{M,p}$  satisfying

$$(6.3) \quad \begin{aligned} B_{0,K}(\bar{e}_{Y\mathfrak{F}}|_K, v) &= F_K(v) + \int_{\Gamma} \int_K \nabla \cdot (a(\mathbf{x}, \mathbf{y}) \nabla u_{X\mathfrak{F}}(\mathbf{x}, \mathbf{y})) v(\mathbf{x}, \mathbf{y}) \, d\mathbf{x} \, d\pi(\mathbf{y}) \\ &\quad - \frac{1}{2} \int_{\Gamma} \int_{\partial K \setminus \partial D} a(s, \mathbf{y}) \left[ \left[ \frac{\partial u_{X\mathfrak{F}}}{\partial n} \right] \right] v(s, \mathbf{y}) \, ds \, d\pi(\mathbf{y}) \end{aligned}$$

for any  $v \in Y_h|_K \otimes \mathcal{P}_{M,p}$ . Here,  $B_{0,K}(\cdot, \cdot)$  and  $F_K(\cdot)$  are the elementwise bilinear form and linear functional, respectively;  $Y_h|_K$  is the restriction of the finite element detail space to the element  $K$ ; and  $\left[ \left[ \frac{\partial u_{X\mathfrak{F}}}{\partial n} \right] \right]$  denotes the flux jump in the approximate solution  $u_{X\mathfrak{F}}$  across interelement edges.

An important feature of our strategy is that the linear algebra associated with solving (6.3) is elementary. The coefficient matrix in the linear system associated with (6.3) has a very simple structure: it is the Kronecker product of a  $5 \times 5$  reduced stiffness matrix (with entries associated with the five spatial basis functions in  $Y_h|_K$ ) and an identity matrix of dimension  $|P| = \dim(\mathcal{P}_{M,p})$ . As a result, the action of the inverse of the coefficient matrix representing the left-hand side of (4.10) can be effected by a block  $LDL^T$  factorization of the element stiffness matrices followed by a sequence of  $|P|$  backward and forward substitutions.<sup>2</sup>

A similar localization procedure can be performed to compute  $\bar{e}_{Y\Omega}$  from (4.11). In this case the matrix of the resulting linear system (at an element level) is also block diagonal with exactly the same  $5 \times 5$  blocks as described above for  $\bar{e}_{Y\mathfrak{F}}$ . The only difference is the number of these blocks, which in this case is equal to  $|Q| = \dim(\mathcal{Q}_{M,p})$ .

In contrast to  $\bar{e}_{Y\mathfrak{F}}$  and  $\bar{e}_{Y\Omega}$ , the computation of the estimator  $\bar{e}_{X\Omega}$  involves the solution of a nonlocal sparse system associated with (4.9). The good news is that this nonlocal system is block diagonal with each block representing the assembled stiffness matrix corresponding to the mean coefficient  $\mathbb{E}[a](\mathbf{x})$  on  $\square_h$ . Once the stiffness matrix has been factorized, the action of the inverse of the coefficient matrix on the left-hand side of (4.9) simply requires  $|Q|$  forward and backward substitutions. Recall that once  $\bar{e}_{Y\mathfrak{F}}$ ,  $\bar{e}_{X\Omega}$ , and  $\bar{e}_{Y\Omega}$  have been computed, the total error estimate  $\eta$  is then given by the combination (4.8).

Note that the solution of the element residual problems (6.3) is a fundamental feature of our energy estimation strategy. Other strategies are also possible: equilibrating the flux jumps before solving the local problem in order to guarantee that

---

<sup>2</sup>The factorizations and triangular solves are logically independent, and thus the entire computation can be vectorized or parallelized over the finite elements that define the spatial subdivision.

the local estimator  $\bar{e}_{Y\mathfrak{P}}$  is a strict upper bound on the spatial error and the use of alternative definitions for the correction space  $Y_h$ . We have not investigated these nor the simple alternative of using scaled residual norms as a surrogate for  $\bar{e}_{Y\mathfrak{P}}$ .

**6.3. Numerical results.** First we test the efficiency of the total error estimate  $\eta$ . As the exact solution to the model problem (which is a function of infinitely many parameters  $y_m$ ) is not available, we compute a reference solution  $u_{\text{ref}} \in X_{h_{\text{ref}}} \otimes \mathcal{P}_{M,p_{\text{ref}}}$  on a very fine spatial mesh  $\square_{h_{\text{ref}}}$  with  $h_{\text{ref}} = 2^{-9}$  in combination with  $M$ -variate polynomials of total degree  $p_{\text{ref}} = 7$ . Recall that we fixed  $M = 3$  for the approximation. We use the same value of  $M$  to compute the reference solution. (Note that this means that  $u_{\text{ref}}$  is actually the solution to a problem with a reduced version of the covariance in (6.1).) Then, for combinations of finite element spaces  $X_h$  and polynomial spaces  $\mathcal{P}_{M,p}$  with various choices of  $h$  and  $p$ , we compute the total estimate  $\eta = \eta(u_{X\mathfrak{P}})$  of the energy error in the sGFEM solution  $u_{X\mathfrak{P}} \in X_h \otimes \mathcal{P}_{M,p}$ . We compare this with the reference error  $\|u_{\text{ref}} - u_{X\mathfrak{P}}\|_B$  and report the effectivity index

$$\theta_{\text{eff}} := \frac{\eta(u_{X\mathfrak{P}})}{\|u_{\text{ref}} - u_{X\mathfrak{P}}\|_B}.$$

The computed results are presented in Table 1 (for fixed  $p = 2$  and varying  $h$ ) and in Table 2 (for fixed  $h = 2^{-5}$  and varying  $p$ ). In all cases the effectivity index can be seen to be very close to unity (more precisely,  $\theta_{\text{eff}}$  varies between 0.88 and 1.27).

TABLE 1

The total error estimates, reference errors, and effectivity indices for the sGFEM solutions  $u_{X\mathfrak{P}} \in X_h \otimes \mathcal{P}_{M,p}$  with fixed  $M = 3$  and fixed  $p = 2$ .

$h$	$N$	$\eta(u_{X\mathfrak{P}})$	$\ u_{\text{ref}} - u_{X\mathfrak{P}}\ _B$	$\theta_{\text{eff}}$
0.2500000	490	1.8411e-02	1.8882e-02	0.98
0.1250000	2250	8.7125e-03	9.4502e-03	0.92
0.0625000	9610	4.3394e-03	4.7752e-03	0.91
0.0312500	39690	2.3500e-03	2.4889e-03	0.94
0.0156250	161290	1.5192e-03	1.4297e-03	1.06
0.0078125	650250	1.2321e-03	1.0032e-03	1.23

TABLE 2

The total error estimates, reference errors, and effectivity indices for the sGFEM solutions  $u_{X\mathfrak{P}} \in X_h \otimes \mathcal{P}_{M,p}$  with fixed  $h = 2^{-5}$  and fixed  $M = 3$ .

$p$	$N$	$\eta(u_{X\mathfrak{P}})$	$\ u_{\text{ref}} - u_{X\mathfrak{P}}\ _B$	$\theta_{\text{eff}}$
1	15876	6.2228e-03	4.8980e-03	1.27
2	39690	2.3500e-03	2.4889e-03	0.94
3	79380	2.0770e-03	2.3539e-03	0.88
4	138915	2.0650e-03	2.3482e-03	0.88
5	222264	2.0645e-03	2.3479e-03	0.88

In the next series of experiments we test the behavior of the three error estimates  $\|\bar{e}_{X\Omega}\|_{B_0}$ ,  $\|\bar{e}_{Y\mathfrak{P}}\|_{B_0}$ ,  $\|\bar{e}_{Y\Omega}\|_{B_0}$  and their combination  $\eta$ , as the finite-dimensional space  $V_{X\mathfrak{P}}$  is enriched by either refining the spatial mesh  $\square_h$  or else by increasing the degree of the polynomial space  $\mathcal{P}_{M,p}$ . In each case we use a log–log scale to plot the resulting error estimates as a function of the total number of degrees of freedom  $N = \dim(V_{X\mathfrak{P}})$ .

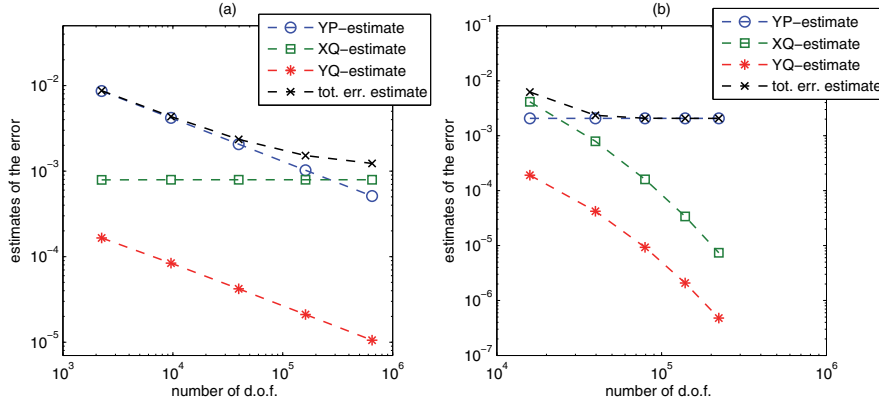


FIG. 1. Error estimates  $\|\bar{e}_{Y\mathfrak{P}}\|_{B_0}$ ,  $\|\bar{e}_{X\Omega}\|_{B_0}$ ,  $\|\bar{e}_{Y\Omega}\|_{B_0}$  and the total estimate  $\eta$  of the energy error in the sGFEM solutions  $u_{X\mathfrak{P}} \in X_h \otimes \mathcal{P}_{M,p}$  computed with (a) fixed  $M = 3$ , fixed  $p = 2$ , and varying  $h = 2^{-k}$ ,  $k = 3, 4, \dots, 7$ ; (b) fixed  $h = 2^{-5}$ , fixed  $M = 3$ , and varying  $p = 1, 2, \dots, 5$ .

Figure 1(a) shows the behavior of the error estimates as we refine the spatial mesh (by choosing  $h = 2^{-3}, 2^{-4}, \dots, 2^{-7}$ ) while having fixed degree  $p = 2$  for polynomial approximations on  $\Gamma$ . We observe that estimates  $\|\bar{e}_{Y\mathfrak{P}}\|_{B_0}$  and  $\|\bar{e}_{Y\Omega}\|_{B_0}$  (that both incorporate the detail finite element space  $Y_h$ ) decay algebraically. In contrast, the error estimate  $\|\bar{e}_{X\Omega}\|_{B_0}$  (the one that does not incorporate  $Y_h$ ) remains constant. Hence, after several mesh refinements we reach the stage when contributions of the  $X\Omega$ -estimate to the total estimate  $\eta$  become dominant. Further mesh refinement from this point onwards will not significantly reduce the total error (the algebraic decay of  $\eta$  is no longer visible). Thus we hypothesize that the polynomial space on  $\Gamma$  needs to be enriched if the error is to be reduced further. (Indeed, we can see from Figure 1 that enriching the polynomial space on  $\Gamma$  when  $h = 2^{-4}$  would be beneficial.) We also note the fact that  $\|\bar{e}_{Y\Omega}\|_{B_0}$  (the estimate that incorporates both detail spaces  $Y_h$  and  $\mathcal{Q}_{M,p}$ ) is relatively small compared to the estimates  $\|\bar{e}_{Y\mathfrak{P}}\|_{B_0}$  and  $\|\bar{e}_{X\Omega}\|_{B_0}$ .

In Figure 1(b), we again plot our error estimates as functions of  $N = \dim(V_{X\mathfrak{P}})$ . In this case, we increase  $N$  by enriching the polynomial space  $\mathcal{P}_{M,p}$  on  $\Gamma$  (we vary  $p$  from 1 to 5) while keeping the spatial mesh (and hence the finite element space  $X_h$ ) fixed (we set  $h = 2^{-5}$ ). In this plot we observe the exponential decay of  $\|\bar{e}_{X\Omega}\|_{B_0}$  and  $\|\bar{e}_{Y\Omega}\|_{B_0}$  (both estimates that incorporate the detail polynomial space  $\mathcal{Q}_{M,p}$ ). As expected, the estimate  $\|\bar{e}_{Y\mathfrak{P}}\|_{B_0}$  (the one that does not incorporate  $\mathcal{Q}_{M,p}$ ) stays constant. Note that for  $p = 1$ , the  $X\Omega$ -estimate has the largest contribution to  $\eta$ , whereas already for  $p = 2$  the  $Y\mathfrak{P}$ -estimate dominates. As before, the  $Y\Omega$ -estimates have the smallest contribution to  $\eta$ , as they are at least one order of magnitude less than other estimates. As a result, the total estimate  $\eta$  decays algebraically when  $p$  is increased from  $p = 1$  to  $p = 2$  and then quickly stagnates with further growth of  $p$ . This gives clear indication that in order to maintain the algebraic decay of  $\eta$ , spatial mesh refinement is needed when  $p = 2$ . Elementwise error estimates  $\|\bar{e}_{X\Omega}\|_{B_{0,K}}$ ,  $\|\bar{e}_{Y\mathfrak{P}}\|_{B_{0,K}}$ , and  $\|\bar{e}_{Y\Omega}\|_{B_{0,K}}$  are plotted in Figure 2 for the case  $h = 2^{-7}$ ,  $M = 3$ , and  $p = 2$ .

We emphasize that conclusions drawn from the second series of experiments are in perfect agreement with the result of Theorem 5.1. Our theory and experiments clearly show that the error estimates  $\|\bar{e}_{Y\mathfrak{P}}\|_{B_0}$  and  $\|\bar{e}_{X\Omega}\|_{B_0}$  should be used for adap-

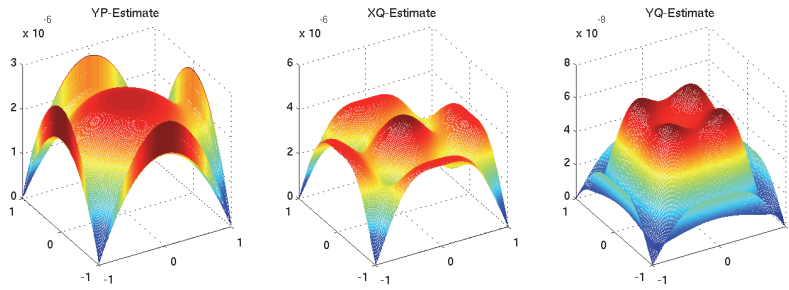


FIG. 2. The elementwise error estimates  $\|\bar{e}_{Y\mathfrak{P}}\|_{B_{0,K}}$ ,  $\|\bar{e}_{X\Omega}\|_{B_{0,K}}$ , and  $\|\bar{e}_{Y\Omega}\|_{B_{0,K}}$  for the sGFEM solution  $u_{X\mathfrak{P}} \in X_h \otimes \mathcal{P}_{M,p}$  with  $h = 2^{-7}$ ,  $M = 3$ , and  $p = 2$ .

tive enrichment of the approximation space in the sGFEM: a dominant  $Y\mathfrak{P}$ -estimate indicates that spatial mesh refinement is needed, whereas a dominant  $X\Omega$ -estimate calls for enrichment of the polynomial space on  $\Gamma$ . These conclusions underpin the generic adaptive algorithm that is presented in the final section.

**7. Adaptive refinement.** In this final section we look to the future and discuss how the error estimate from Theorem 4.1 and the error reduction estimates from Theorem 5.1 may be combined to give a generic adaptive algorithm for approximating solutions of (2.2). To control the energy error  $\|u - u_{X\mathfrak{P}}\|_B$ , the estimate  $\eta$  defined by (4.8)–(4.11) must be compared to a specified tolerance,  $tol$ . The goal of our proposed adaptive algorithm is to generate a sequence of finite-dimensional subspaces

$$X_0 \subset X_1 \subset X_2 \subset \dots \subset X_n \subset H$$

and a sequence of index sets

$$\mathfrak{P}_0 \subset \mathfrak{P}_1 \subset \mathfrak{P}_2 \subset \dots \subset \mathfrak{P}_n \subset \mathfrak{I}$$

such that the tolerance  $tol$  is met by the Galerkin solution  $u_{\text{Gal}} \in X_n \otimes \mathcal{P}_{\mathfrak{P}_n}$ . To control the computational cost, we also insist that the enrichment strategy selected at each step ensures that the total number of degrees of freedom is minimal. More precisely, we start with an initial subspace  $X_0 \subset H$  of low dimension (for example, a finite element space associated with a coarse mesh) and an initial index set  $\mathfrak{P}_0$  (e.g.,  $\mathfrak{P}_0 = \{(0, 0, 0, \dots)\}$ ). For candidate detail spaces  $Y_0$  such that  $X_1 = X_0 \oplus Y_0$  and detail index sets  $\Omega_0$  such that  $\mathfrak{P}_1 = \mathfrak{P}_0 \cup \Omega_0$ , the corresponding error reduction estimates  $\|\bar{e}_{Y_0\mathfrak{P}_0}\|_{B_0}$  and  $\|\bar{e}_{X_0\Omega_0}\|_{B_0}$  from Theorem 5.1 are computed and then divided by the corresponding number of new degrees of freedom that would be introduced if that particular enrichment strategy were implemented (by  $\dim(Y_0)$  for  $X_0$ -enrichment and by  $\dim(\Omega_0)$  for  $\mathfrak{P}_0$ -enrichment). By selecting the largest ratio, we identify the enrichment strategy that leads to the largest error reduction per extra degree of freedom. The process is then repeated. At each step, if the largest ratio corresponds to  $X$ -refinement, then  $X_n$  is updated (that is,  $X_{n+1} := X_n \oplus Y_n$ ) and the index set for the polynomial space on  $\Gamma$  remains the same ( $\mathfrak{P}_{n+1} := \mathfrak{P}_n$  and so  $\mathcal{P}_{\mathfrak{P}_{n+1}} := \mathcal{P}_{\mathfrak{P}_n}$ ). Otherwise, only the index set is updated (i.e.,  $\mathfrak{P}_{n+1} := \mathfrak{P}_n \cup \Omega_n$  so  $\mathcal{P}_{\mathfrak{P}_{n+1}} := \mathcal{P}_{\mathfrak{P}_n} \oplus \mathcal{P}_{\Omega_n}$  and  $X_{n+1} := X_n$ ).

At each step one may wish to test several options for  $X$ - and particularly for  $\mathfrak{P}$ -enrichment. For example, if  $\mathfrak{P}_n = \{(0, 0, 0, \dots), (1, 0, 0, \dots)\}$  for some  $n \in \mathbb{N}_0$ , then



$\mathcal{P}_{\mathfrak{P}_n} = \text{span}\{1, y_1\}$ , and it is natural to test at least two possibilities for the detail index set  $\Omega_n$ , namely,

$$\Omega_n^{(1)} = \{(2, 0, 0, \dots)\} \quad \text{and} \quad \Omega_n^{(2)} = \{(0, 1, 0, \dots)\}.$$

The first choice corresponds to adding  $y_1^2$  to the basis for  $\mathcal{P}_{\mathfrak{P}_n}$  (a higher degree polynomial in the only currently active parameter), while the second choice corresponds to adding the polynomial  $y_2$ . This makes active a new parameter but does not lead to an increase in the polynomial degree of the existing set. Our error reduction estimate can be computed for multiple choices of  $\Omega_n$  (if desired), and the most efficient option can then be selected.

A generic adaptive algorithm (written in the style of [12]) is given in Algorithm 7.1. This algorithm has the following building blocks:

- **Solve** $[A, f, X, \mathfrak{P}]$ —a subroutine that generates the Galerkin approximation  $u_{X\mathfrak{P}} \in X \otimes \mathcal{P}_{\mathfrak{P}}$  satisfying (3.4);
- **Error\_Estimate** $[A, f, u_{X\mathfrak{P}}, Y, \Omega, \Lambda, \beta, \gamma]$ —a subroutine that computes the estimate for the norm of the error in the Galerkin approximation  $u_{X\mathfrak{P}}$  (see (4.27));
- **Error\_Reduction\_1** $[A, f, u_{X\mathfrak{P}}, Y, \Lambda, \gamma]$ —a subroutine that computes the error reduction estimate for  $X$ -enrichment (see (5.4));
- **Error\_Reduction\_2** $[A, f, u_{X\mathfrak{P}}, \Omega, \Lambda]$ —a subroutine that computes the error reduction estimate for  $\mathfrak{P}$ -enrichment (see (5.5)).

ALGORITHM 7.1. **Adaptive\_Galerkin** $[tol, A, f, \Lambda] \rightarrow u_{\text{Gal}}$

**input**  $X_0, \mathfrak{P}_0$

**for**  $n = 0, 1, 2, \dots$  **do**

$u_n \leftarrow \text{Solve}[A, f, X_n, \mathfrak{P}_n]$

$\eta_n \leftarrow \text{Error\_Estimate}[A, f, u_n, Y_n, \Omega_n, \Lambda, \beta_n, \gamma_n]$

**if**  $\eta_n < tol$  **then break**

**for**  $i = 1, 2, \dots, I_n$  **do**

$\delta_X^{(i)} \leftarrow \text{Error\_Reduction\_1}[A, f, u_n, Y_n^{(i)}, \Lambda, \gamma_n^{(i)}]$

$\delta_X^{(i)} := \delta_X^{(i)} / \dim(Y_n^{(i)})$

**end**

**for**  $j = 1, 2, \dots, J_n$  **do**

$\delta_{\mathfrak{P}}^{(j)} \leftarrow \text{Error\_Reduction\_2}[A, f, u_n, \Omega_n^{(j)}, \Lambda]$

$\delta_{\mathfrak{P}}^{(j)} := \delta_{\mathfrak{P}}^{(j)} / \dim(\Omega_n^{(j)})$

**end**

$\delta_X := \max\{\delta_X^{(i)}; i = 1, 2, \dots, I_n\}; \hat{I} := \text{argmax}\{\delta_X^{(i)}; i = 1, 2, \dots, I_n\};$

$\delta_{\mathfrak{P}} := \max\{\delta_{\mathfrak{P}}^{(j)}; j = 1, 2, \dots, J_n\}; \hat{J} := \text{argmax}\{\delta_{\mathfrak{P}}^{(j)}; j = 1, 2, \dots, J_n\};$

**if**  $\delta_X \geq \delta_{\mathfrak{P}}$  **then**  $X_{n+1} := X_n \oplus Y_n^{(\hat{I})}, \mathfrak{P}_{n+1} := \mathfrak{P}_n$

**else**  $X_{n+1} := X_n, \mathfrak{P}_{n+1} := \mathfrak{P}_n \cup \Omega_n^{(\hat{J})}$

**end**

$u_{\text{Gal}} \leftarrow u_n$



In the above,  $\Lambda$  is the constant in (2.16),  $\gamma \in [0, 1)$  is the constant in (3.6) ( $\gamma$  depends only on the choice of the detail space  $Y$ ), and  $\beta \in [0, 1)$  is the constant in the saturation assumption (3.17). Note that  $\beta$  depends on the choice of both  $Y$  and  $\Omega$ .

The subroutines in Algorithm 7.1 are generic in the sense that we have not specified the detail spaces  $Y$  or the detail index sets  $\Omega$ . Indeed, we place no restriction on these inputs. We note that the choice of  $Y$  (resp.,  $\Omega$ ) in `Error_Reduction_1` (resp., in `Error_Reduction_2`) may, in general, differ from the choice of the detail space and the detail index set in the subroutine `Error_Estimate`. Moreover, we emphasize again that several options for  $Y$  (resp.,  $\Omega$ ) may be tested using `Error_Reduction_1` (resp., `Error_Reduction_2`). We have already mentioned several strategies for enhancing the polynomial spaces and hence for selecting  $\Omega$ . As for the finite element detail space  $Y$ , the best choice will largely depend on the following: the problem data (that is, the operator  $A$  and the right-hand side  $f$  in (2.1)); the available information about the regularity of the solution; and the linear algebra solver that is available. For example, for finite element approximations in  $H = H_0^1(D)$ , suppose  $X \subset H$  is the space of piecewise linear polynomials defined over a triangular mesh  $\Delta_h$  on a polygonal domain  $D \subset \mathbb{R}^2$ . Then the detail subspace  $Y$  could be chosen such that the enriched space  $X^* = X \oplus Y$  contains piecewise linear polynomials defined over the *uniformly* refined mesh  $\Delta_{h/2}$  obtained by regular subdivisions of all the triangles in  $\Delta_h$ . Note that for the Laplace operator  $A_0$  and for the choice of  $X$  and  $Y$  just described the constant  $\gamma$  in (3.6) satisfies  $\gamma^2 = 1/2$  (see [4], [1, p. 99]). On the other hand, if it is known that for each  $\mathbf{y} \in \Gamma$ , the solution  $u(\mathbf{y})$  to (2.1) exhibits singular behavior (for example, due to the geometry of  $D$ ), then it is natural to opt for enriched subspaces  $X^*$  corresponding to *locally* refined meshes. Implementing such strategies will be the focus of future work.

## REFERENCES

- [1] M. AINSWORTH AND J. T. ODEN, *A Posteriori Error Estimation in Finite Element Analysis*, Wiley, New York, 2000.
- [2] I. BABUŠKA, R. TEMPONE, AND G. E. ZOURARIS, *Galerkin finite element approximations of stochastic elliptic partial differential equations*, SIAM J. Numer. Anal., 42 (2004), pp. 800–825.
- [3] M. BIERI, R. ANDREEV, AND C. SCHWAB, *Sparse tensor discretization of elliptic SPDEs*, SIAM J. Sci. Comput., 31 (2009), pp. 4281–4304.
- [4] D. BRAESS, *The contraction number of a multigrid method for solving the Poisson equation*, Numer. Math., 37 (1981), pp. 387–404.
- [5] 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.
- [6] T. BUTLER, C. DAWSON, AND T. WILDEY, *A posteriori error analysis of stochastic differential equations using polynomial chaos expansions*, SIAM J. Sci. Comput., 33 (2011), pp. 1267–1291.
- [7] M. K. DEB, I. M. BABUŠKA, AND J. T. ODEN, *Solution of stochastic partial differential equations using Galerkin finite element techniques*, Comput. Methods Appl. Mech. Engrg., 190 (2001), pp. 6359–6372.
- [8] M. EIGEL, C. GITTELSON, C. SCHWAB, AND E. ZANDER, *Adaptive stochastic Galerkin FEM*, Comput. Methods Appl. Mech. Engrg., 270 (2014), pp. 247–269.
- [9] V. ELJKHOUT AND P. VASSILEVSKI, *The role of the strengthened Cauchy–Buniakowski–Schwarz inequality in multilevel methods*, SIAM Rev., 33 (1991), pp. 405–419.
- [10] W. GAUTSCHI, *Orthogonal Polynomials: Computation and Approximation*, Numer. Math. Sci. Comput., Oxford University Press, New York, 2004.
- [11] R. G. GHANEM AND P. D. SPANOS, *Stochastic Finite Elements: A Spectral Approach*, Springer-Verlag, New York, 1991.
- [12] C. J. GITTELSON, *An adaptive stochastic Galerkin method for random elliptic operators*, Math. Comp., 82 (2013), pp. 1515–1541.

- [13] A. KEESE, *A Review of Recent Developments in the Numerical Solution of Stochastic Partial Differential Equations (Stochastic Finite Elements)*, Technical Report 2003–06, Institute of Scientific Computing, TU Braunschweig, 2003.
- [14] O. P. LE MAÎTRE AND O. M. KNIO, *Spectral Methods for Uncertainty Quantification: With Applications to Computational Fluid Dynamics*, Sci. Comput., Springer, New York, 2010.
- [15] O. P. LE MAÎTRE, O. M. KNIO, H. N. NAJM, AND R. G. GHANEM, *Uncertainty propagation using Wiener-Haar expansions*, J. Comput. Phys., 197 (2004), pp. 28–57.
- [16] O. P. LE MAÎTRE, H. N. NAJM, R. G. GHANEM, AND O. M. KNIO, *Multi-resolution analysis of Wiener-type uncertainty propagation schemes*, J. Comput. Phys., 197 (2004), pp. 502–531.
- [17] O. P. LE MAÎTRE, H. N. NAJM, P. P. PÉBAY, R. G. GHANEM, AND O. M. KNIO, *Multi-resolution-analysis scheme for uncertainty quantification in chemical systems*, SIAM J. Sci. Comput., 29 (2007), pp. 864–889.
- [18] 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.
- [19] C. E. POWELL AND H. C. ELMAN, *Block-diagonal preconditioning for spectral stochastic finite-element systems*, IMA J. Numer. Anal., 29 (2009), pp. 350–375.
- [20] A. SCHMIDT AND K. G. SIEBERT, *A posteriori estimators for the h-p version of the finite element method in 1D*, Appl. Numer. Math., 35 (2000), pp. 43–66.
- [21] C. SCHWAB AND C. J. GITTELSON, *Sparse tensor discretizations of high-dimensional parametric and stochastic PDEs*, Acta Numer., 20 (2011), pp. 291–467.
- [22] P. K. SUETIN, *Classical Orthogonal Polynomials*, Nauka, Moscow, 1979 (in Russian).
- [23] X. WAN AND G. E. KARNIADAKIS, *An adaptive multi-element generalized polynomial chaos method for stochastic differential equations*, J. Comput. Phys., 209 (2005), pp. 617–642.
- [24] 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.