UNIFORMLY CONVERGENT ADAPTIVE METHODS FOR A CLASS OF PARAMETRIC OPERATOR EQUATIONS *

CLAUDE JEFFREY GITTELSON[1,2]

Abstract. We derive and analyze adaptive solvers for boundary value problems in which the differential operator depends affinely on a sequence of parameters. These methods converge uniformly in the parameters and provide an upper bound for the maximal error. Numerical computations indicate that they are more efficient than similar methods that control the error in a mean square sense.

1. INTRODUCTION

Boundary value problems with unknown coefficients can be interpreted as parametric equations, in which the unknown coefficients are permitted to depend on a sequence of scalar parameters. It may be possible to interpret these parameters as random variables, in which case the solution to the boundary value problem is a random field. In this probabilistic setting, Galerkin methods have been developed for approximating this random field in a parametric form, see [2,5,6,15,18,27,36–38,43]. Other approaches include collocation methods, see [3, 30, 39, 40, 42] and sampling methods such as quasi-Monte Carlo, see [25]. We refer to [21, 32, 41] for overviews.

These methods generally require strong assumptions on the probability distribution of the random coefficients. In particular, it is often assumed that the scalar parameters, *e.g.* coming from a series expansion of the unknown coefficients, are independent. This assumption is fundamental to the construction of polynomial chaos bases. To cover the more realistic setting of non-independent parameters, an auxiliary measure is introduced *e.g.* in [3,30], but this still requires quite elusive assumptions on the probability distribution.

Our goal is to compute a parametric representation of the solution that is reliable to a given accuracy on the entire parameter domain. This is particularly useful if insufficient statistical data is available to model unknown coefficients as random variables.

[1] Seminar for Applied Mathematics, ETH Zurich, Rämistrasse 101, 8092 Zurich, Switzerland.
claude.gittelson@sam.math.ethz.ch

[2] Department of Mathematics, Purdue University, 150 N. University Street, West Lafayette, 47907 IN, USA

Also, uniform convergence precludes any elusive assumptions on a probability distribution since it implies mean square convergence with respect to any probability measure on the parameter domain, in particular with respect to whatever distribution is deemed physical. A reliable parametric representation could be combined with Monte Carlo sampling in order to compute probabilistic quantities. Instead of solving the boundary value problem independently at every sample point, one can evaluate the parametric representation of the solution, which is generally much faster, see *e.g.* [39].

We consider equations with linear operators that depend in an affine manner on a sequence of scalar parameters. Such parametric operators arise if one or multiple coefficients in a boundary value problem are expanded in a series. We assume that each scalar parameter lies in a bounded interval. For unbounded parameter domains, uniform convergence of polynomial approximations can generally not be expected.

The main difficulty in applying stochastic Galerkin and other spectral methods is the construction of suitable spaces in which to compute approximate solutions. In [23,24], we suggest adaptive methods based on techniques from the adaptive wavelet algorithms [9,10,16,19]. We use an orthonormal polynomial basis on the parameter domain in place of wavelets. An arbitrary discretization of the physical domain can be used to approximate the coefficients of the random solution with respect to this basis.

In order to ensure uniform convergence in the parameter, we deviate a bit further from adaptive wavelet methods, which are formulated in a Hilbert space setting. We follow the approach in [10,24], which is based on applying an iterative method directly to the full parametric boundary value problem. Individual substeps of this iteration, such as application of the parametric operator, are replaced by approximate counterparts, realized by suitable adaptive algorithms. These keep track of errors entering the computation, ensuring convergence of the algorithm, and providing an upper bound on the error of the approximate solution.

In Section 2, we study parametric operator equations in an abstract setting. We show that the parametric solution depends continuously on the parameter for arbitrary parameter domains under mild continuity assumptions on the operator and right hand side. Consequently, the solution is uniformly bounded if the parameter domain is compact.

In the setting that the operator has a dominant nonparametric component, we derive a perturbed stationary linear iteration, which forms the basis for our adaptive method. A similar iteration is proposed in [1,21]. We also present an illustrative example for a parametric boundary value problem which motivates the affine dependence on a sequence of parameters that we later assume.

Our method is formulated on the level of coefficients with respect to a polynomial basis. In Section 3, we apply the Stone–Weierstrass theorem to show that continuous functions can be approximated uniformly by polynomials in an infinite dimensional setting. We construct suitable polynomial bases, and represent a class of parametric operators in these bases.

We present our adaptive method in Section 4. A vital component is an adaptive routine for applying the parametric operator, which is discussed in Section 4.1. In Section 5, we present a variant of our adaptive solver which has the potential to reduce the computational cost while maintaining the same accuracy.

In Section 6, we apply these adaptive solvers to a simple model problem. Numerical computations demonstrate the convergence of the algorithms and compare them to the adaptive methods from [23,24]. We compare observed convergence rates to a revised form of the approximation results [11,12].

2. Parametric operator equations

2.1. Continuous parameter dependence

Let V and W be Banach spaces over $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. We denote by W^* the space of bounded antilinear maps from W to $\mathbb{K}$, which are just the linear maps if $\mathbb{K} = \mathbb{R}$, and by $\mathcal{L}(V, W^*)$ the Banach space of bounded linear maps from V to W^* with the operator norm $\|\cdot\|_{V \to W^*}$. Similarly, $\mathcal{L}(V)$ denotes the space of bounded linear maps from V to itself, which constitutes a Banach algebra whose multiplicative group consists of all invertible bounded linear maps on V.

Let Γ be a nonempty topological space, such as any nonempty subset of $\mathbb{R}^n$, or, as we shall later assume, the infinite-dimensional cube $[-1,1]^\infty$. A parametric linear operator from V to W^* with parameter domain Γ is a continuous map

$$A\colon \Gamma \to \mathcal{L}(V, W^*), \quad y \mapsto A(y). \tag{2.1}$$

For a given $f\colon \Gamma \to W^*$, we are interested in determining $u\colon \Gamma \to V$ such that

$$A(y)u(y) = f(y) \qquad \forall y \in \Gamma. \tag{2.2}$$

Assumption 2.A. *$A(y)$ is bijective for all $y \in \Gamma$.*

By the open mapping theorem, bijective continuous linear operators have continuous inverses, and thus Assumption 2.A implies that $A(y)$ is boundedly invertible for all $y \in \Gamma$. The map $y \mapsto A(y)$ is continuous by definition, but no continuity assumptions are made on $y \mapsto A^{-1}(y)$. The following theorem hinges on the fact that continuity of $y \mapsto A^{-1}(y)$ follows from continuity of $y \mapsto A(y)$ with no further assumptions.

Theorem 2.1. *Equation* (2.2) *has a unique solution* $u\colon \Gamma \to V$. *It is continuous if and only if* $f\colon \Gamma \to W^*$ *is continuous.*

Proof. By Assumption 2.A, (2.2) has the unique solution $u(y) = A(y)^{-1}f(y)$. If $y \mapsto u(y)$ is continuous, then since $y \mapsto A(y)$ is continuous by definition, it follows that $y \mapsto f(y)$ is continuous because

$$\mathrm{mult}\colon \mathcal{L}(V, W^*) \times V \to W^*, \quad \mathrm{mult}(T, z) := Tz,$$

is continuous, and $f(y) = \mathrm{mult}(A(y), u(y))$ is a composition of continuous maps.

If $y \mapsto f(y)$ is continuous, then the same argument shows that $y \mapsto u(y) = \mathrm{mult}(A(y)^{-1}, f(y))$ is continuous, provided that $y \mapsto A^{-1}(y)$ is continuous. To show this, we select a boundedly invertible $D \in \mathcal{L}(V, W^*)$; for example, D could be equal to $A(y)$ for some $y \in \Gamma$. Then $y \mapsto D^{-1}A(y)$ is a continuous map from Γ into $\mathcal{L}(V)$. By the abstract property [26], Proposition 3.1.6, of Banach algebras, the map $T \mapsto \mathrm{inv}(T) := T^{-1}$ defined on the multiplicative group of $\mathcal{L}(V)$ is continuous in the topology of $\mathcal{L}(V)$. Therefore, $y \mapsto \mathrm{inv}(D^{-1}A(y)) = A(y)^{-1}D$ is continuous, and multiplying from the right by the constant D^{-1}, it follows that $y \mapsto A(y)^{-1}$ is a continuous map from Γ to $\mathcal{L}(W^*, V)$. $\qquad \square$

Example 2.2. Assumption 2.A is assured to hold if $A(y)$ is a perturbation of a boundedly invertible $D \in \mathcal{L}(V, W^*)$, *i.e.*

$$A(y) = D + R(y), \quad y \in \Gamma, \tag{2.3}$$

with a continuous $y \mapsto R(y) \in \mathcal{L}(V, W^*)$ satisfying

$$\left\| D^{-1}R(y) \right\|_{V \to V} \le \gamma < 1 \qquad \forall y \in \Gamma. \tag{2.4}$$

Then $A(y)$ can be decomposed as

$$A(y) = D(\mathrm{id}_V + D^{-1}R(y)), \quad y \in \Gamma, \tag{2.5}$$

and consequently, using a Neumann series in $\mathcal{L}(V)$ to invert the second factor,

$$A(y)^{-1} = \left(\sum_{n=0}^{\infty} \left(-D^{-1}R(y) \right)^n \right) D^{-1}, \quad y \in \Gamma. \tag{2.6}$$

In this setting, due to (2.4)–(2.6), the parametric operators $A(y)$ and $A(y)^{-1}$ are uniformly bounded,

$$\|A(y)\|_{V \to W^*} \le \|D\|_{V \to W^*} (1 + \gamma) \qquad \forall y \in \Gamma, \tag{2.7}$$

$$\left\| A(y)^{-1} \right\|_{W^* \to V} \le \frac{\left\| D^{-1} \right\|_{W^* \to V}}{1 - \gamma} \qquad \forall y \in \Gamma. \tag{2.8}$$

A sufficient condition for (2.4) is

$$\|R(y)\|_{V \to W^*} \le \frac{\gamma}{\|D^{-1}\|_{W^* \to V}} \qquad \forall y \in \Gamma \tag{2.9}$$

with $\gamma < 1$. Equation (2.9) does not depend on the precise structure of the operators $D^{-1}R(y)$. Therefore, Assumption 2.A is always satisfied if the parametric component $R(y)$ of $A(y)$ is sufficiently small.

Assumption 2.B. *Γ is a compact Hausdorff space.*

For example, Γ may be any compact metric space, any closed bounded subset of $\mathbb{R}^n$, or any cartesian product of such sets.

Lemma 2.3. *There exist constants $\hat{c}, \check{c} \in \mathbb{R}$ such that*

$$\|A(y)\|_{V \to W^*} \le \hat{c} \qquad and \qquad \left\|A(y)^{-1}\right\|_{W^* \to V} \le \check{c} \qquad \forall y \in \Gamma. \tag{2.10}$$

Proof. By assumption, the map $y \mapsto A(y)$ is continuous. As shown in the proof of Theorem 2.1, $y \mapsto A(y)^{-1}$ is also continuous. Consequently, the maps $y \mapsto \|A(y)\|_{V \to W^*}$ and $y \mapsto \left\|A(y)^{-1}\right\|_{W^* \to V}$ are continuous maps from Γ into $\mathbb{R}$. Since Γ is compact by Assumption 2.B, the ranges of these maps are compact in $\mathbb{R}$, and therefore bounded. $\qquad\square$

For any Banach space X, let $C(\Gamma; X)$ denote the Banach space of continuous maps from Γ to X with norm

$$\|v\|_{C(\Gamma;X)} \coloneqq \max_{y \in \Gamma} \|v(y)\|_X, \quad v \in C(\Gamma; X). \tag{2.11}$$

For any Borel probability measure π on Γ, $p \ge 1$, and any $v \in C(\Gamma; X)$,

$$\|v\|_{L_\pi^p(\Gamma;X)} \le \|v\|_{C(\Gamma;X)} \tag{2.12}$$

and equality holds in (2.12) *e.g.* if π is a Dirac measure at a maximum of $\|v\|_X$. Consequently, estimates in $C(\Gamma; X)$ carry over to $L_\pi^p(\Gamma; X)$ for all π, although they may be too conservative for any particular π. In what follows, we abbreviate $C(\Gamma) \coloneqq C(\Gamma; \mathbb{K})$.

Corollary 2.4. *The operators*

$$\mathcal{A} \colon C(\Gamma; V) \to C(\Gamma; W^*), \quad v \mapsto [y \mapsto A(y)v(y)] \qquad and \tag{2.13}$$
$$\mathcal{A}^{-1} \colon C(\Gamma; W^*) \to C(\Gamma; V), \quad g \mapsto [y \mapsto A(y)^{-1}g(y)] \tag{2.14}$$

are well-defined, inverse to each other, and bounded with norms $\|\mathcal{A}\| \le \hat{c}$ and $\left\|\mathcal{A}^{-1}\right\| \le \check{c}$.

Proof. The assertion is a direct consequence of Theorem 2.1 and Lemma 2.3. $\qquad\square$

2.2. A perturbed linear iteration

We consider the setting of Example 2.2, *i.e.* $\mathcal{A}$ is a sum

$$\mathcal{A} = \mathcal{D} + \mathcal{R} \tag{2.15}$$

with $\mathcal{D} \colon C(\Gamma; V) \to C(\Gamma; W^*)$ of the form $(\mathcal{D}v)(y) = Dv(y)$ for a boundedly invertible $D \in \mathcal{L}(V, W^*)$, and $\mathcal{R} \colon C(\Gamma; V) \to C(\Gamma; W^*)$ satisfies

$$\left\|\mathcal{D}^{-1}\mathcal{R}\right\|_{C(\Gamma;V) \to C(\Gamma;V)} \le \gamma < 1. \tag{2.16}$$

Condition (2.16) implies that $\mathcal{D}^{-1}\mathcal{A}$ can be inverted by a Neumann series in $C(\Gamma;V)$, and

$$\left\|(\mathcal{D}^{-1}\mathcal{A})^{-1}\right\|_{C(\Gamma;V)\to C(\Gamma;V)} \leq \frac{1}{1-\gamma}. \tag{2.17}$$

Therefore, for any $f \in C(\Gamma;W^*)$, the solution u of the operator equation

$$\mathcal{A}u = f \tag{2.18}$$

is the limit of the sequence $(u_k)_{k=0}^{\infty}$ with arbitrary $u_0 \in C(\Gamma;V)$ and

$$u_k := \mathcal{D}^{-1}(f - \mathcal{R}u_{k-1}), \quad k \in \mathbb{N}. \tag{2.19}$$

This iteration effectively computes truncated Neumann expansions of $\mathcal{D}^{-1}\mathcal{A}$ applied to $\mathcal{D}^{-1}f$. We generalize (2.19) by allowing errors in the computation of f, the evaluation of $\mathcal{R}$, and the inversion of $\mathcal{D}$.

Let $\tilde{u}_0 \in C(\Gamma;V)$ be an arbitrary approximation of u with a known upper bound δ_0 for the error $\|u - \tilde{u}_0\|_{C(\Gamma;V)}$. For example, if $\tilde{u}_0 = 0$, we may set

$$\delta_0 := \frac{\left\|D^{-1}\right\|_{W^*\to V}}{1-\gamma}\|f\|_{C(\Gamma;W^*)}. \tag{2.20}$$

We define the approximations $\tilde{u}_k$ together with error bounds δ_k recursively through

$$\delta_k := (\alpha + \beta + \gamma)\delta_{k-1} \tag{2.21}$$

with parameters $\alpha, \beta \geq 0$, and $\tilde{u}_k$ may be any element of $C(\Gamma;V)$ with

$$\left\|\tilde{u}_k - \mathcal{D}^{-1}g_k\right\|_{C(\Gamma;V)} \leq \alpha\delta_{k-1} \tag{2.22}$$

for some $g_k \in C(\Gamma;W^*)$ satisfying

$$\|g_k - (f - \mathcal{R}\tilde{u}_{k-1})\|_{C(\Gamma;W^*)} \leq \beta\delta_{k-1}\left\|D^{-1}\right\|_{W^*\to V}^{-1}. \tag{2.23}$$

Theorem 2.5. *If $\alpha + \beta < 1 - \gamma$, then $\tilde{u}_k \to u$ in $C(\Gamma;V)$, and*

$$\|u - \tilde{u}_k\|_{C(\Gamma;V)} \leq \delta_k = (\alpha + \beta + \gamma)^k \delta_0 \qquad \forall k \in \mathbb{N}_0. \tag{2.24}$$

Proof. Since $\mathcal{D}u = f - \mathcal{R}u$,

$$u - \tilde{u}_k = \mathcal{D}^{-1}(f - \mathcal{R}u) - \mathcal{D}^{-1}(f - \mathcal{R}\tilde{u}_{k-1}) + \mathcal{D}^{-1}(f - \mathcal{R}\tilde{u}_{k-1} - g_k) + \mathcal{D}^{-1}g_k - \tilde{u}_k.$$

By triangle inequality,

$$\begin{aligned}
\|u - \tilde{u}_k\|_{C(\Gamma;V)} &\leq \left\|\mathcal{D}^{-1}\mathcal{R}(u - \tilde{u}_{k-1})\right\|_{C(\Gamma;V)} + \left\|D^{-1}\right\|_{W^*\to V}\|g_k - (f - \mathcal{R}\tilde{u}_{k-1})\|_{C(\Gamma;W^*)} \\
&\quad + \left\|\tilde{u}_k - \mathcal{D}^{-1}g_k\right\|_{C(\Gamma;V)} \\
&\leq \gamma\|u - \tilde{u}_{k-1}\|_{C(\Gamma;V)} + \beta\delta_{k-1} + \alpha\delta_{k-1},
\end{aligned}$$

and the claim follows by induction using $\alpha + \beta + \gamma < 1$. $\qquad\square$

Remark 2.6. Theorem 2.5 uses a priori known quantities $\delta_k = (\alpha + \beta + \gamma)^k \delta_0$ as upper bounds for the error at iteration $k \in \mathbb{N}_0$. However, better estimates may be available or computable during an iteration. The residual at iteration $k \in \mathbb{N}_0$ is given by

$$r_k := f - \mathcal{A}\tilde{u}_k = \mathcal{A}(u - \tilde{u}_k) \in C(\Gamma;W^*). \tag{2.25}$$

Since $\mathcal{A}$ is invertible by a Neumann series,

$$\|u - \tilde{u}_k\|_{C(\Gamma;V)} \le \|\mathcal{A}^{-1}\|_{C(\Gamma;W^*)\to C(\Gamma;V)} \|r_k\|_{C(\Gamma;W^*)} \le \frac{1}{1-\gamma} \|D^{-1}\|_{W^*\to V} \|r_k\|_{C(\Gamma;W^*)}. \tag{2.26}$$

Therefore, if it is known that $\|r_k\|_{C(\Gamma;W^*)} \le \rho_k$, we also have the upper bound

$$\bar{\delta}_k := \frac{1}{1-\gamma} \|D^{-1}\|_{W^*\to V} \rho_k \tag{2.27}$$

of $\|u - \tilde{u}_k\|_{C(\Gamma;V)}$ for all $k \in \mathbb{N}$.

2.3. The parametric diffusion equation

As an illustrative example, we consider the isotropic diffusion equation on a bounded Lipschitz domain $G \subset \mathbb{R}^d$ with homogeneous Dirichlet boundary conditions. For any uniformly positive $a \in L^\infty(G)$ and any $f \in L^2(G)$, we have

$$-\nabla \cdot (a(x)\nabla u(x)) = f(x), \quad x \in G, \quad u(x) = 0, \quad x \in \partial G. \tag{2.28}$$

We model a as a parametric coefficient, depending affinely on a sequence of scalar parameters in $[-1,1]$. For the compact parameter domain $\Gamma := [-1,1]^\infty$, we have

$$a(y,x) := \bar{a}(x) + \sum_{m=1}^{\infty} y_m a_m(x), \quad y = (y_m)_{m=1}^{\infty} \in \Gamma. \tag{2.29}$$

Thus the parameters y_m are coefficients in a series expansion of $a(y,x) - \bar{a}(x)$. We note that the essential assumption on the parameters y_m is that they are bounded; any bounded parameters can be shifted and scaled to be in $[-1,1]$, preserving the structure of (2.29).

We define the parametric operator

$$A(y)\colon H_0^1(G) \to H^{-1}(G), \quad v \mapsto -\nabla \cdot (a(y)\nabla v), \quad y \in \Gamma. \tag{2.30}$$

By linearity, we can expand $A(y)$ as

$$A(y) = D + R(y), \quad R(y) := \sum_{m=1}^{\infty} y_m R_m \qquad \forall y \in \Gamma, \tag{2.31}$$

for

$$\begin{aligned} D\colon H_0^1(G) \to H^{-1}(G), \quad &v \mapsto -\nabla \cdot (\bar{a}\nabla v), \\ R_m\colon H_0^1(G) \to H^{-1}(G), \quad &v \mapsto -\nabla \cdot (a_m\nabla v), \quad m \in \mathbb{N}. \end{aligned}$$

Note that $\|R_m\|_{H_0^1(G)\to H^{-1}(G)} \le \|a_m\|_{L^\infty(G)}$, and thus convergence in (2.31) and (2.29) is assured if the sequence $(\|a_m\|_{L^\infty(G)})_{m=1}^{\infty}$ is summable.

Assuming that $\bar{a}$ is bounded and uniformly positive, the operator D is invertible with

$$\|D^{-1}\|_{H^{-1}(G)\to H_0^1(G)} \le \left(\operatorname*{ess\,inf}_{x\in G} \bar{a}(x) \right)^{-1}. \tag{2.32}$$

3. Polynomial expansion

3.1. Uniform approximation by polynomials in infinite dimensions

We consider polynomials on the compact domain $\Gamma := [-1,1]^\infty$. We denote a generic element of Γ by $y = (y_m)_{m=1}^\infty$.

For any finite set $F \subset \mathbb{N}$, let $\mathcal{P}_F(\Gamma)$ denote the vector space of polynomials in the variables $(y_m)_{m \in F}$. Then

$$\mathcal{P}(\Gamma) := \bigcup_{\substack{F \subset \mathbb{N} \\ \#F < \infty}} \mathcal{P}_F(\Gamma) \tag{3.1}$$

is the vector space of polynomials on the infinite dimensional domain Γ. The following statement is a direct consequence of the Stone–Weierstrass theorem, see *e.g.* [31, 34].

Theorem 3.1. *The space $\mathcal{P}(\Gamma)$ is dense in $C(\Gamma)$.*

A Banach space X is said to have the approximation property if, for every compact set $K \subset X$ and every $\epsilon > 0$, there is a finite rank operator $T \in \mathcal{L}(X)$ such that

$$\|x - Tx\|_X \le \epsilon \quad \forall x \in K. \tag{3.2}$$

We recall that every space with a Schauder basis has the approximation property since T can be chosen as

$$Tx = \sum_{n=1}^N x_n e_n \qquad \text{for} \qquad x = \sum_{n=1}^\infty x_n e_n, \tag{3.3}$$

for a sufficiently large N depending on K, where $(e_n)_{n=1}^\infty$ denotes the Schauder basis of X, *i.e.* every $x \in X$ has a unique expansion of the form (3.3). In particular, every separable Hilbert space has the approximation property since orthonormal bases are Schauder bases.

Theorem 3.1 extends to functions with values in X under the assumption that X has the approximation property. For any finite set $F \subset \mathbb{N}$, let $\mathcal{P}_F(\Gamma; X)$ denote the vector space of polynomials in the variables $(y_m)_{m \in F}$ with coefficients in X. As in (3.1), we define

$$\mathcal{P}(\Gamma; X) := \bigcup_{\substack{F \subset \mathbb{N} \\ \#F < \infty}} \mathcal{P}_F(\Gamma; X). \tag{3.4}$$

Theorem 3.2. *If X has the approximation property, then $\mathcal{P}(\Gamma; X)$ is dense in $C(\Gamma; X)$.*

Proof. Let $f \in C(\Gamma; X)$ and $\epsilon > 0$. Since Γ is compact, $K := f(\Gamma) \subset X$ is compact, and thus there is a finite rank operator $T \in \mathcal{L}(X)$ such that (3.2) holds. We write T as

$$Tx = \sum_{i=1}^n \psi_i(x) x_i$$

with $\psi_i \in X^*$ and $x_i \in X$, scaled such that $\|x_i\|_X = 1$. Since each of the maps $\psi_i \circ f$ is in $C(\Gamma)$, Theorem 3.1 implies that there is a polynomial $p_i \in \mathcal{P}(\Gamma)$ with $|p_i(y) - \psi(f(y))| \le \epsilon/n$ for all $y \in \Gamma$. Consequently, for all $y \in \Gamma$,

$$\left\| f(y) - \sum_{i=1}^n p_i(y) x_i \right\|_X \le \|f(y) - Tf(y)\|_X + \sum_{i=1}^n |\psi_i(f(y)) - p_i(y)| \, \|x_i\|_X \le 2\epsilon. \qquad \square$$

3.2. Polynomial systems in infinite dimensions

Let $(P_n)_{n=0}^\infty$ be a sequence of polynomials on $[-1,1]$ satisfying $P_0(\xi) = 1$, $P_1(\xi) = \xi$ and

$$\xi P_n(\xi) = \pi_n^+ P_{n+1}(\xi) + \pi_n^- P_{n-1}(\xi) \qquad \forall n \in \mathbb{N} \tag{3.5}$$

for all $\xi \in [-1,1]$. In particular, it follows by induction that P_n is a polynomial of degree n. We define $\pi_0^+ := 1$ in order to achieve $\xi P_0(\xi) = \pi_0^+ P_1(\xi)$.

For example, $P_n(\xi) = \xi^n$ if $\pi_n^+ = 1$ and $\pi_n^- = 0$ for all $n \in \mathbb{N}$. If π_n^+ and π_n^- are given by

$$\pi_n^+ := \frac{1}{2} \qquad \text{and} \qquad \pi_n^- := \frac{1}{2}, \tag{3.6}$$

then $(P_n)_{n=0}^\infty$ are Chebyshev polynomials of the first kind. Alternatively, the values

$$\pi_n^+ := \frac{n+1}{2n+1} \qquad \text{and} \qquad \pi_n^- := \frac{n}{2n+1} \tag{3.7}$$

lead to Legendre polynomials. More generally, identities of the type (3.5) follow from recursion formulas for families of orthonormal polynomials with respect to symmetric measures on $[-1,1]$, see *e.g.* [20, 35].

In all of the above examples,

$$|P_n(\xi)| \leq 1 \quad \forall \xi \in [-1,1], \quad \forall n \in \mathbb{N}_0. \tag{3.8}$$

We assume that the polynomials $(P_n)_{n=0}^\infty$ are scaled in such a way that (3.8) holds.

We define the set of finitely supported sequences in $\mathbb{N}_0$ as

$$\Lambda := \left\{ \mu \in \mathbb{N}_0^{\mathbb{N}} \,;\, \# \operatorname{supp} \mu < \infty \right\}, \tag{3.9}$$

where the support is defined by

$$\operatorname{supp} \mu := \{ m \in \mathbb{N} \,;\, \mu_m \neq 0 \}, \quad \mu \in \mathbb{N}_0^{\mathbb{N}}. \tag{3.10}$$

Then countably infinite tensor product polynomials are given by

$$(P_\mu)_{\mu \in \Lambda}, \quad P_\mu := \bigotimes_{m=1}^\infty P_{\mu_m}, \quad \mu \in \Lambda. \tag{3.11}$$

Note that each of these functions depends on only finitely many dimensions,

$$P_\mu(y) = \prod_{m=1}^\infty P_{\mu_m}(y_m) = \prod_{m \in \operatorname{supp} \mu} P_{\mu_m}(y_m), \quad \mu \in \Lambda, \tag{3.12}$$

since $P_0 = 1$.

Proposition 3.3. *If X is a Banach space with the approximation property, then for any $f \in C(\Gamma; X)$ and any $\epsilon > 0$, there is a finite set $\Xi \subset \Lambda$ and $x_\mu \in X$, $\mu \in \Xi$, such that*

$$\max_{y \in \Gamma} \left\| f(y) - \sum_{\mu \in \Xi} x_\mu P_\mu(y) \right\|_X \leq \epsilon. \tag{3.13}$$

Proof. The assertion follows from Theorem 3.2 since $(P_\mu)_{\mu \in \Lambda}$ is an algebraic basis of the vector space $\mathcal{P}(\Gamma; X)$. $\qquad \square$

3.3. Representation of a class of parametric operators

Let V and W be Banach spaces. Motivated by Example 2.2 and (2.31), we consider $\Gamma = [-1,1]^\infty$ and operators $\mathcal{A}\colon C(\Gamma;V) \to C(\Gamma;W^*)$ of the form $\mathcal{A} = \mathcal{D} + \mathcal{R}$ with $(\mathcal{D}v)(y) = Dv(y)$ for a boundedly invertible $D \in \mathcal{L}(V,W^*)$ and

$$(\mathcal{R}v)(y) := \sum_{m=1}^\infty y_m R_m v(y), \quad y \in \Gamma, \quad v \in C(\Gamma;V), \tag{3.14}$$

for $R_m \in \mathcal{L}(V,W^*)$ satisfying

$$\sum_{m=1}^\infty \left\| D^{-1}R_m \right\|_{V\to V} \le \gamma < 1. \tag{3.15}$$

We note that these assumptions imply (2.16). Truncating the series in (3.14), we approximate $\mathcal{R}$ by

$$(\mathcal{R}_{[M]}v)(y) := \sum_{m=1}^M y_m R_m v(y), \quad y \in \Gamma, \quad v \in C(\Gamma;V), \tag{3.16}$$

for $M \in \mathbb{N}$, and $\mathcal{R}_{[0]} := 0$.

Lemma 3.4. *For all $M \in \mathbb{N}_0$,*

$$\left\| \mathcal{R} - \mathcal{R}_{[M]} \right\|_{C(\Gamma;V)\to C(\Gamma;W^*)} \le \sum_{m=M+1}^\infty \|R_m\|_{V\to W^*}. \tag{3.17}$$

In particular, $\mathcal{R}_{[M]} \to \mathcal{R}$ in $\mathcal{L}(C(\Gamma;V), C(\Gamma;W^*))$.

Proof. For any $M \in \mathbb{N}_0$, $v \in C(\Gamma;V)$ and $y \in \Gamma$, since $|y_m| \le 1$,

$$\left\| (\mathcal{R}v)(y) - (\mathcal{R}_{[M]}v)(y) \right\|_{W^*} \le \sum_{m=M+1}^\infty |y_m| \, \|R_m v(y)\|_{W^*} \le \sum_{m=M+1}^\infty \|R_m\|_{V\to W^*} \, \|v\|_{C(\Gamma;V)}.$$

Furthermore, (3.15) implies that $(R_m)_{m\in\mathbb{N}} \in \ell^1(\mathbb{N}; \mathcal{L}(V,W^*))$. $\qquad\square$

According to the following statement, $\mathcal{R}_{[M]}$ maps $\mathcal{P}(\Gamma;V)$ into $\mathcal{P}(\Gamma;W^*)$. We determine the coefficients of $\mathcal{R}_{[M]}v$ in terms of those of $v \in \mathcal{P}(\Gamma;V)$ with respect to a polynomial basis $(P_\mu)_{\mu\in\Lambda}$ from Section 3.2.

Lemma 3.5. *For any $M \in \mathbb{N}$ and any $v \in \mathcal{P}(\Gamma;V)$, represented as*

$$v(y) = \sum_{\mu\in\Xi} v_\mu P_\mu(y), \quad y \in \Gamma, \tag{3.18}$$

for a finite set $\Xi \subset \Lambda$, $\mathcal{R}_{[M]}v \in \mathcal{P}(\Gamma;W^*)$ has the form

$$(\mathcal{R}_{[M]}v)(y) = \sum_{\mu\in\Xi} \sum_{m=1}^M R_m v_\mu \big(\pi_{\mu_m}^+ P_{\mu+\epsilon_m}(y) + \pi_{\mu_m}^- P_{\mu-\epsilon_m}(y) \big), \quad y \in \Gamma, \tag{3.19}$$

where $\epsilon_m \in \Lambda$ is the Kronecker sequence $(\epsilon_m)_n := \delta_{mn}$, and we set $P_\mu := 0$ if any $\mu_m < 0$.

Proof. By the definitions (3.16) and (3.18),

$$(\mathcal{R}_{[M]}v)(y) = \sum_{\mu\in\Xi} \sum_{m=1}^M R_m v_\mu y_m P_\mu(y).$$

Equation (3.5) implies

$$y_m P_\mu(y) = \pi_{\mu_m}^+ P_{\mu+\epsilon_m}(y) + \pi_{\mu_m}^- P_{\mu-\epsilon_m}(y). \qquad\square$$

Combining Proposition 3.3, Lemmas 3.4 and 3.5, one can represent $\mathcal{R}v$ as a limit of terms of the form (3.19) for any $v \in C(\Gamma;V)$, provided that V has the approximation property.

4. A UNIFORMLY CONVERGENT ADAPTIVE SOLVER

4.1. Adaptive application of parametric operators

We consider operators $\mathcal{R}$ of the form (3.14). For all $M \in \mathbb{N}$, let $\bar{e}_{\mathcal{R},M}$ be given such that

$$\left\| \mathcal{R} - \mathcal{R}_{[M]} \right\|_{C(\Gamma;V) \to C(\Gamma;W^*)} \leq \bar{e}_{\mathcal{R},M}. \tag{4.1}$$

For example, by Lemma 3.4, these bounds can be chosen as

$$\bar{e}_{\mathcal{R},M} := \sum_{m=M+1}^{\infty} \|R_m\|_{V \to W^*}, \tag{4.2}$$

or as estimates for these sums. We assume that $(\bar{e}_{\mathcal{R},M})_{M=0}^{\infty}$ is nonincreasing and converges to 0, and also that the sequence of differences $(\bar{e}_{\mathcal{R},M} - \bar{e}_{\mathcal{R},M+1})_{M=0}^{\infty}$ is nonincreasing. For (4.2), the latter property holds if R_m are arranged in decreasing order of $\|R_m\|_{V \to W^*}$.

Alternative values for $\bar{e}_{\mathcal{R},M}$ are provided by the following elementary estimate, which is a direct consequence of Lemma 3.4, see [23], Proposition 4.4.

Proposition 4.1. *Let $s > 0$. If either*

$$\|R_m\|_{V \to W^*} \leq s\delta_{\mathcal{R},s}(m+1)^{-s-1} \qquad \forall m \in \mathbb{N} \tag{4.3}$$

or the sequence $(\|R_m\|_{V \to W^*})_{m=1}^{\infty}$ is nonincreasing and

$$\left(\sum_{m=1}^{\infty} \|R_m\|_{V \to W^*}^{\frac{1}{s+1}} \right)^{s+1} \leq \delta_{\mathcal{R},s}, \tag{4.4}$$

then

$$\left\| \mathcal{R} - \mathcal{R}_{[M]} \right\|_{C(\Gamma;V) \to C(\Gamma;W^*)} \leq \delta_{\mathcal{R},s}(M+1)^{-s} \qquad \forall M \in \mathbb{N}_0. \tag{4.5}$$

A fundamental concept in the construction of the present adaptive application routine is that, given a vector, more effort should be invested into coefficients of large magnitude than into coefficients with small magnitude. This suggests sorting the coefficients as a first step, but exact sorting is an unnecessary luxury. We allow for an approximate sorting routine.

Given a vector $\boldsymbol{w} = (w_\mu)_{\mu \in \Lambda} \in \ell^1(\Lambda)$, let

$$\Lambda_p := \left\{ \mu \in \Lambda \, ; \, |w_\mu| \in (2^{-p} \|\boldsymbol{w}\|_{\ell^\infty}, 2^{-(p-1)} \|\boldsymbol{w}\|_{\ell^\infty}] \right\} \tag{4.6}$$

for all $p \in \mathbb{N}$, and let $\boldsymbol{w}_{\{p\}} := (w_\mu)_{\mu \in \Lambda_p}$ denote the restriction of $\boldsymbol{w}$ to Λ_p. The original vector $\boldsymbol{w}$ can be reconstructed as the sum of all $\boldsymbol{w}_{\{p\}}$, where each $\boldsymbol{w}_{\{p\}}$ is extended to Λ by zero on $\Lambda \setminus \Lambda_p$. The first few of these sets are constructed by the routine

$$\texttt{BucketSort}[\boldsymbol{w}, \epsilon] \mapsto (\Lambda_p)_{p=1}^{P}, \tag{4.7}$$

which ensures a tolerance of ϵ in $\ell^1(\Lambda)$ for the approximation $\boldsymbol{w} \approx \boldsymbol{w}_{\{1\}} + \cdots + \boldsymbol{w}_{\{P\}}$ by choosing P as the smallest integer with

$$2^{-P} \|\boldsymbol{w}\|_{\ell^\infty(\Lambda)} \# \operatorname{supp} \boldsymbol{w} \leq \epsilon, \tag{4.8}$$

see [4, 16, 19, 28]. By [19], Remark 2.3 or [16], Proposition 4.4, the number of operations and storage locations required by a call of $\texttt{BucketSort}[\boldsymbol{w}, \epsilon]$ is bounded by a multiple of

$$\# \operatorname{supp} \boldsymbol{w} + \max(1, \lceil \log(\|\boldsymbol{w}\|_{\ell^\infty(\Lambda)} (\# \operatorname{supp} \boldsymbol{w})/\epsilon) \rceil), \tag{4.9}$$

which is less than the computational cost of exact comparison-based sorting algorithms.

Let $\boldsymbol{v} = (v_\mu)_{\mu \in \Lambda}$ be a finitely supported sequence in V, indexed by Λ. Such a vector represents a polynomial $v \in \mathcal{P}(\Gamma; V)$ by

$$v(y) = \sum_{\mu \in \operatorname{supp} \boldsymbol{v}} v_\mu P_\mu(y), \quad y \in \Gamma, \tag{4.10}$$

where $\operatorname{supp} \boldsymbol{v} = \{\mu \in \Lambda \,;\, v_\mu \neq 0\}$ is a finite subset of Λ by assumption. Due to the normalization (3.8),

$$\|v\|_{C(\Gamma;V)} = \max_{y \in \Gamma} \|v(y)\|_V \leq \sum_{\mu \in \Lambda} \|v_\mu\|_V = \|\boldsymbol{v}\|_{\ell^1(\Lambda;V)}. \tag{4.11}$$

The routine $\mathtt{Apply}_{\mathcal{R}}[\boldsymbol{v}, \epsilon]$ adaptively approximates $\mathcal{R}v$ for $v \in \mathcal{P}(\Gamma; V)$ in three distinct steps. First, the elements of the coefficient vector $\boldsymbol{v}$ of v are grouped according to their norm using $\mathtt{BucketSort}$. This already discards indices $\mu \in \Lambda_p$ for $p > P$ with P as above. As this truncation may be too conservative, further sets Λ_p are discarded, and the final partitioning of $\boldsymbol{v}$ has a truncation error of at most $\delta \leq \epsilon/2$.

Next, an approximation $\mathcal{R}_{[M_p]}$ of $\mathcal{R}$ is assigned to each segment $\boldsymbol{v}_{\{p\}} = (v_\mu)_{\mu \in \Lambda_p}$ of $\boldsymbol{v}$. The construction of the parameters $(M_p)_{p=1}^\ell$ is iterative and employs a greedy strategy to minimize the cost while ensuring an accuracy of $\epsilon - \delta$. Until an estimate of the error reaches this tolerance, the algorithm repeatedly selects an M_p with a maximal ratio between the reduction in the error bound caused by incrementing M_p, and the additional cost of this refinement, and increments this M_p by one.

Finally, the operations selected in the previous two steps are performed, *i.e.* for each p, $\mathcal{R}_{[M_p]}$ is applied to $\boldsymbol{v}_{\{p\}}$. Each multiplication $R_m v_\mu$ is performed just once, and copied to the appropriate entries of $\boldsymbol{z}$. Then the polynomial

$$z(y) := \sum_{\mu \in \operatorname{supp} \boldsymbol{z}} z_\mu P_\mu(y), \quad y \in \Gamma, \tag{4.12}$$

is an approximation of $\mathcal{R}v$ with error at most ϵ.

Algorithm 1. $\mathtt{Apply}_{\mathcal{R}}[\boldsymbol{v}, \epsilon] \mapsto \boldsymbol{z}$

$(\Lambda_p)_{p=1}^P \longleftarrow \mathtt{BucketSort}\left[(\|v_\mu\|_V)_{\mu \in \Lambda}, \dfrac{\epsilon}{2\bar{e}_{\mathcal{R},0}} \right]$

for $p = 1, \ldots, P$ **do** $\boldsymbol{v}_{\{p\}} \longleftarrow (v_\mu)_{\mu \in \Lambda_p}$

Compute the minimal $\ell \in \{0, 1, \ldots, P\}$ s.t. $\delta := \bar{e}_{\mathcal{R},0} \left\| \boldsymbol{v} - \sum_{p=1}^\ell \boldsymbol{v}_{\{p\}} \right\|_{\ell^1(\Lambda;V)} \leq \dfrac{\epsilon}{2}$

for $p = 1, \ldots, \ell$ **do** $M_p \longleftarrow 0$

while $\sum_{p=1}^\ell \bar{e}_{\mathcal{R}, M_p} \left\| \boldsymbol{v}_{\{p\}} \right\|_{\ell^1(\Lambda;V)} > \epsilon - \delta$ **do**

$\quad\quad q \longleftarrow \operatorname{argmax}_{p=1,\ldots,\ell} (\bar{e}_{\mathcal{R}, M_p} - \bar{e}_{\mathcal{R}, M_p+1}) \left\| \boldsymbol{v}_{\{p\}} \right\|_{\ell^1(\Lambda;V)} / \# \Lambda_p$

$\quad\quad M_q \longleftarrow M_q + 1$

$\boldsymbol{z} = (z_\nu)_{\nu \in \Lambda} \longleftarrow \boldsymbol{0}$

for $p = 1, \ldots, \ell$ **do**

$\quad$ **forall** $\mu \in \Lambda_p$ **do**

$\quad\quad$ **for** $m = 1, \ldots, M_p$ **do**

$\quad\quad\quad w \longleftarrow R_m v_\mu$

$\quad\quad\quad z_{\mu + \epsilon_m} \longleftarrow z_{\mu + \epsilon_m} + \pi_{\mu_m}^+ w$

$\quad\quad\quad$ **if** $\mu_m \geq 1$ **then** $z_{\mu - \epsilon_m} \longleftarrow z_{\mu - \epsilon_m} + \pi_{\mu_m}^- w$

Proposition 4.2. *For any $\epsilon > 0$ and any $v \in \mathcal{P}(\Gamma; V)$ with coefficient vector $\boldsymbol{v}$ as in (4.10), $\mathtt{Apply}_\mathcal{R}[\boldsymbol{v}, \epsilon]$ produces a finitely supported $\boldsymbol{z} \in \ell^1(\Lambda; W^*)$ such that*

$$\# \operatorname{supp} \boldsymbol{z} \leq 2 \sum_{p=1}^{\ell} M_p \# \Lambda_p \tag{4.13}$$

and the polynomial $z \in \mathcal{P}(\Gamma; W^*)$ from (4.12) satisfies

$$\|\mathcal{R}v - z\|_{C(\Gamma; W^*)} \leq \delta + \sum_{p=1}^{\ell} \bar{e}_{\mathcal{R}, M_p} \left\|\boldsymbol{v}_{\{p\}}\right\|_{\ell^1(\Lambda; V)} \leq \epsilon, \tag{4.14}$$

where $\delta \leq \epsilon/2$ and M_p refers to the final value of this variable in the call of $\mathtt{Apply}_\mathcal{R}$. The total number of products $R_m v_\mu$ computed in $\mathtt{Apply}_\mathcal{R}[\boldsymbol{v}, \epsilon]$ is $\sum_{p=1}^{\ell} M_p \# \Lambda_p$.

Proof. For each $\mu \in \Lambda_p$, $R_m v_\mu$ is computed for $m = 1, \ldots, M_p$, and passed on to at most two coefficients of $\boldsymbol{z}$. This shows (4.13) and the bound on the number of multiplications. Since $\|\mathcal{R}\|_{C(\Gamma; V) \to C(\Gamma; W^*)} \leq \bar{e}_{\mathcal{R}, 0}$, using (4.11),

$$\|\mathcal{R}v - \mathcal{R}w\|_{C(\Gamma; W^*)} \leq \bar{e}_{\mathcal{R}, 0} \|v - w\|_{C(\Gamma; W)} \leq \bar{e}_{\mathcal{R}, 0} \|\boldsymbol{v} - \boldsymbol{w}\|_{\ell^1(\Lambda; V)} = \delta \leq \frac{\epsilon}{2},$$

where $\boldsymbol{w} := \sum_{p=1}^{\ell} \boldsymbol{v}_{\{p\}}$ and w is the polynomial (4.10) with coefficients $\boldsymbol{w}$. For all $p = 1, \ldots, \ell$, let $v_{\{p\}} \in \mathcal{P}(\Gamma; V)$ denote the polynomial with coefficients $\boldsymbol{v}_{\{p\}}$. Due to (4.1) and the termination criterion of the greedy subroutine in $\mathtt{Apply}_\mathcal{R}$,

$$\sum_{p=1}^{\ell} \left\|\mathcal{R}v_{\{p\}} - \mathcal{R}_{[M_p]} v_{\{p\}}\right\|_{C(\Gamma; W^*)} \leq \sum_{p=1}^{\ell} \bar{e}_{\mathcal{R}, M_p} \left\|v_{\{p\}}\right\|_{C(\Gamma; V)} \leq \sum_{p=1}^{\ell} \bar{e}_{\mathcal{R}, M_p} \left\|\boldsymbol{v}_{\{p\}}\right\|_{\ell^1(\Lambda; V)} \leq \epsilon - \delta.$$

The assertion follows since $z = \sum_{p=1}^{\ell} \mathcal{R}_{[M_p]} v_{\{p\}}$. $\qquad \square$

Remark 4.3. By Proposition 4.2, the cost of $\mathtt{Apply}_\mathcal{R}$ is described by $\sum_{p=1}^{\ell} M_p \# \Lambda_p$, and up to the term δ from the truncation of $\boldsymbol{v}$, the error is bounded by

$$\sum_{p=1}^{\ell} \bar{e}_{\mathcal{R}, M_p} \left\|\boldsymbol{v}_{\{p\}}\right\|_{\ell^1(\Lambda; V)}. \tag{4.15}$$

Due to the assumption that $(\bar{e}_{\mathcal{R}, M} - \bar{e}_{\mathcal{R}, M+1})_{M=0}^{\infty}$ is nonincreasing, the greedy algorithm used in $\mathtt{Apply}_\mathcal{R}$ to determine M_p is guaranteed to minimize $\sum_{p=1}^{\ell} M_p \# \Lambda_p$ under the constraint that (4.15) is at most $\epsilon - \delta$.

4.2. Formulation of the method

The adaptive application routine from Section 4.1 efficiently realizes the approximate application routine of the operator $\mathcal{R}$, which is a crucial component of the perturbed linear iteration from Section 2.2. We assume that polynomial approximations of the right hand side $f \in C(\Gamma; W^*)$ in (2.18) are available with arbitrary precision. By Theorem 3.2, such approximations are guaranteed to exist if W^* has the approximation property. We assume that a routine

$$\mathtt{RHS}_f[\epsilon] \mapsto \tilde{\boldsymbol{f}} \tag{4.16}$$

is available which, for any $\epsilon > 0$, returns a finitely supported $\tilde{\boldsymbol{f}} = (\tilde{f}_\nu)_{\nu \in \Lambda} \in \ell^1(\Lambda; W^*)$ with

$$\left\|f - \tilde{f}\right\|_{C(\Gamma; W^*)} \leq \epsilon \quad \text{for} \quad \tilde{f}(y) := \sum_{\nu \in \Lambda} \tilde{f}_\nu P_\nu(y), \quad y \in \Gamma. \tag{4.17}$$

Of course, $\mathtt{RHS}_f$ is trivial if f does not depend on $y \in \Gamma$.

Furthermore, let $\texttt{Solve}_D$ be a solver for D such that for any $g \in W^*$ and any $\epsilon > 0$,

$$\texttt{Solve}_D[g, \epsilon] \mapsto v, \quad \left\| v - D^{-1}g \right\|_V \leq \epsilon. \tag{4.18}$$

For example, $\texttt{Solve}_D$ could be an adaptive wavelet method, see *e.g.* [9, 10, 19], an adaptive frame method, see *e.g.* [13, 14, 33], or a finite element method with *a posteriori* error estimation, see *e.g.* [7, 17, 29].

A realization of the iteration from Section 2.2 using the above approximations is given in $\texttt{SolveDirect}_{\mathcal{A},f}$. We write $\tilde{u}_k$, u_ϵ and g_k for the polynomials with coefficients $\tilde{\boldsymbol{u}}_k$, $\boldsymbol{u}_\epsilon$ and $\boldsymbol{g}_k$, respectively. The initial values can be set to

$$\tilde{\boldsymbol{u}}_0 := \boldsymbol{0} \quad \text{and} \quad \delta_0 := (1 - \gamma)^{-1} \left\| D^{-1} \right\|_{W^* \to V} \| f \|_{C(\Gamma; W^*)}. \tag{4.19}$$

Note that δ_0 is an upper bound for the initial error $\| u - \tilde{u}_0 \|_{C(\Gamma; V)}$. The argument ϵ is the target accuracy of the algorithm, and γ is the upper bound on $\mathcal{R}$ from (3.15). The parameters β_0, β_1 and α distribute the admissible error per iteration among the various approximations within the algorithm, such as the application of $\mathcal{R}$ and the inversion of D. Ranges for these parameters are given in Theorem 4.4.

Theorem 4.4. *For any $\epsilon > 0$ and any $\tilde{u}_0 \in \mathcal{P}(\Gamma; V)$, if $\| u - \tilde{u}_0 \|_{C(\Gamma; V)} \leq \delta_0$, $\alpha > 0$, $\beta_0, \beta_1 > 0$ and $\alpha + \beta_0 + \beta_1 + \gamma < 1$, then $\texttt{SolveDirect}_{\mathcal{A},f}[\tilde{\boldsymbol{u}}_0, \delta_0, \epsilon, \alpha, \beta_0, \beta_1, \gamma]$ terminates with u_ϵ such that*

$$\| u - u_\epsilon \|_{C(\Gamma; V)} \leq \bar{\epsilon} \leq \epsilon. \tag{4.20}$$

Furthermore, for all $k \in \mathbb{N}$ reached in the iteration,

$$\| u - \tilde{u}_k \|_{C(\Gamma; V)} \leq \min(\delta_k, \bar{\delta}_k) \leq (\alpha + \beta_0 + \beta_1 + \gamma)^k \delta_0. \tag{4.21}$$

Algorithm 2. $\texttt{SolveDirect}_{\mathcal{A},f}[\tilde{\boldsymbol{u}}_0, \delta_0, \epsilon, \alpha, \beta_0, \beta_1, \gamma] \mapsto [\boldsymbol{u}_\epsilon, \bar{\epsilon}]$

for $k = 1, 2, \ldots$ **do**
$\quad \eta_k \longleftarrow \delta_{k-1} \left\| D^{-1} \right\|_{W^* \to V}^{-1}$
$\quad \boldsymbol{g}_k = (g_{k,\mu})_{\mu \in \Lambda} \longleftarrow \texttt{RHS}_f[\beta_0 \eta_k] - \texttt{Apply}_{\mathcal{R}}[\tilde{\boldsymbol{u}}_{k-1}, \beta_1 \eta_k]$
$\quad \zeta_k \longleftarrow \alpha \delta_{k-1} (\# \operatorname{supp} \boldsymbol{g}_k)^{-1}$
$\quad \textbf{forall } \mu \in \operatorname{supp} \boldsymbol{g}_k \textbf{ do } \tilde{u}_{k,\mu} \longleftarrow \texttt{Solve}_D[g_{k,\mu}, \zeta_k]$
$\quad \tilde{\boldsymbol{u}}_k \longleftarrow (\tilde{u}_{k,\mu})_{\mu \in \Lambda}$
$\quad \bar{\delta}_{k-1} \longleftarrow (1 - \gamma)^{-1} \left(\| \tilde{\boldsymbol{u}}_k - \tilde{\boldsymbol{u}}_{k-1} \|_{\ell^1(\Lambda; V)} + (\alpha + \beta_0 + \beta_1)\delta_{k-1} \right)$
$\quad \delta_k \longleftarrow (\alpha + \beta_0 + \beta_1)\delta_{k-1} + \gamma \min(\delta_{k-1}, \bar{\delta}_{k-1})$
$\quad \textbf{if } \delta_k \leq \epsilon \textbf{ then break}$
$\boldsymbol{u}_\epsilon \longleftarrow \tilde{\boldsymbol{u}}_k$
$\bar{\epsilon} \longleftarrow \delta_k$

Proof. We show that for all $k \in \mathbb{N}$,
$$\| u - \tilde{u}_k \|_{C(\Gamma; V)} \leq \min(\delta_k, \bar{\delta}_k).$$

Let $\| u - \tilde{u}_{k-1} \|_{C(\Gamma; V)} \leq \min(\delta_{k-1}, \bar{\delta}_{k-1})$. Then as in the proof of Theorem 2.5, since $\mathcal{D}u = f - \mathcal{R}u$,

$$u - \tilde{u}_k = \mathcal{D}^{-1}(f - \mathcal{R}u) - \mathcal{D}^{-1}(f - \mathcal{R}\tilde{u}_{k-1}) + \mathcal{D}^{-1}(f - \mathcal{R}\tilde{u}_{k-1} - g_k) + \mathcal{D}^{-1}g_k - \tilde{u}_k.$$

Due to (2.16),

$$\left\| \mathcal{D}^{-1}(f - \mathcal{R}u) - \mathcal{D}^{-1}(f - \mathcal{R}\tilde{u}_{k-1}) \right\|_{C(\Gamma; V)} = \left\| \mathcal{D}^{-1}\mathcal{R}(u - \tilde{u}_{k-1}) \right\|_{C(\Gamma; V)} \leq \gamma \min(\delta_{k-1}, \bar{\delta}_{k-1}).$$

Furthermore, using $\left\|\mathcal{D}^{-1}\right\|_{C(\Gamma;W^*)\to C(\Gamma;V)} = \left\|D^{-1}\right\|_{W^*\to V}$, Proposition 4.2 and (4.17),

$$\left\|\mathcal{D}^{-1}(f - \mathcal{R}\tilde{u}_{k-1} - g_k)\right\|_{C(\Gamma;V)} \le \left\|D^{-1}\right\|_{W^*\to V}(\beta_0\eta_k + \beta_1\eta_k) = (\beta_0 + \beta_1)\delta_{k-1}.$$

Finally, due to (4.11), (4.18) and $\zeta_k = \alpha\delta_{k-1}(\#\operatorname{supp}\boldsymbol{g}_k)^{-1}$,

$$\left\|\mathcal{D}^{-1}g_k - \tilde{u}_k\right\|_{C(\Gamma;V)} \le \sum_{\mu\in\operatorname{supp}\boldsymbol{g}_k}\alpha\delta_{k-1}(\#\operatorname{supp}\boldsymbol{g}_k)^{-1} = \alpha\delta_{k-1}.$$

By triangle inequality, these estimates imply

$$\|u - \tilde{u}_k\|_{C(\Gamma;V)} \le \gamma\min(\delta_{k-1}, \bar{\delta}_{k-1}) + (\beta_0 + \beta_1)\delta_{k-1} + \alpha\delta_{k-1} = \delta_k.$$

The residual at iteration k is $r_k := f - \mathcal{A}\tilde{u}_k = A(u - \tilde{u}_k)$. We observe that by (2.17),

$$\|u - \tilde{u}_k\|_{C(\Gamma;V)} \le \left\|(\mathcal{D}^{-1}\mathcal{A})^{-1}\right\|\left\|\mathcal{D}^{-1}r_k\right\|_{C(\Gamma;V)} \le \frac{1}{1-\gamma}\left\|\mathcal{D}^{-1}r_k\right\|_{C(\Gamma;V)}.$$

Furthermore, $\mathcal{D}^{-1}r_k$ can be approximated by known quantities since, similarly to above,

$$\left\|\tilde{u}_{k+1} - \tilde{u}_k - \mathcal{D}^{-1}r_k\right\|_{C(\Gamma;V)} = \left\|\tilde{u}_{k+1} - \mathcal{D}^{-1}(f - \mathcal{R}\tilde{u}_k)\right\|_{C(\Gamma;V)} \le (\alpha + \beta_0 + \beta_1)\delta_k.$$

Consequently, using (4.11),

$$\|u - \tilde{u}_k\|_{C(\Gamma;V)} \le \frac{1}{1-\gamma}\left(\|\tilde{u}_{k+1} - \tilde{u}_k\|_{C(\Gamma;V)} + (\alpha + \beta_0 + \beta_1)\delta_k\right) \le \bar{\delta}_k.$$

Equation (4.21) follows since $\delta_k \le (\alpha + \beta_0 + \beta_1 + \gamma)\delta_{k-1}$. $\square$

Remark 4.5. The error bounds in $\mathtt{SolveDirect}_{\mathcal{A},f}$ can be improved if each of the subroutines $\mathtt{RHS}_f$, $\mathtt{Apply}_{\mathcal{R}}$ and $\mathtt{Solve}_D$ returns an estimate of the error it attains. These values can replace $\alpha\delta_{k-1}$, $\beta_0\delta_{k-1}$ and $\beta_1\delta_{k-1}$ in the definitions of $\bar{\delta}_{k-1}$ and δ_k. For better legibility, we refrain from making this explicit.

4.3. Comparison to existing methods

Other methods based on the iteration from Section 2.2 have been suggested. The approach is mentioned in [21] and was analyzed in [1] for the parametric diffusion equation from Section 2.3. In [1], the series (2.29) is truncated to K terms, and subsequently M steps of the iteration (2.19) are performed. The errors due to these two approximations are analyzed individually, neglecting spatial approximation errors. Because its computational cost scales as K^M, this method is only feasible for very small M if K is moderately large. In order to accelerate convergence, the authors suggest to decompose the parameter domain Γ into multiple regions, and to apply (2.19) separately in each.

The method $\mathtt{SolveDirect}_{\mathcal{A},f}$ follows a different strategy. The perturbation in Section 2.2 of the iteration (2.19) allows for the very flexible truncation strategy of (2.29) in $\mathtt{Apply}_{\mathcal{R}}$. Compared to a static truncation at K terms, this avoids unimportant components R_m of $\mathcal{R}$ early in the iteration, but activates them as they become relevant. This leads to a much sparser representation of u, and also prevents superfluous iterations from being performed after the iteration has converged beyond the truncation error.

In conjunction with its adaptive structure, $\mathtt{SolveDirect}_{\mathcal{A},f}$ provides an upper bound for the error in $C(\Gamma;V)$. A similar method with control of the error in $L^2_\pi(\Gamma;V)$ for a suitable probability measure π on Γ is analyzed in [24], and other adaptive strategies are presented in [8, 23]. Another adaptive method is presented in the following section.

5. Alternating subspace correction

5.1. Motivation

By (3.14) and (3.16), if $v \in C(\Gamma; V)$ is an even function, *i.e.* $v(-y) = v(y)$, then $\mathcal{R}v$ and $\mathcal{R}_{[M]}v$ are odd functions, *i.e.* $v(-y) = -v(y)$. Similarly, if v is odd, then $\mathcal{R}v$ and $\mathcal{R}_{[M]}v$ are even. Since $\mathcal{D}$ does not depend on y, $\mathcal{D}v$ is even if v is even, and odd if v is odd.

Let $[n] := n + 2\mathbb{Z}$ denote the equivalence class modulo two of $n \in \mathbb{Z}$, *i.e.* $[n] = [m]$ if $n - m$ is even.

The right hand side $f \in C(\Gamma; W^*)$ of (2.18) can be divided into even and odd parts as $f = f^{[0]} + f^{[1]}$ for

$$f^{[0]}(y) := \frac{1}{2}\big(f(y) + f(-y)\big) \qquad \text{and} \qquad f^{[1]}(y) := \frac{1}{2}\big(f(y) - f(-y)\big). \tag{5.1}$$

Then the iteration (2.19) is equivalent to $u_k = u_k^{[k]} + u_k^{[k-1]}$ for

$$u_k^{[k]} := \mathcal{D}^{-1}\big(f^{[k]} - \mathcal{R}u_{k-1}^{[k-1]}\big) \qquad \text{and} \qquad u_k^{[k-1]} := \mathcal{D}^{-1}\big(f^{[k-1]} - \mathcal{R}u_{k-1}^{[k]}\big). \tag{5.2}$$

We note that $u_k^{[k]}$ only depends on $u_{k-1}^{[k-1]}$ and $u_k^{[k-1]}$ only depends on $u_{k-1}^{[k]}$. We can therefore perform just one of these iterations, say

$$u_k^{[k]} := \mathcal{D}^{-1}\big(f^{[k]} - \mathcal{R}u_{k-1}^{[k-1]}\big), \tag{5.3}$$

and approximate u by $u_k^{[k]} + u_{k-1}^{[k-1]}$ instead of u_k.

For polynomials $v \in \mathcal{P}(\Gamma; V)$, the separation into even and odd parts carries over to the coefficients of v in V. We define the index sets

$$\Lambda^{[n]} := \{\mu \in \Lambda \,;\, [|\mu|] = [n]\}, \quad n \in \mathbb{Z}, \tag{5.4}$$

where $|\mu| = \|\mu\|_{\ell^1(\mathbb{N})}$. Then

$$\Lambda = \Lambda^{[0]} \sqcup \Lambda^{[1]}. \tag{5.5}$$

We call $\mu \in \Lambda$ even if $\mu \in \Lambda^{[0]}$ and odd if $\mu \in \Lambda^{[1]}$.

Remark 5.1. A finitely supported sequence $(v_\mu)_{\mu \in \Lambda}$ defines a polynomial function

$$v(y) = \sum_{\mu \in \Lambda} v_\mu P_\mu(y), \quad y \in \Gamma. \tag{5.6}$$

The function v is even if and only if $v_\mu = 0$ for all $\mu \in \Lambda^{[1]}$ and odd if and only if $v_\mu = 0$ for all $\mu \in \Lambda^{[0]}$ since P_μ is even for $\mu \in \Lambda^{[0]}$ and odd for $\mu \in \Lambda^{[1]}$, and the representation (5.6) is unique.

5.2. Formulation of the method

We assume that routines $\mathtt{RHS}_f^{[0]}$ and $\mathtt{RHS}_f^{[1]}$ are available similar to $\mathtt{RHS}_f$ from (4.16) to construct approximations of $f^{[0]}$ and $f^{[1]}$ from (5.1), such that the approximations of $f^{[0]}$ are even and those of $f^{[1]}$ are odd.

The method $\mathtt{Apply}_{\boldsymbol{R}}$ from Section 4.1 already respects even and odd functions in the sense that if $\boldsymbol{v}$ in $\boldsymbol{z} := \mathtt{Apply}_{\mathcal{R}}[\boldsymbol{v}, \epsilon]$ is supported in $\Lambda^{[n]}$, then $\boldsymbol{z}$ is supported in $\Lambda^{[n+1]}$.

Algorithm 3. $\texttt{SolveAlternate}_{\mathcal{A},f}[\tilde{\boldsymbol{u}}_{-1}^{[-1]}, \delta_{-1}, \epsilon, \alpha, \beta_0, \beta_1, \gamma] \mapsto [\boldsymbol{u}_\epsilon, \bar{\epsilon}]$

> **for** $k = 0, 1, 2, \ldots$ **do**
> > $\eta_k \longleftarrow \delta_{k-1} \left\| D^{-1} \right\|_{W^* \to V}^{-1}$
> > $\boldsymbol{g}_k^{[k]} = (g_{k,\mu})_{\mu \in \Lambda^{[k]}} \longleftarrow \texttt{RHS}_f^{[k]}[\beta_0 \eta_k] - \texttt{Apply}_{\mathcal{R}}[\tilde{\boldsymbol{u}}_{k-1}^{[k-1]}, \beta_1 \eta_k]$
> > $\zeta_k \longleftarrow \alpha \delta_{k-1} (\# \operatorname{supp} \boldsymbol{g}_k^{[k]})^{-1}$
> > **forall** $\mu \in \operatorname{supp} \boldsymbol{g}_k^{[k]}$ **do** $\tilde{u}_{k,\mu} \longleftarrow \texttt{Solve}_D[g_{k,\mu}, \zeta_k]$
> > $\tilde{\boldsymbol{u}}_k^{[k]} \longleftarrow (\tilde{u}_{k,\mu})_{\mu \in \Lambda^{[k]}}$
> > $\delta_k \longleftarrow (\alpha + \beta_0 + \beta_1 + \gamma) \delta_{k-1}$
> > **if** $\delta_{k-1} + \delta_k \leq \epsilon$ **then break**
> $\boldsymbol{u}_\epsilon \longleftarrow \tilde{\boldsymbol{u}}_{k-1}^{[k-1]} + \tilde{\boldsymbol{u}}_k^{[k]}$
> $\bar{\epsilon} \longleftarrow \delta_{k-1} + \delta_k$

As in (5.1), let

$$u^{[k]}(y) := \frac{1}{2}\big(u(y) + (-1)^k u(-y)\big), \tag{5.7}$$

such that $u = u^{[k-1]} + u^{[k]}$ for any $k \in \mathbb{Z}$.

Theorem 5.2. *For any $\epsilon > 0$ and any finitely supported $\tilde{\boldsymbol{u}}_{-1}^{[-1]} \in \ell^1(\Lambda^{[-1]}; V)$, if $\alpha + \beta_0 + \beta_1 + \gamma < 1$ and $\left\| u^{[-1]} - \tilde{u}_{-1}^{[-1]} \right\|_{C(\Gamma;V)} \leq \delta_{-1}$, then $\texttt{SolveAlternate}_{\mathcal{A},f}[\tilde{\boldsymbol{u}}_{-1}^{[-1]}, \delta_{-1}, \epsilon, \alpha, \beta_0, \beta_1, \gamma]$ terminates with u_ϵ such that*

$$\|u - u_\epsilon\|_{C(\Gamma;V)} \leq \bar{\epsilon} \leq \epsilon. \tag{5.8}$$

Furthermore, for all $k \in \mathbb{N}$ reached in the iteration,

$$\left\| u^{[k]} - \tilde{u}_k^{[k]} \right\|_{C(\Gamma;V)} \leq \delta_k \leq (\alpha + \beta_0 + \beta_1 + \gamma)^{k+1} \delta_{-1}. \tag{5.9}$$

Proof. Since $\mathcal{D}u^{[k]} = f^{[k]} - \mathcal{R}u^{[k-1]}$,

$$u^{[k]} - \tilde{u}_k^{[k]} = \mathcal{D}^{-1}(f^{[k]} - \mathcal{R}u^{[k-1]}) - \mathcal{D}^{-1}(f^{[k]} - \mathcal{R}\tilde{u}_{k-1}^{[k-1]}) + \mathcal{D}^{-1}(f^{[k]} - \mathcal{R}\tilde{u}_{k-1}^{[k-1]} - g_k^{[k]}) + \mathcal{D}^{-1}g_k^{[k]} - \tilde{u}_k^{[k]}.$$

Due to (2.16),

$$\left\| \mathcal{D}^{-1}(f^{[k]} - \mathcal{R}u^{[k-1]}) - \mathcal{D}^{-1}(f^{[k]} - \mathcal{R}\tilde{u}_{k-1}^{[k-1]}) \right\|_{C(\Gamma;V)} \leq \gamma \left\| u^{[k-1]} - \tilde{u}_{k-1}^{[k-1]} \right\|_{C(\Gamma;V)}.$$

By definition of $\boldsymbol{g}_k^{[k]}$, using $\left\| \mathcal{D}^{-1} \right\|_{C(\Gamma;W^*) \to C(\Gamma;V)} = \left\| D^{-1} \right\|_{W^* \to V}$,

$$\left\| \mathcal{D}^{-1}(f^{[k]} - \mathcal{R}\tilde{u}_{k-1}^{[k-1]} - g_k^{[k]}) \right\|_{C(\Gamma;V)} \leq \left\| D^{-1} \right\|_{W^* \to V} (\beta_0 \eta_k + \beta_1 \eta_k) = (\beta_0 + \beta_1)\delta_{k-1}.$$

Also, by (4.18),

$$\left\| \mathcal{D}^{-1}g_k^{[k]} - \tilde{u}_k^{[k]} \right\|_{C(\Gamma;V)} \leq \left\| (D^{-1}g_{k,\mu})_\mu - \tilde{u}_k^{[k]} \right\|_{\ell^1(\Lambda;V)} \leq \sum_{\mu \in \operatorname{supp} \boldsymbol{g}_k^{[k]}} \alpha \delta_{k-1} (\# \operatorname{supp} \boldsymbol{g}_k^{[k]})^{-1} = \alpha \delta_{k-1}.$$

Combining these estimates leads to

$$\left\| u^{[k]} - \tilde{u}_k^{[k]} \right\|_{C(\Gamma;V)} \leq (\alpha + \beta_0 + \beta_1)\delta_{k-1} + \gamma \left\| u^{[k-1]} - \tilde{u}_{k-1}^{[k-1]} \right\|_{C(\Gamma;V)}.$$

Consequently, if $\left\| u^{[k-1]} - \tilde{u}_{k-1}^{[k-1]} \right\|_{C(\Gamma;V)} \leq \delta_{k-1}$, then $\left\| u^{[k]} - \tilde{u}_k^{[k]} \right\|_{C(\Gamma;V)} \leq \delta_k$, and (5.9) follows by induction.

By triangle inequality, since $u_\epsilon = \tilde{u}_{k-1}^{[k-1]} + \tilde{u}_k^{[k]}$ for the final value of k in $\texttt{SolveAlternate}_{\mathcal{A},f}$, u_ϵ satisfies

$$\|u - u_\epsilon\|_{C(\Gamma;V)} \leq \left\|u^{[k-1]} - \tilde{u}_{k-1}^{[k-1]}\right\|_{C(\Gamma;V)} + \left\|u^{[k]} - \tilde{u}_k^{[k]}\right\|_{C(\Gamma;V)} \leq \delta_{k-1} + \delta_k = \bar{\epsilon} \leq \epsilon. \qquad \square$$

Remark 5.3. As in Remark 2.6 the error bounds δ_k can be refined using an approximation of the residual analogously to $\texttt{SolveDirect}_{\mathcal{A},f}$. As in the proof of Theorem 4.4, it follows that

$$\left\|u - (\tilde{u}_{k-1}^{[k-1]} + \tilde{u}_k^{[k]})\right\|_{C(\Gamma;V)} \leq \frac{1}{1-\gamma}\left(\left\|\tilde{\boldsymbol{u}}_{k+1}^{[k+1]} - \tilde{\boldsymbol{u}}_{k-1}^{[k-1]}\right\|_{\ell^1(\Lambda;V)} + (\alpha + \beta_0 + \beta_1)(\delta_{k-1} + \delta_k)\right). \qquad (5.10)$$

This term can be used as an alternative upper bound for each of the error components $\left\|u^{[k]} - \tilde{u}_k^{[k]}\right\|_{C(\Gamma;V)}$ and $\left\|u^{[k-1]} - \tilde{u}_{k-1}^{[k-1]}\right\|_{C(\Gamma;V)}$. However, since it applies to the total error instead of directly to the even or odd part, we expect it to be less useful than the bound $\bar{\delta}_k$ in $\texttt{SolveDirect}_{\mathcal{A},f}$.

Remark 5.4. Comparing the convergence estimates (4.21) and (5.9), it appears that the two numerical solvers $\texttt{SolveDirect}_{\mathcal{A},f}$ and $\texttt{SolveAlternate}_{\mathcal{A},f}$ converge at the same rate. Therefore, since the latter method updates only half of the solution vector in each iteration, it should be roughly twice as efficient. However, Remark 5.3 suggests that $\texttt{SolveDirect}_{\mathcal{A},f}$ may provide a sharper bound for the error. It is not clear a priori which of these effects is more significant; numerical computations presented in Section 6.3 indicate that the two solvers are equally efficient.

6. Numerical computations

6.1. A model problem

We consider as a model problem the diffusion equation (2.28) on the one dimensional domain $G = (0,1)$. For two parameters k and γ, the diffusion coefficient has the form

$$a(y,x) = 1 + \frac{1}{c}\sum_{m=1}^{\infty} y_m \frac{1}{m^k}\sin(m\pi x), \quad x \in (0,1), \quad y \in \Gamma = [-1,1]^\infty, \qquad (6.1)$$

where c is chosen as

$$c = \gamma \sum_{m=1}^{\infty} \frac{1}{m^k}, \qquad (6.2)$$

such that $|a(y,x) - 1|$ is always less than γ. We set the parameters to $k = 2$ and $\gamma = 1/2$. A few realizations of $a(y)$ and the resulting solutions $u(y)$ of (2.28) are plotted in Figure 1.

On the parameter domain, we consider Chebyshev polynomials of the first kind and Legendre polynomials. We use a multilevel finite element discretization with piecewise linear basis functions on uniform meshes. The residual-based a posteriori error estimator from [23] is used to estimate the error in $\texttt{Solve}_D$. In order to isolate the discretization of the parameter domain, we also consider a fixed spatial discretization, using linear finite elements on a uniform mesh of $(0,1)$ with 1024 elements to approximate all coefficients. We refer to these simpler versions of the numerical methods as single level discretizations. All computations were performed in Matlab on a workstation with an AMD Athlon$^{\text{TM}}$ 64 X2 5200+ processor and 4GB of memory.

6.2. Approximation rates

Approximation rates of u by polynomials on Γ are proven in [11, 12] for the parametric diffusion equation (2.28). These results are formulated for monomials and for Legendre polynomial bases, but extend to other choices such as Chebyshev polynomials. The uniform ellipticity assumptions in [12] follow from (3.15).
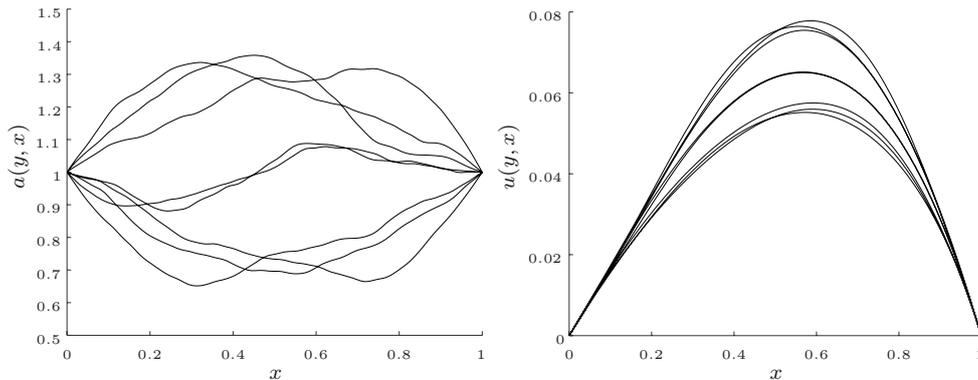
FIGURE 1. Realizations of $a(y,x)$ (left) and $u(y,x)$ (right).

The main result of [11, 12] is that if $(a_m)_{m=1}^\infty \in \ell^p(\mathbb{N}; L^\infty(G))$ with $0 < p < 1$, then $\boldsymbol{u} \in \ell^p(\Lambda; V)$ for the same p. This implies

$$\|u - u_N\|_{C(\Gamma;V)} \le C N^{-s}, \quad s = \frac{1}{p} - 1, \tag{6.3}$$

where u_N is a best approximation of u in the span of just N basis functions P_μ with coefficients in V and $C = \|\boldsymbol{u}\|_{\ell^p(\Lambda;V)}$, see [12], Theorem 4.1. The proof is not constructive, however, so (6.3) provides a benchmark for the convergence of numerical algorithms, but does not suggest a method.

For the model problem from Section 6.1, $\|a_m\|_{L^\infty(G)} = c^{-1} m^{-k}$, and thus $p > k^{-1}$ in (6.3). This implies an approximation rate of $s < k - 1 = 1$ for $k = 2$ with respect to the number of active polynomial basis functions.

Spatial regularity must be taken into account in order to show convergence with respect to the total number of degrees of freedom. By [12], Theorem 5.5, if $(a_m)_{m=1}^\infty \in \ell^p(\mathbb{N}; W^{1,\infty}(G))$, then (6.3) holds with N denoting the total number of degrees of freedom used to represent u_N with $e.g.$ piecewise linear finite elements on G. For our model problem, this imples $p > (k-1)^{-1}$ and therefore $s < k - 2 = 0$ for $k = 2$, making the statement void.

Remark 6.1. The argument in [12] can be refined under the combined assumption $(a_m)_{m=1}^\infty \in \ell^p(\mathbb{N}; L^\infty(G))$ and $(a_m)_{m=1}^\infty \in \ell^q(\mathbb{N}; W^{1,\infty}(G))$ with q different from p. Following the proof of Theorem 5.5 [12], but using q only where the additional spatial regularity is required leads to a convergence rate of

$$s = \frac{\frac{1}{p} - 1}{1 + \frac{1}{p} - \frac{1}{q}} \tag{6.4}$$

instead of $s = \frac{1}{q} - 1$. For our model problem, $p > k^{-1}$ and $q > (k-1)^{-1}$, implying $s < \frac{1}{2}(k-1) = 1/2$ for $k = 2$. In general, the exponents $\min(s,t)$ in [12], Theorem 5.5, can be improved to

$$\min\left(\frac{st}{t + \frac{1}{p} - \frac{1}{q}}, t\right) \tag{6.5}$$

with s as in the statement of the theorem, $i.e.$ $s = \frac{1}{p} - 1$ for uniform convergence and $s = \frac{1}{p} - \frac{1}{2}$ for mean square convergence, and t denotes the spatial convergence rate, $e.g.$ $t = 1$ for linear finite elements.

6.3. Convergence of solvers with uniform error control

The convergence of $\mathtt{SolveDirect}_{\mathcal{A},f}$ and $\mathtt{SolveAlternate}_{\mathcal{A},f}$ is plotted in Figure 2. We use Chebyshev polynomials on the parameter domain Γ, and the parameters of both methods are set to $\alpha = 1/20$, $\beta_0 = 0$ and $\beta_1 = 1/10$.
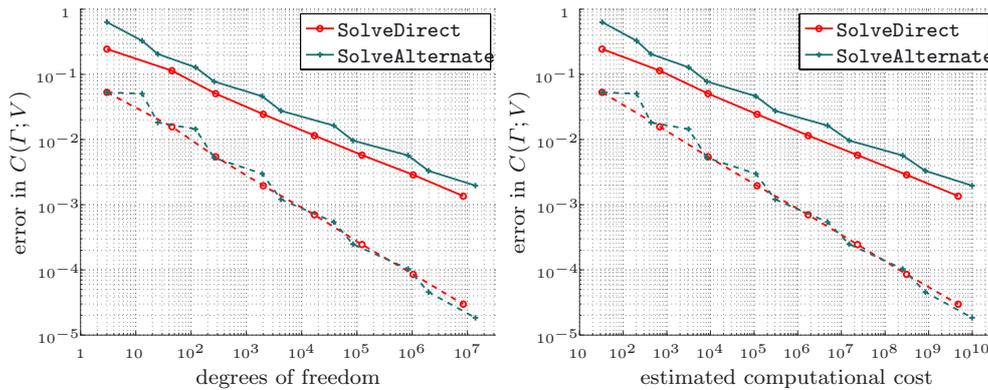
FIGURE 2. Convergence of $\texttt{SolveDirect}_{\mathcal{A},f}$ and $\texttt{SolveAlternate}_{\mathcal{A},f}$. Solid lines refer to the error estimate δ_k, dashed lines are maximal errors on Γ_S.

The solid lines in Figure 2 refer to the error estimate δ_k, which is an upper bound for the error in $C(\Gamma; V)$. We use Remark 5.3 to take advantage of information on the residual when determining this upper bound for $\texttt{SolveAlternate}_{\mathcal{A},f}$. The original formulation of $\texttt{SolveDirect}_{\mathcal{A},f}$ already makes use of this. Markers indicate the positions along the convergence curves at which approximations are computed.

The dashed lines represent the maximal error on a sample set Γ_S of 64 points y in Γ, which is a lower bound for the actual error. The sample set Γ_S is chosen as a subset of the boundary of Γ. Each $y \in \Gamma_S$ consists of randomly chosen $y_m \in \{-1, 0, 1\}$ for $m \leq 250$ and $y_m = 0$ for $m > 250$. Realizations of $a(y)$ and $u(y)$ for the first eight points $y \in \Gamma_S$ are given in Figure 1. We suppose that the maximal error on Γ_S is a good approximation of the maximal error on all of Γ.

On the left, the errors are plotted against the total number of basis functions used in the discretization. On the right, we plot the errors against an estimate of the computational cost given by

$$7 \sum_j (2^j - 1)S_j + 3 \sum_j (2^j - 1)M_j + \sum_j (2^j - 1)P_j, \tag{6.6}$$

where S_j, M_j and P_j denote the total number of linear solves, matrix-vector multiplications and scalar products, respectively, computed on discretization level j. The factor $2^j - 1$ is the dimension of the finite element space on level j. The weights 7, 3 and 1 in (6.6) were determined empirically by timing the operations for tridiagonal sparse matrices in Matlab.

We can see in Figure 2 that δ_k is a coarse upper bound for the actual error. Furthermore, the convergence rate of δ_k in this example is only $1/3$, compared to a rate of $1/2$ for the estimate of the error. This leads to an over-estimation of the error by almost two orders of magnitude towards the end of the computation. This effect is not visible in [22], where a larger k is used in the model problem, *i.e.* the series in (6.1) converges faster.

As anticipated, the error bound for $\texttt{SolveAlternate}_{\mathcal{A},f}$ is slightly coarser than that of $\texttt{SolveDirect}_{\mathcal{A},f}$. However, the convergence of the two methods is very similar.

Figure 3 compares the Chebyshev basis used in Figure 2 to the Legendre basis. There does not seem to be much of a difference between these two choices.

Figure 4 shows the convergence of the single level variants of $\texttt{SolveDirect}_{\mathcal{A},f}$ and $\texttt{SolveAlternate}_{\mathcal{A},f}$ that use a fixed finite element discretization. The spatial discretization error of approximately 2×10^{-4} is suppressed in the convergence plot. The dashed lines refer to the maximal difference between the adaptively computed parametric solutions on the sample set, and the Galerkin projections computed individually for each point $y \in \Gamma_S$ in this set. The solid lines represent the error bounds δ_k, which are only an upper bound for the parametric error for single level methods, and do not capture the spatial discretization error.
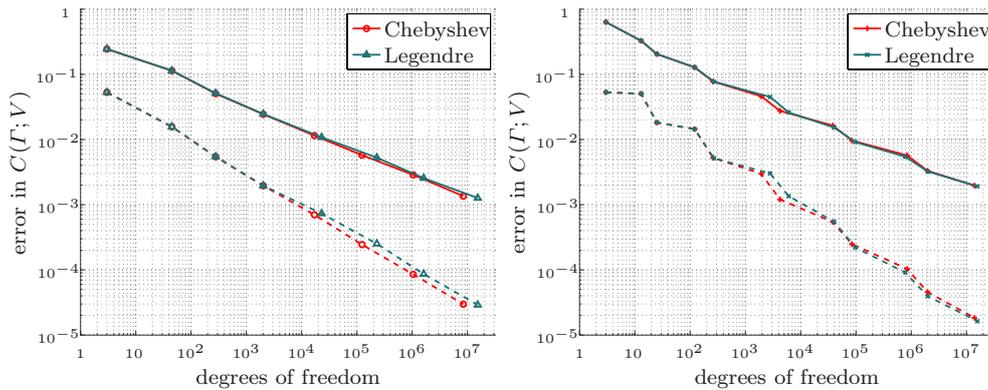
FIGURE 3. Convergence of SolveDirect$_{\mathcal{A},f}$ (left) and SolveAlternate$_{\mathcal{A},f}$ (right) with Chebyshev and Legendre polynomial bases. Solid lines refer to the error estimate δ_k, dashed lines are maximal errors on Γ_S.
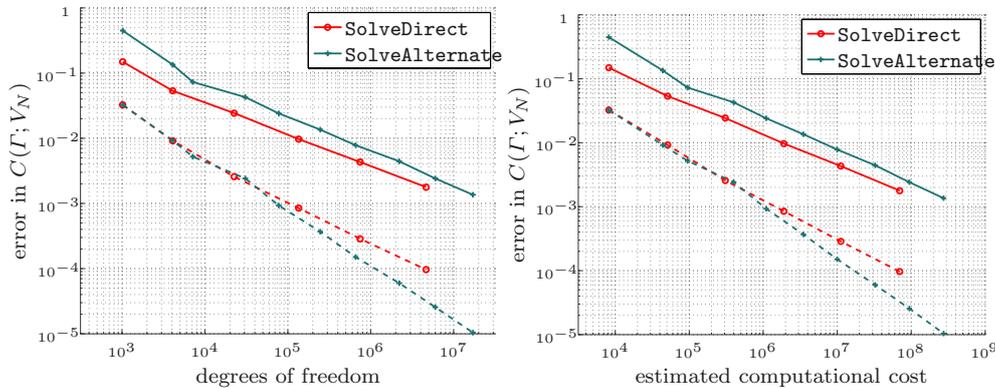

FIGURE 4. Convergence of SolveDirect$_{\mathcal{A},f}$ and SolveAlternate$_{\mathcal{A},f}$ with a fixed finite element discretization. Solid lines refer to the error estimate δ_k, dashed lines are maximal errors on Γ_S, compared with a finite element solution on the same level.

The single level solvers in Figure 4 simulate SolveDirect$_{\mathcal{A},f}$ and SolveAlternate$_{\mathcal{A},f}$ with no spatial discretization. In this setting, a theoretical asymptotic approximation rate of 1 is shown in [11,12]. We observe a rate of approximately $1/2$ for our adaptive solvers, although the convergence rate of the error of SolveAlternate$_{\mathcal{A},f}$ on the sample set approaches one.

For the fully discrete system, *i.e.* Figures 2 and 3, the assumptions of the approximation results in [11,12] are not satisfied, and thus no convergence is shown there. However, by Remark 6.1, [12], Theorem 5.5, can be strengthened to show an approximation rate of $1/2$. This agrees with the observed rate for the actual errors attained by the numerical algorithms, but the computable error bounds δ_k converge only with rate $1/3$.

6.4. Comparison to other adaptive methods

We compare the convergence of SolveDirect$_{\mathcal{A},f}$ to that of similar methods from [23,24], which control the error in $L_\pi^2(\Gamma;V)$ for a probability measure π on the parameter domain Γ. In the following, this probability measure is always chosen in such a way that the polynomial basis $(P_\mu)_{\mu \in \Lambda}$ is orthogonal. For example, for Legendre polynomials, π is a countable product of uniform distributions on $[-1,1]$.
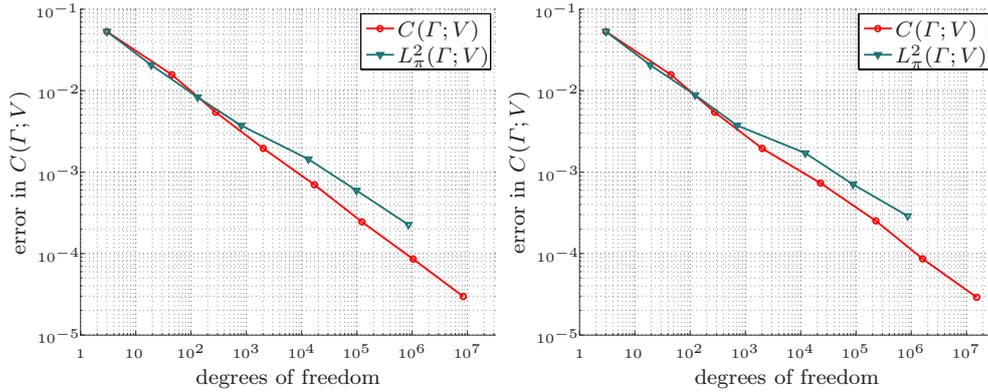
FIGURE 5. Convergence in $C(\Gamma; V)$ of $\texttt{SolveDirect}_{\mathcal{A},f}$ with Chebyshev polynomials (left) and Legendre polynomials (right). The two versions of $\texttt{SolveDirect}_{\mathcal{A},f}$ control the error in $C(\Gamma; V)$ and $L^2_\pi(\Gamma; V)$, respectively.
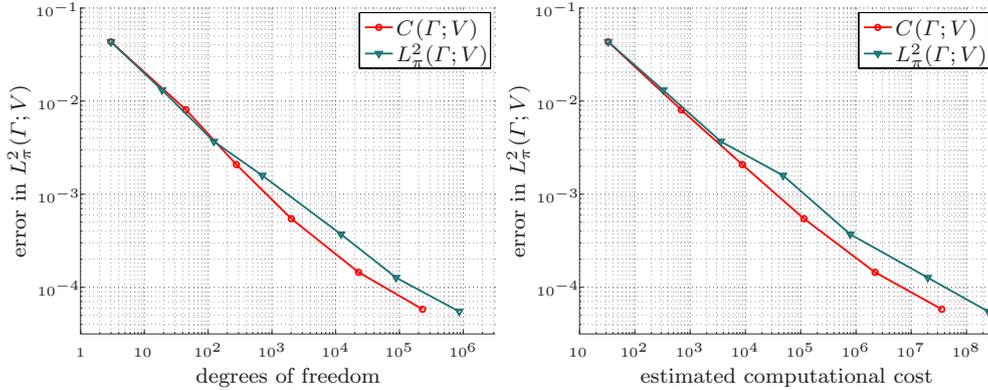

FIGURE 6. Convergence in $L^2_\pi(\Gamma; V)$ of $\texttt{SolveDirect}_{\mathcal{A},f}$ with Legendre polynomials. The two versions of $\texttt{SolveDirect}_{\mathcal{A},f}$ control the error in $C(\Gamma; V)$ and $L^2_\pi(\Gamma; V)$, respectively.

Since the solvers from [23, 24] do not provide bounds for the error in $C(\Gamma; V)$, we do not consider computationally accessible error estimates such as δ_k. We approximate the error in $C(\Gamma; V)$ by the maximal error on the finite sample set Γ_S, as in Section 6.3, and errors in $L^2_\pi(\Gamma; V)$ refer to the difference to a reference solution, evaluated using Parseval's identity.

In Figure 5, the convergence of $\texttt{SolveDirect}_{\mathcal{A},f}$ in $C(\Gamma; V)$ is compared to an analogous method from [24], which is set in $L^2_\pi(\Gamma; V)$ instead of $C(\Gamma; V)$. We observe that the method which controls the error in $C(\Gamma; V)$ converges slightly faster, although the other method also appears to converge uniformly in the parameter $y \in \Gamma$.

Surprisingly, $\texttt{SolveDirect}_{\mathcal{A},f}$ with error control in $C(\Gamma; V)$ also converges faster in $L^2_\pi(\Gamma; V)$, as shown in Figure 6. Here, the reference solution has an error of approximately 5×10^{-5}, which may explain the slight flattening of the convergence curves.

In Figure 7 the convergence in $C(\Gamma; V)$ of $\texttt{SolveDirect}_{\mathcal{A},f}$ with error control in $C(\Gamma; V)$ is compared to that of the adaptive method $\texttt{SolveGalerkin}_{\mathcal{A},f}$ from [23], which controls the error in the energy norm on $L^2_\pi(\Gamma; V)$. The parameters of $\texttt{SolveGalerkin}_{\mathcal{A},f}$ are chosen as in [23]. The latter method includes a coarsening step, which ensures that the approximate solutions are sparse, *i.e.* for a given error tolerance in $L^2_\pi(\Gamma; V)$, the approximate solution constructed by $\texttt{SolveGalerkin}_{\mathcal{A},f}$ should contain a minimal number of degrees of freedom, up to a
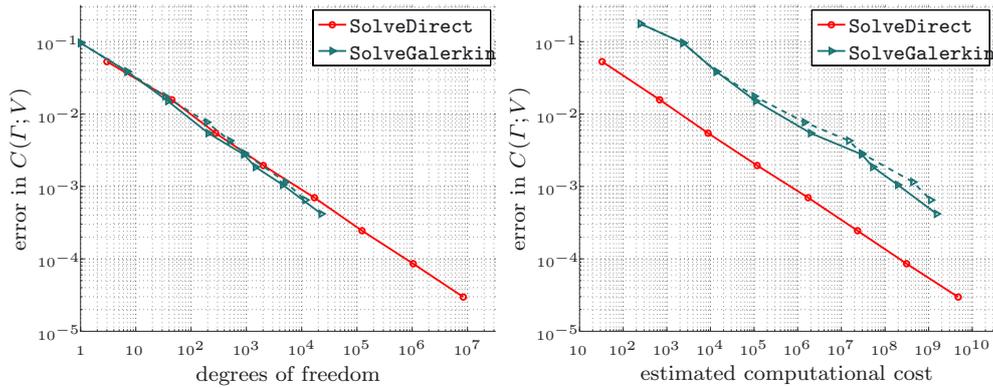
FIGURE 7. Convergence in $C(\Gamma; V)$ of $\texttt{SolveDirect}_{\mathcal{A},f}$ with Chebyshev polynomials and error control in $C(\Gamma; V)$ compared to that of $\texttt{SolveGalerkin}_{\mathcal{A},f}$ with Chebyshev polynomials (solid) and Legendre polynomials (dashed).
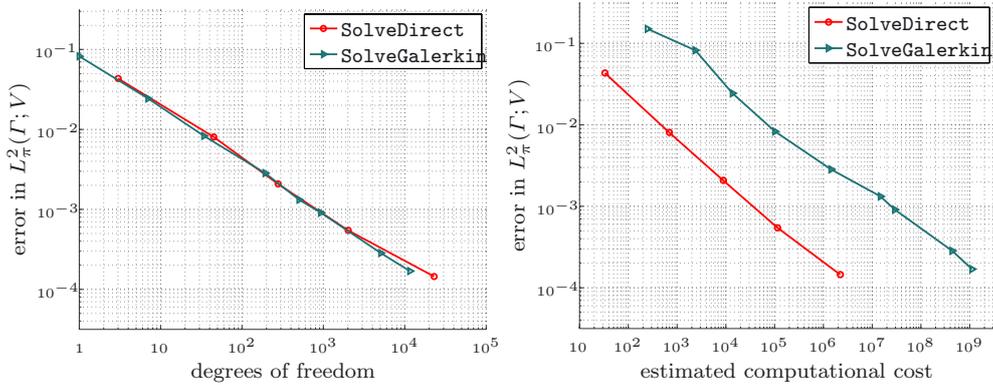

FIGURE 8. Convergence in $L^2_\pi(\Gamma; V)$ of $\texttt{SolveDirect}_{\mathcal{A},f}$ with error control in $C(\Gamma; V)$ compared to that of $\texttt{SolveGalerkin}_{\mathcal{A},f}$ with Legendre polynomials for the uniform distribution π on Γ.

constant factor. Without such a coarsening procedure, $\texttt{SolveDirect}_{\mathcal{A},f}$ produces approximate solutions with almost identical sparsity for a given error. However, the computational cost of $\texttt{SolveDirect}_{\mathcal{A},f}$ is two orders of magnitude lower than that of $\texttt{SolveGalerkin}_{\mathcal{A},f}$. The situation is similar if the error is measured in mean square instead of uniformly in the parameter, as shown in Figure 8.

7. CONCLUSION AND OUTLOOK

Our adaptive methods are proven to converge uniformly in the parameter, which is assumed to be in an infinite dimensional cube. The convergence rates we observe in numerical computations presented in Section 6.3 agree with an extension of the approximation results in [11,12] for the fully discrete algorithms. In a semidiscrete setting, the observed convergence rates are slightly lower than the approximation rates.

The comparisons in Section 6.4 indicate that, for constructing a reliable parametric representation of the solution to a parametric boundary value problem, the adaptive methods presented here are more efficient than similar methods that control the error in $L^2_\pi(\Gamma; V)$ rather than $C(\Gamma; V)$.

Our methods provide a reliable upper bound for the error in $C(\Gamma; V)$. We observed that this bound may overestimate the actual error and may even converge with a lower rate. It would be desirable to have a less

conservative upper bound. Also, the addition of a coarsening step may improve the efficiency of the algorithm. These points are the subject of ongoing research.

References

[1] I. Babuška and P. Chatzipantelidis, On solving elliptic stochastic partial differential equations. *Comput. Methods Appl. Mech. Eng.* **191** (2002) 4093–4122.

[2] I.M. Babuška, R. Tempone and G.E. Zouraris, Galerkin finite element approximations of stochastic elliptic partial differential equations. *SIAM J. Numer. Anal.* **42** (2004) 800–825 (electronic).

[3] I.M. Babuška, F. Nobile and R. Tempone, A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM J. Numer. Anal.* **45** (2007) 1005–1034 (electronic).

[4] A. Barinka, *Fast Evaluation Tools for Adaptive Wavelet Schemes*. Ph.D. thesis, RWTH Aachen (2005).

[5] M. Bieri and C. Schwab, Sparse high order FEM for elliptic sPDEs. *Comput. Methods Appl. Mech. Eng.* **198** (2009) 1149–1170.

[6] M. Bieri, R. Andreev and C. Schwab, Sparse tensor discretization of elliptic SPDEs. *SIAM J. Sci. Comput.* **31** (2009/2010) 4281–4304.

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

[8] A. Chkifa, A. Cohen, R. DeVore and C. Schwab, *Sparse adaptive Taylor approximation algorithms for parametric and stochastic elliptic PDEs.* Technical Report 44, SAM, ETHZ (2011).

[9] A. Cohen, W. Dahmen and R.A. DeVore, Adaptive wavelet methods for elliptic operator equations: convergence rates. *Math. Comput.* **70** (2001) 27–75 (electronic).

[10] A. Cohen, W. Dahmen and R.A. DeVore, Adaptive wavelet methods. II. Beyond the elliptic case. *Found. Comput. Math.* **2** (2002) 203–245.

[11] A. Cohen, R.A. DeVore and C. Schwab, Convergence rates of best N-term Galerkin approximations for a class of elliptic sPDEs. *Found. Comput. Math.* **10** (2010) 615–646.

[12] A. Cohen, R. DeVore and C. Schwab, Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDE's. *Anal. Appl. (Singap.)* **9** (2011) 11–47.

[13] S. Dahlke, M. Fornasier and T. Raasch, Adaptive frame methods for elliptic operator equations. *Adv. Comput. Math.* **27** (2007) 27–63.

[14] S. Dahlke, T. Raasch, M. Werner, M. Fornasier and R. Stevenson, Adaptive frame methods for elliptic operator equations: the steepest descent approach. *IMA J. Numer. Anal.* **27** (2007) 717–740.

[15] M.K. Deb, I.M. Babuška and J.T. Oden, Solution of stochastic partial differential equations using Galerkin finite element techniques. *Comput. Methods Appl. Mech. Eng.* **190** (2001) 6359–6372.

[16] T.J. Dijkema, C. Schwab and R. Stevenson, An adaptive wavelet method for solving high-dimensional elliptic PDEs. *Constr. Approx.* **30** (2009) 423–455.

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

[18] P. Frauenfelder, C. Schwab and R.A. Todor, Finite elements for elliptic problems with stochastic coefficients. *Comput. Methods Appl. Mech. Eng.* **194** (2005) 205–228.

[19] T. Gantumur, H. Harbrecht and R. Stevenson, An optimal adaptive wavelet method without coarsening of the iterands. *Math. Comput.* **76** (2007) 615–629 (electronic).

[20] W. Gautschi, Orthogonal polynomials: computation and approximation, in *Numer. Math. Sci. Comput.* Oxford University Press, Oxford Science Publications, New York (2004).

[21] R.G. Ghanem and P.D. Spanos, *Stochastic finite elements: a spectral approach.* Springer-Verlag, New York (1991).

[22] C.J. Gittelson, *Adaptive Galerkin Methods for Parametric and Stochastic Operator Equations.* Ph.D. thesis, ETH Dissertation No. 19533. ETH Zürich (2011).

[23] C.J. Gittelson, An adaptive stochastic Galerkin method for random elliptic operators. *Math. Comput.* (2011). To appear.

[24] C.J. Gittelson, Convergence Rates of Multilevel and Sparse Tensor Approximations for a Random Elliptic PDE (2012). Submitted.

[25] I.G. Graham, F.Y. Kuo, D. Nuyens, R. Scheichl and I.H. Sloan, Quasi-Monte Carlo methods for elliptic PDEs with random coefficients and applications. *J. Comput. Phys.* **230** (2011) 3668–3694.

[26] R.V. Kadison and J.R. Ringrose, Fundamentals of the theory of operator algebras I, Elementary theory, Reprint of the 1983 original, in *Graduate Studies in Mathematics. Amer. Math. Soc.* **15** (1997).

[27] H.G. Matthies and A. Keese, Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations. *Comput. Methods Appl. Mech. Eng.* **194** (2005) 1295–1331.

[28] A. Metselaar, *Handling Wavelet Expansions in Numerical Methods.* Ph.D. thesis, University of Twente (2002).

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

[30] F. Nobile, R. Tempone and C.G. Webster, An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J. Numer. Anal.* **46** (2008) 2411–2442.

[31] W. Rudin, Functional analysis, 2nd edition. International Series in *Pure Appl. Math.* McGraw-Hill Inc., New York (1991).

[32] C. Schwab and C.J. Gittelson, Sparse tensor discretization of high-dimensional parametric and stochastic PDEs. *Acta Numer.* **20** (2011) 291–467.

[33] R. Stevenson, Adaptive solution of operator equations using wavelet frames. *SIAM J. Numer. Anal.* **41** (2003) 1074–1100 (electronic).

[34] M.H. Stone, The generalized Weierstrass approximation theorem. *Math. Mag.* **21** (1948) 237–254.

[35] G. Szegő, Orthogonal polynomials, 4th edition, in *Colloq. Publ.* XXIII. *Amer. Math. Soc.* (1975).

[36] R.A. Todor and C. Schwab, Convergence rates for sparse chaos approximations of elliptic problems with stochastic coefficients. *IMA J. Numer. Anal.* **27** (2007) 232–261.

[37] X. Wan and G.E. Karniadakis, An adaptive multi-element generalized polynomial chaos method for stochastic differential equations. *J. Comput. Phys.* **209** (2005) 617–642.

[38] X. Wan and G.E. Karniadakis, Multi-element generalized polynomial chaos for arbitrary probability measures. *SIAM J. Sci. Comput.* **28** (2006) 901–928 (electronic).

[39] X. Wan and G.E. Karniadakis, Solving elliptic problems with non-Gaussian spatially-dependent random coefficients. *Comput. Methods Appl. Mech. Eng.* **198** (2009) 1985–1995.

[40] D. Xiu, Efficient collocational approach for parametric uncertainty analysis. *Commun. Comput. Phys.* **2** (2007) 293–309.

[41] D. Xiu, *Numerical methods for stochastic computations*: *A spectral method approach.* Princeton University Press, Princeton, NJ (2010).

[42] D. Xiu and J.S. Hesthaven, High-order collocation methods for differential equations with random inputs. *SIAM J. Sci. Comput.* **27** (2005) 1118–1139 (electronic).

[43] D. Xiu and G.E. Karniadakis, The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comput.* **24** (2002) 619–644 (electronic).