OPTIMALITY OF ADAPTIVE GALERKIN METHODS FOR RANDOM PARABOLIC PARTIAL DIFFERENTIAL EQUATIONS

CLAUDE JEFFREY GITTELSON, ROMAN ANDREEV, AND CHRISTOPH SCHWAB

ABSTRACT. Galerkin discretizations of a class of parametric and random parabolic partial differential equations (PDEs) are considered. The parabolic PDEs are assumed to depend on a vector $y = (y_1, y_2, ...)$ of possibly countably many parameters y_j which are assumed to take values in $[-1, 1]$. Well-posedness of weak formulations of these parametric equation in suitable Bochner spaces is established. Adaptive Galerkin discretizations of the equation based on a tensor product of a generalized polynomial chaos in the parameter domain $\Gamma = [-1, 1]^{\mathbb{N}}$, and of suitable wavelet bases in the time interval $I = [0, T]$ and the spatial domain $D \subset \mathbb{R}^d$ are proposed and their optimality is established.

INTRODUCTION

In recent years, based on the pioneering works [4, 5], and the subsequent refinements [23, 13, 8, 21, 22], a rigorous theory of optimal (in the sense that convergence rates which are afforded by best N-term approximations from a biorthogonal expansion of the unknown solution in some a-priori given Riesz basis are achieved) adaptive Galerkin approximation methods has emerged. After initial applications to linear elliptic partial differential equations in [4, 5] using isotropically supported multiresolution bases, extensions to integrodifferential operators have been considered in [23, 13], first applications to elliptic multiscale problems using anisotropic tensor product Riesz bases have been considered in [8, 21] and, subsequently, to the space-time compressive discretization of linear parabolic (integro)differential equations have been considered in [22].

In recent years, in particular in connection with the numerical solution of partial differential equations with random inputs, for example with random coefficients given by Karhunen–Loève expansions, initial boundary value problems of parametric, deterministic partial differential operators which depend on a sequence of countably many parameters have been considered. Various discretization approaches, for example collocation and Monte Carlo sampling techniques, have been considered (see, *e.g.*, [20] and the references therein).

While affording convenient implementation, the analysis of sampling methods currently leaves open the question of optimality. Here, the situation for the so-called stochastic Galerkin discretizations is quite different: since the discretization consists in a mean-square projection onto a polynomial chaos, *i.e.* onto a finite span from

Date: March 19, 2013.

2010 *Mathematics Subject Classification.* 35R60,47B80,60H35,65C20,65N12,65N22,65J10,65Y20.

Key words and phrases. partial differential equations with random coefficients, parabolic differential equations, uncertainty quantification, stochastic finite element methods, operator equations, adaptive methods, wavelets.

Research supported in part by the Swiss National Science Foundation grant No. 200021-120290/1.

Research supported in part by the Swiss National Science Foundation grant No. PDFMP2-127034/1.

Research supported in part by the ERC Advanced Grant grant STAHDPDE 247277.

a countable ensemble of tensorized orthogonal polynomials, in principle techniques for establishing optimality of Galerkin projection methods for the approximate solution of operator equations can be brought to bear. This programme has been implemented in [15] and the references therein for parametric operator equations.

In the present paper, we adapt these techniques to prove optimality of an adaptive Galerkin scheme for linear, parametric and parabolic equations. Here, we use a Legendre generalized polynomial chaos in the parameter space, and a space-time tensor product wavelet basis that was shown to lead to a an optimal Galerkin approximation for the non-parametric, parabolic initial boundary problems in [22]. Based on the approach and the tensorized space-time Riesz bases for the Bochner space in these references, we develop in the present paper a family of adaptive Galerkin discretizations which are based on tensorizing the generalized polynomial chaos and the space-time tensor product wavelet bases, resulting in discretization schemes which are simultaneously adaptive in space-time and in the parameter space. We establish here optimality of the resulting algorithm, which implies that the best N-term approximation rates which are afforded by the exact solution from the tensorized basis are, indeed, realized by the sequence of finitely supported approximations generated by the proposed adaptive Galerkin discretization.

The outline of this note is as follows. In Section 1.1, we present an abstract class of parametric, parabolic problems which may depend on a countable number of parameters. We elaborate the specific class of *affine parameter dependence* of the parameteric operator equations.

In Section 2, we introduce a space-time weak formulation which also includes a weak form of the parameter dependence.

Sections 3 and 4 introduce the requirement for polynomial chaos type Riesz bases in the parameter domain, and for the multiresolultion (wavelet) Riesz bases on the space and time domains.

Section 5 introduces an equivalent bi-finite matrix equation which, in particular, allows for suitable compressibility results.

Section 6 present elements of the general adaptive Galerkin framework, based on the general reference [4, 5, 13] where adaptive wavelet methods were developed in the context of wavelet discretizations of elliptic operator equations, to the extent required by the ensuing developments.

Section 7 recapitulates from [15] general results on the optimality of adaptive Galerkin approximations of deterministic operator equations. Finally, Section 8 contains statements and proofs of the main result of the present paper, the optimality of the proposed adaptive Galerkin approximations in space, time and parameter domain by sparse, tensorized bases consisting of tensor products of Riesz bases $\boldsymbol{\Theta}$, $\boldsymbol{\Sigma}$ and of $\boldsymbol{P}$.

1. Random and parametric parabolic equations

1.1. **Abstract setting.** Let V and H be real or complex separable Hilbert spaces. We denote by V^* the dual space of V, which consists of all bounded antilinear functionals on V. Assuming a dense embedding $V \hookrightarrow H$, we obtain a Gelfand triple $V \hookrightarrow H \hookrightarrow V^*$, where H is canonically identified with its dual.

We shall consider equations in V that depend on a temporal variable $t \in I := [0,T]$ and also on a parameter sequence $y \in \Gamma := [-1,1]^{\mathbb{N}}$. On Γ, we define a probability measure

$$\boldsymbol{\pi} = \bigotimes_{m\in\mathbb{N}} \pi_m \,, \tag{1.1}$$

where each π_m is assumed to be a probability measure on $[-1,1]$ with the Borel σ-algebra. Although the product structure of the domain Γ and the measure π

is irrelevant for the abstract problem formulation, it is pivotal to the subsequent construction of a basis on the parameter domain in Section 3.

For a.e. $t \in I$ and π-a.e. $y \in \Gamma$, we denote by $A(t,y;\cdot,\cdot)$ a sesquilinear form on $V \times V$ such that for any $v, w \in V$, the map $(t,y) \mapsto A(t,y;v,w)$ is Borel-measurable on $I \times \Gamma$, and such that for a.e. t and y

$$|A(t,y;v,w)| \le c_{\max}\|v\|_V\|w\|_V \qquad \forall v,w \in V\,, \tag{1.2}$$

$$\Re A(t,y;v,v) + c_0\|v\|_H^2 \ge c_{\min}\|v\|_V^2 \qquad \forall v \in V\,, \tag{1.3}$$

with fixed constants $c_{\max} > 0$, $c_{\min} > 0$ and $c_0 \ge 0$. For any $v \in V$, the antilinear functional $A(t,y;v,\cdot)$ is an element of V^*. This allows us to interpret $A(t,y)$ as a bounded linear map from V to V^* for a.e. t and y.

We are interested in solving the parametric parabolic equation

$$\begin{cases} \partial_t u(t,y) + A(t,y)u(t,y) = g(t,y)\,, & t \in I\,, \\ u(0,y) = h(y) \end{cases} \tag{1.4}$$

for given $g(\cdot,y) \in L^2(I;V^*)$ and $h(y) \in H$, for a.e. $y \in \Gamma$.

1.2. **The heat equation.** Let $D \subset \mathbb{R}^d$ be a bounded Lipschitz domain. In $I \times D \times \Gamma$, we consider the random parabolic initial boundary value problem

$$\partial_t u(t,x,y) - \nabla_x \cdot (a(t,x,y)\nabla_x u(t,x,y)) = g(t,x,y)\,, \quad (t,x,y) \in I \times D \times \Gamma\,, \tag{1.5}$$

with boundary condition $u|_{\partial D} = 0$ and initial condition $u(0,x,y) = h(x,y)$. Here, a may be interpreted as a random field on the space-time cylinder $I \times D$. Due to the frequently used Karhunen–Loève expansion to parametrize random field coefficients, we are particularly interested in the case that this field is expanded in a series as

$$a(t,x,y) = \tilde{a}(t,x) + \sum_{m=1}^{\infty} y_m a_m(t,x)\,, \quad (t,x) \in I \times D\,, \quad y = (y_m)_{m=1}^{\infty} \in \Gamma\,. \tag{1.6}$$

Our assumption that the measure $\boldsymbol{\pi}$ be a product measure (1.1) on Γ is equivalent to the condition that the coefficients y_m in this series expansion correspond to independent random variables.

For this example, the spaces V and H are $H = L^2(D)$ and, due to the homogeneous Dirichlet boundary conditions $V = H_0^1(D)$, with dual $V^* = H^{-1}(D)$ (the ensuing analysis will remain valid also for mixed or even Neumann boundary conditions with the obvious modifications of V). Furthermore,

$$A(t,y;v,w) \coloneqq \int_D a(t,x,y)\nabla_x v(x) \cdot \overline{\nabla_x w(x)}\,\mathrm{d}x\,, \quad v,w \in H_0^1(D)\,. \tag{1.7}$$

When interpreted as an operator from V to V^*, this is $A(t,y)v = -\nabla_x\cdot(a(t,\cdot,y)\nabla_x v)$. Due to (1.6), the corresponding parametric operator A can be expanded into a series

$$A(t,y) = \tilde{A}(t) + \sum_{m=1}^{\infty} y_m A_m(t)\,, \tag{1.8}$$

where $\tilde{A}(t)v = -\nabla_x \cdot (\tilde{a}(t,\cdot)\nabla_x v)$ and $A_m(t)v = -\nabla_x \cdot (a_m(t,\cdot)\nabla_x v)$ for $v \in V$, with unconditional convergence in $\mathcal{L}(V,V^*)$ under suitable decay assumptions on a_m.

We specialize to operators of the form (1.8) in the following subsection. For this example, the assumptions (1.10) and (1.11) below are satisfied with $c_0 = 0$ if

$$0 < \tilde{c}_{\min} \le \Re\tilde{a}(t,x) \le |\tilde{a}(t,x)| \le \tilde{c}_{\max} < \infty \qquad \forall (t,x) \in I \times D\,. \tag{1.9}$$

Furthermore, $\|A_m(t)\|_{V\to V^*} = \|a_m\|_{L^\infty(I\times D)}$ in (1.12).

1.3. **Affine dependence on the parameter.** We consider abstract operators of the form (1.8), with

$$|\tilde{A}(t;v,w)| \le \tilde{c}_{\max}\|v\|_V\|w\|_V \qquad \forall v,w \in V\,, \tag{1.10}$$

$$\Re\tilde{A}(t;v,v) + c_0\|v\|_H^2 \ge \tilde{c}_{\min}\|v\|_V^2 \qquad \forall v \in V\,, \tag{1.11}$$

for a.e. t and constants $\tilde{c}_{\max} > 0$, $\tilde{c}_{\min} > 0$ and $c_0 \ge 0$. Under the assumption

$$\sum_{m=1}^{\infty} \|A_m(t)\|_{V\to V^*} \le \gamma\tilde{c}_{\min} \qquad \forall t \in I \tag{1.12}$$

with $\gamma \in [0,1)$, since $|y_m| \le 1$ for all m, (1.2) and (1.3) hold with constants

$$c_{\min} = (1-\gamma)\tilde{c}_{\min} \quad \text{and} \quad c_{\max} = \tilde{c}_{\max} + \gamma\tilde{c}_{\min}\,. \tag{1.13}$$

2. Weak formulation

2.1. **Weak interpretation of the parameter dependence.** We abbreviate $\mathcal{V} \coloneqq L^2_\pi(\Gamma;V)$ and $\mathcal{H} \coloneqq L^2_\pi(\Gamma;H)$. Identifying $\mathcal{H}$ with its dual, and $\mathcal{V}^*$ with the Bochner space $L^2_\pi(\Gamma;V^*)$, we obtain a Gelfand triple $\mathcal{V} \hookrightarrow \mathcal{H} \hookrightarrow \mathcal{V}^*$ of separable Hilbert spaces with dense injections. Due to (1.2) and (1.3), the π-averaged sesquilinear form

$$\mathcal{A}(t;v,w) \coloneqq \int_\Gamma A(t,y;v(y),w(y))\,\mathrm{d}\pi(y)\,, \quad v,w \in \mathcal{V}\,, \tag{2.1}$$

is bounded and coercive on $\mathcal{V}$ with

$$|\mathcal{A}(t;v,w)| \le c_{\max}\|v\|_{\mathcal{V}}\|w\|_{\mathcal{V}} \qquad \forall v,w \in \mathcal{V}\,, \tag{2.2}$$

$$\Re\mathcal{A}(t;v,v) + c_0\|v\|_{\mathcal{H}}^2 \ge c_{\min}\|v\|_{\mathcal{V}}^2 \qquad \forall v \in \mathcal{V}\,, \tag{2.3}$$

for a.e. $t \in I$. Omitting the arguments v and w in (2.1), we interpret $\mathcal{A}(t)$ also as a bounded linear operator from $\mathcal{V}$ to $\mathcal{V}^*$ mapping $v \in \mathcal{V}$ to $\mathcal{A}(t;v,\cdot)$.

For $g \in L^2(I;\mathcal{V}^*)$ and $h \in \mathcal{H}$, we arrive at the parabolic equation

$$\begin{cases} \partial_t u(t) + \mathcal{A}(t)u(t) = g(t)\,, \quad t \in I\,, \\ u(0) = h\,. \end{cases} \tag{2.4}$$

It can be shown as in *e.g.* [20, Thm. 2.18] by testing (2.4) with $v \in V$ multiplied by the indicator function of an arbitrary measurable subset of Γ that the solutions of (2.4) and (1.4) coincide for π-a.e. $y \in \Gamma$.

2.2. **Variational formulation of the parabolic equation.** In the form (2.4), the random parabolic equation fits the setting of [22]. Consequently, the variational formulation derived there applies. It is based on the spaces

$$\mathcal{X} \coloneqq L^2(I;\mathcal{V}) \cap H^1(I;\mathcal{V}^*) = L^2(I;L^2_\pi(\Gamma;V)) \cap H^1(I;L^2_\pi(\Gamma;V^*)) \tag{2.5}$$

and

$$\mathcal{Y} \coloneqq L^2(I;\mathcal{V}) \times \mathcal{H} = L^2(I;L^2_\pi(\Gamma;V)) \times L^2_\pi(\Gamma;H)\,. \tag{2.6}$$

Due to the tensor product structure of Bochner spaces, $\mathcal{X}$ and $\mathcal{Y}$ can be identified with the Hilbert tensor product spaces

$$\begin{aligned} \mathcal{X} &= (L^2(I) \otimes L^2_\pi(\Gamma) \otimes V) \cap (H^1(I) \otimes L^2_\pi(\Gamma) \otimes V^*) \\ &= L^2_\pi(\Gamma) \otimes [(L^2(I) \otimes V) \cap (H^1(I) \otimes V^*)] \end{aligned} \tag{2.7}$$

and

$$\mathcal{Y} = (L^2(I) \otimes L^2_\pi(\Gamma) \otimes V) \times (L^2_\pi(\Gamma) \otimes H) = L^2_\pi(\Gamma) \otimes [(L^2(I) \otimes V) \times H]\,. \tag{2.8}$$

In particular, $\mathcal{X}$ and $\mathcal{Y}$ are tensor products of $L^2_\pi(\Gamma)$ with spaces $X \coloneqq L^2(I;V) \cap H^1(I;V^*)$ and $Y \coloneqq L^2(I;V) \times H$, respectively, that do not depend on π and Γ.

By *e.g.* [7, Ch. 18], $X \hookrightarrow C(I;H)$, and consequently

$$\mathcal{X} = L^2_\pi(\Gamma;X) \hookrightarrow L^2_\pi(\Gamma;C(I;H)) \hookrightarrow C(I;\mathcal{H})\,, \tag{2.9}$$

with $\mathcal{H} = L^2_\pi(\Gamma;H)$. Of course, the weaker statement $\mathcal{X} \hookrightarrow C(I;\mathcal{H})$ follows directly from [7].

Following [22], we define the sesquilinear form $\mathcal{B}(\cdot,\cdot)$ on $\mathcal{X}\times\mathcal{Y}$ as

$$\mathcal{B}(w,(v_1,v_2)) \coloneqq \int_I \langle \partial_t w(t), v_1(t)\rangle + \mathcal{A}(t;w(t),v_1(t))\,\mathrm{d}t + (w(0),v_2)_{\mathcal{H}}\,, \tag{2.10}$$

and interpret this also as a bounded linear operator $\mathcal{B}$ from $\mathcal{X}$ to $\mathcal{Y}^*$. Then for the right hand side $f\in\mathcal{Y}^*$ given by

$$f(v_1,v_2) \coloneqq \int_I \langle g(t), v_1(t)\rangle\,\mathrm{d}t + (h,v_2)_{\mathcal{H}}\,,\quad (v_1,v_2)\in\mathcal{Y}\,, \tag{2.11}$$

we formulate (2.4) as the operator equation

$$\mathcal{B}u = f\,. \tag{2.12}$$

Theorem 2.1. *The operator $\mathcal{B}\colon \mathcal{X}\to\mathcal{Y}^*$ is boundedly invertible.*

For a proof of Thm. 2.1 with bounds on the norm of $\mathcal{B}$ and its inverse, we refer to [22, Thm. 5.1]; see also [7, 25].

3. A polynomial basis on the parameter domain

3.1. **Univariate orthonormal polynomials.** For all $m\in\mathbb{N}$, let $(P^m_n)_{n=0}^\infty$ be an orthonormal polynomial basis of $L^2_{\pi_m}([-1,1])$ where P^m_n is a polynomial of degree n. Such a basis satisfies the three term recursion $P^m_{-1}\coloneqq 0$, $P^m_0\coloneqq 1$ and

$$\beta^m_n P^m_n(\xi) \coloneqq (\xi-\alpha^m_{n-1})P^m_{n-1}(\xi) - \beta^m_{n-1}P^m_{n-2}(\xi)\,,\quad n\in\mathbb{N}\,, \tag{3.1}$$

with

$$\alpha^m_n \coloneqq \int_{-1}^1 \xi P^m_n(\xi)^2\,\mathrm{d}\pi_m(\xi) \qquad\text{and}\qquad \beta^m_n \coloneqq \frac{c^m_{n-1}}{c^m_n}\,, \tag{3.2}$$

where c^m_n is the leading coefficient of P^m_n, $\beta^m_0\coloneqq 1$, and P^m_n is chosen as normalized to unit norm in $L^2_{\pi_m}([0,1])$. This basis is unique *e.g.* if c^m_n is chosen to be positive. We refer to *e.g.* [14, 24] for details.

3.2. **Countable tensor products of orthonormal polynomials.** We define the set of finitely supported sequences in $\mathbb{N}_0$ as

$$\Lambda \coloneqq \{\mu\in\mathbb{N}_0^{\mathbb{N}}\,;\,\#\operatorname{supp}\mu<\infty\}\,, \tag{3.3}$$

where the support of a sequence μ in $\mathbb{N}_0$ is

$$\operatorname{supp}\mu \coloneqq \{m\in\mathbb{N}\,;\,\mu_m\neq 0\}\,,\quad \mu\in\mathbb{N}_0^{\mathbb{N}}\,. \tag{3.4}$$

Then countably infinite tensor product polynomials are given by

$$\boldsymbol{P} \coloneqq (P_\mu)_{\mu\in\Lambda}\,,\quad P_\mu \coloneqq \bigotimes_{m=1}^\infty P^m_{\mu_m}\,,\quad \mu\in\Lambda\,. \tag{3.5}$$

Note that each of these functions depends nontrivially on only finitely many variables; for all $y=(y_m)_{m=1}^\infty\in\Gamma$,

$$P_\mu(y) = \prod_{m=1}^\infty P^m_{\mu_m}(y_m) = \prod_{m\in\operatorname{supp}\mu} P^m_{\mu_m}(y_m)\,,\quad \mu\in\Lambda\,, \tag{3.6}$$

since $P^m_0=1$ for all $m\in\mathbb{N}$.

Theorem 3.1. $\boldsymbol{P}$ *is an orthonormal basis of $L^2_\pi(\Gamma)$.*

We refer to *e.g.* [16, Theorem 2.8] for a proof of Theorem 3.1; see also [12].

4. Wavelet bases on the spatial and temporal domains

In this section we construct Riesz bases $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ of $X = L^2(I;V) \cap H^1(I;V^*)$ and $Y = L^2(I;V) \times H$, respectively. These will consist of suitable bases for the spaces $L^2(I)$ and V. We remark that in general, $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ can be chosen separately. Here we forego that generality and follow the construction given in [22].

Let $\boldsymbol{\Theta} = \{\vartheta_\lambda\,;\,\lambda \in \nabla_t\} \subset H^1(I)$. We assume that $\boldsymbol{\Theta}$ is a (multi-)wavelet basis constructed using dyadic refinements, that is, to each index $\lambda \in \nabla_t$ a level $|\lambda| \in \mathbb{N}_0$ is associated such that $\{\lambda \in \vartheta\,;\,|\lambda| = k\} \sim 2^k$. We assume that

(1) $\boldsymbol{\Theta}$ is a normalized Riesz basis in $L^2(I)$,
(2) $\boldsymbol{\Theta}$ can be rescaled to a Riesz basis in $H^1(I)$
(3) $\boldsymbol{\Theta}$ is of order $d_t > 1$ (in the sense of *e.g.* [21, 22]).

Concerning the wavelet nature of the basis $\boldsymbol{\Theta}$ we further assume that it is

(1) *local*, that is $\sup_{t\in I, k\in\mathbb{N}_0} \#\{\lambda \in \nabla_t\,;\,|\lambda| = k \wedge t \in \operatorname{supp}\vartheta_\lambda\} < \infty$,
(2) *piecewise polynomial of order* d_t, in particular, the singular support of each ϑ_λ is a subset of the closure of I with uniformly bounded finite cardinality,
(3) *continuous*, and $\|\vartheta_\lambda\|_{L^\infty(I)} + 2^{-|\lambda|}\|\vartheta'_\lambda\|_{L^\infty(I)} \lesssim 2^{\frac{1}{2}|\lambda|}$,
(4) such that ϑ_λ has $\tilde{d}_t \geq d_t$ vanishing moments for all $|\lambda| \geq 1$.

With these assumptions, it was shown in [22] that the matrices

$$[(\|\vartheta_{\lambda'}\|^{-1}_{H^1(I)}\vartheta'_{\lambda'}, \vartheta_\lambda)_{L^2(I)}]_{\lambda',\lambda\in\nabla_t}, \quad [(\vartheta_{\lambda'}, \vartheta_\lambda)_{L^2(I)}]_{\lambda',\lambda\in\nabla_t} \tag{4.1}$$

and their adjoints are s^*-computable for any $s^* > 0$.

Bases $\boldsymbol{\Theta}$ with properties required above are well-known. Examples include biorthogonal spline wavelets, see *e.g.* [6, 18, 19] and references therein.

Let $\boldsymbol{\Sigma} = \{\sigma_\nu\,;\,\nu \in \nabla_x\} \subset V$ be a basis for V. We begin by assuming that $\boldsymbol{\Sigma}$

(1) is a normalized Riesz basis for H,
(2) can be rescaled to a Riesz basis for V, and properly rescaled constitutes a Riesz basis for V^*.

This already greatly restricts the choice of bases $\boldsymbol{\Sigma}$. We will moreover need s^*-computability of the matrices

$$[(\|\sigma_{\nu'}\|^{-1}_{V^*}\sigma_{\nu'}, \|\sigma_\nu\|^{-1}_{V}\sigma_\nu)_H]_{\nu',\nu\in\nabla_x}, \quad [(\sigma_{\nu'}, \sigma_\nu)_H]_{\nu',\nu\in\nabla_x} \tag{4.2}$$

and their adjoints.

In the situation that $A(t,y)$ is a linear integro-differential operator of order $2m$ on the spatial domain D, sufficient conditions for s^*-computability of the matrices (4.2) were discussed in [22]. Here we are primarily interested in the case that $A(t,y)$ is given by (1.7), *i.e.* $m = 1$, which shall be therefore discussed in more detail. We further focus on the case that $D \subset \mathbb{R}^d$ is a cube, *i.e.* $D = \prod_{i=1}^d (a_i, b_i)$ with $a_i < b_i$. The coefficient $a(t,x,y)$ in (1.7) is, however, not required to be a separable as a function $D \ni x \to a(t,x,y)$. We thus assume that $\boldsymbol{\Sigma} = \bigotimes_{i=1}^d \boldsymbol{\Sigma}_i$ with wavelet bases $\boldsymbol{\Sigma}_i$ satisfying similar assumptions as for $\boldsymbol{\Theta}$. We state these first in the case $d = 1$. For some $r_x \in \mathbb{N}_0$ with $m - 1 \leq r_x \leq d_x - 2$, and $\tilde{d}_x \in \mathbb{N}_0$, we assume that σ_ν

(1) are local and piecewise polynomial of order d_x,
(2) are globally C^{r_x}, specifically $\sum_{k=0}^{r_x+1} 2^{-k|\nu|}\|\sigma_\nu^{(k)}\|_{L^\infty(a_1,b_1)} \lesssim 2^{\frac{1}{2}|\nu|}$
(3) for $|\nu| \geq 1$ have the cancellation property of order $\tilde{d}_x$, that is

$$\left|\int_{(a_1,b_1)} w(x)\sigma_\nu(x)\,\mathrm{d}x\right| \lesssim 2^{-|\nu|(\frac{1}{2}+k)}\|w\|_{W^{k,\infty}(a_1,b_1)} \tag{4.3}$$

for all $k \in \{0, \tilde{d}_x\}$, $w \in W^{k,\infty}(a_1,b_1) \cap H^1_0(a_1,b_1)$.

When $d \geq 1$, we assume that these properties hold for each $\boldsymbol{\Sigma}_i$, $i = 1, \dots, d$.

Piecewise polynomial (L^2-orthonormal) wavelets on an interval satisfying these assumptions were constructed in [9, 10, 17] (see also [11]), and used in *e.g.* [8] for the adaptive tensor product wavelet method for an elliptic problem. These are constructions based on [9], where it was shown that given a multiresolution analysis (MRA) $V_\ell \subset V_{\ell+1} \subset L^2(\mathbb{R})$, $\ell \in \mathbb{Z}$, generated by finitely many compactly supported scaling functions, there exist $q \in \mathbb{N}_0$, $m \in \mathbb{N}$ and an orthogonal MRA $\check{V}_\ell \subset \check{V}_{\ell+1} \subset L^2(\mathbb{R})$, $\ell \in \mathbb{Z}$, generated by finitely many compactly supported *orthogonal* scaling functions with $V_q \subset \check{V}_0 \subset V_{q+m}$. Moreover, an orthogonal MRA on a bounded interval can be derived, see [10, Theorem 4.4].

The Riesz bases $\boldsymbol{\Phi} \subset X$ and $\boldsymbol{\Psi} \subset Y$ are then defined analogously to [21] by

$$\boldsymbol{\Phi} = \left\{ \frac{\vartheta_\lambda \otimes \sigma_\nu}{\sqrt{\|\vartheta_\lambda\|^2_{L^2(I)}\|\sigma_\nu\|^2_V + \|\vartheta_\lambda\|^2_{H^1(I)}\|\sigma_\nu\|^2_{V^*}}}; (\lambda,\nu) \in \nabla_t \times \nabla_x \right\} \tag{4.4}$$

and

$$\boldsymbol{\Psi} = \left\{ \left(\frac{\vartheta_\lambda \otimes \sigma_\nu}{\|\vartheta_\lambda\|_{L^2(I)}\|\sigma_\nu\|_V}, 0 \right); (\lambda,\nu) \in \nabla_t \times \nabla_x \right\} \cup \{(0,\sigma_\nu); \nu \in \nabla_x\}. \tag{4.5}$$

5. Equivalent bi-infinite matrix equation

5.1. **Derivation.** Due to Parseval's identity, Theorem 3.1 states that the synthesis operator

$$T_{\boldsymbol{P}} \colon \ell^2(\Lambda) \to L^2_\pi(\Gamma), \quad (c_\mu)_{\mu\in\Lambda} \mapsto \sum_{\mu\in\Lambda} c_\mu P_\mu \tag{5.1}$$

is unitary. The property that $\boldsymbol{\Phi}$ is a Riesz basis of X can be expressed as bounded invertibility of the synthesis operator

$$T_{\boldsymbol{\Phi}} \colon \ell^2(\Xi) \to X, \quad (c_\iota)_{\iota\in\Xi} \mapsto \sum_{\iota\in\Xi} c_\iota \varphi_\iota. \tag{5.2}$$

In particular, the synthesis operator of $\boldsymbol{P} \otimes \boldsymbol{\Phi} = (P_\mu \otimes \varphi_\iota)_{(\mu,\iota)\in\Lambda\otimes\Xi}$,

$$T_{\boldsymbol{P}\otimes\boldsymbol{\Phi}} = T_{\boldsymbol{P}} \otimes T_{\boldsymbol{\Phi}} \colon \ell^2(\Lambda\times\Xi) \to \mathcal{X} = L^2_\pi(\Gamma; X) \tag{5.3}$$

is an isomorphism of Hilbert spaces, *i.e.* $\boldsymbol{P} \times \boldsymbol{\Phi}$ is a Riesz basis of $\mathcal{X}$. Similarly, since $\boldsymbol{\Psi}$ is a Riesz basis of Y, the products $\boldsymbol{P} \otimes \boldsymbol{\Psi} = (P_\mu \otimes \psi_\eta)_{(\mu,\eta)\in\Lambda\times\Upsilon}$ form a Riesz basis of $\mathcal{Y} = L^2_\pi(\Gamma; Y)$.

We reformulate (2.12) equivalently in sequence spaces using the above synthesis operators. Let

$$\boldsymbol{u} \coloneqq T^{-1}_{\boldsymbol{P}\times\boldsymbol{\Phi}} u \in \ell^2(\Lambda\times\Xi), \quad \boldsymbol{f} \coloneqq T^*_{\boldsymbol{P}\times\boldsymbol{\Psi}} f \in \ell^2(\Lambda\times\Upsilon) \tag{5.4}$$

and

$$\boldsymbol{B} \coloneqq T^*_{\boldsymbol{P}\times\boldsymbol{\Psi}} \mathcal{B} T_{\boldsymbol{P}\times\boldsymbol{\Phi}} \colon \ell^2(\Lambda\times\Xi) \to \ell^2(\Lambda\times\Upsilon). \tag{5.5}$$

Then (2.12) is equivalent to the bi-infinite scalar linear system

$$\boldsymbol{B}\boldsymbol{u} = \boldsymbol{f}. \tag{5.6}$$

The solution u can be reconstructed from $\boldsymbol{u} = (u_{\mu\iota})$ as

$$u = T_{\boldsymbol{P}\times\boldsymbol{\Phi}}\boldsymbol{u} = \sum_{(\mu,\iota)\in\Lambda\times\Xi} u_{\mu\iota} P_\mu \otimes \varphi_\iota. \tag{5.7}$$

5.2. **Structure of the discrete operator.** We separate the spatial and temporal discretization from that in the parameter $y \in \Gamma$. Here and in the following, we restrict to the setting of Section 1.3, *i.e.* $A(t,y)$ has the form (1.8) with affine dependence on $y \in \Gamma$.

We construct an approximation to the parabolic sesquilinear form $\mathcal{B}$ independent of the parameter $y \in \Gamma$,

$$\tilde{B}(w,(v_1,v_2)) \coloneqq \int_I \langle \partial_t w(t), v_1(t)\rangle + \tilde{A}(t; w(t), v_1(t))\,\mathrm{d}t + (w(0), v_2)_H\,, \tag{5.8}$$

for $w \in X$ and $(v_1,v_2) \in Y$, and interpret $\tilde{B}$ also as an operator $\tilde{B} \in \mathcal{L}(X, Y^*)$. Similarly, we interpret A_m as a bounded linear map from X to Y^*. Then due to the series expansion (1.8),

$$\mathcal{B}(w,(v_1,v_2)) = \int_\Gamma \tilde{B}(w,(v_1,v_2)) + \sum_{m=1}^\infty y_m A_m(w,v_1)\,\mathrm{d}\pi(y)\,, \tag{5.9}$$

for $w \in \mathcal{X}$ and $(v_1,v_2) \in \mathcal{Y}$.

Let $\tilde{\boldsymbol{B}} \coloneqq T_{\boldsymbol{\Psi}}^* \tilde{B} T_{\boldsymbol{\Phi}}$ and $\boldsymbol{A}_m \coloneqq T_{\boldsymbol{\Psi}}^* A_m T_{\boldsymbol{\Phi}}$. These are bounded linear maps from $\ell^2(\Xi)$ to $\ell^2(\Upsilon)$ and, as such, can be interpreted as bi-infinite matrices.

To capture the dependence on the parameter sequence $y \in \Gamma$, we consider the representation of

$$K_m \colon L^2_\pi(\Gamma) \to L^2_\pi(\Gamma)\,, \quad v(y) \mapsto y_m v(y) \tag{5.10}$$

with respect to the polynomial basis $\boldsymbol{P}$. Solving for the term ξP_{n-1}^m in (3.1), it follows that for all $m \in \mathbb{N}$, $\boldsymbol{K}_m \coloneqq T_{\boldsymbol{P}}^* K_m T_{\boldsymbol{P}} \in \mathcal{L}(\ell^2(\Lambda))$ has the form

$$(\boldsymbol{K}_m \boldsymbol{c})_\mu = \beta^m_{\mu_m+1} c_{\mu+\epsilon_m} + \alpha^m_{\mu_m} c_\mu + \beta^m_{\mu_m} c_{\mu-\epsilon_m}\,, \quad \mu \in \Lambda\,, \tag{5.11}$$

for $\boldsymbol{c} = (c_\mu)_{\mu\in\Lambda} \in \ell^2(\Lambda)$, where $c_\mu \coloneqq 0$ if $\mu_m < 0$ for any $m \in \mathbb{N}$, and ϵ_m denotes the Kronecker sequence $(\epsilon_m)_n = \delta_{mn}$. Also, let $\boldsymbol{I}$ be the identity on $\ell^2(\Lambda)$.

Combining the above basis representations leads to

$$\boldsymbol{B} = \boldsymbol{I} \otimes \tilde{\boldsymbol{B}} + \sum_{m=1}^\infty \boldsymbol{K}_m \otimes \boldsymbol{A}_m\,, \tag{5.12}$$

with convergence in $\mathcal{L}(\ell^2(\Lambda\times\Xi), \ell^2(\Lambda\times\Upsilon))$, see [15, Prop. 2.3].

As discussed in [22, Sec. 6], the matrix representation $\tilde{\boldsymbol{B}}$ of the discrete deterministic parabolic operator $\tilde{B}$ has the form

$$\begin{bmatrix} \|\vartheta_\lambda\sigma_\nu\|_X^{-1}\|\sigma_{\nu_1'}\|_V^{-1}\big((\vartheta_\lambda', \vartheta_{\lambda'})_{L^2(I)}(\sigma_\nu,\sigma_{\nu_1'})_H + \int_I \tilde{A}(t;\vartheta_\lambda(t)\sigma_\nu, \vartheta_{\lambda'}(t)\sigma_{\nu_1'})\,\mathrm{d}t\big) \\ \|\vartheta_\lambda\sigma_\nu\|_X^{-1}(\vartheta_\lambda(0)\sigma_\nu, \sigma_{\nu_2'})_H \end{bmatrix} \tag{5.13}$$

with column indices $(\lambda,\nu) \in \nabla_t \times \nabla_x$ and row indices $(\lambda',\nu_1',\nu_2') \in \nabla_t \times \nabla_x \times \nabla_x$. The matrix representations $\boldsymbol{A}_m$ of A_m are

$$\begin{bmatrix} \|\vartheta_\lambda\sigma_\nu\|_X^{-1}\|\sigma_{\nu_1'}\|_V^{-1}\int_I A_m(t;\vartheta_\lambda(t)\sigma_\nu, \vartheta_{\lambda'}(t)\sigma_{\nu_1'})\,\mathrm{d}t \\ 0 \end{bmatrix}\,, \tag{5.14}$$

with the same indexation.

5.3. **The discrete adjoint operator and the normal equations.** Since $\boldsymbol{B}$ is not symmetric, we consider the discrete normal equations

$$\boldsymbol{B}^*\boldsymbol{B}\boldsymbol{u} = \boldsymbol{B}^*\boldsymbol{f}\,, \tag{5.15}$$

where $\boldsymbol{B}^*$ is the discrete adjoint operator

$$\boldsymbol{B}^* = T^*_{\boldsymbol{P}\times\boldsymbol{\Phi}} \mathcal{B}^* T_{\boldsymbol{P}\times\boldsymbol{\Psi}} \colon \ell^2(\Lambda\times\Upsilon) \to \ell^2(\Lambda\times\Lambda)\,. \tag{5.16}$$

Since $\boldsymbol{B}^*$ is bijective, (5.15) is equivalent to (5.6).

The operator $\boldsymbol{B}^*$ has the same tensor product structure as $\boldsymbol{B}$, with $\tilde{\boldsymbol{B}}$ and $\boldsymbol{A}_m$ replaced by their repsective adjoints. Since $\boldsymbol{I}$ and $\boldsymbol{K}_m$ are symmetric,

$$\boldsymbol{B}^* = \boldsymbol{I} \otimes \tilde{\boldsymbol{B}}^* + \sum_{m=1}^{\infty} \boldsymbol{K}_m \otimes \boldsymbol{A}_m^* \,. \tag{5.17}$$

The bi-infinite matrices $\tilde{\boldsymbol{B}}^*$ and $\boldsymbol{A}_m^*$ are given by the transpositions of (5.13) and (5.14), respectively.

6. Adaptive Galerkin methods

6.1. **Overview.** Adaptive Galerkin discretization methods are set on the abstract level of coefficients with respect to a Riesz basis, where a linear operator equation is recast as a bi-infinite linear system of the form (5.6). As such, these methods can be applied in our setting, with the Riesz bases $\boldsymbol{P} \times \boldsymbol{\Phi}$ and $\boldsymbol{P} \times \boldsymbol{\Psi}$, provided that certain subroutines discussed below are available. To keep the exposition general, we consider an arbitrary linear system $\boldsymbol{B}\boldsymbol{u} = \boldsymbol{f}$, and enumerate both index sets such that this equation is set in ℓ^2.

In [4, 13], for elliptic problems, the bi-infinite linear system is approximated by a sequence of adaptively generated finite sections, the solutions of which are Galerkin projections onto finite dimensional spaces. The refinement of the active set of indices is governed by approximations of the discrete residuals. Sufficient accuracy in the computation of each residual is shown in [4, 13] to ensure a reduction in the error of the Galerkin solution on the refined set of active indices by a fixed factor, independent of the support size of the "active" components of the operator section.

A different approach is followed in [5]. There, an iterative solver is applied directly to the full bi-infinite linear system. Operations in each step are replaced by approximate counterparts with sufficient accuracy to ensure convergence of the method. We note in passing that an a priori selection of stable space-time sparse trial and test spaces for space-time Petrov-Galerkin formulations of parabolic evolution equations is also possible [1, 2], which would lead to an a priori known finite section of the bi-infinite linear system.

Although originally formulated for positive symmetric systems, both approaches extend to nonsymmetric linear equations by solving the normal equations (5.15), as noted in [5]. This is developed explicitely for parablic problems in [22].

The adaptive wavelet methods described above can be applied quite generically as black-box solvers, provided that the application of the bi-infinite matrix $\boldsymbol{B}$ to any finitely supported vector can be approximated to any desired accuracy. This is achieved by a routine

$$\mathtt{Apply}_{\boldsymbol{B}}[\boldsymbol{w}, \epsilon] \mapsto \boldsymbol{z} \tag{6.1}$$

which, for any $\epsilon > 0$ and any finitely supported vector $\boldsymbol{w}$, constructs a finitely supported vector $\boldsymbol{z}$ with $\|\boldsymbol{B}\boldsymbol{w} - \boldsymbol{z}\|_{\ell^2} \leq \epsilon$. The discrete operator $\boldsymbol{B}$ is called s^*-admissible for a given $s^* \in (0, \infty]$ if, for all $s \in (0, s^*)$, $\#\operatorname{supp} \boldsymbol{z} \lesssim \epsilon^{-1/s} \|w\|^{1/s}_{\mathcal{A}^s_\infty(\ell^2)}$ and the number of arithmetic operations and storage locations used by $\mathtt{Apply}_{\boldsymbol{B}}[\boldsymbol{w}, \epsilon]$ is bounded by

$$\epsilon^{-1/s} \|\boldsymbol{w}\|^{1/s}_{\mathcal{A}^s_\infty(\ell^2)} + \#\operatorname{supp} \boldsymbol{w} + 1 \tag{6.2}$$

up to a fixed constant independent of $\boldsymbol{w}$. Here $\mathcal{A}^s_\infty(\ell^2)$ denotes the subspace of ℓ^2 defined by the quasi-norm

$$\|\boldsymbol{v}\|_{\mathcal{A}^s_\infty(\ell^2)} \coloneqq \sup_{N \in \mathbb{N}_0} (N+1)^s \|\boldsymbol{v} - P_N(\boldsymbol{v})\|_{\ell^2} \,, \tag{6.3}$$

where $P_N(\boldsymbol{v})$ is any best N-term approximation of $\boldsymbol{v}$ in ℓ^2.

In the case of nonsymmetric problems, since we consider the normal equations (5.15), we also require a routine $\texttt{Apply}_{\boldsymbol{B}^*}$ for approximating the application of the adjoint operator.

Similarly, we need to approximate the right hand side $\boldsymbol{f}$ to arbitrary precision by finitely supported vectors. We assume that a routine $\texttt{RHS}_{\boldsymbol{f}}[\epsilon]$ is available which, for any $\epsilon > 0$, constructs a finitely supported vector $\boldsymbol{f}_\epsilon$ with $\|\boldsymbol{f} - \boldsymbol{f}_\epsilon\|_{\ell^2} \leq \epsilon$ and

$$\#\operatorname{supp} \boldsymbol{f}_\epsilon \leq \inf\{N \in \mathbb{N}\,;\, \|\boldsymbol{f} - P_N(\boldsymbol{f})\| \leq C\epsilon\} \tag{6.4}$$

for a fixed constant $C \geq 1$.

Theorem 6.1. *If $\boldsymbol{B}$ is symmetric positive definite and s^*-admissible, then for any $\epsilon > 0$, the adaptive wavelet methods from* [4, 5, 13] *construct an approximation $\boldsymbol{u}_\epsilon$ of $\boldsymbol{u}$ with $\|\boldsymbol{u} - \boldsymbol{u}_\epsilon\|_{\ell^2} \leq \epsilon$. If $\boldsymbol{u} \in \mathcal{A}^s_\infty(\ell^2)$ for any $s > 0$, then $\#\operatorname{supp} \boldsymbol{u}_\epsilon \lesssim \epsilon^{-1/s}\|\boldsymbol{u}\|_{\mathcal{A}^s_\infty(\ell^2)}$. If $s < s^*$, then the number of arithmetic operations and storage locations used to compute $\boldsymbol{u}_\epsilon$ is bounded by an affine function of $\epsilon^{-1/s}\|\boldsymbol{u}\|_{\mathcal{A}^s_\infty(\ell^2)}$.*

We refer to [4, 5, 13] for proofs of Theorem 6.1 for each of these adaptive solvers.

Remark 6.2. The estimate $\#\operatorname{supp} \boldsymbol{u}_\epsilon \lesssim \epsilon^{-1/s}\|\boldsymbol{u}\|_{\mathcal{A}^s_\infty(\ell^2)}$ can be interpreted as follows: if $\|\boldsymbol{u} - P_N(\boldsymbol{u})\|_{\ell^2} \lesssim N^{-s}$, then $\|\boldsymbol{u}\|_{\mathcal{A}^s_\infty(\ell^2)}$ is finite, and consequently $\|\boldsymbol{u} - \boldsymbol{u}_\epsilon\|_{\ell^2} \leq \epsilon \lesssim (\#\operatorname{supp} \boldsymbol{u}_\epsilon)^{-s}$. Thus adaptive Galerkin methods recover the optimal convergence rate in terms of the support size, albeit with a larger constant in the error estimate. In the case $s < s^*$, the same estimate holds for the computational cost and, in this sense, the solvers have optimal complexity.

Remark 6.3. Replacing $\boldsymbol{B}$ by $\boldsymbol{B}^*\boldsymbol{B}$ and $\boldsymbol{f}$ by $\boldsymbol{B}^*\boldsymbol{f}$, Theorem 6.1 applies to the normal equations for nonsymmetric systems. Valid `Apply` and `RHS` routines are given by

$$\texttt{Apply}_{\boldsymbol{B}^*\boldsymbol{B}}[\boldsymbol{w}, \epsilon] \coloneqq \texttt{Apply}_{\boldsymbol{B}^*}[\texttt{Apply}_{\boldsymbol{B}}[\boldsymbol{w}, \epsilon/(2\|\boldsymbol{B}\|)], \epsilon/2] \tag{6.5}$$

and

$$\texttt{RHS}_{\boldsymbol{B}^*\boldsymbol{f}}[\epsilon] \coloneqq \texttt{Apply}_{\boldsymbol{B}^*}[\texttt{RHS}_{\boldsymbol{f}}[\epsilon/(2\|\boldsymbol{B}\|)], \epsilon/2]\,, \tag{6.6}$$

where $\|\boldsymbol{B}\|$ denotes the operator norm of $\boldsymbol{B}$. The product $\boldsymbol{B}^*\boldsymbol{B}$ is s^*-admissible if both $\boldsymbol{B}$ and $\boldsymbol{B}^*$ are s^*-admissible, and a slightly weaker but sufficient variant of (6.4) holds for $\texttt{RHS}_{\boldsymbol{B}^*\boldsymbol{f}}$. We note that, in principle, (6.4) can be satisfied be lowering the tolerances in (6.6) and appending a thresholding step, but the practical merit of this procedure is questionable.

6.2. **Approximate application of discrete operators.** The construction of a routine $\texttt{Apply}_{\boldsymbol{B}}$ hinges on the ability to approximate $\boldsymbol{B}$ by sparse matrices. We call a bi-infinite matrix $\boldsymbol{B} \in \mathcal{L}(\ell^2)$ *n-sparse* if each column of $\boldsymbol{B}$ contains at most n nonzero entries. It is *s^*-compressible* for an $s^* \in (0, \infty]$ if there exists a sequence $(\boldsymbol{B}_j)_{j\in\mathbb{N}}$ in $\mathcal{L}(\ell^2)$ such that $\boldsymbol{B}_j$ is n_j-sparse with $(n_j)_{j\in\mathbb{N}} \in \mathbb{N}^{\mathbb{N}}$ strictly increasing and satisfying

$$\sup_{j\in\mathbb{N}} \frac{n_{j+1}}{n_j} < \infty \tag{6.7}$$

and such that for every $s \in (0, s^*)$,

$$\sup_{j\in\mathbb{N}} n_j^s \|\boldsymbol{B} - \boldsymbol{B}_j\|_{\ell^2\to\ell^2} < \infty\,. \tag{6.8}$$

This last condition states that the sparse operators $\boldsymbol{B}_j$ converge to $\boldsymbol{B}$ with a rate of essentially s^* with respect to n_j.

By (6.7), the sequence $(n_j)_{j\in\mathbb{N}}$ grows at most geometrically. Consequently, for any $r > 0$, there is a $j(r)$ such that $n_{j(r)} \leq r$ and $\sup_{r>0} r^s\|\boldsymbol{B} - \boldsymbol{B}_{j(r)}\| < \infty$ for all $s \in (0, s^*)$. Here, we extend the sequence of approximations by $\boldsymbol{B}_0 \coloneqq \boldsymbol{0}$ with $n_0 = 0$. In particular, we may assume without loss of generality that $n_j = j$, as is done *e.g.* in [13, 22].

In the definition of s^*-compressibility, n_j is proportional to the cost of accessing one column of $\boldsymbol{B}_j$. In order to capture also the assembly cost, we introduce following [23, 13], a somewhat stronger notion of s^*-computability: a bi-infinite matrix $\boldsymbol{B} \in \mathcal{L}(\ell^2)$ is *s^*-computable* if it is s^*-compressible and if the number of arithmetic operations and storage locations used to construct an arbitrary column of $\boldsymbol{B}_j$ is bounded by a multiple of n_j for all $j \in \mathbb{N}$.

Proposition 6.4. *Any s^*-computable $\boldsymbol{B} \in \mathcal{L}(\ell^2)$ is s^*-admissible.*

Proposition 6.4 is proven by constructing a suitable routine $\texttt{Apply}_{\boldsymbol{B}}$. This is done in [4]; see also [8, 15] for a variant of this method with quantitative improvements. All of these algorithms partition the argument $\boldsymbol{w}$ in (6.1) according to the modulus of its entries. Then approximations $\boldsymbol{B}_j$ with large j are used for the most significant entries of $\boldsymbol{w}$, and coarser approximations suffice for smaller entries of $\boldsymbol{w}$.

7. Approximations of deterministic operators

7.1. **Compressibility of discrete parabolic operators.** The wavelets in Section 4 were chosen to ensure that the deterministic operators appearing in the series expansions (5.12) and (5.17) of $\boldsymbol{B}$ and $\boldsymbol{B}^*$ be s^*-computable.

Proposition 7.1. *For sufficiently smooth $\tilde{a}$ and a_m, $m \in \mathbb{N}$, the bi-infinite matrices $\tilde{\boldsymbol{B}}$, $\tilde{\boldsymbol{B}}^*$, $\boldsymbol{A}_m$ and $\boldsymbol{A}_m^*$ are s^*-computable with $s^* = \min(\tilde{d}_t, \tilde{d}_x)$.*

A proof of Proposition 7.1 is given in [22, Sec. 8]; see also [21] for time-independent operators. We refer to [22] for compressibility properties in more general settings, for example if D is not a product domain.

Proposition 7.1 implies that there is a sequence $(\tilde{\boldsymbol{B}}_j)_{j\in\mathbb{N}}$ of bi-infinite matrices such that $\tilde{\boldsymbol{B}}_j$ is $n_{0,j}$-sparse with $(n_{0,j})_{j\in\mathbb{N}}$ increasing and satisfying (6.7), and

$$\|\tilde{\boldsymbol{B}} - \tilde{\boldsymbol{B}}_j\|_{\ell^2(\Xi)\to\ell^2(\Upsilon)} \lesssim n_{0,j}^{-s} \qquad \forall s \in (0, s^*)\,. \tag{7.1}$$

Furthermore, the number of arithmetic operations and storage locations required to compute any column of $\tilde{\boldsymbol{B}}_j$ is an affine function of $n_{0,j}$. We extend these sequences by $\tilde{\boldsymbol{B}}_0 \coloneqq \boldsymbol{0}$ and $n_{0,0} \coloneqq 0$.

Analogous properties hold for $\tilde{\boldsymbol{B}}^*$, $\boldsymbol{A}_m$ and $\boldsymbol{A}_m^*$. We denote the sequences of sparse approximations by $(\tilde{\boldsymbol{B}}_j^*)_{j\in\mathbb{N}_0}$, $(\boldsymbol{A}_{m,j})_{j\in\mathbb{N}_0}$ and $(\boldsymbol{A}_{m,j}^*)_{j\in\mathbb{N}_0}$, and the corresponding sparsities by $(n_{0,j}^*)_{j\in\mathbb{N}_0}$, $(n_{m,j})_{j\in\mathbb{N}_0}$ and $(n_{m,j}^*)_{j\in\mathbb{N}_0}$, respectively. Although $\tilde{\boldsymbol{B}}_j^*$ may be the adjoint of $\tilde{\boldsymbol{B}}_j$ in some situations, as suggested by the notation, this is not assumed; similarly, $\boldsymbol{A}_{m,j}^*$ need not be the adjoint of $\boldsymbol{A}_{m,j}$.

7.2. **Numerical approximation of error bounds.** In order to construct sparse approximations of $\boldsymbol{B}$, we require explicit knowledge of the constants in the estimates (7.1) and similar estimates for $\boldsymbol{A}_m$ or, more precisely, we require numerical sequences $(\tilde{e}_{m,j})_{j\in\mathbb{N}_0}$, $m \in \mathbb{N}_0$, such that

$$\|\tilde{\boldsymbol{B}} - \tilde{\boldsymbol{B}}_j\|_{\ell^2(\Xi)\to\ell^2(\Upsilon)} \le \tilde{e}_{0,j} \quad \text{and} \quad \|\boldsymbol{A}_m - \boldsymbol{A}_{m,j}\|_{\ell^2(\Xi)\to\ell^2(\Upsilon)} \le \tilde{e}_{m,j}\,. \tag{7.2}$$

Optimal values of $\tilde{e}_{0,j}$ are given by the square roots of the spectral radii of the positive symmetric operators $(\tilde{\boldsymbol{B}} - \tilde{\boldsymbol{B}}_j)^*(\tilde{\boldsymbol{B}} - \tilde{\boldsymbol{B}}_j)$ since

$$\|\tilde{\boldsymbol{B}} - \tilde{\boldsymbol{B}}_j\|_{\ell^2(\Xi)\to\ell^2(\Upsilon)}^2 = \sup_{\|\boldsymbol{v}\|_{\ell^2(\Xi)}=1} |\boldsymbol{v}^H(\tilde{\boldsymbol{B}} - \tilde{\boldsymbol{B}}_j)^*(\tilde{\boldsymbol{B}} - \tilde{\boldsymbol{B}}_j)\boldsymbol{v}|\,, \tag{7.3}$$

and similarly for $\tilde{e}_{m,j}$. Following [15, Sec. 6], we approximate these bounds by a power iteration with suitably approximated matrix-vector multiplies.

The primary component of this power iteration is the repeated approximate application of the operators $(\tilde{\boldsymbol{B}} - \tilde{\boldsymbol{B}}_j)^*$ and $\tilde{\boldsymbol{B}} - \tilde{\boldsymbol{B}}_j$ to finitely supported vectors.

This is achieved using the sparse approximations $\tilde{\boldsymbol{B}}_{j+k} - \tilde{\boldsymbol{B}}_j$, $k \in \mathbb{N}$, of $\tilde{\boldsymbol{B}} - \tilde{\boldsymbol{B}}_j$ and $\tilde{\boldsymbol{B}}_{j+k}^* - (\tilde{\boldsymbol{B}}_j)^*$ of $(\tilde{\boldsymbol{B}} - \tilde{\boldsymbol{B}}_j)^*$ in routines

$$\texttt{NApply}_{\boldsymbol{B}-\boldsymbol{B}_j}[\boldsymbol{v}, N] \mapsto \boldsymbol{z} \quad \text{and} \quad \texttt{NApply}_{(\boldsymbol{B}-\boldsymbol{B}_j)^*}[\boldsymbol{w}, N] \mapsto \boldsymbol{z} \tag{7.4}$$

similar to $\texttt{Apply}_{\boldsymbol{B}}$ from (6.1), but with a prescribed maximal support size $\#\operatorname{supp} \boldsymbol{z} \leq N$ instead of a target accuracy ϵ. These routines combine to

$$\texttt{NApply}_{(\boldsymbol{B}-\boldsymbol{B}_j)^*(\boldsymbol{B}-\boldsymbol{B}_j)}[\boldsymbol{v}, N] \coloneqq \texttt{NApply}_{(\boldsymbol{B}-\boldsymbol{B}_j)^*}[\texttt{NApply}_{\boldsymbol{B}-\boldsymbol{B}_j}[\boldsymbol{v}, N], N]\,. \tag{7.5}$$

All vectors appearing within these routines are ensured to have support size not larger than N.

The approximate power iteration for the computation of $\tilde{e}_{0,j}$ consists of repeated application of $\texttt{NApply}_{(\boldsymbol{B}-\boldsymbol{B}_j)^*(\boldsymbol{B}-\boldsymbol{B}_j)}$ and normalization of the resulting vector. The approximations of $\tilde{e}_{0,j}$ are given by the scalar products

$$\tilde{e}_{0,j} \approx \tilde{e}_{0,j}^n \coloneqq \frac{\boldsymbol{v}_n^H \boldsymbol{v}_{n+1}}{\boldsymbol{v}_n^H \boldsymbol{v}_n}\,, \quad \boldsymbol{v}_{n+1} \coloneqq \texttt{NApply}_{(\boldsymbol{B}-\boldsymbol{B}_j)^*(\boldsymbol{B}-\boldsymbol{B}_j)}[\boldsymbol{v}_n, N]\,, \tag{7.6}$$

where $\boldsymbol{v}_0$ is chosen randomly and N is fixed. We use analogous iterations to approximate $\tilde{e}_{m,j}$ as well as the bounds in

$$\|\tilde{\boldsymbol{B}}^* - \tilde{\boldsymbol{B}}_j^*\|_{\ell^2(\Xi)\to\ell^2(\Upsilon)} \leq \tilde{e}_{0,j}^* \quad \text{and} \quad \|\boldsymbol{A}_m^* - \boldsymbol{A}_{m,j}^*\|_{\ell^2(\Xi)\to\ell^2(\Upsilon)} \leq \tilde{e}_{m,j}^* \tag{7.7}$$

used to construct sparse approximations of $\boldsymbol{B}^*$.

Convergence of a somewhat idealized variant of (7.6) is shown in [15, Thm. 6.3]. The analysis differs substantially from the standard analysis of the power iteration for matrices since, in the present infinite-dimensional setting, no gap between the largest and second-largest eigenvalues is assumed. We refer to [15] for further details.

8. Approximations of discrete random parabolic operators

8.1. **Sparse approximation of discrete random operators.** We construct sparse approximations of $\boldsymbol{B}$ by truncating the series (5.12) and by replacing the remaining bi-infinite matrices $\tilde{\boldsymbol{B}}$ and $\boldsymbol{A}_m$ by appropriate sparse approximations $\tilde{\boldsymbol{B}}_{j_0}$ and $\boldsymbol{A}_{m,j_m}$.

To this end, we assume that sequences $(\tilde{e}_{m,j})_{j\in\mathbb{N}_0}$ are available for all $m \in \mathbb{N}_0$ such that (7.2) holds. These can be computed numerically as described in Section 7.2 and [15, Sec. 6] or derived analytically as in [23, 3]. By switching to a subsequence, we assume without loss of generality that $(\tilde{e}_{m,j})_{j\in\mathbb{N}_0}$ is nonincreasing for all $m \in \mathbb{N}_0$ and, if $i \geq j$, then

$$\frac{-(\tilde{e}_{m,i+1} - \tilde{e}_{m,i})}{n_{m,i+1} - n_{m,i}} \leq \frac{-(\tilde{e}_{m,j+1} - \tilde{e}_{m,j})}{n_{m,j+1} - n_{m,j}}\,, \tag{8.1}$$

where $\tilde{\boldsymbol{B}}_{j,0}$ is $n_{0,j}$-sparse and $\boldsymbol{A}_{m,j}$ is $n_{m,j}$-sparse.

For all finitely supported sequences $\boldsymbol{j} \coloneqq (j_m)_{m\in\mathbb{N}_0}$ in $\mathbb{N}_0$, we define

$$\boldsymbol{B}_{\boldsymbol{j}} \coloneqq \boldsymbol{I} \otimes \tilde{\boldsymbol{B}}_{j_0} + \sum_{m=1}^{\infty} \boldsymbol{K}_m \otimes \boldsymbol{A}_{m,j_m}\,. \tag{8.2}$$

Since $\boldsymbol{j}$ is finitely supported and since $\boldsymbol{A}_{m,0} = \boldsymbol{0}$ for all m, the sum in (8.2) is finite, and no convergence issues arise. By the triangle inequality,

$$\|\boldsymbol{B} - \boldsymbol{B}_{\boldsymbol{j}}\|_{\ell^2(\Lambda\times\Xi)\to\ell^2(\Lambda\times\Upsilon)} \leq \sum_{m=0}^{\infty} \tilde{e}_{m,j_m} =: \tilde{e}_{\boldsymbol{j}}\,. \tag{8.3}$$

By (5.11), $\boldsymbol{K}_m$ is σ_m-sparse with $\sigma_m = 2$ if the distribution π_m is symmetric, and $\sigma_m = 3$ in general. Consequently, $\boldsymbol{K}_m \otimes \boldsymbol{A}_{m,j}$ is $\sigma_m n_{m,j}$-sparse. We set $\sigma_0 \coloneqq 1$

such that $\boldsymbol{I} \otimes \tilde{\boldsymbol{B}}_j$ is $\sigma_0 n_{0,j}$-sparse for all $j \in \mathbb{N}_0$. Then the total number of nonzero elements in any column of $\boldsymbol{B}_{\boldsymbol{j}}$ is at most

$$N_{\boldsymbol{j}} \coloneqq \sum_{m=0}^{\infty} \sigma_m n_{m,j_m}\,, \tag{8.4}$$

and, assuming that entries of $\tilde{\boldsymbol{B}}_j$ and $\boldsymbol{A}_{m,j}$ can be computed in unit time, $N_{\boldsymbol{j}}$ is also a bound for the cost of constructing any column of $\boldsymbol{B}_{\boldsymbol{j}}$.

We use a greedy algorithm to select a sequence $(\boldsymbol{j}_k)_{k\in\mathbb{N}_0}$, and define $\boldsymbol{B}_k \coloneqq \boldsymbol{B}_{\boldsymbol{j}_k}$, which is an approximation of $\boldsymbol{B}$ with error at most $\tilde{e}_k \coloneqq \tilde{e}_{\boldsymbol{j}_k}$, and containing at most $N_k \coloneqq N_{\boldsymbol{j}_k}$ nonzero elements per column.

As usual, the initial approximation is $\boldsymbol{B}_0 = \boldsymbol{0}$, with $\boldsymbol{j}_0 \coloneqq \boldsymbol{0}$. Going from $\boldsymbol{j}_k = (j_{k,m})_{m\in\mathbb{N}_0}$ to $\boldsymbol{j}_{k+1}$, the entry $j_{k,m}$ for which m maximizes

$$\frac{-(\tilde{e}_{m,j_{k,m}+1} - \tilde{e}_{m,j_{k,m}})}{\sigma_m(n_{m,j_{k,m}+1} - n_{m,j_{k,m}})} \tag{8.5}$$

is incremented by one. If this m is not unique, one maximum is selected, *e.g.* the smallest m that maximizes (8.5).

In order to ensure optimality of this greedy algorithm, we assume that the sequence $(\tilde{e}_{m,0})_{m\in\mathbb{N}}$ is in ℓ^1 and $\sigma_m^{-1} n_{m,1}^{-1}(\tilde{e}_{m,1} - \tilde{e}_{m,0})$ is nonincreasing in m. The following optimality property of the sparse approximations $\boldsymbol{B}_k$ is [15, Cor. 7.2].

Theorem 8.1. *For all $k \in \mathbb{N}_0$, $\boldsymbol{j}_k$ minimizes the error bound $\tilde{e}_{\boldsymbol{j}}$ among all finitely supported sequences $\boldsymbol{j}$ in $\mathbb{N}_0$ with $N_{\boldsymbol{j}} \le N_k$. Furthermore, if $\tilde{e}_k \neq 0$, then $\boldsymbol{j}_k$ minimizes $N_{\boldsymbol{j}}$ among all $\boldsymbol{j}$ with $\tilde{e}_{\boldsymbol{j}} \le \tilde{e}_k$.*

Remark 8.2. Since the structure of $\boldsymbol{B}^*$ is identical to that of $\boldsymbol{B}$, the discussion in this section, including Sec. 8.2 below, applies verbatim to the adjoint operator $\boldsymbol{B}^*$, with $\tilde{\boldsymbol{B}}$ replaced by $\tilde{\boldsymbol{B}}^*$ and $\boldsymbol{A}_m$ replaced by $\boldsymbol{A}_m^*$.

8.2. **Compressibility and computability.** In order to derive s^*-compressibility of $\boldsymbol{B}$, we assume that the estimate (6.7) holds uniformly for all $(n_{m,j})_{j\in\mathbb{N}}$, *i.e.*

$$\sup_{m\in\mathbb{N}_0} \sup_{j\in\mathbb{N}} \frac{n_{m,j+1}}{n_{m,j}} < \infty\,. \tag{8.6}$$

The following theorem is the first case of [15, Thm. 8.4]. All unspecified norms refer to operator norms between sequence spaces ℓ^2 for the appropriate index sets.

Theorem 8.3. *If* (8.6) *holds, $\tilde{a}$ and a_m, $m \in \mathbb{N}$, are sufficiently smooth, and*

$$\sum_{m=1}^{\infty} \big(\sup_{j\in\mathbb{N}} n_{m,j}^s \|\boldsymbol{A}_m - \boldsymbol{A}_{m,j}\|\big)^{\frac{1}{s+1}} < \infty \tag{8.7}$$

for all $s \in (0, \bar{s})$, then $\boldsymbol{B}$ is s^*-compressible for $s^* = \min(\tilde{d}_t, \tilde{d}_x, \bar{s})$ and $(\boldsymbol{B}_k)_{k\in\mathbb{N}_0}$ from Sec. 8.1 is a valid sequence of sparse approximations, satisfying

$$N_k^s \|\boldsymbol{B} - \boldsymbol{B}_k\| \le \left(\big(\sup_{j\in\mathbb{N}} n_{0,j}^s \|\tilde{\boldsymbol{B}} - \tilde{\boldsymbol{B}}_j\|\big)^{\frac{1}{s+1}} + \sum_{m=1}^{\infty} \big(\sup_{j\in\mathbb{N}} n_{m,j}^s \|\boldsymbol{A}_m - \boldsymbol{A}_{m,j}\|\big)^{\frac{1}{s+1}} \right)^{s+1}, \tag{8.8}$$

for all $s \in (0, s^*)$, where $\boldsymbol{B}_k$ is N_k-sparse.

Compressibility of $\boldsymbol{B}$ can also be derived if (8.7) does not hold, as in the following theorem, which is the second case of [15, Thm. 8.4].

Theorem 8.4. *If* (8.6) *holds, $\tilde{a}$ and a_m, $m \in \mathbb{N}$, are sufficiently smooth,*

$$\sum_{m=1}^{\infty} \|\boldsymbol{A}_m\|^{\frac{1}{s+1}} < \infty \tag{8.9}$$

for all $s \in (0, \bar{s}_0)$*, and*

$$\sup_{M\in\mathbb{N}} M^{-\tau} \sum_{m=1}^{M} \big(\sup_{j\in\mathbb{N}} n_{m,j}^{s} \|\boldsymbol{A}_m - \boldsymbol{A}_{m,j}\|_{\ell^2(\Xi)\to\ell^2(\Upsilon)}\big)^{\frac{1}{s+1}} < \infty \tag{8.10}$$

for all $s \in (0, \hat{s})$*, then* $\boldsymbol{B}$ *is* s^**-compressible for*

$$s^* := \frac{\min(\tilde{d}_t, \tilde{d}_x, \hat{s})}{1 + \tau/\bar{s}_0} \tag{8.11}$$

and $(\boldsymbol{B}_k)_{k\in\mathbb{N}_0}$ *from Sec. 8.1 is a valid sequence of sparse approximations.*

Remark 8.5. A numerical algorithm for constructing an arbitrary column of $\boldsymbol{B}_k$ is provided in [15, Sec. 7.2]. It assumes that either $(\boldsymbol{j}_k)$ are precomputed, or the operators $\boldsymbol{B}_k$ are accessed sequentially, such that only one step of the greedy optimization needs to be performed the first time $\boldsymbol{B}_k$ is accessed. With this small caveat, s^*-computability, and thus s^*-admissibility, of $\boldsymbol{B}$ follow from Theorems 8.3 and 8.4.

The above discussion carries over to show s^*-computability of $\boldsymbol{B}^*$, and s^*-admissibility of $\boldsymbol{B}^*\boldsymbol{B}$ follows as in Remark 6.3. In particular, Theorem 6.1 applies, showing optimality of adaptive wavelet methods applied to (5.15).

References

[1] R. Andreev, *Stability of space-time Petrov-Galerkin discretizations for parabolic evolution equations*, PhD thesis, ETH Zürich, 2012. Dissertation No. 20842, Examiner Prof. Dr. Christoph Schwab.

[2] ———, *Stability of sparse space-time finite element discretizations of linear parabolic evolution equations*, IMA Journal of Numerical Analysis, (2012).

[3] A. Barinka, S. Dahlke, and W. Dahmen, *Adaptive application of operators in standard representation*, Adv. Comput. Math., 24 (2006), pp. 5–34.

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

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

[6] W. Dahmen, A. Kunoth, and K. Urban, *Biorthogonal spline wavelets on the interval—stability and moment conditions*, Appl. Comput. Harmon. Anal., 6 (1999), pp. 132–196.

[7] R. Dautray and J.-L. Lions, *Mathematical analysis and numerical methods for science and technology. Vol. 5*, Springer-Verlag, Berlin, 1992. Evolution problems. I, With the collaboration of Michel Artola, Michel Cessenat and Hélène Lanchon, Translated from the French by Alan Craig.

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

[9] G. C. Donovan, J. S. Geronimo, and D. P. Hardin, *Intertwining multiresolution analyses and the construction of piecewise-polynomial wavelets*, SIAM J. Math. Anal., 27 (1996), pp. 1791–1815.

[10] G. C. Donovan, J. S. Geronimo, and D. P. Hardin, *Orthogonal polynomials and the construction of piecewise polynomial smooth wavelets*, SIAM J. Math. Anal., 30 (1999), pp. 1029–1056.

[11] ———, *Compactly supported, piecewise affine scaling functions on triangulations*, Constr. Approx., 16 (2000), pp. 201–219.

[12] O. G. Ernst, A. Mugler, H.-J. Starkloff, and E. Ullmann, *On the convergence of generalized polynomial chaos expansions*, Tech. Rep. 60, DFG Schwerpunktprogramm 1324, 2010.

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

[14] W. Gautschi, *Orthogonal polynomials: computation and approximation*, Numerical Mathematics and Scientific Computation, Oxford University Press, New York, 2004. Oxford Science Publications.

[15] C. J. Gittelson, *Adaptive wavelet methods for elliptic partial differential equations with random operators*, Tech. Rep. 2011-37, Seminar for Applied Mathematics, ETH Zürich, 2011.

[16] ———, *Stochastic Galerkin approximation of operator equations with infinite dimensional noise*, Tech. Rep. 2011-10, Seminar for Applied Mathematics, ETH Zürich, 2011.
[17] T. N. T. Goodman, *A class of orthogonal refinable functions and wavelets*, Constr. Approx., 19 (2003), pp. 525–540.
[18] B. Han and Z. Shen, *Wavelets with short support*, SIAM J. Math. Anal., 38 (2006), pp. 530–556 (electronic).
[19] M. Primbs, *New stable biorthogonal spline-wavelets on the interval*, Results Math., 57 (2010), pp. 121–162.
[20] C. Schwab and C. J. Gittelson, *Sparse tensor discretization of high-dimensional parametric and stochastic PDEs*, in Acta Numerica, vol. 20 of Acta Numer., Cambridge Univ. Press, Cambridge, 2011, pp. 291–467.
[21] C. Schwab and R. Stevenson, *Adaptive wavelet algorithms for elliptic PDE's on product domains*, Math. Comp., 77 (2008), pp. 71–92 (electronic).
[22] ———, *Space-time adaptive wavelet methods for parabolic evolution problems*, Math. Comp., 78 (2009), pp. 1293–1318.
[23] R. Stevenson, *On the compressibility of operators in wavelet coordinates*, SIAM J. Math. Anal., 35 (2004), pp. 1110–1132 (electronic).
[24] G. Szegő, *Orthogonal polynomials*, American Mathematical Society, Providence, R.I., fourth ed., 1975. American Mathematical Society, Colloquium Publications, Vol. XXIII.
[25] J. Wloka, *Partielle Differentialgleichungen*, B. G. Teubner, Stuttgart, 1982. Sobolevräume und Randwertaufgaben. [Sobolev spaces and boundary value problems], Mathematische Leitfäden. [Mathematical Textbooks].

Department of Mathematics, Purdue University, 150 N. University Street, West Lafayette, IN 47907-2067, USA
E-mail address: cgittels@purdue.edu

CSCAMM, University of Maryland, CSIC Building #406, Paint Branch Dr., College Park, MD 20742, USA
E-mail address: andreevr@umd.edu

Seminar for Applied Mathematics, ETH Zürich, Rämistrasse 101, CH-8092 Zürich, Switzerland
E-mail address: schwab@sam.math.ethz.ch