Weierstraß-Institut
für Angewandte Analysis und Stochastik

Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint ISSN 2198-5855

A convergent adaptive finite element stochastic Galerkin method based on multilevel expansions of random fields

Markus Bachmayr[1], Martin Eigel[2], Henrik Eisenmann[1], Igor Voulis[3]

submitted: June 3, 2024

[1] Institut für Geometrie und Praktische Mathematik
RWTH Aachen University
Templergraben 55
52062 Aachen
Germany
E-Mail: bachmayr@igpm.rwth-aachen.de
eisenmann@igpm.rwth-aachen.de

[2] Weierstrass Institute
Mohrenstr. 39
10117 Berlin
Germany
E-Mail: martin.eigel@wias-berlin.de

[3] Institute for Numerical and Applied Mathematics
University of Göttingen
Lotzestr. 16-18
37083 Göttingen
Germany
E-Mail: i.voulis@math.uni-goettingen.de

No. 3112
Berlin 2024

2020 *Mathematics Subject Classification.* 35J25, 35R60, 41A10, 41A63, 42C10, 65N50, 65N30.

Key words and phrases. Stochastic Galerkin method, finite elements, frame-based error estimation, multilevel expansions of random fields.

M.B. acknowledges funding by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project numbers 442047500, 501389786. M.E. acknowledges funding by the DFG SPP 2298 and ANR-DFG project "COFNET". The work of H.E. was funded by Deutsche Forschungsgemeinschaft – project number 501389786. I.V. acknowledges funding by the DFG through grant 432680300 – SFB 1456.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Leibniz-Institut im Forschungsverbund Berlin e. V.
Mohrenstraße 39
10117 Berlin
Germany

Fax: +49 30 20372-303
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

A convergent adaptive finite element stochastic Galerkin method based on multilevel expansions of random fields

Markus Bachmayr, Martin Eigel, Henrik Eisenmann, Igor Voulis

Abstract

The subject of this work is an adaptive stochastic Galerkin finite element method for parametric or random elliptic partial differential equations, which generates sparse product polynomial expansions with respect to the parametric variables of solutions. For the corresponding spatial approximations, an independently refined finite element mesh is used for each polynomial coefficient. The method relies on multilevel expansions of input random fields and achieves error reduction with uniform rate. In particular, the saturation property for the refinement process is ensured by the algorithm. The results are illustrated by numerical experiments, including cases with random fields of low regularity.

1 Introduction

Elliptic partial differential equations with coefficients depending on countably many parameters arise in particular in problems of uncertainty quantification. In this context, they result from expansions of random fields on the computational domain as function series with scalar random coefficients corresponding to the parametric variables. The method constructed and analyzed here yields approximations of the parameter-dependent solutions using adaptive finite elements in the spatial variables combined with a sparse polynomial expansion in the parametric variables.

1.1 Problem statement

On a polygonal domain $D \subset \mathbb{R}^d$, where typically $d \in \{1, 2, 3\}$, we consider the elliptic model problem

$$-\nabla \cdot (a \nabla u) = f \quad \text{on } D, \qquad u = 0 \quad \text{on } \partial D, \tag{1}$$

in weak formulation with $f \in L_2(D)$. Assuming $\mathcal{M}_0$ to be a countable index set with $0 \in \mathcal{M}_0$ and taking $\mathcal{M} = \mathcal{M}_0 \setminus \{0\}$, the parameter-dependent coefficient a is assumed to be given by an affine parameterization

$$a(y) = \theta_0 + \sum_{\mu \in \mathcal{M}} y_\mu \theta_\mu, \quad y = (y_\mu)_{\mu \in \mathcal{M}} \in Y = [-1, 1]^{\mathcal{M}} \tag{2}$$

with $\theta_\mu \in L_\infty(D)$ for $\mu \in \mathcal{M}_0$ and $\operatorname{ess\,inf}_D \theta_0 > 0$. Well-posedness is ensured by the *uniform ellipticity condition* [17]

$$c_B = \operatorname*{ess\,inf}_D \Big\{ \theta_0 - \sum_{\mu \in \mathcal{M}} |\theta_\mu| \Big\} > 0. \tag{3}$$

DOI 10.20347/WIAS.PREPRINT.3112

Note that as a consequence, for $C_B = \sup_{y\in Y} \|a(y)\|_{L_\infty(D)}$ we have

$$C_B \leq 2\|\theta_0\|_{L_\infty} - c_B. \tag{4}$$

To fix a probability distribution of the random coefficients $y \in Y$ in (2), we now introduce a product measure σ on Y. For simplicity, we take σ to be the uniform measure on Y, where $\sigma = \bigotimes_{\mu\in\mathcal{M}} \sigma_1$ with σ_1 the uniform measure on $[-1, 1]$. With $V = H_0^1(D)$, for each given $y \in Y$ let $u(y) \in V$ be defined by

$$\int_D a(y)\nabla u(y) \cdot \nabla v \,\mathrm{d}x = \int_D f v \,\mathrm{d}x \quad \text{for all } v \in V. \tag{5}$$

Then by (3), the mapping $Y \ni y \mapsto u(y) \in V$ can be regarded as an element of the Bochner space

$$\mathcal{V} := L_2(Y, V, \sigma) \simeq V \otimes L_2(Y, \sigma).$$

From the univariate Legendre polynomials $\{L_k\}_{k\in\mathbb{N}}$ that are orthonormal with respect to the uniform measure on $[-1, 1]$, we obtain the orthonormal basis $\{L_\nu\}_{\nu\in\mathcal{F}}$ of product Legendre polynomials for $L_2(Y, \sigma)$, which for $y \in Y$ are given by

$$L_\nu(y) = \prod_{\mu\in\mathcal{M}} L_{\nu_\mu}(y_\mu), \quad \nu \in \mathcal{F} = \{\nu \in \mathbb{N}_0^{\mathcal{M}} \colon \nu_\mu \neq 0 \text{ for finitely many } \mu \in \mathcal{M}\};$$

see, for example, [15, 25]. For $u \in \mathcal{V}$ solving (1) with coefficient (2) in the sense of (5), we thus have the basis expansion

$$u(y) = \sum_{\nu\in\mathcal{F}} u_\nu L_\nu(y), \quad u_\nu = \int_Y u(y)\, L_\nu(y) \,\mathrm{d}\sigma(y) \in V. \tag{6}$$

By restricting the summation over ν in (6) to a finite subset $F \subset \mathcal{F}$, we obtain *semidiscrete* best approximations by elements of $V \otimes \mathrm{span}\{L_\nu\}_{\nu\in F}$. To obtain fully discrete computable approximations, for each ν the coefficient $u_\nu \in V$ needs to be replaced by a further approximation in some finite-dimensional subspace $V_\nu \subset V$. We thus aim to find an approximation of u from a subspace

$$\tilde{\mathcal{V}} = \left\{ \sum_{\nu\in F} v_\nu L_\nu \colon v_\nu \in V_\nu, \nu \in F \right\} \subset \mathcal{V} \tag{7}$$

of total dimension $\sum_{\nu\in F} \dim V_\nu$. In the present work, each V_ν is chosen as a suitable finite element subspace of V.

1.2 Parameter expansions and approximability

Different types of expansion (2) can be used for the parametrized coefficient a. A typical choice are expansions with similar properties as Karhunen-Loève representations of random fields, where the functions θ_μ, $\mu \in \mathcal{M}$, oscillate with increasing frequencies on all of D. In the case $d = 1$ with $D = (0, 1)$, a popular test case of this type is

$$a(y)(x) = 1 + c \sum_{j=1}^{\infty} j^{-\beta} y_j \sin(j\pi x), \quad x \in D = (0, 1),\ y \in Y, \tag{8}$$

with $\beta > 1$ and sufficiently small $c > 0$, corresponding to the choice $\theta_0 = 1$, $\mathcal{M} = \mathbb{N}$, and $\theta_j(x) = cj^{-\beta}\sin(j\pi x)$ for each $j \in \mathcal{M}$. In the limiting case $\beta = 1$, the functions θ_j result from the Karhunen-Loève decomposition of the Brownian bridge on $(0,1)$, and with the random coefficients $y \in Y$ distributed according to σ in (8) one obtains analogous smoothness properties of random draws of $a(y)$.

However, coefficients a with very similar features can also be obtained by different expansions with multilevel structure. A basic example, again on $D = (0,1)$, are hierarchical piecewise linear hat functions: let $\theta(x) = \max\{1 - 2|x - \frac{1}{2}|, 0\}$ and

$$\theta_{\ell,k}(x) = \theta(2^\ell x - k), \quad (\ell,k) \in \mathcal{M} = \big\{(\ell,k)\colon \ell \in \mathbb{N}_0,\ k \in \{0,\dots,2^\ell - 1\}\big\}.$$

Then for any $\alpha > 0$,

$$a(y)(x) = 1 + c \sum_{(\ell,k)\in\mathcal{M}} 2^{-\alpha\ell} y_{\ell,k}\theta_{\ell,k}(x), \quad x \in D = (0,1),\ y \in Y, \tag{9}$$

yields a random field a with very similar features as (8), but expanded in terms of the locally supported functions $\theta_{\ell,}$ with $|\operatorname{supp}\theta_{\ell,k}| = 2^{-\ell}$. The choice $\alpha = \frac{1}{2}$ corresponds to the well-known Lévy-Ciesielski representation of the Brownian bridge [14].

One advantage of the multilevel representation (9) in view of the uniform ellipticity condition (3) is that coefficients with lower smoothness than in (8) can be realized: for any $\alpha > 0$, the series in (9) converges absolutely in $L_\infty(0,1)$ uniformly in $y \in Y$, and thus one can parameterize coefficients with arbitrarily low Hölder regularity.

However, representations with multilevel structure as in (9) are also advantageous for the convergence rates of adaptive sparse approximations of u. One observes that the localization in the hat functions $\theta_{\ell,k}$ in (9) translates to highly localized features in the Legendre coefficients u_ν, depending on the coordinates that are activated in ν. This is illustrated in the case $\alpha = 1$ in Figure 1. As one can recognize, these coefficients have efficient approximations by piecewise linear finite elements on *adaptive* grids. However, these grids clearly need to be chosen with a different local refinement for each ν, so that the corresponding subspaces V_ν in (7) differ accordingly.

Similar results can be obtained for more general domains D in dimensions $d > 1$ when the diffusion coefficient a is given in terms of an expansion with an analogous multilevel structure. The precise assumptions that we use on such multilevel structure for general d are as follows.

Assumption 1. We assume $\theta_\mu \in L_\infty(D)$ for $\mu \in \mathcal{M}_0$ such that the following hold for all $\mu \in \mathcal{M}$:

(i) $\operatorname{diam}\operatorname{supp}\theta_\mu \sim 2^{-|\mu|}$,

(ii) $\#\{\mu \in \mathcal{M}\colon |\mu| = \ell\} \lesssim 2^{d\ell}$ for each $\ell \in \mathbb{N}_0$ and there exists $M > 0$ such that for each μ,

$$\#\{\mu' \in \mathcal{M}\colon |\mu| = |\mu'|,\, \operatorname{supp}\theta_\mu \cap \operatorname{supp}\theta_{\mu'} \neq \emptyset\} \leq M,$$

(iii) for some $\alpha > 0$, one has $\|\theta_\mu\|_{L_\infty(\Omega)} \lesssim 2^{-\alpha|\mu|}$.

As in the above one-dimensional example, representing a in terms of such multilevel basis functions with localized supports leads to favorable sparse approximability of u. As a consequence of [2, Cor. 8.8], for sufficiently regular f and D and expansions of a according to Assumption 1, for $\alpha \in (0,1]$ and with $\tilde{\mathcal{V}}$ as in (7) we have

$$\min_{v\in\tilde{\mathcal{V}}} \|u - v\|_{\mathcal{V}} \leq C\Big(\sum_{\nu\in F} \dim V_\nu\Big)^{-s} \quad \text{for any } s < \begin{cases} \frac{\alpha}{d}, & d \geq 2, \\ \frac{2}{3}\alpha, & d = 1. \end{cases} \tag{10}$$

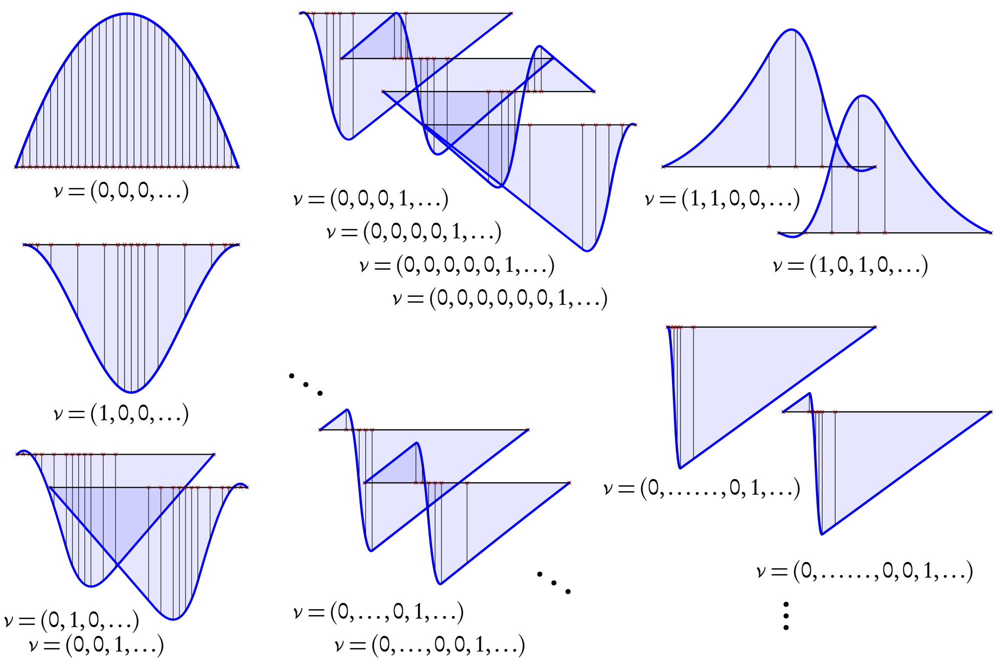

Figure 1: Plots of Legendre coefficients $u_\nu \in H_0^1(0,1)$ (normalized to equal $\|u_\nu\|_{L\infty}$) in the expansion (6) with a given by the hierarchical hat function expansion (9) with $\alpha = 1$, and nodes of adaptively generated piecewise linear approximations on dyadic subintervals.

The numerical tests in [6] indicate that with spatial basis functions of sufficiently high approximation order, this also holds for $\alpha > 1$.

Remarkably, the limiting convergence rate α/d in (10) for $d \geq 2$ is the same as for approximating $u(y)$ by finite elements in V for randomly chosen $y \sim \sigma$ (see [1]), and also the same as the rate with respect to $\#F$ for semidiscrete approximation of u from $V \otimes \operatorname{span}\{L_\nu\}_{\nu\in F}$ obtained in [3]. In other words, when a is given as a multilevel expansion, when using independent adaptive approximations for each Legendre coefficient u_ν, $\nu \in F$, one can achieve sufficiently strong sparsification that combined spatial-parametric approximations converge at the same rate as only spatial or only parametric approximations. In contrast to results obtained using expansions of a without further structure, these results hold even when convergence is limited by the decay rate α in the coefficient expansion, rather than by the spatial approximation rate of the finite elements.

1.3 Stochastic Galerkin discretization

We define $B\colon \mathcal{V} \to \mathcal{V}'$ and $\Phi \in \mathcal{V}'$ by

$$\langle Bv, w\rangle = \int_Y \int_D \Big(\theta_0 + \sum_{\mu\in\mathcal{M}} y_\mu \theta_\mu\Big) \nabla v(y)\cdot\nabla w(y)\,\mathrm{d}x\,\mathrm{d}\sigma(y), \quad \langle \Phi, w\rangle = \int_Y f\big(w(y)\big)\,\mathrm{d}\sigma(y)\,. \tag{11}$$

The operator B induces an inner product $\langle\cdot,\cdot\rangle_B = \langle B\cdot,\cdot\rangle$ with corresponding norm $\|\cdot\|_B$ on $\mathcal{V}$. Note that

$$\sqrt{c_B}\|v\|_{\mathcal{V}} = \sqrt{\inf_{y\in Y} a(y)}\,\|v\|_{\mathcal{V}} \leq \|v\|_B \leq \sqrt{\sup_{y\in Y} a(y)}\,\|v\|_{\mathcal{V}} = \sqrt{C_B}\|v\|_{\mathcal{V}}. \tag{12}$$

DOI 10.20347/WIAS.PREPRINT.3112

As a consequence,

$$\sqrt{c_B}\|v\|_B \le \frac{\langle Bv, v\rangle}{\|v\|_{\mathcal{V}}} \le \|Bv\|_{\mathcal{V}'} = \sup_{w\in\mathcal{V}} \frac{\langle Bv, w\rangle}{\|w\|_{\mathcal{V}}} \le \sup_{w\in\mathcal{V}} \frac{\langle Bv, w\rangle}{C_B^{-1/2}\|w\|_B} = \sqrt{C_B}\|v\|_B. \tag{13}$$

The stochastic variational formulation of (1) with coefficient a given by (2) then reads: find $u \in \mathcal{V}$ such that

$$Bu = \Phi \quad \text{in } \mathcal{V}'. \tag{14}$$

Inserting product Legendre expansions as in (6) of u, v into (14) leads to the *semidiscrete form* of the stochastic Galerkin problem for the coefficient functions u_ν, $\nu \in \mathcal{F}$,

$$\sum_{\mu\in\mathcal{M}_0} \sum_{\nu'\in\mathcal{F}} (\mathbf{M}_\mu)_{\nu,\nu'} A_\mu u_{\nu'} = \delta_{0,\nu} f,\ \nu \in \mathcal{F}, \tag{15}$$

where $A_\mu \colon V \to V'$ are defined, for $\mu \in \mathcal{M}_0$, by

$$\langle A_\mu v, w\rangle = \int_D \theta_\mu \nabla v \cdot \nabla w \,\mathrm{d}x \quad \text{for all } v, w \in V,$$

and the mappings $\mathbf{M}_\mu \colon \ell_2(\mathcal{F}) \to \ell_2(\mathcal{F})$ are given by

$$\mathbf{M}_0 = \left(\int_Y L_\nu(y) L_{\nu'}(y)\,\mathrm{d}\sigma(y)\right)_{\nu,\nu'\in\mathcal{F}}, \qquad \mathbf{M}_\mu = \left(\int_Y y_\mu L_\nu(y) L_{\nu'}(y)\,\mathrm{d}\sigma(y)\right)_{\nu,\nu'\in\mathcal{F}},\ \mu \in \mathcal{M}.$$

Since the $L_2([-1,1],\sigma_1)$-orthonormal Legendre polynomials $\{L_k\}_{k\in\mathbb{N}}$ satisfy the three-term recurrence relation

$$yL_k(y) = \sqrt{\beta_{k+1}} L_{k+1}(y) + \sqrt{\beta_k} L_{k-1}(y), \quad \beta_k = (4 - k^{-2})^{-1},$$

with $L_0 = 1$, $L_{-1} = 0$, $\beta_0 = 0$, we have

$$\mathbf{M}_0 = \left(\delta_{\nu,\nu'}\right)_{\nu,\nu'\in\mathcal{F}}, \qquad \mathbf{M}_\mu = \left(\sqrt{\beta_{\nu_\mu+1}}\,\delta_{\nu+e_\mu,\nu'} + \sqrt{\beta_{\nu_\mu}}\,\delta_{\nu-e_\mu,\nu'}\right)_{\nu,\nu'\in\mathcal{F}},\ \mu \in \mathcal{M},$$

with the Kronecker vectors $e_\mu = (\delta_{\mu,\mu'})_{\mu'\in\mathcal{M}}$.

In the remainder of this work, for simplicity we formulate our method and its analysis for $d = 2$. The results carry over immediately to the cases $d = 1$ and $d > 2$. We assume a fixed conforming simplicial coarsest triangulation $\hat{\mathcal{T}}_0$ of D. If a second (not necessarily conforming) triangulation $\mathcal{T}$ can be generated from $\hat{\mathcal{T}}_0$ by steps of newest vertex bisection, we write $\mathcal{T} \ge \hat{\mathcal{T}}_0$. For such triangulations that are in addition conforming, we consider the standard Lagrange finite element spaces

$$V(\mathcal{T}) = \mathbb{P}_1(\mathcal{T}) \cap V,$$

where $\mathbb{P}_1(\mathcal{T})$ denotes the functions on D that are piecewise affine on each element of $\mathcal{T}$.

Assume a family of triangulations $\mathbb{T} = (\mathcal{T}_\nu)_{\nu\in F}$ with finite $F \subset \mathcal{F}$ and conforming $\mathcal{T}_\nu \ge \hat{\mathcal{T}}_0$ for each $\nu \in F$. We consider stochastic Galerkin discretization subspaces $\mathcal{V}(\mathbb{T})$ given in terms of (7) by

$$\mathcal{V}(\mathbb{T}) = \left\{\sum_{\nu\in F} v_\nu L_\nu \colon v_\nu \in V(\mathcal{T}_\nu),\ \nu \in F\right\} \subset \mathcal{V}. \tag{16}$$

The total number of degrees of freedom for representing each element of $\mathcal{V}(\mathbb{T})$ is then $\dim \mathcal{V}(\mathbb{T}) = \sum_{\nu \in F} \dim V(\mathcal{T}_\nu)$. We use the abbreviation

$$N(\mathbb{T}) = \sum_{\nu \in F} \#\mathcal{T}_\nu, \tag{17}$$

so that $N(\mathbb{T}) \eqsim \dim \mathcal{V}(\mathbb{T})$.

For the finite-dimensional subspaces $\mathcal{V}(\mathbb{T}) \subset \mathcal{V}$, we consider the stochastic Galerkin variational formulation for $u_\mathbb{T} \in \mathcal{V}(\mathbb{T})$,

$$\langle B u_\mathbb{T}, v \rangle = \Phi(v) \quad \text{for all } v \in \mathcal{V}(\mathbb{T}). \tag{18}$$

As a consequence of (3), the bilinear form given by the left hand side of (18) is elliptic and bounded on $\mathcal{V}$, and by Céa's lemma we obtain

$$\|u_\mathbb{T} - u\|_\mathcal{V} \le \frac{2\|\theta_0\|_{L_\infty} - c_B}{c_B} \min_{v \in \mathcal{V}(\mathbb{T})} \|v - u\|_\mathcal{V} .$$

1.4 Adaptive scheme and novel contributions

The objective of this paper is to construct an adaptive stochastic Galerkin finite element method that can make full use of the approximability result (10) by performing independent refinements of meshes for each Legendre coefficient; in other words, for each u_ν we use a potentially completely different finite element subspace $V(\mathcal{T}_\nu)$. Our discretization refinement indicators are obtained from residuals in a stochastic Galerkin formulation of the problem. The basic strategy for doing so follows our earlier results in [6], where we used adaptive wavelet discretizations rather than finite elements for the spatial variable. As there, the first step for obtaining residual approximations in the present work is a semidiscrete adaptive operator compression only in the parametric variables, which identifies the relevant interactions between different Legendre coefficients.

Based on this information, the errors in the individual spatial discretizations of the Legendre coefficients need to be handled. As an initial step towards results on optimal computational costs, as achieved with spatial discretizations by wavelets in [6], in the present work we propose and analyze an adaptive method using standard $\mathbb{P}_1$ finite elements that in each refinement step ensures a fixed error reduction factor, a feature that is also known as *saturation property*. This guarantee is in contrast to previous works using independent finite element subspaces $V(\mathcal{T}_\nu)$ with potentially different meshes $\mathcal{T}_\nu$ for the different coefficients u_ν, such as [7, 18, 19]: in particular, the convergence analysis given in [18] and [7] is based on *assuming* the saturation property to hold for a certain hierarchical error estimation strategy.

One can distinguish two separate difficulties in proving the saturation property for refinements based on the residual in a stochastic Galerkin method. First, in the stochastic variables, we need to identify product polynomial indices that should be added to the approximation subset $F \subset \mathcal{F}$. Strategies that ensure an error reduction have previously been considered, for example, in [20, 21]; these, however, rely on summability of the norms $\|\theta_\mu\|_{L_\infty}$, which for multilevel expansions of random fields is too strong a requirement. To exploit the multilevel structure in the expansion (2) and obtain a scheme that enables linear scaling of computational costs, we instead adapt the approach of [6] based on operator compression in the stochastic variables. This amounts to adaptively dropping summands in the semidiscrete operator representation $\left(\sum_{\mu \in \mathcal{M}_0} \sum_{\nu' \in \mathcal{F}} (\mathbf{M}_\mu)_{\nu,\nu'} A_\mu\right)_{\nu,\nu' \in \mathcal{F}}$ as in (15) that acts on sequences of Legendre coefficients.

The second and more subtle difficulty arises in the refinement of spatial discretizations. In the case of discretizations with a single spatial mesh (that is, where the meshes $\mathcal{T}_\nu$ are identical for all $\nu \in F$), standard residual error estimation techniques can be applied as in [20]. Since we use *independent* meshes for each ν, each spatial component of the residual contains sums of jump functionals corresponding to different meshes, which cannot be treated using standard techniques based on Galerkin orthogonality.

Moreover, with such residual components that are proper functionals in $H^{-1}(D)$ arising from integrals over the edges (or facets) of different meshes, it is not clear *a priori* whether a method using a solve-mark-refine cycle can actually achieve a reduction of the error (or a suitable notion of a quasi-error [12]) by a fixed factor in every iteration: as considered in detail in [16], this may require several refinement steps. In [16], which addresses adaptive finite elements for standard scalar elliptic problems with right-hand sides that are not in $L_2(D)$, an additional inner refinement loop is introduced. However, this approach is not easily generalizable to our setting, since in the present case the functionals in question depend on the approximate solution.

We thus use a different strategy for error estimation and mesh refinement that is based on residual indicators obtained with a BPX finite element frame [11,22,24]. Such a frame can be obtained by taking the union over all standard finite element basis functions associated to a hierarchy of grid refinements. Frames of this type have been used in [23] in the construction of sparse tensor products and in [22] in the analysis of hierarchical error estimation. We approximate H^{-1}-norms via frame coefficients similarly as outlined in [22, Rem. 6.4], where in contrast to the graded quadrilateral meshes assumed in [22], we work with meshes refined via standard newest vertex bisection.

1.5 Outline

We begin by recapitulating basic properties of newest vertex bisection and of finite element frames in Section 2. In Section 3, we describe the adaptive stochastic Galerkin solver and collect the relevant properties of its constituent procedures, which we use in Section 4 to prove convergence of the method with a fixed error reduction in each step. Section 5 is devoted to numerical tests that, in view of the expected convergence rates (10), hint at optimality properties of the method. In Section 6, we summarize our findings and point out several directions for further work.

2 Finite element frames

Our adaptive scheme is driven by error estimates derived from the norm in $\mathcal{V}' = L_2(Y, V', \sigma)$ of the residual for a given approximation $v = \sum_{\nu\in\mathcal{F}} v_\nu L_\nu \in L_2(Y, V, \sigma)$ in the stochastic variational formulation (14),

$$\|Bv - \Phi\|_{\mathcal{V}'} = \left(\sum_{\nu\in\mathcal{F}} \Big\| \sum_{\mu\in\mathcal{M}_0} \sum_{\nu'\in\mathcal{F}} (\mathbf{M}_\mu)_{\nu,\nu'} A_\mu v_{\nu'} - \delta_{0,\nu} f \Big\|_{V'}^2 \right)^{\frac{1}{2}}. \tag{19}$$

Since we are mainly interested in the effect of the parameter-dependent diffusion coefficient, to avoid technicalities, in what follows we assume the right-hand side f to be piecewise polynomial on the initial spatial mesh $\hat{\mathcal{T}}_0$, from which all meshes arising in the spaces (16) are generated by newest vertex bisection. The treatment of more general f is discussed in Remark 10.

Our main task in the approximation of residuals is thus to identify the most relevant indices in the summation over ν in (19), and for each such ν to approximate the V'-norm of linear combinations of terms of the form $A_\mu v_{\nu'}$, where $v_{\nu'} \in V(\mathcal{T}_{\nu'})$ for some conforming $\mathcal{T}_{\nu'} \geq \hat{\mathcal{T}}_0$. For the latter task, we use finite element frames derived from a hierarchy of uniform refinements of $\hat{\mathcal{T}}_0$.

For newest vertex bisection, we assume the initial triangulation $\hat{\mathcal{T}}_0$ to have an edge labelling that is admissible in the sense of [9, Lem. 2.1]; that is, each edge in $\hat{\mathcal{T}}_0$ is labelled either 0 or 1 such that each triangle in $\hat{\mathcal{T}}_0$ has exactly two edges labelled 1 and one edge labelled 0.

Newest vertex bisection is then applied to a triangle with labels $(i, i, i-1)$ by bisecting the edge with the lowest label $i-1$ and assigning the label $i+1$ to both halves of the bisected edge and to the newly added bisecting edge, so that the two newly created triangles both have labels $(i+1, i+1, i)$ and the edges opposite the newly added vertex will be the next to be bisected. The meshes generated by newest vertex bisection are uniform shape regular, dependent only on the initial triangulation $\hat{\mathcal{T}}_0$.

Note that after two applications of newest vertex bisection to a mesh with admissible initial labelling, every edge has been bisected once. This gives rise to a hierarchy of meshes $\hat{\mathcal{T}}_1, \hat{\mathcal{T}}_2, \ldots$, where $\hat{\mathcal{T}}_{j+1}$ is obtained from $\hat{\mathcal{T}}_j$ by applying two passes of newest vertex bisection to the full mesh, such that a new node is added to each edge in $\hat{\mathcal{T}}_j$. For each j, we define ψ_λ with $\lambda = (j,k)$ as the enumeration of piecewise affine hat functions in $V(\hat{\mathcal{T}}_j)$, normalized such that $\|\psi_\lambda\|_V = 1$, where $k = 1, \ldots, N_j = \dim V(\hat{\mathcal{T}}_j)$. We assume that $\hat{\mathcal{T}}_0$ gives rise to a nontrivial finite element space, i.e., $N_0 > 0$, and set

$$\Theta = \big\{(j,k) \colon j = 0, 1, 2, \ldots,\ k = 1, \ldots, N_j\big\}.$$

The family ψ_λ, $\lambda \in \Theta$, of hat functions on all levels of the uniformly refined grid hierarchy (that is, the function system underlying the classical BPX preconditioner [11]) is then a *frame* of V, which means that for $\xi \in V'$, we have a proportionality with uniform constants between $\|(\langle \xi, \psi_\lambda\rangle)_{\lambda\in\Theta}\|_{\ell_2(\Theta)}$ and $\|\xi\|_{V'}$. Under the given assumptions on the mesh hierarchy $(\hat{\mathcal{T}}_j)_{j\geq 0}$, we obtain the result of [23, Thm. 5] in the following form.

Lemma 1. *The family $\Psi = (\psi_\lambda)_{\lambda\in\Theta}$ is a frame of V, that is, there exist $c_\Psi, C_\Psi > 0$ depending only on $\hat{\mathcal{T}}_0$ such that*

$$c_\Psi^2 \|\xi\|_{V'}^2 \leq \sum_{\lambda\in\Theta} |\langle \xi, \psi_\lambda\rangle|^2 \leq C_\Psi^2 \|\xi\|_{V'}^2,\ \xi \in V'. \tag{20}$$

We assume each function θ_μ, $\mu \in \mathcal{M}$, to be piecewise polynomial on a triangulation $\hat{\mathcal{T}}_j$ for some j depending on $|\mu|$ in the following sense.

Assumption 2. There exists $m \in \mathbb{N}_0$, $k \in \mathbb{N}_0$ and $K \in \mathbb{N}$ such that for all $\mu \in \mathcal{M}$, $\theta_\mu \in \mathbb{P}_m(\hat{\mathcal{T}}_{|\mu|+k})$ and $\#\{T \in \mathcal{T}_{|\mu|+k} \colon \operatorname{supp}\theta_\mu \cap T \neq \emptyset\} \leq K$.

The case of given non-polynomial functions θ_ν can be treated by replacing them by polynomial approximations. For example, for functions in wavelet-type expansions of Matérn random fields constructed in [4], it is not difficult to obtain suitable efficient polynomial approximations with uniform accuracy that satisfy Assumption 2, see [4, Eq. (139)].

For approximating dual norms in V' of sums of functionals of the form $A_\mu v_\nu$ with $v_\nu \in V(\mathcal{T}_\nu)$, with $\mathcal{T}$ the joint refinement of $\mathcal{T}_\nu$ and the mesh on which θ_μ is piecewise polynomial, we use integration by parts to obtain

$$\langle A_\mu v_\nu, w\rangle = \int_D \theta_\mu \nabla v_\nu \cdot \nabla w \,\mathrm{d}x = \sum_{K\in\mathcal{T}} \left\{ \int_{\partial K \cap D} \theta_\mu n \cdot \nabla v_\nu \, w \,\mathrm{d}s - \int_K \nabla \cdot (\theta_\mu \nabla v_\nu)\, w \,\mathrm{d}x \right\} \tag{21}$$

for $w \in V$. Here the integrands $\theta_\mu n \cdot \nabla v_\nu$ and $\nabla \cdot (\theta_\mu \nabla v_\nu)$ are polynomials on the respective subdomains by **??** 2. We can thus apply the following result on the decay of ℓ_2-tails of frame coefficients. A related estimate is obtained in [22, Sec. 6] with frames on graded quadrilateral meshes.

We define $\operatorname{level}(E)$ for an edge E and $\operatorname{level}(T)$ for a triangular element T as the unique j such that the uniform mesh $\hat{\mathcal{T}}_j$ contains E or either T or a bisection of T, respectively. Similarly, for the enumeration indices $\lambda = (j,k) \in \Theta$, we define $|\lambda| = j$.

Lemma 2. *Let $d_1, d_2 \in \mathbb{N}_0$. There exist $C > 0$ and $J_0 \in \mathbb{N}$ depending only on $\hat{\mathcal{T}}_0$ and d_1, d_2 such that for all $J \geq J_0$ the following holds: For any $\mathcal{T} \geq \hat{\mathcal{T}}_0$ with interior edges $\mathcal{E}$ and any $\xi \in V'$ of the form*

$$\langle \xi, v \rangle = \sum_{K \in \mathcal{T}} \int_K \xi_K v \,\mathrm{d}x + \sum_{E \in \mathcal{E}} \int_E \xi_E v \,\mathrm{d}s, \quad v \in V, \tag{22}$$

where $\xi_K \in \mathbb{P}_{d_1}(K)$, $K \in \mathcal{T}$, and $\xi_E \in \mathbb{P}_{d_2}(E)$, $E \in \mathcal{E}$, we have

$$\Big(\sum_{\lambda \in \Theta_J(\mathcal{T})} \big| \langle \xi, \psi_\lambda \rangle \big|^2 \Big)^{\frac{1}{2}} \leq C 2^{-J} \Big(\sum_{\lambda \in \Theta} \big| \langle \xi, \psi_\lambda \rangle \big|^2 \Big)^{\frac{1}{2}}$$

with

$$\begin{aligned} \Theta_J(\mathcal{T}) = \big\{ \lambda \in \Theta \colon & (\forall K \in \mathcal{T} \colon \operatorname{meas}_2(\operatorname{supp} \psi_\lambda \cap K) > 0 \implies |\lambda| > \operatorname{level}(K) + J) \\ & \wedge (\forall E \in \mathcal{E} \colon \operatorname{meas}_1(\operatorname{supp} \psi_\lambda \cap E) > 0 \implies |\lambda| > \operatorname{level}(E) + 2J) \big\}. \end{aligned} \tag{23}$$

Proof. For $\lambda \in \Theta_J(\mathcal{T})$ we have, by uniform shape regularity and the definition of $\Theta_J(\mathcal{T})$, that ψ_λ have support on a uniformly bounded number of elements $K \in \mathcal{T}$. Thus

$$\Big| \sum_{K \in \mathcal{T}} \big| \langle \xi_K, \psi_\lambda \rangle \big| \Big|^2 \lesssim \sum_{K \in \mathcal{T}} \big| \langle \xi_K, \psi_\lambda \rangle \big|^2, \qquad \Big| \sum_{E \in \mathcal{E}} \big| \langle \xi_E, \psi_\lambda \rangle \big| \Big|^2 \lesssim \sum_{E \in \mathcal{E}} \big| \langle \xi_E, \psi_\lambda \rangle \big|^2 \tag{24}$$

with uniform constants. Let

$$h_K = 2^{-\operatorname{level}(K)}, \quad h_E = 2^{-\operatorname{level}(E)}.$$

By (24) and using that by the normalization of ψ_λ, with $j = |\lambda|$ we have that $\|\psi_\lambda\|_{L^2(K)} \lesssim 2^{-j}$ and $\|\psi_\lambda\|_{L^2(E)} \lesssim 2^{-\frac{1}{2}j}$. Thus

$$\begin{aligned} \sum_{\lambda \in \Theta_J(\mathcal{T})} \big| \langle \xi, \psi_\lambda \rangle \big|^2 &\lesssim \sum_{K \in \mathcal{T}} \sum_{\lambda \in \Theta_J(\mathcal{T})} \big| \langle \xi_K, \psi_\lambda \rangle \big|^2 + \sum_{E \in \mathcal{E}} \sum_{\lambda \in \Theta_J(\mathcal{T})} \big| \langle \xi_E, \psi_\lambda \rangle \big|^2 \\ &\lesssim \sum_{K \in \mathcal{T}} \sum_{j > \operatorname{level}(K) + J} 2^{-2j} \|\xi_K\|^2_{L_2(K)} + \sum_{E \in \mathcal{E}} \sum_{j > \operatorname{level}(E) + 2J} 2^{-j} \|\xi_E\|^2_{L_2(E)} \\ &\lesssim 2^{-2J} \sum_{K \in \mathcal{T}} h_K^2 \|\xi_K\|^2_{L_2(K)} + 2^{-2J} \sum_{E \in \mathcal{E}} h_E \|\xi_E\|^2_{L_2(E)} . \end{aligned}$$

Since for all $K \in \mathcal{T}$ and $E \in \mathcal{E}$, the components ξ_K and ξ_E are polynomial with degrees bounded by d_1 and d_2, respectively, [27, Thm. 3.59] yields

$$\sum_{K \in \mathcal{T}} h_K^2 \|\xi_K\|^2_{L_2(K)} + \sum_{E \in \mathcal{E}} h_E \|\xi_E\|^2_{L_2(E)} \lesssim \|\xi\|^2_{V'}$$

with a constant depending only on the initial triangulation $\hat{\mathcal{T}}_0$ and on $\max\{d_1, d_2\}$. By (20),

$$\|\xi\|^2_{V'} \lesssim \sum_{\lambda \in \Theta} \big| \langle \xi, \psi_\lambda \rangle \big|^2$$

with a further constant depending only on $\hat{\mathcal{T}}_0$. This concludes the proof. □

We use Lemma 2 to estimate the V'-norms in (19): we first determine the relevant indices $\nu \in \mathcal{F}$ in the summation on the right-hand side of (19). This is done by an operator compression in the parametric degrees of freedom that is independent of the spatial meshes. For each of these ν, we evaluate the frame coefficients

$$\mathbf{r}_\nu = (\mathbf{r}_{\nu,\lambda})_{\lambda\in\Theta_\nu}, \quad \mathbf{r}_{\nu,\lambda} = \Big\langle \sum_{\nu'\in\mathcal{F}} \sum_{\mu\in\mathcal{M}_0(\nu,\nu')} (\mathbf{M}_\mu)_{\nu,\nu'} A_\mu v_{\nu'}, \psi_\lambda \Big\rangle$$

for suitable finite subsets $\mathcal{M}_0(\nu,\nu') \subset \mathcal{M}_0$ and for $\lambda \in \Theta_\nu \subset \Theta$. Here Θ_ν is chosen according to Lemma 2 such that

$$\sum_{\lambda\notin\Theta_\nu} \mathbf{r}_{\nu,\lambda}^2 \le \zeta^2 \sum_{\lambda\in\Theta} \mathbf{r}_{\nu,\lambda}^2$$

with a sufficiently small relative error $\zeta \in (0,1)$.

The indices (ν,λ) corresponding to the largest values $|\mathbf{r}_{\nu,\lambda}|$ are selected for refinement, based on a condition analogous to the Dörfler criterion, by a tree thresholding procedure. Subsequently, for each ν, the associated current mesh $\mathcal{T}_\nu$ is refined such that all selected frame elements are contained in the resulting finite element space. A similar strategy has been outlined for the refinement of a single finite element mesh in [22]. It is especially in our setting, with interactions between many different meshes in elliptic systems of PDEs, that this technique provides crucial flexbility compared to standard approaches of a posteriori error estimation.

3 Adaptive solver

In this section, we describe the adaptive solver and analyze its constituent procedures. As noted above, the adaptive scheme is based on Galerkin discretizations on successively refined meshes. Here the way in which the underlying residual error indicators are obtained and used in the mesh refinement differs from previous approaches to adaptive stochastic Galerkin finite element methods. We combine adaptive operator compression in the stochastic degrees of freedom with frame-based spatial refinement indicators. These are subsequently used in a tree-based selection of refinements that in each iteration of the adaptive scheme permits the application of multiple refinement steps within single mesh elements; this latter property is crucial in ensuring error reduction by a uniform factor.

3.1 Residual estimation

As a first first step in the adaptive scheme, we require a semidiscrete adaptive compression of the operator $B\colon \mathcal{V} \to \mathcal{V}'$. For $\ell \in \mathbb{N}_0$ we define the truncated operators B_ℓ by

$$\langle B_\ell v, w\rangle = \int_Y \int_D \Big(\theta_0 + \sum_{\substack{\mu\in\mathcal{M}\\|\mu|<\ell}} y_\mu \theta_\mu\Big) \nabla v(y)\cdot\nabla w(y)\,\mathrm{d}x\,\mathrm{d}\sigma(y) \quad \text{for all } v,w\in\mathcal{V}. \tag{25}$$

We make use of two particular consequences of Assumption 1: First, for the expansion functions $(\theta_\mu)_{\mu\in\mathcal{M}}$ we have that there exists $C_1 > 0$ such that for all $\ell \ge 0$,

$$\#\{\mu : |\mu| = \ell\} \le C_1 2^{d\ell}. \tag{26}$$

Second, there exists $C_2 > 0$ such that with the $\alpha > 0$ in Assumption 1(iii), for all $\ell \geq 0$,

$$\sum_{|\mu|=\ell} |\theta_\mu| \leq C_2 2^{-\alpha\ell} \quad \text{a.e. in } D. \tag{27}$$

As in [6, Prop. 3.2], we have the following approximation result.

Proposition 3. *With $C_2 > 0$ as in* (27)*, for all $\ell \geq 0$,*

$$\|B - B_\ell\|_{\mathcal{V}\to\mathcal{V}'} = \Big\| \sum_{|\mu|\geq\ell} \mathbf{M}_\mu \otimes A_\mu \Big\|_{\ell_2(\mathcal{F})\otimes V \to \ell_2(\mathcal{F})\otimes V'} \leq \frac{C_2}{1-2^{-\alpha}} 2^{-\alpha\ell}.$$

We introduce the following notation: for $\nu \in \mathcal{F}$ and $v \in \mathcal{V}$,

$$[v]_\nu = \int_Y v L_\nu(y) \, \mathrm{d}\sigma(y),$$

and similarly, for $\xi \in \mathcal{V}' = L_2(Y, V', \sigma)$,

$$[\xi]_\nu = \int_Y \xi L_\nu(y) \, \mathrm{d}\sigma(y)\,, \tag{28}$$

so that

$$\langle \xi, v\rangle_{\mathcal{V}',\mathcal{V}} = \sum_{\nu\in\mathcal{F}} \langle [\xi]_\nu, [v]_\nu \rangle_{V',V}.$$

Note that $\|v\|_{\mathcal{V}} = \|(\|[v]_\nu\|_V)_{\nu\in\mathcal{F}}\|_{\ell_2}$ and thus also $\|\xi\|_{\mathcal{V}'} = \|(\|[\xi]_\nu\|_{V'})_{\nu\in\mathcal{F}}\|_{\ell_2}$.

For sequences $\mathbf{v} \in \ell_2(\mathcal{F})$ and $s > 0$, we define the standard approximation spaces $\mathcal{A}^s = \mathcal{A}^s(\mathcal{F})$ with quasi-norm

$$\|\mathbf{v}\|_{\mathcal{A}^s(\mathcal{F})} = \sup_{N\in\mathbb{N}_0} (1+N)^s \min\{\|\mathbf{v}-\mathbf{w}\|_{\ell_2(\mathcal{F})} \colon \#\operatorname{supp}\mathbf{w} \leq N\}. \tag{29}$$

Thus for $v \in L_2(Y, V, \sigma)$ with $\big\| \big(\|[v]_\nu\|_V\big)_{\nu\in\mathcal{F}} \big\|_{\mathcal{A}^s} < \infty$, for each $N \in \mathbb{N}$ there exists $F_N \subset \mathcal{F}$ with $\#F_N \leq N$ such that

$$\Big\| v - \sum_{\nu\in F_N} [v]_\nu L_\nu \Big\|_{L_2(Y,V,\sigma)} \leq (N+1)^{-s} \big\| \big(\|[v]_\nu\|_V\big)_{\nu\in\mathcal{F}} \big\|_{\mathcal{A}^s}.$$

Similarly as in [6], we construct a routine Apply in Algorithm 3.1 taking in a tolerance $\eta > 0$ and $v \in \mathcal{V}$ with finite stochastic support $\operatorname{supp}([v]_\nu)_{\nu\in\mathcal{F}} < \infty$ that produces a blockwise operator compression, adapted to v and encoded by subsets of $\mathcal{M}_0 \times \operatorname{supp}([v]_\nu)_{\nu\in\mathcal{F}}$. These subsets of indices specify which truncation of B as in (25) should be applied to which subset of Legendre coefficients of v. The following result is obtained by minor modifications of the proof of [6, Prop. 4.8].

Proposition 4. *Let $s > 0$ with $s < \frac{\alpha}{d}$, let B be as in* (11)*, let v satisfy $\#\operatorname{supp}([v]_\nu)_{\nu\in\mathcal{F}} < \infty$, and let ℓ_i, $i = 0, \ldots, I$, and g be as defined in Algorithm 3.1. Then $\|Bv - g\|_{\mathcal{V}'} \leq \eta$, for $F = \operatorname{supp}([g]_\nu)_{\nu\in\mathcal{F}}$ we have*

$$\#F \leq \sum_{\nu\in F} \#M(\nu) \lesssim \sum_{j=0}^{J} 2^{d\ell_j} \#F_j \lesssim \eta^{-\frac{1}{s}} \big\| \big(\|[v]_\nu\|_V\big)_{\nu\in\mathcal{F}} \big\|_{\mathcal{A}^s}^{\frac{1}{s}}, \tag{30}$$

and

$$\max_{i=0,\ldots,I} \ell_i \lesssim 1 + |\log\eta| + \log \big\| \big(\|[v]_\nu\|_V\big)_{\nu\in\mathcal{F}} \big\|_{\mathcal{A}^s}. \tag{31}$$

The constants in the inequalities depend on C as in Algorithm 3.1, C_B, d, α, s, and on C_1 from (26)*.*

Algorithm 3.1 $(M(\nu))_{\nu\in F}, (F_i,\ell_i)_{i=0}^I = \textsc{Apply}(v;\eta)$, for $N := \#\operatorname{supp}([v]_\nu)_{\nu\in\mathcal{F}} < \infty$, $\eta > 0$.

(i) If $\|B\|_{\mathcal{V}\to\mathcal{V}'}\|v\|_{\mathcal{V}} \le \eta$, return the empty tuple with $F = \emptyset$; otherwise, with $\bar{I} := \lceil \log_2 N\rceil$, for $i = 0,\dots,\bar{I}$, determine $F_i \subset \mathcal{F}$ such that $\#F_i \le 2^i$ and $P_{F_i}v = \sum_{\nu\in F_i}[v]_\nu L_\nu$ satisfies

$$\|v - P_{F_i}v\|_{\mathcal{V}} \le C \min_{\#\tilde{F}\le 2^i}\|v - P_{\tilde{F}}v\|_{\mathcal{V}}$$

with an absolute constant $C > 0$. Choose I as the minimal integer such that

$$\delta = \|B\|_{\mathcal{V}\to\mathcal{V}'}\|v - P_{F_I}v\|_{\mathcal{V}} \le \frac{\eta}{2}.$$

(ii) With $d_0 = P_{F_0}v$, $d_i = (P_{F_i} - P_{F_{i-1}})v$, $i = 1,\dots,\bar{I}$, and $N_i = \#F_i$, set

$$\ell_i = \left\lceil \alpha^{-1}\log_2\left(\frac{C_B}{\eta-\delta}\left(\frac{\|d_i\|_{\mathcal{V}}}{N_i}\right)^{\frac{\alpha}{\alpha+d}}\left(\sum_{k=0}^{I}\|d_k\|_{\mathcal{V}}^{\frac{d}{\alpha+d}}N_k^{\frac{\alpha}{\alpha+d}}\right)\right)\right\rceil, \quad i = 0,\dots,I.$$

(iii) With g defined by

$$g = \sum_{i=0}^{I} B_{\ell_i}d_i,$$

for each $\nu\in F = \operatorname{supp}([g]_\nu)_{\nu\in\mathcal{F}}$, collect the sets $M(\nu)\subset\mathcal{M}_0\times\operatorname{supp}([v]_\nu)_{\nu\in\mathcal{F}}$ of minimal size such that

$$[g]_\nu = \sum_{(\mu,\nu')\in M(\nu)}(\mathbf{M}_\mu)_{\nu,\nu'}\,A_\mu\,v_{\nu'}, \quad \nu\in F,$$

and return $(M(\nu))_{\nu\in F}$ as well as $(F_i,\ell_i)_{i=0}^I$.

As a next step, for g and F as defined in Algorithm 3.1, for each $\nu\in F$ we need to evaluate

$$[g]_\nu = \sum_{(\mu,\nu')\in M(\nu)}(\mathbf{M}_\mu)_{\nu,\nu'}\,A_\mu\,v_{\nu'} \,\in\, V'. \tag{32}$$

We represent each $[g]_\nu$ as a piecewise polynomial on a (not necessarily conforming) triangulation as in Theorem 2. To this end, we first compute each summand

$$[g]_{\nu,\nu'} = (\mathbf{M}_\mu)_{\nu,\nu'}\,A_\mu\,v_{\nu'}$$

in (32). Note that the index μ in the definition $[g]_{\nu,\nu'}$ is the unique index such that $\nu' = \nu\pm e_\mu$. For fixed ν', we can compute all $[g]_{\nu,\nu'}$ for $\nu\in F$ by traversing the mesh of $v_{\nu'}$ once. Let $\mathcal{K}_{\nu'}$ denote the elements in the mesh of $v_{\nu'}$ and let $\ell_{\nu'} = \max_{\nu'\in F_i}\ell_i$. **??** 1 (ii) guarantees that each element $K\in\mathcal{K}_{\nu'}$ is required for at most $C\ell_{\nu'}(2^{d\ell_{\nu'}-\mathrm{level}(K)})$ summands, with a uniform constant C. Hence, computing all $[g]_{\nu,\nu'}$ for $\nu\in F$ has a complexity of order $O(\ell_{\nu'}\#\mathcal{K}_{\nu'} + 2^{d\ell_{\nu'}})$. However, for fixed ν, the summands $[g]_{\nu,\nu'}$ are on different meshes.

To efficiently evaluate the sum, we use the natural tree structure on the triangles that are generated by newest vertex bisection, that is, the children of a triangle are the two triangles generated by bisection. The routine $\textsc{Sum}$ in Algorithm 3.2 yields a representation of $[g]_\nu$ on a joint grid, with a computational complexity that is linear in the sum of the sizes of the triangulations of $[g]_{\nu,\nu'}$.

Algorithm 3.2 $\bar{\xi} = \textsc{Sum}((\xi_\nu)_{\nu\in F})$ with finite $F \subset \mathcal{F}$ and ξ_ν is of the form (22) on a mesh $\mathcal{T}_\nu$ with edges $\mathcal{E}_\nu$ for each $\nu \in F$.

Let $\xi_{\nu,K}$ and $\xi_{\nu,E}$ be the polynomials of this representation for each $K \in \mathcal{T}_\nu$ and $E \in \mathcal{E}_\nu$, respectively. For each triangle K let $E(K)$ denote the edge that is bisected by newest vertex bisection of K, and K_i and $E_i(K)$ for $i = 1, 2$ be the corresponding bisected elements and edges.

(i) Initialize $\bar{\xi}$: Let $\bar{\xi}_K = 0$ and $\bar{\xi}_E = 0$ for every element and every edge;

(ii) For every K and E in the elements and edges of $\mathcal{T}_\nu$ and every ν, add the respective polynomial to $\bar{\xi}$:

$$\bar{\xi}_K \leftarrow \bar{\xi}_K + \xi_{\nu,K}; \quad \bar{\xi}_E \leftarrow \bar{\xi}_E + \xi_{\nu,E}.$$

(iii) Set $\mathfrak{T} = \emptyset$ and for each K such that $\bar{\xi}_K \neq 0$ or $\bar{\xi}_{E(K)} \neq 0$, insert all ancestors of K into $\mathfrak{T}$, so that $\mathfrak{T}$ becomes a tree with root elements $\mathfrak{R} \subseteq \hat{\mathcal{T}}_0$;

(iv) While $\mathfrak{T}$ is not empty:

For all K in $\mathfrak{R}$ add polynomials to the corresponding children in $\mathfrak{T}$:

$$\begin{aligned} \bar{\xi}_{K_i} &\leftarrow \bar{\xi}_{K_i} + \bar{\xi}_K \text{ for } i = 1, 2; & \bar{\xi}_K &\leftarrow 0; \\ \bar{\xi}_{E_i(K)} &\leftarrow \bar{\xi}_{E_i(K)} + \bar{\xi}_{E(K)} \text{ for } i = 1, 2; & \bar{\xi}_{E(K)} &\leftarrow 0; \end{aligned}$$

remove K from $\mathfrak{T}$ and $\mathfrak{R}$ and insert K_1, K_2 into $\mathfrak{R}$;

(v) Return $\bar{\xi}$.

Remark 5. The step (ii) in Algorithm 3.2 can be executed while computing $[g]_{\nu,\nu'}$ without explicitly storing $[g]_{\nu,\nu'}$.

3.1.1 Approximate dual norm evaluation

By Theorem 1, we have the equivalent expression for the dual norm of the approximate residual

$$\|w\|_{\mathcal{V}'}^2 = \|(\|[g]_\nu\|_{V'})_{\nu\in\mathcal{F}}\|_{\ell_2}^2 \eqsim \sum_{\nu\in F}\sum_{\lambda\in\Theta} |\langle [g]_\nu, \psi_\lambda\rangle|^2.$$

With the help of Theorem 2 we can estimate the latter expression by evaluating the coefficients $\langle [g]_\nu, \psi_\lambda\rangle$ for all $\lambda \in \Theta \setminus \Theta_J(\mathcal{T})$ for some J. We now ensure that the costs for computing these coefficients are linear in the size of $\Theta \setminus \Theta_J(\mathcal{T})$. For any frame element ψ_λ, we define $\mathcal{K}(\psi_\lambda)$ as the supporting triangles, $\mathcal{E}(\psi_\lambda)$ as the corresponding interior edges, and

$$\mathcal{R}(\psi_\lambda) = \{\psi_\mu \colon |\mu| = |\lambda| + 1 \text{ and } \operatorname{supp}(\psi_\mu) \subset \operatorname{supp}(\psi_\lambda)\}$$

as the set of frame elements in terms of which ψ_λ can be represented on the next higher level of refinement, that is, there are coefficients $h_{\lambda,mu}$ such that $\psi_\lambda = \sum_{\psi_\mu\in\mathcal{R}(\psi_\lambda)} h_{\lambda,\mu}\psi_\mu$.

Lemma 6. *Let $J \in \mathbb{N}$ and let $\mathcal{T} \geq \hat{\mathcal{T}}_0$ be a triangulation with edges $\mathcal{E}$. Then for $\Theta_J(\mathcal{T})$ as in* (23)*, we have $\#(\Theta \setminus \Theta_J(\mathcal{T})) \lesssim 4^J \#\mathcal{T}$.*

Proof. By construction of ψ_λ, for $K \in \mathcal{T}$ we have

$$\#\big\{\lambda \in \Theta\colon\ \mathrm{meas}_2(\operatorname{supp}\psi_\lambda \cap K) > 0 \text{ and } |\lambda| \le \mathrm{level}(K) + J\big\} \lesssim 4^J,$$

and for any edge $E \in \mathcal{E}$,

$$\#\big\{\lambda \in \Theta\colon\ \mathrm{meas}_1(\operatorname{supp}\psi_\lambda \cap E) > 0 \text{ and } |\lambda| \le \mathrm{level}(E) + 2J\big\} \lesssim 4^J.$$

Since

$$\begin{aligned}\Theta \setminus \Theta_J(\mathcal{T}) = \big\{\lambda \in \Theta\colon (\exists K \in \mathcal{T}\colon\ &\mathrm{meas}_2(\operatorname{supp}\psi_\lambda \cap K) > 0 \wedge |\lambda| \le \mathrm{level}(K) + J)\\ &\vee (\exists E \in \mathcal{E}\colon\ \mathrm{meas}_1(\operatorname{supp}\psi_\lambda \cap E) > 0 \wedge |\lambda| \le \mathrm{level}(E) + 2J)\big\},\end{aligned}$$

we have $\#\big(\Theta \setminus \Theta_J(\mathcal{T})\big) \lesssim 4^J \#\mathcal{T}$. □

Lemma 7. *Let $J \in \mathbb{N}$. For any not necessarily conforming triangulation $\mathcal{T}$ with interior edges $\mathcal{E}$ and any $\xi \in V'$ of the form* (22)*, where $\xi_K \in \mathbb{P}_{d_1}(K)$, $K \in \mathcal{T}$, and $\xi_E \in \mathbb{P}_{d_2}(E)$, $E \in \mathcal{E}$, the evaluation of the coefficients $\langle \xi, \psi_\lambda\rangle$ for all $\lambda \in \Theta \setminus \Theta_J(\mathcal{T})$ requires at most $C\#(\Theta \setminus \Theta_J(\mathcal{T}))$ basic operations, where C depends only on d_1, d_2 and on $\max_{\lambda\in\Theta} \#\mathcal{K}(\psi_\lambda)$, $\max_{\lambda\in\Theta} \#\mathcal{E}(\psi_\lambda)$, and $\max_{\lambda\in\Theta} \#\mathcal{R}(\psi_\lambda)$.*

Proof. First, if ξ is polynomial on $K \in \mathcal{K}(\psi_\lambda)$ and $E \in \mathcal{E}(\psi_\lambda)$, then $\langle \xi, \psi_\lambda\rangle$ can be evaluated in $c\,(\#\mathcal{K}(\psi_\lambda) + \#\mathcal{E}(\psi_\lambda))$ basic operations using quadrature, where c depends only on the polynomial degrees d_1 and d_2. Otherwise, we evaluate

$$\langle \xi, \psi_\lambda\rangle = \sum_{\psi_\mu \in \mathcal{R}(\psi_\lambda)} h_{\lambda,\mu} \langle \xi, \psi_\mu\rangle,$$

which requires $\#\mathcal{R}(\psi_\lambda)$ operations. It remains to estimate the size of

$$\Theta_\xi = \{\lambda' \in \Theta : \psi_{\lambda'} \in \mathcal{R}(\psi_\lambda) \text{ for } \lambda \in \Theta \text{ with } \xi \text{ not polynomial on } K \in \mathcal{K}(\psi_\lambda) \text{ or } E \in \mathcal{E}(\psi_\lambda)\}.$$

Note that if ξ is not polynomial on all $K \in \mathcal{K}(\psi_\lambda)$ and $E \in \mathcal{E}(\psi_\lambda)$, then $\psi_\lambda \notin \Theta_0(\mathcal{T})$ by definition. Hence,

$$\Theta_\xi \subseteq \{\lambda' \in \Theta : \psi_{\lambda'} \in \mathcal{R}(\psi_\lambda) \text{ for some } \lambda \in \Theta \setminus \Theta_0(\mathcal{T})\}$$

and $\#\Theta_\xi \le \#(\Theta \setminus \Theta_{\xi,0}) \max_{\lambda\in\Theta} \#\mathcal{R}(\psi_\lambda) \le \#(\Theta \setminus \Theta_J(\mathcal{T})) \max_{\lambda\in\Theta} \#\mathcal{R}(\psi_\lambda)$. The number of basic operations thus does not exceed

$$c\,(\#\mathcal{K}(\psi_\lambda) + \#\mathcal{E}(\psi_\lambda) + \max_{\lambda\in\Theta} \#\mathcal{R}(\psi_\lambda))\ (\#\Theta_\xi + \#\Theta \setminus \Theta_J(\mathcal{T})) \le C\#\Theta \setminus \Theta_J(\mathcal{T}). \qquad \square$$

A practical method implementing this result is given with the method DUALNORMINDICATORS in Algorithm 3.3. Finally, this and the previous algorithms are used to estimate the residual similarly as in [6] with Algorithm 3.4 for prescribed tolerances of the approximations.

Remark 8*.* For computational purposes one can avoid explicit treatment of edge terms: instead of applying integration by parts as in (21) to each term of the form $A_\mu v_\nu$, one can also use for each given ψ_λ that

$$\langle A_\mu v_\nu, \psi_\lambda\rangle = \int_D q_{\mu,\nu} \cdot \nabla\psi_\lambda \,\mathrm{d}x$$

with a piecewise polynomial vector field $q_{\mu,\nu}$. One can thus apply the same algorithmic considerations to the components of these vector fields on the triangles and then form inner products with the gradients of the frame elements.

Algorithm 3.3 $(\Theta^+, (\langle\xi,\psi_\lambda\rangle)_{\lambda\in\Theta^+}) = \textsc{DualNormIndicators}(\xi, J)$ with ξ as in (22), $J \in \mathbb{N}$.

Let $\mathcal{T}$ and $\mathcal{E}$ be as in Theorem 2, with $\mathcal{T}$ the smallest triangulation such that (22) holds for ξ. Initialize $L = 0$ and $\tilde\Theta_L = \{\lambda\colon |\lambda| = 0\}$.

(i) Set $L \leftarrow L + 1$ and

$$\tilde\Theta_L = \Big\{\lambda'\colon \psi_{\lambda'} \in \mathcal{R}(\psi_\lambda) \text{ for } \lambda \in \tilde\Theta_{L-1} \wedge \Big(\big(\xi \text{ not polynomial on } K \text{ or } E \text{ for some } K \in \mathcal{K}(\psi_\lambda), E \in \mathcal{E}(\psi_\lambda)\big) \vee \big(\operatorname{supp}\psi_{\lambda'} \cap K' \neq \emptyset \text{ and } L \le \operatorname{level}(K') + J \text{ for some } K' \in \mathcal{K}\big) \vee \big(\operatorname{supp}\psi_{\lambda'} \cap E' \neq \emptyset \text{ and } L \le \operatorname{level}(E') + 2J \text{ for some } E' \in \mathcal{E}\big)\Big)\Big\};$$

(ii) If $\tilde\Theta_L \neq \emptyset$ go to **(i)**, else go to **(iii)**;

(iii) For $j = L-1, L-2, \ldots, 0$: For all $\lambda \in \tilde\Theta_j$ evaluate $\langle\xi,\psi_\lambda\rangle$ via

$$\langle\xi,\psi_\lambda\rangle = \sum_{\psi_{\lambda'}\in\mathcal{R}(\psi_\lambda)} r_{\lambda,\lambda'}\langle\xi,\psi_{\lambda'}\rangle \quad \text{if } \mathcal{R}(\psi_\lambda) \subset \tilde\Theta_{j+1}, \qquad \text{or by quadrature otherwise;}$$

(iv) Set $\Theta^+ = \bigcup_{j=0}^{L-1} \tilde\Theta_j$ and return $\big(\Theta^+,\ \big(\langle\xi,\psi_\lambda\rangle\big)_{\lambda\in\Theta^+}\big)$.

Note that with sequences $(\mathbf{r}_\nu)_{\nu\in\mathcal{F}}$ such that $\mathbf{r}_\nu \in \ell_2(\Theta)$ for $\nu \in \mathcal{F}$ and $\sum_{\nu\in\mathcal{F}}\|\mathbf{r}_\nu\|_{\ell_2}^2 < \infty$, we associate $\mathbf{r} = (\mathbf{r}_{\nu,\lambda})_{\nu\in\mathcal{F},\lambda\in\Theta} \in \ell_2(\mathcal{F}\times\Theta)$. We write $\|\mathbf{r}\| = \|\mathbf{r}\|_{\ell_2}$, so that in particular

$$\|\mathbf{r}\| = \big\|\big(\|\mathbf{r}_\nu\|_{\ell_2(\Theta)}\big)_{\nu\in\mathcal{F}}\big\|_{\ell_2(\mathcal{F})}.$$

Proposition 9. *Let $((\Theta_\nu^+)_{\nu\in F^+}, (\hat{\mathbf{r}}_\nu)_{\nu\in F^+}, \eta, b)$ be the return values of Algorithm 3.4 and let*

$$\Lambda^+ = \big\{(\nu,\lambda) \in \mathcal{F}\times\Theta\colon \nu \in F^+, \lambda \in \Theta_\nu^+\big\}\,. \tag{33}$$

Set $\hat{\mathbf{r}}_\nu = 0$ for $\nu \notin F^+$. Furthermore, let $\mathbf{z} = (\langle[Bv]_\nu,\psi_\lambda\rangle)_{\nu\in\mathcal{F},\lambda\in\Theta}$ and $\mathbf{f} = (\langle[f]_\nu,\psi_\lambda\rangle)_{\nu\in\mathcal{F},\lambda\in\Theta}$. Then $\|\mathbf{z}-\mathbf{f}\| \le b$ and either $b \le \varepsilon$, or $\hat{\mathbf{r}}$ satisfies

$$\|\hat{\mathbf{r}} - (\mathbf{f}-\mathbf{z})\| \le \zeta\|\mathbf{f}-\mathbf{z}\|, \tag{34}$$

where we have $\#\operatorname{supp}\hat{\mathbf{r}} \le \#\Lambda^+ = \sum_{\nu\in F^+}\#\Theta_\nu^+$ and

$$\#\Lambda^+ \lesssim \#\mathcal{T}(f) + (1 + |\log\eta| + \log\|(\|[v]_\nu\|)_{\nu\in\mathcal{F}}\|_{\mathcal{A}^s})^2 N(\mathbb{T}) + (1 + |\log\eta| + \log\|(\|[v]_\nu\|)_{\nu\in\mathcal{F}}\|_{\mathcal{A}^s})\eta^{-\frac{1}{s}}\|(\|[v]_\nu\|)_{\nu\in\mathcal{F}}\|_{\mathcal{A}^s}^{\frac{1}{s}}\,. \tag{35}$$

The number of operations in Algorithm 3.4 is bounded by a fixed multiple of

$$(1 + \log_2(\eta_0/\eta))\Big(\#F\log\#F + \#\mathcal{T}(f) + (1 + |\log\eta| + \log\|(\|[v]_\nu\|)_{\nu\in\mathcal{F}}\|_{\mathcal{A}^s})^2 N(\mathbb{T}) + (1 + |\log\eta| + \log\|(\|[v]_\nu\|)_{\nu\in\mathcal{F}}\|_{\mathcal{A}^s})\eta^{-\frac{1}{s}}\|(\|[v]_\nu\|)_{\nu\in\mathcal{F}}\|_{\mathcal{A}^s}^{\frac{1}{s}}\Big). \tag{36}$$

Algorithm 3.4 $((\Theta^+_\nu)_{\nu\in F^+}, (\hat{\mathbf{r}}_\nu)_{\nu\in F^+}, ([r]_\nu)_{\nu\in F^+}, \eta, b) = \textsc{ResEstimate}(v;\zeta,\eta_0,\varepsilon)$, for $\#\operatorname{supp}([v]_\nu)_\nu < \infty$, relative tolerance $\zeta > 0$, initial tolerance η_0, target tolerance ε.

Set $\eta = \eta_0$; choose $\hat{J}$ such that $\zeta_{\hat{J}} := C2^{-\hat{J}} < \zeta$.

(i) Set $(M(\nu))_{\nu\in F^+}, (F_i, \ell_i)_{i=0}^I = \textsc{Apply}(v; \eta/C_\Psi)$ by Algorithm 3.1;

(ii) For each $\nu' \in \operatorname{supp}([v]_\nu)_\nu$

Let $\ell_{\nu'} = \max_{\nu'\in F_i} \ell_i$;

For each μ such that $|\mu| \le \ell_{\nu'}$ and ν such that $(\mu,\nu') \in M(\nu)$ compute

$$[g]_{\nu,\nu'} = (\mathbf{M}_\mu)_{\nu,\nu'} A_\mu [v]_{\nu'}$$

by traversing the mesh of $[v]_{\nu'}$ once;

(iii) For each $\nu \in F^+$, use Algorithm 3.2 to evaluate

$$[r]_\nu = \delta_{0,\nu} f - \sum_{(\mu,\nu')\in M(\nu)} [g]_{\nu,\nu'} = \textsc{Sum}\big(\delta_{0,\nu} f, (-[g]_{\nu,\nu'})_{(\mu,\nu')\in M(\nu)}\big);$$

(iv) For each $\nu \in F^+$ set $(\Theta^+_\nu, \hat{\mathbf{r}}_\nu) = \textsc{DualNormIndicators}([r]_\nu, \hat{J})$ by Algorithm 3.3;

(v) Let $b = \eta + (1 + \frac{\zeta_{\hat{J}}}{\sqrt{1-\zeta_{\hat{J}}^2}})\|(\|\hat{\mathbf{r}}_\nu\|_{\ell_2})_{\nu\in F^+}\|$. If $\eta \le \frac{\zeta-\zeta_{\hat{J}}}{1+\zeta}\|(\|\hat{\mathbf{r}}_\nu\|_{\ell_2})_{\nu\in F^+}\|$ or $b \le \varepsilon$,

return $((\Theta^+_\nu)_{\nu\in F^+}, (\hat{\mathbf{r}}_\nu)_{\nu\in F^+}, ([r]_\nu)_{\nu\in F^+}, \eta, b)$;

otherwise, set $\eta \leftarrow \eta/2$ and go to **(i)**;

Proof. First, we show the residual error bounds similarly to [6, Thm. 4.15]. As a consequence of Theorem 1 and Theorem 4, we have $\|\mathbf{z}-\mathbf{g}\| \le C_\Psi \|Bv-g\|_{\mathcal{V}'} \le \eta$, where with $[g]_\nu$ as in (32),

$$\mathbf{g}_\nu = \big(\langle [g]_\nu, \psi_\lambda\rangle\big)_{\lambda\in\Theta} \quad\text{and}\quad \big(\langle [g]_\nu - \delta_{0,\nu} f, \psi_\lambda\rangle\big)_{\lambda\in\Theta_\nu} = \hat{\mathbf{r}}_\nu.$$

By Theorem 2, we can bound the truncation error by

$$\sum_{\nu\in F^+}\sum_{\lambda\notin\Theta_\nu} |\langle [g]_\nu - \delta_{0,\nu} f, \psi_\lambda\rangle|^2 \le \zeta_{\hat{J}}^2 \sum_{\nu\in F^+}\sum_{\lambda\in\Theta} |\langle [g]_\nu - \delta_{0,\nu} f, \psi_\lambda\rangle|^2.$$

Thus

$$\|\hat{\mathbf{r}}\|^2 = \sum_{\nu\in F^+}\sum_{\lambda\in\Theta_\nu} |\langle [g]_\nu - \delta_{0,\nu} f, \psi_\lambda\rangle|^2 \ge (1-\zeta_{\hat{J}}^2)\sum_{\nu\in F^+}\sum_{\lambda\in\Theta} |\langle [g]_\nu - \delta_{0,\nu} f, \psi_\lambda\rangle|^2 = (1-\zeta_{\hat{J}}^2)\|\mathbf{f}-\mathbf{g}\|^2.$$

Together with $\|\mathbf{f}-\mathbf{z}\| \ge \|\mathbf{f}-\mathbf{g}\| - \eta$, this first results in

$$\|\hat{\mathbf{r}} - (\mathbf{f}-\mathbf{z})\| \le \|\mathbf{w}-\mathbf{z}\| + \|\hat{\mathbf{r}} - (\mathbf{f}-\mathbf{g})\| \le \eta + \zeta_{\hat{J}}\|\mathbf{f}-\mathbf{g}\| \le \eta + \frac{\zeta_{\hat{J}}}{\sqrt{1-\zeta_{\hat{J}}^2}}\|\hat{\mathbf{r}}\| \tag{37}$$

and hence $\|\mathbf{f}-\mathbf{z}\| \le \eta + (1+\zeta_{\hat{J}}(1-\zeta_{\hat{J}}^2)^{-\frac{1}{2}})\|\hat{\mathbf{r}}\| = b$. If additionally $\eta \le \frac{\zeta-\zeta_{\hat{J}}}{1+\zeta}\|\hat{\mathbf{r}}\|$, it results in

$$\|\hat{\mathbf{r}} - (\mathbf{f}-\mathbf{z})\| \le \eta + \zeta_{\hat{J}}\|\mathbf{f}-\mathbf{g}\| \le (\zeta-\zeta_{\hat{J}})\|\hat{\mathbf{r}}\| + \|\mathbf{f}-\mathbf{g}\| - \zeta\eta \le \zeta(\|\mathbf{f}-\mathbf{g}\| - \eta) \le \zeta\|\mathbf{f}-\mathbf{z}\|.$$

This shows the prescribed error accuracy.

We now estimate $\#\Lambda^+ = \sum_{\nu\in F^+}\#\Theta_\nu^+$. Let $\tilde{\mathcal{T}}_\nu \ge \hat{\mathcal{T}}_0$ be a triangulation, such that $[r]_\nu$ is polynomial on its triangles and edges. Then by Theorem 6 and Theorem 7, we have

$$\#\Theta_\nu^+ \le c(\hat{J})\#\tilde{\mathcal{T}}_\nu.$$

We thus estimate $\sum_{\nu\in F^+}\#\tilde{\mathcal{T}}_\nu$. To this end, we also denote by

$$\tilde{\mathcal{T}}_{\nu,\nu'} = \{T\in\tilde{\mathcal{T}}_\nu\colon\ \operatorname{supp}[g]_{\nu,\nu'}\cap\overline{T}\neq\emptyset\}$$

the supporting triangles of $[g]_{\nu,\nu'}$. Let T be a triangle in the triangulation $\mathcal{T}_{\nu'}$ of $[v]_{\nu'}$. If $\operatorname{level}(T) = \ell + k$ for some $\ell \le \ell_{\nu'}$, then all θ_μ with $|\mu|\le\ell$ are polynomial on T by **??** 2. On the one hand, by **??** 1 (ii) we have

$$\#\big\{(\tilde{T},\nu)\colon\ \tilde{T}\in\hat{\mathcal{T}}_{\nu,\nu'} \text{ for some } \nu=\nu'\pm\mu,\ |\mu|\le\ell \text{ and } T\cap\tilde{T}\neq\emptyset\big\} \le 2M\ell \le 2M\ell_{\nu'}.$$

On the other hand, $\#\{\tilde{T}\in\mathcal{T}_{\nu'\pm e_\mu,\nu'}\colon\ \operatorname{level}(\tilde{T}) = |\mu|+k\} \le 2K$ by **??** 2. Hence, we have

$$\sum_{|\mu|\le\ell_\nu}\big(\#\mathcal{T}_{\nu'+e_\mu,\nu'} + \#\mathcal{T}_{\nu'-e_\mu,\nu'}\big) \le 2M\ell_{\nu'}\#\mathcal{T}_{\nu'} + \sum_{|\mu|\le\ell_{\nu'}} 2K \lesssim \#\mathcal{T}_{\nu'}\ell_{\nu'} + 2^{d\ell_{\nu'}}$$

and

$$\#\Lambda^+ = \sum_{\nu\in F^+}\#\Theta_\nu^+ \lesssim \#\mathcal{T}(f) + N(\mathbb{T})\max_{\nu'\in F}\ell_{\nu'}^2 + \sum_{j=0}^{J}\#F_j 2^{d\ell_j}\max_{\nu'\in F}\ell_{\nu'}\,,$$

where an additional factor of $\max_{\nu'\in F}\ell_{\nu'}$ occurs by completing the supports $\tilde{\mathcal{T}}_{\nu,\nu'}$ to a (not necessarily conforming) triangulation. The estimate (35) then follows with Theorem 4.

It remains to estimate the number of operations. As in [6], the cost of Algorithm 3.1 is of order

$$N(\mathbb{T}) + \#F \log \#F + \eta^{-\frac{1}{s}} \|(\|[v]_\nu\|)_{\nu\in\mathcal{F}}\|_{\mathcal{A}^s}^{\frac{1}{s}}.$$

For step (ii) in Algorithm 3.4, we have to count how often each triangle in $\mathbb{T}$ is required, leading to an order of

$$(1 + |(|\log \eta + \log \|(\|[v]_\nu\|)_{\nu\in\mathcal{F}}\|_{\mathcal{A}^s}))N(\mathbb{T}) + \eta^{-\frac{1}{s}} \|(\|[v]_\nu\|)_{\nu\in\mathcal{F}}\|_{\mathcal{A}^s}^{\frac{1}{s}}.$$

The application of SUM in Algorithm 3.2 is linear in the size of triangulations. This results in the complexity

$$\begin{aligned} \#\mathcal{T}(f) + (1 + |\log \eta + \log \|(\|[v]_\nu\|)_{\nu\in\mathcal{F}}\|_{\mathcal{A}^s}|)^2 N(\mathbb{T}) \\ + \eta^{-\frac{1}{s}} \|(\|[v]_\nu\|)_{\nu\in\mathcal{F}}\|_{\mathcal{A}^s}^{\frac{1}{s}} (1 + |\log \eta| + \log \|(\|[v]_\nu\|)_{\nu\in\mathcal{F}}\|_{\mathcal{A}^s}). \end{aligned}$$

The number of basic operations in Algorithm 3.3 is linear in the size of Λ^+, which is again of the same order by Theorem 7. Finally, we have at most $1 + \log_2(\eta_0/\eta)$ outer loops, which leads to the bound (36). □

Remark 10. For simplicity, we have assumed the parameter-independent source term f to be piecewise polynomial on the initial triangulation $\hat{\mathcal{T}}_0$. However, the above strategy for computing residual indicators can easily be modified to accommodate more general f. In particular for $f \in L_2(D)$, it suffices to adapt the computation of the single residual component $[r]_\nu$ for $\nu = 0$ in step **(iii)** of Algorithm 3.4, where f can be replaced by an approximation on a sufficiently fine grid according to the current error tolerance. For $f \in H^{-1}(D)$ that are proper functionals, one may instead directly modify step **(iv)** using the frame coefficients $\langle f, \psi_\lambda\rangle$, $\lambda \in \Theta$, and knowledge on their decay properties. Since such modifications are problem-dependent, we do not go into further detail here.

3.2 Refining the triangulations

Our refinement strategy is based on selecting a subset of the residual indicators produced by RESESTIMATE (Algorithm 3.4) according to a bulk chasing criterion and subsequently refining the individual meshes in the approximation such that they resolve the selected frame elements. We assume that RESESTIMATE is performed for an approximation on the given conforming meshes $\mathbb{T} = (\mathcal{T}_\nu)_{\nu\in F^0}$ with finite $F^0 \subset \mathcal{F}$.

Let us first consider the selection of residual indicators. Recall that for all $\nu \in F^+ \subset \mathcal{F}$, RESESTIMATE produces vectors of residual indicators $\hat{\mathbf{r}}_\nu$ corresponding to indices Θ_ν^+, with associated spatial-parametric index set $\Lambda^+ \subset \mathcal{F} \times \Theta$ as in (33). Since the computational costs of the operations that we perform depend on tree structure in the frame index sets, we use the strategy based on tree coarsening described in [6, Sec. 4.5] that preserves such structures in the selection.

To this end we first fix a tree structure on the frame elements ψ_λ, $\lambda \in \Theta$, and thus on the index set Θ. This tree structure is determined by choosing a unique parent $\psi_{\lambda'}$ for each frame element ψ_λ with $|\lambda| > 0$ such that $|\lambda| = |\lambda'| + 1$ and $\operatorname{meas}_2(\operatorname{supp}\psi_\lambda \cap \operatorname{supp}\psi_{\lambda'}) > 0$. The tree structure on Θ induces a natural tree structure on $\Theta^{\mathcal{F}}$ (and thus on $\mathcal{F} \times \Theta$) with roots $(\{\psi_\lambda : |\lambda| = 0\})_{\nu\in\mathcal{F}}$.

We use the procedure $\Lambda =$ TREEAPPROX$(\Lambda^0, \Lambda^+, \hat{\mathbf{r}}, \eta)$ from [6, Alg. 4.4] to obtain a coarsening Λ of Λ^+ such that $\Lambda^0 \subseteq \Lambda \subseteq \Lambda^+$, where Λ^0, Λ are subsets with the chosen tree structure and $\eta > 0$; see also [8, 10]. Here we take $\Lambda^0 = \{(\nu, \lambda) \in F^0 \times \Theta : \psi_\lambda \in V(\mathcal{T}_\nu)\}$.

For Λ generated in this manner, TREEAPPROX ensures

$$\|\hat{\mathbf{r}}|_{\Lambda}\|^2 \geq \|\hat{\mathbf{r}}\|^2 - \eta.$$

For any prescribed $\omega_0 \in (0,1)$, taking $\eta = (1-\omega_0^2)\|\mathbf{r}\|^2$ we obtain the desired bulk chasing condition

$$\|\hat{\mathbf{r}}|_{\Lambda}\| \geq \omega_0 \|\hat{\mathbf{r}}\|\,. \tag{38}$$

Remark 11. As a consequence of [6, Cor. 4.20], the resulting Λ has the following optimality property: for each $\omega_1 \in [\omega_0, 1)$, there exists $\tilde{C} > 0$ such that $\#(\Lambda \setminus \Lambda^0) \leq \tilde{C}\#(\tilde{\Lambda} \setminus \Lambda^0)$ for all tree subsets $\tilde{\Lambda} \supseteq \Lambda^0$ such that $\|\hat{\mathbf{r}}|_{\tilde{\Lambda}}\| \geq \omega_1\|\hat{\mathbf{r}}\|$.

For Λ satisfying (38) selected in this manner, let $F = \{\nu \in \mathcal{F} \colon \exists \lambda \in \Theta \colon (\nu, \lambda) \in \Lambda\} \subseteq F^+$ and $\Theta_\nu = \{\lambda \in \Theta \colon (\nu,\lambda) \in \Lambda\} \subseteq \Theta_\nu^+$ for $\nu \in F$, where Θ_ν inherits the tree structure of Λ. We now define a procedure

$$(\tilde{\mathcal{T}}_\nu)_{\nu \in F} = \text{MESH}(\Lambda) \tag{39}$$

that outputs the componentwise smallest sequence of meshes such that for each $\nu \in F$, $\tilde{\mathcal{T}}_\nu$ is conforming and $\operatorname{span}\{\psi_\lambda\}_{\lambda \in \Theta_\nu} \subseteq V(\tilde{\mathcal{T}}_\nu)$. Note that in a setting where $\Lambda^0 \subseteq \Lambda$, the meshes $\tilde{\mathcal{T}}_\nu$ are refinements of the initial meshes $\mathcal{T}_\nu$ that are given by the frame elements in Λ^0: $(\mathcal{T}_\nu)_{\nu \in F^0} = \text{MESH}(\Lambda^0)$.

Proposition 12. *The meshes $(\tilde{\mathcal{T}}_\nu)_{\nu \in F}$ produced by* MESH(Λ) *in* (39) *satisfy $\#\tilde{\mathcal{T}}_\nu \lesssim \#\Theta_\nu$ for each $\nu \in F$ and can be obtained using a number of operations proportional to $\#\Theta_\nu$, with constants depending only on $\hat{\mathcal{T}}_0$.*

The proof relies on the following bound on the complexity of conforming meshes created by newest vertex bisection, which is a consequence of [9, Thm. 2.4], see also [26, Thm. 3.2] and [12, Lem. 2.3].

Theorem 13. *Let $\mathcal{T}_0 = \hat{\mathcal{T}}_0$ and for $k = 1, \ldots, n$, let $\mathcal{T}_k$ be defined as the smallest conforming refinement of $\mathcal{T}_{k-1}$ such that the elements $\mathcal{M}_{k-1} \subseteq \mathcal{T}_{k-1}$ are bisected. Then with $C_0 > 0$ depending only on $\hat{\mathcal{T}}_0$,*

$$\#\mathcal{T}_n - \#\mathcal{T}_0 \leq C_0 \sum_{k=0}^{n-1} \#\mathcal{M}_k.$$

Proof of Proposition 12. For each $\nu \in F$, given a tree Θ_ν, we can apply Theorem 13 to the construction of a suitable triangulation $\tilde{\mathcal{T}}_\nu$ with $\#\tilde{\mathcal{T}}_\nu \lesssim \#\Theta_\nu$ in the following way. We may assume that Θ_ν contains the roots $\{\lambda \colon |\lambda| = 0\}$ as

$$\#(\{\lambda : \lambda \text{ is an ancestor of some } \lambda' \in \Theta_\nu\} \cup \Theta_\nu) \lesssim \#\Theta_\nu$$

by the tree structure of Θ_ν. Now let the sequence of triangulations in Theorem 13 be defined by $\mathcal{T}_0 = \hat{\mathcal{T}}_0$ and

$$\mathcal{M}_k = \{T \in \mathcal{T}_k \colon T \cap \operatorname{supp} \psi_\lambda \text{ and } |\lambda| = k+1\}$$

with the slight adaptation that each marked triangle is bisected twice. We get $\sum_{k=0}^{n-1} \#\mathcal{M}_k \lesssim \Theta_\nu$ as each hat function on a uniform refinement has support on only a bounded number of triangles, and thus the result follows. □

Algorithm 3.5 $w = \textsc{GalerkinSolve}(\mathbb{T}, v, r, \ell, \varepsilon_0)$ given a family of triangulations $\mathbb{T}$, an approximate solution $v \in \mathcal{V}(\mathbb{T})$, an approximated residual $r \in \mathcal{V}'$, accuracy parameter ℓ of the operator compression, and a target tolerance ϵ_0.

(i) Assemble linear operators $\mathbf{B}_\ell, \mathbf{P}$ on the coefficients of the nodal basis of $\mathbb{T}$ such that $\mathbf{u}^\intercal \mathbf{B}_\ell \mathbf{v} = \langle B_\ell v, u\rangle$ for all $u, v \in \mathcal{V}(\mathbb{T})$, $\mathbf{P}$ satisfies (41), and $\mathbf{r}$ such that $\langle \mathbf{v}, \mathbf{r}\rangle = \langle r, v\rangle$.

(ii) Use the preconditioned conjugated gradient method to find q with nodal coefficients $\mathbf{q}$ and $\langle \mathbf{P}(\mathbf{B}_\ell \mathbf{q} - \mathbf{r}), (\mathbf{B}_\ell \mathbf{q} - \mathbf{r})\rangle \le \varepsilon_0^2$.

(iii) Return $w = v + q$;

3.3 Solving Galerkin systems

As a starting point for solving the Galerkin systems, we recall the frame property Theorem 1 and assume an approximated residual r computed by Algorithm 3.4 with tolerance η, which satisfies $\|r - (f - Bv)\|_{\mathcal{V}'} \le \eta / C_\Psi$. We use the preconditioned conjugated gradient (PCG) method to find a correction q of the current solution approximation such that for a $\rho > 0$,

$$\|r - B_\ell q\|_{\mathcal{V}(\mathbb{T})'} \le \frac{\rho}{c_\Psi} \|\hat{\mathbf{r}}\|. \tag{40}$$

Here, the discrete dual norm is given by

$$\|g\|_{\mathcal{V}(\mathbb{T})'} = \sup_{v \in \mathcal{V}(\mathbb{T})} \frac{\langle g, v\rangle}{\|v\|_{\mathcal{V}}}.$$

The condition (40) can be verified directly in the PCG method as follows. Let $\mathbf{A} \in \mathbb{R}^{N\times N}$ with $N = \dim \mathcal{V}(\mathbb{T})$ be the coefficient matrix such that for nodal coordinate vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^N$ of functions $u, v \in \mathcal{V}(\mathbb{T})$ we have $\langle \mathbf{A}\mathbf{u}, \mathbf{v}\rangle = \int_Y \int_D \nabla u(x,y) \cdot \nabla v(x,y) \,\mathrm{d}x \,\mathrm{d}y$. We use the modification of the BPX preconditioner to general meshes generated by newest vertex bisection analyzed in [13]. For this preconditioner with matrix representation $\mathbf{P}$, there exist $c_p, C_P > 0$ such that we have the spectral equivalence

$$c_P \langle \mathbf{P}^{-1}\mathbf{u}, \mathbf{u}\rangle \le \langle \mathbf{A}\mathbf{u}, \mathbf{u}\rangle \le C_P \langle \mathbf{P}^{-1}\mathbf{u}, \mathbf{u}\rangle \quad \text{for all } \mathbf{u} \in \mathbb{R}^N. \tag{41}$$

For g in $\mathcal{V}'$, we thus have

$$\|g\|^2_{\mathcal{V}(\mathbb{T})'} = \sup_{v \in \mathcal{V}(\mathbb{T})} \frac{\langle g, v\rangle^2}{\|v\|^2_{\mathcal{V}}} = \sup_{\mathbf{v} \in \mathbb{R}^N} \frac{|\langle \mathbf{g}, \mathbf{v}\rangle|^2}{\langle \mathbf{A}\mathbf{v}, \mathbf{v}\rangle} \le \sup_{\mathbf{h} \in \mathbb{R}^N} \frac{|\langle \mathbf{g}, \mathbf{P}\mathbf{h}\rangle|^2}{c_P \langle \mathbf{h}, \mathbf{P}\mathbf{h}\rangle} \le \frac{1}{c_P} \langle \mathbf{P}\mathbf{g}, \mathbf{g}\rangle,$$

where we substituted $\mathbf{v} = \mathbf{P}\mathbf{h}$, used the spectral equivalence and the Cauchy-Schwarz inequality. Similarly, we have $\|g\|^2_{\mathcal{V}(\mathbb{T})'} \ge \frac{1}{C_P} \langle \mathbf{P}\mathbf{g}, \mathbf{g}\rangle$.

Proposition 14. *Assume an initial approximation v on a triangulation $\tilde{\mathbb{T}} \le \mathbb{T}$, and with sufficiently small global tolerance ε, let*

$$((\Theta^+_\nu)_{\nu \in F^+}, (\hat{\mathbf{r}}_\nu)_{\nu \in F^+}, ([r]_\nu)_{\nu \in F^+}, \eta, b) = \textsc{ResEstimate}(v; \zeta, \eta_0, \varepsilon)$$

be the output of Algorithm 3.4 and $w = \textsc{GalerkinSolve}(\mathbb{T}, v, r, \ell, c_P{}^{-1/2} c_\Psi^{-1} \rho \|\hat{\mathbf{r}}\|)$ be the output of Algorithm 3.5. Then the approximation w of the Galerkin solution $u_\mathbb{T}$ satisfies the error bound

$$\|u_\mathbb{T} - w\|_B \le \gamma(\zeta, \ell, \rho, \hat{J}) \|\hat{\mathbf{r}}\|,$$

DOI 10.20347/WIAS.PREPRINT.3112

where

$$\gamma(\zeta,\ell,\rho,\hat{J}) = \frac{1}{c_\Psi}\frac{1}{\sqrt{c_B}}\left(\frac{\zeta-\zeta_{\hat{J}}}{1+\zeta} + \frac{\zeta_{\hat{J}}}{\sqrt{1-\zeta_{\hat{J}}^2}} + \rho + \frac{C_2}{1-2^{-\alpha}}2^{-\alpha\ell}\frac{1}{c_B}\left(\frac{1}{1-\zeta_{\hat{J}}}+\rho\right)\right).$$

Proof. We assume that ε is sufficiently small so that ResEstimate does not terminate with the condition $b \le \varepsilon$ in step **(v)**. Recall from (13) that $\|v\|_B \le c_B^{-1/2}\|Bv\|_{\mathcal{V}(\mathbb{T})'}$ for $v \in \mathcal{V}(\mathbb{T})$. We consider a decomposition of the total error for $w = v + q$,

$$\begin{aligned} \|u_\mathbb{T} - w\|_B &\le \frac{1}{\sqrt{c_B}}\|f - B(v+q)\|_{\mathcal{V}(\mathbb{T})'} \\ &\le \frac{1}{\sqrt{c_B}}\left(\|r-(f-Bv)\|_{\mathcal{V}(\mathbb{T})'} + \|r - B_\ell q\|_{\mathcal{V}(\mathbb{T})'} + \|(B_\ell - B)q\|_{\mathcal{V}(\mathbb{T})'}\right). \end{aligned} \tag{42}$$

For the first term on the right in (42), using (37) in the proof of Theorem 9 and Theorem 1

$$\|r-(f-Bv)\|_{\mathcal{V}(\mathbb{T})'} \le \frac{1}{c_\Psi}\left(\frac{\zeta-\zeta_{\hat{J}}}{1+\zeta} + \frac{\zeta_{\hat{J}}}{\sqrt{1-\zeta_{\hat{J}}^2}}\right)\|\hat{\mathbf{r}}\|.$$

Note that Theorem 9 uses the frame norm while the estimate above is with respect to the discrete dual norm, which is obviously smaller than the full dual norm used in Theorem 1.

By assumption, the second term on the right in (42) satisfies the bound (40) with relative solver error ρ as a result of Algorithm 3.5. For the last term, first note that

$$\|B_\ell q\|_{\mathcal{V}(\mathbb{T})'} = \|r - (r - B_\ell q)\|_{\mathcal{V}(\mathbb{T})'} \le \|r\|_{\mathcal{V}(\mathbb{T})'} + \frac{\rho}{c_\Psi}\|\hat{\mathbf{r}}\|.$$

Moreover, again with Theorem 9 for the approximation of the full sequence of residual frame coefficients $\bar{\mathbf{r}} = ((\langle [r]_\nu, \psi_\lambda\rangle)_{\lambda\in\Theta})_{\nu\in\mathcal{F}}$, we derive

$$\|r\|_{\mathcal{V}(\mathbb{T})'} \le \frac{1}{c_\Psi}\|\bar{\mathbf{r}}\| \le \frac{1}{c_\Psi}\frac{1}{1-\zeta_{\hat{J}}}\|\hat{\mathbf{r}}\|. \tag{43}$$

By Theorem 3 and using that $c_B\|q\|_{\mathcal{V}(\mathbb{T})} \le \|B_\ell q\|_{\mathcal{V}(\mathbb{T})'}$, we thus obtain

$$\|(B_\ell - B)q\|_{\mathcal{V}(\mathbb{T})'} \le \frac{C_2}{1-2^{-\alpha}}2^{-\alpha\ell}\|q\|_{\mathcal{V}(\mathbb{T})} \le \frac{C_2}{1-2^{-\alpha}}2^{-\alpha\ell}\frac{1}{c_B}\left(\|r\|_{\mathcal{V}(\mathbb{T})'} + \frac{\rho}{c_\Psi}\|\hat{\mathbf{r}}\|\right).$$

With (43) it follows that

$$\|(B_\ell - B)q\|_{\mathcal{V}(\mathbb{T})'} \le \frac{C_2}{1-2^{-\alpha}}2^{-\alpha\ell}\frac{1}{c_\Psi}\frac{1}{c_B}\left(\frac{1}{1-\zeta_{\hat{J}}}+\rho\right)\|\hat{\mathbf{r}}\|.$$

A combination of the preceding estimates yields the result. □

Remark 15. The number of steps of the PCG method in Algorithm 3.5 terminates after a uniform finite number of steps since we require a fixed relative accuracy. Hence, the computational complexity of the algorithm does not deteriorate when successive mesh refinements take place.

Algorithm 3.6 Adaptive Galerkin Method

Set the parameters ℓ, ζ, ω_0 and relative accuracy ρ, set initial $u^0 = 0$ and formally $\|\hat{\mathbf{r}}^{-1}\| = C_\Psi \|f\|_{\mathcal{V}'}$. Let $k = 0$;

(i) $(\Lambda^+, (\hat{\mathbf{r}}^k_\nu)_\nu, ([r^k]_\nu)_\nu, \eta_k, b_k) = \textsc{ResEstimate}(u^k; \zeta, \frac{\zeta}{1+\zeta}\|\hat{\mathbf{r}}^{k-1}\|, \varepsilon)$ by Algorithm 3.4

(ii) If $b_k \le \varepsilon$, return u^k.

(iii) $\Lambda^{k+1} = \textsc{TreeApprox}(\Lambda^k, \Lambda^+, \hat{\mathbf{r}}^k, (1-\omega_0^2)\|\hat{\mathbf{r}}^k\|^2)$;

(iv) $\mathbb{T}^{k+1} = \textsc{Mesh}(\Lambda^{k+1})$;

(v) $u^{k+1} = \textsc{GalerkinSolve}(\mathbb{T}^{k+1}, u^k, r^k, \ell, \frac{1}{\sqrt{c_P}}\frac{\rho}{c_\Psi}\|\hat{\mathbf{r}}^k\|)$ by Algorithm 3.5

(vi) $k \leftarrow k+1$ and go to (i);

4 Convergence analysis

We now come to the main result of this work, where we show error reduction by a uniform factor in each step of the adaptive scheme Algorithm 3.6. This reduction factor depends on a number of parameters and constants from previous sections: on c_B and C_B from (3) and (4), respectively; on the frame bounds c_Ψ, C_Ψ in (20); and ω_0 as in (38).

Let $\mathbb{T} = (\mathcal{T}_\nu)_{\nu\in F}$ with conforming $\mathcal{T}_\nu \ge \hat{\mathcal{T}}_0$ for all $\nu \in F$ with finite $F \subset \mathcal{F}$ be given, and let $\tilde{\mathbb{T}} = (\tilde{\mathcal{T}}_\nu)_{\nu\in\tilde{F}}$ be any refinement with $\tilde{F} \supseteq F$ and conforming $\tilde{\mathcal{T}}_\nu \ge \mathcal{T}_\nu$ for $\nu \in F$ as well as $\tilde{\mathcal{T}}_\nu \ge \mathcal{T}_0$ for $\nu \in F^+ \setminus F$. Note that for the exact solution $u \in \mathcal{V}$, the Galerkin solution $u_{\tilde{\mathbb{T}}} \in \mathcal{V}(\tilde{\mathbb{T}})$ and an arbitrary $w \in \mathcal{V}(\mathbb{T})$, we then have the Galerkin orthogonality relation

$$\|u - w\|_B^2 = \|u - u_{\tilde{\mathbb{T}}}\|_B^2 + \|u_{\tilde{\mathbb{T}}} - w\|_B^2 \,. \tag{44}$$

In what follows, recalling notation (28), let

$$\mathbf{r}(w) = \big(\langle [Bw - f]_\nu, \psi_\lambda\rangle\big)_{\nu\in\mathcal{F},\lambda\in\Theta} \in \ell_2(\mathcal{F}\times\Theta)\,.$$

We denote by $P_{\mathbb{T}}\colon \mathcal{V} \to \mathcal{V}$ the orthogonal projection in $\mathcal{V}$ onto $\mathcal{V}(\mathbb{T})$, which corresponds to the V-orthogonal projection onto $V(\mathcal{T}_\nu)$ for each ν.

Lemma 16. *For $w \in \mathcal{V}(\mathbb{T})$, let $\hat{\mathbf{r}} \in \ell_2(\mathcal{F}\times\Theta)$ be such that*

$$\|\mathbf{r}(w) - \hat{\mathbf{r}}\|_{\ell_2} \le \zeta \|\mathbf{r}(w)\|_{\ell_2}$$

with $\zeta \in (0, \frac{1}{2})$, and let $\tilde{\Lambda} = (\tilde{\Theta}_\nu)_{\nu\in\mathcal{F}}$ with $\tilde{\Theta}_\nu \subset \Theta$, $\nu \in \mathcal{F}$, be chosen such that

$$\|\hat{\mathbf{r}}|_{\tilde{\Lambda}}\|_{\ell_2} \ge \omega_0 \|\hat{\mathbf{r}}\|_{\ell_2}$$

with $\omega_0 \in (0,1]$, where $(1+\omega_0)\zeta < \omega_0$. Let $\tilde{\mathbb{T}} = \textsc{Mesh}(\tilde{\Lambda})$, then with $C_{B,\Psi} = C_\Psi^2 C_B/(c_\Psi^2 c_B) \ge 1$, we have

$$\|u - u_{\tilde{\mathbb{T}}}\|_B \le \left(1 - \frac{\big(\omega_0 - (1+\omega_0)\zeta\big)^2}{C_{B,\Psi}}\right)^{\frac{1}{2}} \|u - w\|_B.$$

Proof. Using (20), we obtain

$$\begin{aligned}
\|w-u_{\tilde{\mathbb{T}}}\|_B &\geq \frac{1}{\sqrt{C_B}}\|B(w-u_{\tilde{\mathbb{T}}})\|_{\mathcal{V}'} \geq \frac{1}{\sqrt{C_B}}\|P'_{\tilde{\mathbb{T}}}B(w-u_{\tilde{\mathbb{T}}})\|_{\mathcal{V}'} = \frac{1}{\sqrt{C_B}}\|P'_{\tilde{\mathbb{T}}}(Bw-f)\|_{\mathcal{V}'}\\
&\geq \frac{1}{\sqrt{C_B}C_\Psi}\Big(\sum_\nu\sum_{\lambda\in\Theta}|\langle[Bw-f]_\nu, P_{\tilde{\mathbb{T}}}\psi_\lambda\rangle|^2\Big)^{\frac12}\\
&= \frac{1}{\sqrt{C_B}C_\Psi}\Big(\sum_\nu\sum_{\lambda\in\tilde{S}_\nu}|\langle[Bw-f]_\nu, \psi_\lambda\rangle|^2 + \sum_\nu\sum_{\lambda\in\Theta\setminus\tilde{S}_\nu}|\langle[Bw-f]_\nu, P_{\tilde{\mathbb{T}}}\psi_\lambda\rangle|^2\Big)^{\frac12}\\
&\geq \frac{1}{\sqrt{C_B}C_\Psi}\Big(\sum_\nu\sum_{\lambda\in\tilde{S}_\nu}|\langle[Bw-f]_\nu, \psi_\lambda\rangle|^2\Big)^{\frac12} = \frac{1}{C_\Psi\sqrt{C_B}}\|\mathbf{r}(w)|_{\tilde{\Lambda}}\|.
\end{aligned}$$

We now note that

$$\|\mathbf{r}(w)|_{\tilde{\Lambda}}\| \geq \|\hat{\mathbf{r}}|_{\tilde{\Lambda}}\| - \|\mathbf{r}(w)-\hat{\mathbf{r}}\| \geq \omega_0\|\hat{\mathbf{r}}\| - \zeta\|\mathbf{r}(w)\| \geq \omega_0\|\mathbf{r}(w)\| - (1+\omega_0)\zeta\|\mathbf{r}(w)\|,$$

and thus

$$\begin{aligned}
\|\mathbf{r}(w)|_{\tilde{\Lambda}}\| &\geq \big(\omega_0-(1+\omega_0)\zeta\big)\|\mathbf{r}(w)\| = \big(\omega_0-(1+\omega_0)\zeta\big)\Big(\sum_\nu\sum_{\lambda\in\Theta}|\langle[Bw-f]_\nu,\psi_\lambda\rangle|^2\Big)^{\frac12}\\
&\geq \big(\omega_0-(1+\omega_0)\zeta\big)c_\Psi\|Bw-f\|_{\mathcal{V}'} \geq \big(\omega_0-(1+\omega_0)\zeta\big)c_\Psi\sqrt{c_B}\|w-u\|_B\,.
\end{aligned}$$

The statement thus follows with Galerkin orthogonality (44). □

Lemma 17. *Let $\|u_{\tilde{\mathbb{T}}}-\tilde{w}\|_B \leq \gamma\|\hat{\mathbf{r}}\|$. Then*

$$\|u-\tilde{w}\|_B \leq \delta\|u-w\|_B \tag{45}$$

with

$$\delta = \left(1 - \frac{\big(\omega_0-(1+\omega_0)\zeta\big)^2}{C_{B,\Psi}} + \gamma^2(1+\zeta)^2C_\Psi^2C_B\right)^{\frac12}$$

and $C_{B,\Psi}$ from Theorem 16.

Proof. Combining

$$\gamma\|\hat{\mathbf{r}}\| \leq \gamma(1+\zeta)\|\mathbf{r}(w)\| \leq \gamma(1+\zeta)C_\Psi\|Bw-f\|_{\mathcal{V}'} \leq \gamma(1+\zeta)C_\Psi\sqrt{C_B}\|w-u\|_B$$

with the Galerkin orthogonality $\|u-\tilde{w}\|_B^2 = \|u-u_{\tilde{\mathbb{T}}}\|_B^2 + \|u_{\tilde{\mathbb{T}}}-\tilde{w}\|_B^2$, the statement follows. □

Considering Theorem 14, there are some requirements for the parameters that enter in γ to ensure an error reduction in Theorem 17. The conditions on ζ and ω_0 read

$$0 < (C+2)\zeta < 1 \qquad \text{with} \qquad C^2 = \frac{C_\Psi^2 C_B C_{B,\Psi}}{c_\Psi^2 c_B} \tag{46}$$

and

$$\omega_0 > (C+1)\frac{\zeta}{1-\zeta}. \tag{47}$$

Then it is possible to also find parameters ℓ, ρ and $\hat{J}$ to achieve the condition on γ

$$\gamma = \gamma(\zeta,\ell,\rho,\hat{J}) < \frac{\omega_0-(1+\omega_0)\zeta}{(1+\zeta)\sqrt{C_{B,\Psi}C_B}C_\Psi}. \tag{48}$$

With the help of Theorem 14 and Theorem 17 we can thus state the following result on error reduction in each step.

Theorem 18. *Assume conditions* (46) *and* (47)*. Then there exist parameters $\ell, \rho, \hat{J}$ such that* (48) *is satisfied. Consequently, the adaptive algorithm achieves the error reduction*

$$\|u - u^{k+1}\|_B \leq \delta \|u - u^k\|_B,$$

where δ from Theorem 17 is independent of k.

Proof. We first note that (48) and

$$1 - \frac{(\omega_0 - (1+\omega_0)\zeta)^2}{C_{B,\Psi}} + \gamma^2(1+\zeta)^2 C_\Psi^2 C_B < 1,$$

are equivalent and thus error reduction is achieved due to (45) in Theorem 17. The terms $2^{-\alpha\ell}$, $\zeta_{\hat{J}}$ and ρ appearing in $\gamma(\zeta, \ell, \rho, \hat{J})$ from Theorem 14 can be made arbitrarily small for sufficiently large $\hat{J}$ and ℓ and sufficiently small ρ. Thus the condition (46) is satisfiable whenever

$$\frac{1}{c_\Psi} \frac{1}{\sqrt{c_B}} \left(\frac{\zeta}{1+\zeta} \right) < \frac{\omega_0 - (1+\omega_0)\zeta}{(1+\zeta)\sqrt{C_{B,\Psi} C_B} C_\Psi},$$

which directly leads to (47), and noting that $\omega_0 < 1$, the condition (46) for ζ follows as well. □

Remark 19*.* Concerning the computational complexity of the method, note that all substeps of the adaptive method Algorithm 3.6 have either linear costs up to logarithmic factors in $N(\mathbb{T})$ or in the target accuracy rate $\eta^{-\frac{1}{s}}$, as seen notably for the most expensive step ResEstimate in Theorem 9. The complexity of GalerkinSolve is linear in $N(\mathbb{T})$ and only deteriorates for larger ℓ, as noted in Theorem 15. For quasi-optimal computational complexity of the adaptive method, it thus only remains to show that the number of degrees of freedom in the approximation that is added in each step of the iterative method remains quasi-optimal.

5 Numerical Experiments

The adaptive Galerkin method Algorithm 3.6 was implemented for spatial dimensions $d = 1$ and $d = 2$ using the Julia programming language, version 1.9.2. The numerical experiments were performed on a Dell PowerEdge R725 workstation with AMD EPYC 7742 64-core processor. For the code to reproduce the tests, see [5].

For $d = 2$ we chose the L-shaped domain $D = (0,1)^2 \backslash (0, 0.5)^2$, with an initial mesh of 24 congruent triangles with 5 interior nodes. For $d = 1$ we simply take $D = (0, 1)$. In Figure 2 we show refinements of different components that appear during the adaptive method.

For the random fields $a(y)$, we use an expansion in terms of hierarchical hat functions formed by dilations and translations of $\theta(x) = (1 - |2x - 1|)_+$. Specifically, for $d = 1$, θ_μ with $\mu = (\ell, k)$ is given by

$$\theta_{\ell,k}(x) := c 2^{-\alpha\ell} \theta(2^\ell x - k), \quad k = 0, \ldots, 2^\ell - 1, \ \ell \in \mathbb{N}_0. \tag{49}$$

This yields a wavelet-type multilevel structure satisfying Assumptions 1 and thus (26) and (27), where

$$\mathcal{M} = \{(\ell, k) \colon k = 0, \ldots, 2^\ell - 1, \ \ell \geq 0\}$$

with level parameters $|(\ell, k)| = \ell$. For $d = 2$, we take the isotropic product hierarchical hat functions

$$\theta_{\ell,k_1,k_2}(x) := c 2^{-\alpha\ell} \theta(2^\ell x - k_1)\, \theta(2^\ell x - k_2), \quad (\ell, k_1, k_2) \in \mathcal{M}, \tag{50}$$

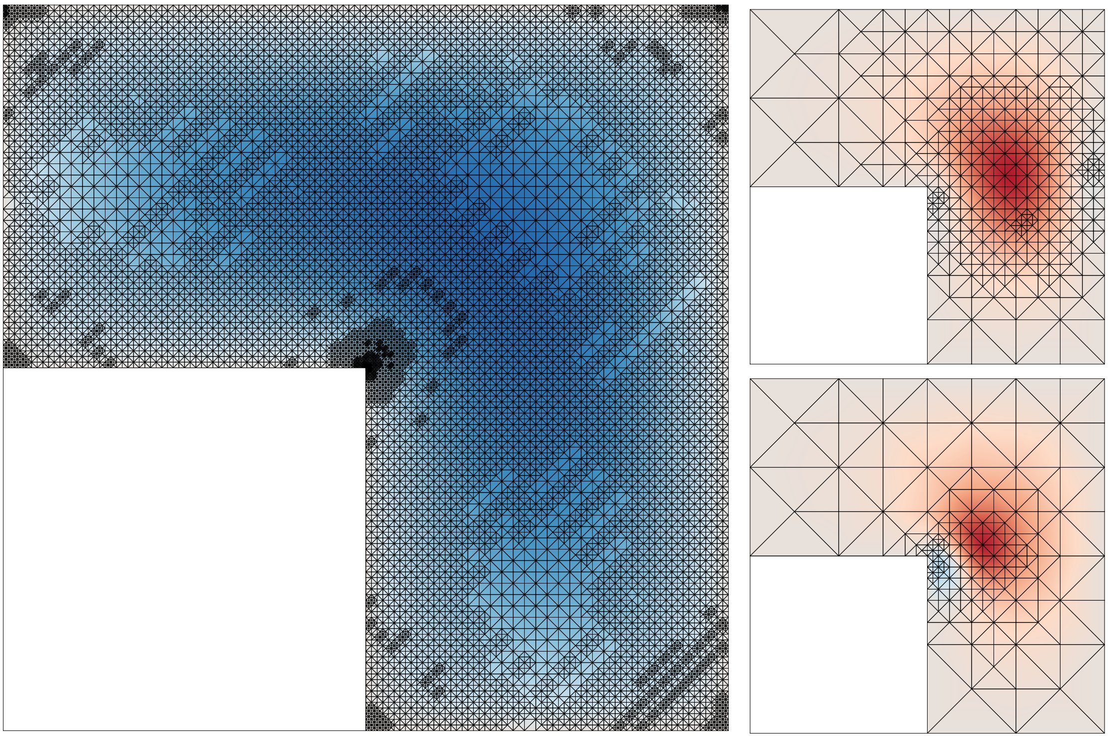

Figure 2: Triangulation of approximations of different components u_ν in (6), analogous to Figure 1, generated by the adaptive method for different ν for $d = 2$. Top: $\nu = 0$, bottom: $\nu = e_\mu$ for two different $\mu \in \mathcal{M}$.

with

$$\mathcal{M} = \big\{(\ell, k_1, k_2) \colon \ \ell \in \mathbb{N}_0, \ k_1, k_2 = 0, \tfrac{1}{2}, \dots, 2^\ell - \tfrac{3}{2}, 2^\ell - 1 \text{ with } k_1 \in \mathbb{N}_0 \text{ or } k_2 \in \mathbb{N}_0, \\ \text{and } \operatorname{supp} \theta_{\ell,k_1,k_2} \subset D\big\}.$$

As in [6], we improved the quantitative performance by choosing parameters not strictly to theory. The adaptive scheme is tested with $\alpha = \frac{1}{2}, \frac{2}{3}, 1, 2$ for both $d = 1$ and $d = 2$. We take $f \equiv 1$ and $c = \frac{1}{10}$ in (49), (50). The parameters of the scheme are chosen as $\omega_0 = \frac{1}{10}$, $C_B = \frac{1}{50}$, $c_\Psi = C_\Psi = c_P = C_P = 1$, and the relative accuracy $\rho = \frac{1}{50}$. To choose an adequate parameter $\hat{J}$, we tested different ranges of refinement, until we found no qualitative differences compared to higher values of $\hat{J}$. This resulted in $\hat{J} = 2$ for $d = 1$ and $\hat{J} = 1$ for $d = 2$. The results of the numerical tests are shown in Figure 3 for $d = 1$ and in Figure 4 for $d = 2$. They are compared to the expected approximation rates (10) seen as dashed lines. Remarkably, for $d = 1$ the rate for degrees of freedom even resembles the expected limiting approximation rates for stochastic variables α instead of $\frac{2}{3}\alpha$. In the case $d = 2$, the limiting approximation rates $\frac{1}{2}\min\{\alpha, 1\}$ expected for piecewise linear spatial approximations are recovered by the adaptive method.

6 Conclusions and Outlook

We have constructed a novel adaptive stochastic Galerkin finite element method that guarantees a reduction of the error in energy norm in every step of the adaptive scheme while using an independently refined spatial mesh for each product orthonormal polynomial coefficient of the solution. All operations

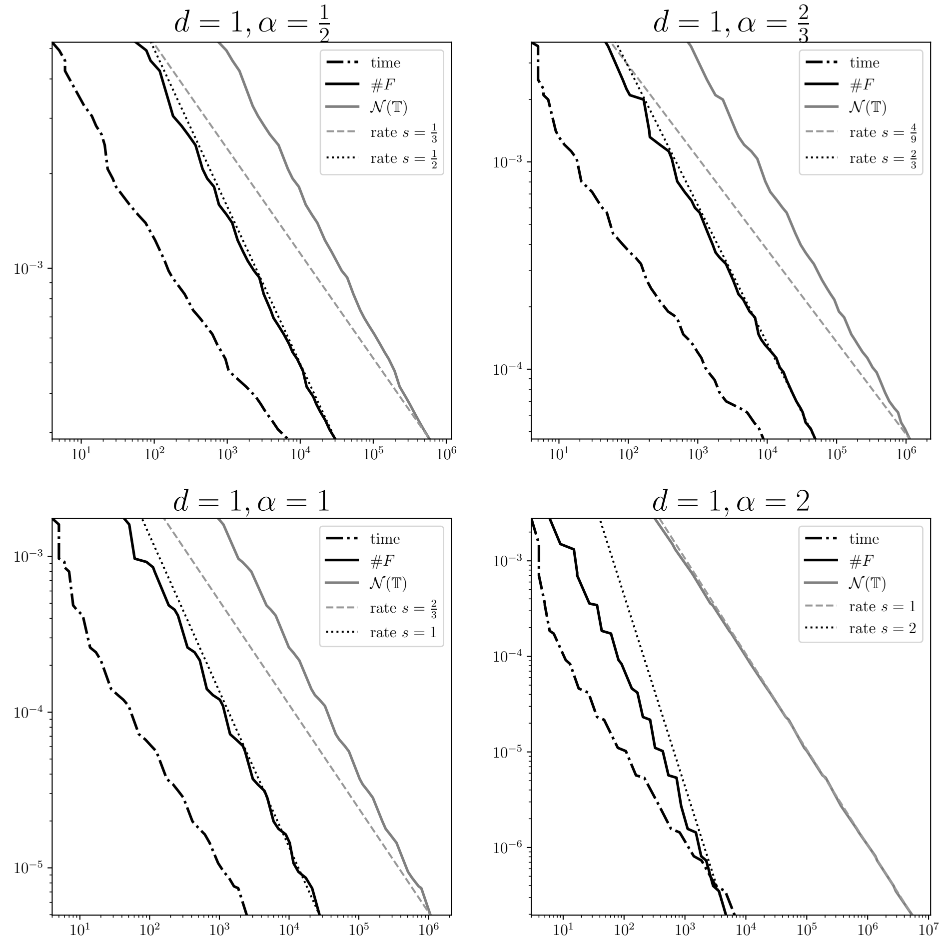

Figure 3: Computed residual bounds for $d = 1$ as a function of total number of degrees of freedom $\dim \mathcal{V}(\mathbb{T}) \eqsim N(\mathbb{T})$ of the current approximation of u (solid grey lines), degrees of stochastic freedom $\#F$ (solid black lines), and elapsed computation time in seconds (dash-dotted line).

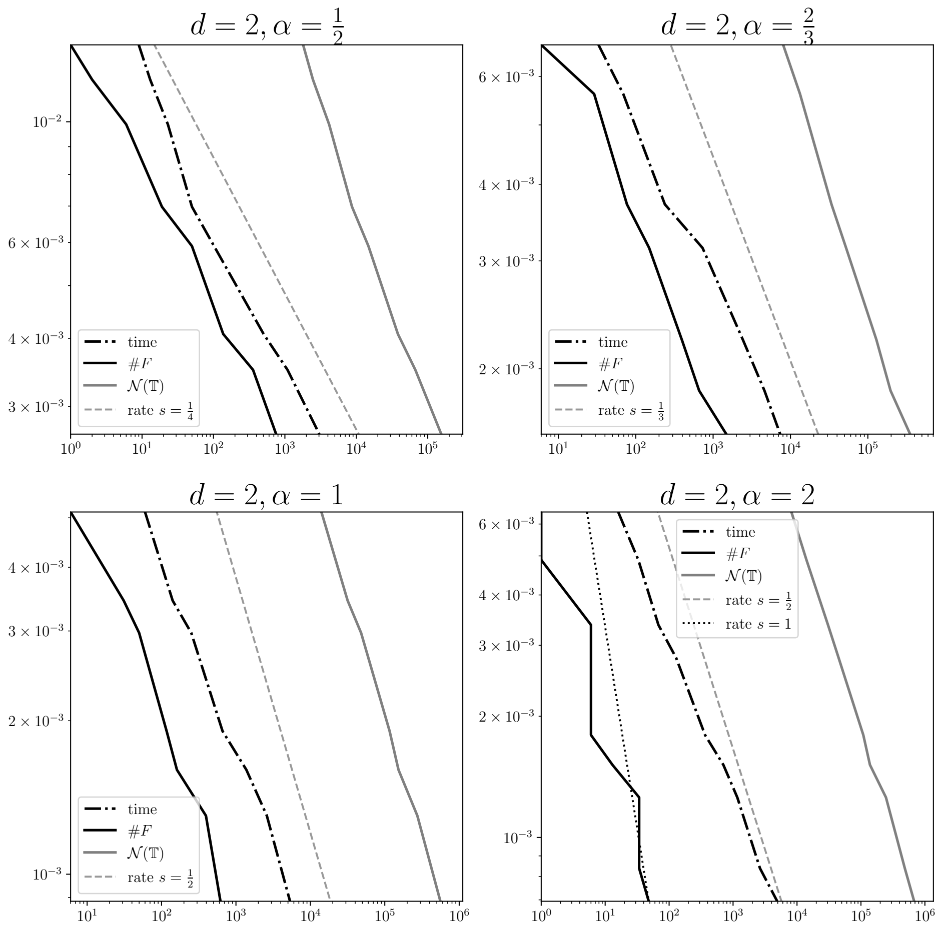

Figure 4: Computed residual bounds for $d = 2$ as a function of total number of degrees of freedom $\dim \mathcal{V}(\mathbb{T}) \eqsim N(\mathbb{T})$ of the current approximation of u (solid grey lines), number of product Legendre polynomials $\#F$ (solid black lines), and elapsed computation time in seconds (dash-dotted line).

that need to be performed also have costs that are consistent with total computational costs scaling linearly (up to logarithmic factors) with respect to the total number of degrees of freedom $N(\mathbb{T})$. Such scaling of the costs is also visible in the numerical experiments. What is thus left for future work is to show that the number of new degrees of freedom that is added in each iteration of the adaptive scheme is quasi-optimal, which will then yield optimal computational complexity of the method.

As for the earlier wavelet-based method in [6], our numerical tests confirm the approximability results (10) obtained in [2]. For $d = 1$ with spatial approximation by piecewise linear finite elements, we observe a new effect going beyond the existing tests, where for the reachable accuracies we obtain a better convergence rate than ensured by the theory (and observed for higher-order wavelets in [6]). This may be related to the particular piecewise linear structures that appear in the exact solution for $d = 1$ (see Figure 1), but this effect is yet to be understood in detail.

Finally, let us note that although we have conducted the analysis of the method for the case $d = 2$ for simplicity, everything that we have done can be generalized immediately to $d > 2$, but the practical implementation in the case $d = 3$ poses some further technical challenges.

References

[1] M. Bachmayr, A. Cohen, and W. Dahmen, *Parametric PDEs: Sparse or low-rank approximations?*, IMA Journal of Numerical Analysis **38** (2018), no. 4, 1661–1708.

[2] M. Bachmayr, A. Cohen, D. Dũng, and Ch. Schwab, *Fully discrete approximation of parametric and stochastic elliptic PDEs*, SIAM Journal on Numerical Analysis **55** (2017), 2151–2186.

[3] M. Bachmayr, A. Cohen, and G. Migliorati, *Sparse polynomial approximation of parametric elliptic PDEs. Part I: affine coefficients*, ESAIM. Mathematical Modelling and Numerical Analysis **51** (2017), no. 1, 321–339.

[4] ———, *Representations of Gaussian random fields and approximation of elliptic PDEs with lognormal coefficients*, J. Fourier Anal. Appl. **24** (2018), 621–649.

[5] M. Bachmayr, H. Eisenmann, and I. Voulis, *Julia implementation of an adaptive stochastic Galerkin finite element method with multilevel representations of random coefficients*, 2024, available at https://git.rwth-aachen.de/bachmayr/asgfemmlrrc.git.

[6] M. Bachmayr and I. Voulis, *An adaptive stochastic Galerkin method based on multilevel expansions of random fields: convergence and optimality*, ESAIM Math. Model. Numer. Anal. **56** (2022), no. 6, 1955–1992.

[7] A. Bespalov, D. Praetorius, and M. Ruggeri, *Convergence and rate optimality of adaptive multilevel stochastic Galerkin FEM*, IMA J. Numer. Anal. **42** (2022), no. 3, 2190–2213.

[8] P. Binev, *Tree approximation for hp-adaptivity*, SIAM Journal on Numerical Analysis **56** (2018), no. 6, 3346–3357.

[9] P. Binev, W. Dahmen, and R. DeVore, *Adaptive finite element methods with convergence rates*, Numerische Mathematik **97** (2004), no. 2, 219–268.

[10] P. Binev and R. DeVore, *Fast computation in adaptive tree approximation*, Numerische Mathematik **97** (2004), no. 2, 193–217.

[11] J. H. Bramble, J. E. Pasciak, and J. Xu, *Parallel multilevel preconditioners*, Math. Comp. **55** (1990), no. 191, 1–22.

[12] 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), no. 5, 2524–2550.

[13] L. Chen, R. H. Nochetto, and J. Xu, *Optimal multilevel methods for graded bisection grids*, Numer. Math. **120** (2012), no. 1, 1–34.

[14] Z. Ciesielski, *Hölder conditions for realizations of Gaussian processes*, Trans. Amer. Math. Soc. **99** (1961), 403–413. MR 132591

[15] A. Cohen and R. DeVore, *Approximation of high-dimensional parametric PDEs*, Acta Numerica **24** (2015), 1–159.

[16] A. Cohen, R. DeVore, and R. H. Nochetto, *Convergence rates of AFEM with H^{-1} data*, Found. Comput. Math. **12** (2012), no. 5, 671–718.

[17] A. Cohen, R. DeVore, and Ch. Schwab, *Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDE's*, Analysis and Applications **9** (2011), no. 1, 11–47.

[18] A. J. Crowder, C. E. Powell, and A. Bespalov, *Efficient adaptive multilevel stochastic Galerkin approximation using implicit a posteriori error estimation*, SIAM Journal on Scientific Computing **41** (2019), no. 3, A1681–A1705.

[19] M. Eigel, C. J. Gittelson, Ch. Schwab, and E. Zander, *Adaptive stochastic Galerkin FEM*, Computer Methods in Applied Mechanics and Engineering **270** (2014), 247–269.

[20] ______, *A convergent adaptive stochastic Galerkin finite element method with quasi-optimal spatial meshes*, ESAIM: Mathematical Modelling and Numerical Analysis **49** (2015), no. 5, 1367–1398.

[21] C. J. Gittelson, *An adaptive stochastic Galerkin method for random elliptic operators*, Mathematics of Computation **82** (2013), 1515–1541.

[22] H. Harbrecht and R. Schneider, *A note on multilevel based error estimation*, Computational Methods in Applied Mathematics **16** (2016), no. 3, 447–458.

[23] H. Harbrecht, R. Schneider, and C. Schwab, *Multilevel frames for sparse tensor product spaces*, Numerische Mathematik **110** (2008), no. 2, 199–220.

[24] P. Oswald, *Multilevel finite element approximation*, Teubner Skripten zur Numerik. [Teubner Scripts on Numerical Mathematics], B. G. Teubner, Stuttgart, 1994, Theory and applications.

[25] Ch. Schwab and C. J. Gittelson, *Sparse tensor discretizations of high-dimensional parametric and stochastic PDEs*, Acta Numerica **20** (2011), 291–467.

[26] R. Stevenson, *Optimality of a standard adaptive finite element method*, Foundations of Computational Mathematics **7** (2007), no. 2, 245–269.

[27] R. Verfürth, *A posteriori error estimation techniques for finite element methods*, Oxford University Press, 2013.