Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Seminar for Applied Mathematics SAM

Compressive Space-Time Galerkin Discretizations of Parabolic Partial Differential Equations

S. Larsson and Ch. Schwab

Research Report No. 2015-04
January 2015

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

Funding ERC: AdG no. 247277

COMPRESSIVE SPACE-TIME GALERKIN DISCRETIZATIONS OF PARABOLIC PARTIAL DIFFERENTIAL EQUATIONS

STIG LARSSON AND CHRISTOPH SCHWAB

Abstract. We study linear parabolic initial-value problems in a space-time variational formulation based on fractional calculus. This formulation uses "time derivatives of order one half" on the bi-infinite time axis. We show that for linear, parabolic initial-boundary value problems on $(0,\infty)$, the corresponding bilinear form admits an inf-sup condition with sparse tensor product trial and test function spaces. We deduce optimality of compressive, space-time Galerkin discretizations, where stability of Galerkin approximations is implied by the well-posedness of the parabolic operator equation. The variational setting adopted here admits more general Riesz bases than previous work; in particular, *no stability in negative order Sobolev spaces on the spatial or temporal domains* is required of the Riesz bases accommodated by the present formulation. The trial and test spaces are based on Sobolev spaces of equal order $1/2$ with respect to the temporal variable. Sparse tensor products of multi-level decompositions of the spatial and temporal spaces in Galerkin discretizations lead to large, non-symmetric linear systems of equations. We prove that their condition numbers are uniformly bounded with respect to the discretization level. In terms of the total number of degrees of freedom, the convergence orders equal, up to logarithmic terms, those of best N-term approximations of solutions of the corresponding elliptic problems.

Date: January 19, 2015.

1991 *Mathematics Subject Classification.* 65N30.

Key words and phrases. Fractional Calculus, Parabolic Problems, Wavelets, Adaptivity, Space-Time Discretization, Compressive Galerkin.

Acknowledgement. The work of Christoph Schwab was supported in part by ERC AdG no. 247277.

Contents

1. Introduction

For a bounded linear and self-adjoint operator $A \in \mathcal{L}(V, V^*)$ in an evolution triplet $V \subset H \simeq H^* \subset V^*$, and a bounded domain $D \subset \mathbb{R}^n$, we consider the initial boundary value problem for abstract, linear parabolic evolution equations

$$Bu := \partial_t u + Au = f \quad \text{in} \quad \mathbb{R}_> = (0, \infty) , \tag{1.1}$$

with homogeneous initial condition

$$u(0) = 0 . \tag{1.2}$$

In (1.1), we think of A as linear, strongly elliptic (pseudo)differential operator of order $2m > 0$, and of V as a closed subspace of $H^m(D)$ supporting homogeneous, essential boundary conditions of the initial boundary value problem (1.1), (1.2).

Optimality of adaptive variational space-time Galerkin discretizations of (1.1), (1.2) on $(0, T)$ for $T < \infty$ were shown for the first time in [20]. There, well-posedness of suitable space-time variational saddle-point formulations of the parabolic initial boundary value problems (1.1), (1.2) were established. By means of tensorized Riesz bases of the Bochner spaces which underlie the space-time variational formulations, the parabolic initial boundary value problems were converted to equivalent bi-infinite matrix problems. These matrix problems were subsequently solved numerically, in optimal complexity, by means of adaptive wavelet discretizations from [6]. We note that *adaptive wavelet techniques* from [6] were essential in the algorithms in [20], since it used the paradigm "stability by adaptivity" from [6]. In particular in [20], no stability result for nonadaptive discretizations could be obtained, but rather followed from the well-posedness of the infinite-dimensional problem, the Riesz basis property and certain optimality properties of the adaptive Galerkin discretizations ("stability by adaptivity").

In the present paper, building on fractional calculus techniques pioneered in variational formulations of parabolic initial boundary value problems by M. Fontes [10, 11], we propose a space-time variational formulation based on bilinear forms, which are, unlike the formulations considered in [20], *"symmetric"* in the sense that trial and test spaces, which arise in the variational formulation, are Sobolev spaces of equal orders with respect to time differentiation. Stability (in the sense that a discrete inf-sup condition holds) of our space-time Galerkin discretization requires that the finite-dimensional trial and test spaces are different.

The presently considered space-time variational formulation admits a unique variational solution in a Bochner space X, which is intermediate to the solution spaces which are obtained by the "classical" approach. Moreover, as shown by M. Fontes in [11, 12], the presently considered solutions can be obtained by monotone operator methods and, therefore, Galerkin approximations are well-defined and stable with any closed subspaces, including in particular sparse tensor products of multilevel hierarchies in space and time. It is interesting to note that time derivatives of order $1/2$ were used already in [2, 17] in order to prove error estimates in the X-norm for finite element approximations of (1.1)–(1.2).

As in [20], we establish in the present paper quasi-optimality of linear and nonlinear space-time adaptive and compressive Galerkin discretizations in the space-time cylinder. To this end, we show a discrete inf-sup condition in the present paper, for a suitable sparse tensor space-time Petrov-Galerkin discretization. The use of wavelet-type Riesz bases in space and time then results in uniformly bounded condition numbers of the finite-dimensional problems; notably, this holds without the Riesz basis property in V^* of the spatial wavelet basis

Σ, which was essential in [20]. In the presently considered variational formulation, we consider in particular long-term evolution, i.e., the time interval $(0,T)$ with $T=\infty$, and analyze space-time compressive and adaptive numerical approximation of long-time integration for these problems. Unlike [1, 20], we obtain stability, multilevel preconditioning and space-time compressibility even without adaptivity, and with trial and test spaces of equal dimension (albeit being possibly different so that we consider a Petrov-Galerkin formulation as in [1]). Moreover, the optimality results in Section 5 entail optimal, adaptive and space-time compressive methods for long-time integration (i.e., $T=\infty$) for parabolic evolution problems.

The outline of this paper is as follows: in Section 2, we present basic definitions and facts from functional analysis and fractional calculus. In Section 3, we present the space-time variational formulation of (1.1), (1.2). Section 4, we consider compressive space-time Galerkin discretization with sparse tensor subspaces. Section 5 addresses the space-time *adaptive* discretization of the variational formulation in Section 3 and establishes optimality. The analysis in Sections 3–5 is developed for long-time integration, i.e., for $T=\infty$.

2. Preliminaries

2.1. **Functional analysis.** We require some tools from functional analysis. Throughout this paper all vector spaces are real unless explicitly stated otherwise. Consider two Banach spaces X and Y and a bilinear form $\mathcal{B}\colon X\times Y\to\mathbb{R}$, which is bounded, i.e., there exists a constant C such that

$$|\mathcal{B}(w,v)|\le C\|w\|_X\|v\|_Y \quad \forall w\in X\ ,\ v\in Y\ . \tag{2.1}$$

We are interested in solving the linear, variational problem: for each $F\in Y^*$, find a unique $u\in X$ such that

$$\mathcal{B}(u,v)=F(v) \quad \forall v\in Y\ . \tag{2.2}$$

The form $\mathcal{B}(\cdot,\cdot)$ induces in a one-to-one fashion a bounded, linear operator $B\in\mathcal{L}(X,Y^*)$ via

$$ {}_{Y^*}\langle Bw,v\rangle_Y=\mathcal{B}(w,v) \quad \forall w\in X\ ,\ v\in Y\ ,$$

so that the unique solvability of (2.2) is related to the question of bounded invertibility of the operator $B\in\mathcal{L}(X,Y^*)$. There holds:

Proposition 2.1. *Let X,Y be Banach spaces; Y reflexive. Let $\mathcal{B}\colon X\times Y\to\mathbb{R}$ be a bounded bilinear form and consider the inf-sup condition:*

$$\inf_{0\neq w\in X}\sup_{0\neq v\in Y}\frac{\mathcal{B}(w,v)}{\|w\|_X\|v\|_Y}\ge\gamma>0\ , \tag{2.3}$$

and the (adjoint) injectivity condition:

$$\sup_{w\in X}\mathcal{B}(w,v)>0 \quad \forall 0\neq v\in Y\ . \tag{2.4}$$

The conditions (2.3)–(2.4) *hold if and only if for each $F\in Y^*$, the variational problem* (2.2) *admits a unique solution $u\in X$ and in this case there holds the estimate*

$$\|u\|_X\le\frac{1}{\gamma}\|F\|_{Y^*}\ .$$

In other words, (2.3)–(2.4) *hold if and only if the corresponding operator $B\in\mathcal{L}(X,Y^*)$ is boundedly invertible, in which case $\|B^{-1}\|_{\mathcal{L}(Y^*,X)}\le\gamma^{-1}$.*

The Proposition 2.1 was used in [20] in verifying that space-time saddle point formulations of (1.1) are well-posed. Below, we shall be interested in the following special case where $X = Y$.

Corollary 2.2. *Assume that X is a reflexive Banach space, and that the bounded bilinear form $\mathcal{B}\colon X \times X \to \mathbb{R}$ is* coercive-equivalent*, i.e., there exists an isomorphism $S \in \mathcal{L}(X, X)$ such that $\mathcal{B}(\cdot, S\cdot)$ is coercive, i.e., there exists $c > 0$ such that*

$$\mathcal{B}(w, Sw) \geq c\|w\|_X^2 \quad \forall w \in X \ . \tag{2.5}$$

Then the corresponding operator $B \in \mathcal{L}(X, X^*)$ is boundedly invertible.

Proof. We assume (2.5) and verify conditions (2.3)–(2.4) in Proposition 2.1. For $0 \neq w \in X$, we have $\|Sw\|_X \leq c_S\|w\|_X$ and $Sw \neq 0$, since S is an isomorphism. Together with (2.5) this leads to

$$\sup_{0\neq v\in X} \frac{\mathcal{B}(w,v)}{\|v\|_X} \geq \frac{\mathcal{B}(w,Sw)}{\|Sw\|_X} \geq c\frac{\|w\|_X^2}{\|Sw\|_X} \geq \frac{c}{c_S}\|w\|_X \ .$$

This proves (2.3). To verify (2.4) we compute

$$\sup_{w\in X} \mathcal{B}(w,v) \geq \mathcal{B}(S^{-1}v, v) = \mathcal{B}(S^{-1}v, S(S^{-1}v)) \geq c\|S^{-1}v\|_X^2 \geq \frac{c}{c_S^2}\|v\|_X^2 > 0$$

for $0 \neq v \in X$. □

2.2. **The elliptic operator.** We let $(H, \langle\cdot,\cdot\rangle_H)$ and $(V, \langle\cdot,\cdot\rangle_V)$ denote two separable Hilbert spaces with dense embedding $V \subset H$ and duals H^* and V^*. We identify $H \simeq H^*$ according to the Riesz representation theorem and obtain the Gel'fand triple

$$V \subset H \simeq H^* \subset V^* \ ,$$

again with dense injections. Let $A \in \mathcal{L}(V, V^*)$ be a bounded self-adjoint linear operator such that the corresponding bilinear form $a(v, w) = {}_{V^*}\langle Av, w\rangle_V$ is coercive and bounded on $V \times V$, i.e., for some $0 < \lambda_- \leq \lambda_+ < \infty$,

$$a(v,v) \geq \lambda_-\|v\|_V^2 \ , \quad |a(v,w)| \leq \lambda_+\|v\|_V\|w\|_V \ . \tag{2.6}$$

Example 2.3. *In a bounded Lipschitz domain $D \subset \mathbb{R}^n$ of dimension $n \geq 1$, we consider the linear, second order divergence form operator given for $v \in C_0^\infty(D)$ by*

$$Av = -\nabla\cdot(a(x)\nabla v) + c(x)v \ .$$

Here, $a \in (L^\infty(D))^{n\times n}_{\rm sym}$ and $c \in L^\infty(D)$ satisfy the ellipticity conditions

$$\exists\gamma > 0 \ \ \forall\xi \in \mathbb{R}^n : \quad \xi^\top a\,\xi \geq \gamma|\xi|^2 \ , \quad \operatorname{ess\,inf}_{x\in D} c(x) \geq 0 \ .$$

In this case $V = H_0^1(D)$, $H = L^2(D)$, $a(v,w) = (a\nabla v, \nabla w) + (cv, w)$, and (2.6) is valid.

Example 2.4. *With D as in Example 2.3, we consider the Stokes equation. Then*

$$\begin{aligned} H &= \{v \in L^2(D)^n : \operatorname{div} v = 0 \ \text{ in } \ L^2(D) \ , \gamma_0(v\cdot n) = 0 \ \text{ in } \ H^{-\frac{1}{2}}(\partial D)\} \ , \\ V &= \{v \in H_0^1(D)^n : \operatorname{div} v = 0 \ \text{ in } \ L^2(D)\} \ , \end{aligned}$$

where γ_0 denotes the trace operator and the bilinear form is given by $a(w,v) = \int_D \nabla w : \nabla v \,\mathrm{d}x$.

2.3. **Bochner spaces.** We require Bochner spaces of vector-valued functions defined on intervals. For an interval I, a Banach space X with norm $\|\cdot\|_X$, and for $1 \le p \le \infty$, we denote by $L^p(I;X)$ the space of strongly measurable functions $u\colon I \mapsto X$ such that

$$\|u\|_{L^p(I;X)} = \left(\int_I \|u(t)\|_X^p \,\mathrm{d}t\right)^{1/p} < \infty$$

for $1 \le p < \infty$ with the usual modification for $p = \infty$. Similarly, we denote by $H^1(I;X)$ the space of functions whose distributional time derivative belongs to $L^2(I;X)$. We also need spaces of continuous functions: for $k \in \mathbb{N}_0$, we denote by $C^k(\bar{I};X)$ the Banach space of k-times continuously differentiable and bounded mappings $u\colon \bar{I} \mapsto X$ endowed with the standard norm $\|\cdot\|_{C^k(\bar{I};X)}$.

2.4. **Interpolation spaces.** We repeatedly use assorted facts from the theory of function space interpolation (see, e.g., [3, 18, 25]). In particular, we use the interpolation spaces $[X,Y]_s$, $0 < s < 1$, between two Hilbert spaces with dense embedding $X \subset Y$, as defined, for example, in [18, Chap. 1, Déf. 2.1].

For $0 < T \le \infty$ we denote by $I = (-T,T)$ the symmetric interval, with $I = \mathbb{R}$ implied if $T = \infty$, and set $I_> = I \cap \{t > 0\}$. For a separable Hilbert space H, we define

$$H^1_{0,\{0\}}(I_>;H) := \{v \in H^1(I_>;H) : v(0) = 0\} .$$

By the continuity of the embedding $H^1(I_>;H) \subset C^0(\overline{I_>};H)$, the set $H^1_{0,\{0\}}(I_>;H)$ is the null space of the trace operator at $t = 0$ and, therefore, a norm-closed, linear subspace of $H^1(I_>;H)$. We introduce the interpolation spaces

$$\begin{aligned}
H^s(I;H) &:= [L^2(I;H), H^1(I;H)]_s , && s \in (0,1) ,\\
H^s(I_>;H) &:= [L^2(I_>;H), H^1(I_>;H)]_s , && s \in (0,1) ,\\
H^s_{0,\{0\}}(I_>;H) &:= [L^2(I_>;H), H^1_{0\{0\}}(I_>;H)]_s , && s \in (0,1) \setminus \{\tfrac12\} ,\\
H^{\frac12}_{00,\{0\}}(I_>;H) &:= [L^2(I_>;H), H^1_{0,\{0\}}(I_>;H)]_{\frac12} .
\end{aligned}$$

Remark 2.5. *With $I_> = (0,T)$ for $0 < T \le \infty$ there holds:*
(1) *Consider the interpolation spaces $[L^2(I_>;H), H^1_{0,\{0\}}(I_>;H)]_s$ for $0 < s < 1$, $s \ne \frac12$. For $0 < s < \frac12$, it holds that $[L^2(I_>;H), H^1_{0,\{0\}}(I_>;H)]_s = H^s(I_>;H) = [L^2(I_>;H), H^1(I_>;H)]_s$, i.e., the homogeneous boundary condition at $\{0\}$ is "not seen" by the interpolation space, whereas for $\frac12 < s < 1$ we have that*

$$[L^2(I_>;H), H^1_{0,\{0\}}(I_>;H)]_s = H^s_{0,\{0\}}(I_>;H) \subset [L^2(I_>;H), H^1(I_>;H)]_s = H^s(I_>;H)$$

is a subspace which is norm-closed in $H^s(I_>;H)$, [18, Chap. 1, Remarque 11.3].
(2) *The space $H^{\frac12}_{00,\{0\}}(I_>;H)$, which will be important in the present paper, is strictly included in $H^{\frac12}(I_>;H) = [L^2(I_>;H), H^1(I_>;H)]_{\frac12}$ with a topology which is strictly finer than that of $H^{\frac12}(I_>;H)$,* [18, Chap. 1, Thm. 11.7].
(3) *$H^{\frac12}_{00,\{0\}}(I_>;H)$ is not closed in the norm of $H^{\frac12}(I_>;H)$. It is a dense subspace (*[16, Theorem 1.4.2.4] *with $p = 2$) and the embedding $H^{\frac12}_{00,\{0\}}(I_>;H) \subset H^{\frac12}(I_>;H)$ is continuous,* [12, Lemma 4.8].

The following intrinsic characterizations of the spaces of order $\frac{1}{2}$ will be useful. We refer to [18, Chap. 1], in particular, for the first one Théorème 9.1 and (10.23) in Section 10.3, and for the second one, Théorème 11.7 and Remarque 11.4.

Proposition 2.6. *Let $I = (-T, T)$, $I_> = (0, T)$ for $T \in (0, \infty]$.*
(1) *The interpolation space $H^{\frac{1}{2}}(I_>; H)$ consists of all $u \in L^2(I_>; H)$ which are equal to the restriction to $I_>$ of some $\tilde{u} \in H^{\frac{1}{2}}(I; H)$. The interpolation norm of $H^{\frac{1}{2}}(I_>; H)$ is equivalent to the intrinsic norm $\|\cdot\|_{H^{\frac{1}{2}}(I_>;H)}$ given by*

$$\|u\|^2_{H^{\frac{1}{2}}(I_>;H)} = \|u\|^2_{L^2(I_>;H)} + \int_{I_>}\int_{I_>} \frac{\|u(s)-u(t)\|^2_H}{|s-t|^2}\,\mathrm{d}s\,\mathrm{d}t\,. \tag{2.7}$$

(2) *The interpolation space $H^{\frac{1}{2}}_{00,\{0\}}(I_>; H)$ consists of all $u \in H^{\frac{1}{2}}(I_>; H)$ such that the function $s \mapsto s^{-\frac{1}{2}}u(s)$ belongs to $L^2(I_>; H)$ with intrinsic norm $\|\cdot\|_{H^{\frac{1}{2}}_{00,\{0\}}(I_>;H)}$ given by*

$$\|u\|^2_{H^{\frac{1}{2}}_{00,\{0\}}(I_>;H)} = \|u\|^2_{L^2(I_>;H)} + \int_{I_>}\int_{I_>} \frac{\|u(s)-u(t)\|^2_H}{|s-t|^2}\,\mathrm{d}s\,\mathrm{d}t + \int_{I_>}\frac{1}{s}\|u(s)\|^2_H\,\mathrm{d}s\,. \tag{2.8}$$

The constants implied by the norm equivalences are independent of $T \in (0, \infty]$.

2.5. **Fractional calculus on the half line.** To render our presentation self-contained, we recapitulate here fractional calculus from [19] as necessary by our subsequent analysis.

For $\phi \in L^1(\mathbb{R}_>; \mathbb{C})$, $\alpha \in (0, 1)$, the Riemann-Liouville fractional integrals [19, Def. 2.1] are

$$(\mathrm{I}^\alpha_+\phi)(t) = \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\phi(s)\,\mathrm{d}s\,, \quad t \in \mathbb{R}_>\,,$$
$$(\mathrm{I}^\alpha_-\phi)(t) = \frac{1}{\Gamma(\alpha)}\int_t^\infty (s-t)^{\alpha-1}\phi(s)\,\mathrm{d}s\,, \quad t \in \mathbb{R}_>\,.$$

Then we have integration by parts [19, (2.20) and Corollary to Theorem 3.5 p. 67]:

$$\int_{\mathbb{R}_>} (\mathrm{I}^\alpha_+\psi)(t)\phi(t)\,\mathrm{d}t = \int_{\mathbb{R}_>} \psi(t)(\mathrm{I}^\alpha_-\phi)(t)\,\mathrm{d}t \tag{2.9}$$

and the semigroup property [19, (2.21)]:

$$\mathrm{I}^{\alpha+\beta}_+\phi = \mathrm{I}^\alpha_+\mathrm{I}^\beta_+\phi\,, \quad \mathrm{I}^{\alpha+\beta}_-\phi = \mathrm{I}^\alpha_-\mathrm{I}^\beta_-\phi\,, \quad \alpha,\beta > 0\,. \tag{2.10}$$

The proofs of (2.9), (2.10) are elementary calculations with integrals.

By D we denote the time derivative of order 1 and we define time derivatives of fractional order $\alpha \in (0, 1)$ for $u \in C^\infty_0(\mathbb{R})$,

$$(\mathrm{D}^\alpha_+u)(t) := (\mathrm{D}\mathrm{I}^{1-\alpha}_+u)(t) = \frac{1}{\Gamma(1-\alpha)}\mathrm{D}\int_0^t (t-s)^{-\alpha}u(s)\,\mathrm{d}s\,, \tag{2.11}$$

$$(\mathrm{D}^\alpha_-u)(t) := -(\mathrm{D}\mathrm{I}^{1-\alpha}_-u)(t) = -\frac{1}{\Gamma(1-\alpha)}\mathrm{D}\int_t^\infty (s-t)^{-\alpha}u(s)\,\mathrm{d}s\,. \tag{2.12}$$

We require a space of test functions, which is closed under the action of D^α_+ and D^α_-; to this end we introduce (cp. [12])

$$\mathcal{F}(\mathbb{R}; \mathbb{C}) := \left\{u \in C^\infty(\mathbb{R}; \mathbb{C}) : \|u\|_{H^s(\mathbb{R};\mathbb{C})} < \infty \ \ \forall s \in \mathbb{R}\right\}\,.$$

The set $\mathcal{F}(\mathbb{R};\mathbb{C})$ is a Fréchet space with respect to the topology induced by the family of norms $\{\|\cdot\|_{H^s(\mathbb{R};\mathbb{C})}\}_{s\in\mathbb{R}}$ and we have the dense embeddings $\mathcal{D}(\mathbb{R};\mathbb{C}) \subset \mathcal{S}(\mathbb{R};\mathbb{C}) \subset \mathcal{F}(\mathbb{R};\mathbb{C}) \subset \mathcal{E}(\mathbb{R};\mathbb{C})$ (where $\mathcal{D}$, $\mathcal{S}$, and $\mathcal{E}$ are the classical test function spaces). We observe that the definitions (2.11), (2.12) remain meaningful for $u \in \mathcal{F}(\mathbb{R};\mathbb{C})$. We further define test function spaces

$$\mathcal{F}(\mathbb{R}_>;\mathbb{C}) = \{u \in C^\infty(\mathbb{R}_>;\mathbb{C}) : \exists \tilde{u} \in \mathcal{F}(\mathbb{R};\mathbb{C}) \text{ such that } u = \tilde{u}|_{\mathbb{R}_>}\}$$

and, with E_0 the "extension by zero" operator,

$$\mathcal{F}_0(\mathbb{R}_>;\mathbb{C}) = \{u \in C^\infty(\mathbb{R}_>;\mathbb{C}) : E_0 u \in \mathcal{F}(\mathbb{R};\mathbb{C})\} \ .$$

The subspaces $\mathcal{F}_0(\mathbb{R}_>;\mathbb{C}) \subset H^{\frac{1}{2}}_{00,\{0\}}(\mathbb{R}_>;\mathbb{C})$, $\mathcal{F}(\mathbb{R}_>;\mathbb{C}) \subset H^{\frac{1}{2}}(\mathbb{R}_>;\mathbb{C})$ are dense, see [12, Lemma 3.7].

We denote the corresponding spaces of distributions by $\mathcal{F}_0'(\mathbb{R}_>;\mathbb{C}) = \mathcal{F}(\mathbb{R}_>;\mathbb{C})^*$ and $\mathcal{F}'(\mathbb{R}_>;\mathbb{C}) = \mathcal{F}_0(\mathbb{R}_>;\mathbb{C})^*$. Then it follows that, see [12, (2.24)–(2.29)],

$$\mathrm{D}_+^\alpha : \mathcal{F}_0(\mathbb{R}_>;\mathbb{C}) \to \mathcal{F}_0(\mathbb{R}_>;\mathbb{C}) \ , \quad \mathrm{D}_-^\alpha : \mathcal{F}(\mathbb{R}_>;\mathbb{C}) \to \mathcal{F}(\mathbb{R}_>;\mathbb{C}) \ ,$$
$$\mathrm{D}_+^\alpha : \mathcal{F}_0'(\mathbb{R}_>;\mathbb{C}) \to \mathcal{F}_0'(\mathbb{R}_>;\mathbb{C}) \ , \quad \mathrm{D}_-^\alpha : \mathcal{F}'(\mathbb{R}_>;\mathbb{C}) \to \mathcal{F}'(\mathbb{R}_>;\mathbb{C}) \ .$$

Here the $\mathcal{F}_0'(\mathbb{R}_>;\mathbb{C})$ distribution derivative D_+^α means

$$\langle \mathrm{D}_+^\alpha u, \phi\rangle = \int_{\mathbb{R}_>} u\, \mathrm{D}_-^\alpha \phi \,\mathrm{d}t \quad \forall \phi \in \mathcal{F}(\mathbb{R}_>;\mathbb{C}) \ ,$$

and the $\mathcal{F}'(\mathbb{R}_>;\mathbb{C})$ distribution derivative D_-^α means

$$\langle \mathrm{D}_-^\alpha u, \phi\rangle = \int_{\mathbb{R}_>} u\, \mathrm{D}_+^\alpha \phi \,\mathrm{d}t \quad \forall \phi \in \mathcal{F}_0(\mathbb{R}_>;\mathbb{C}) \ .$$

We can now prove a relevant integration by parts formula.

Lemma 2.7. *The $\mathcal{F}_0'(\mathbb{R}_>;\mathbb{C})$ distribution derivative $\mathrm{D}w$ of $w \in H^{\frac{1}{2}}_{00,\{0\}}(\mathbb{R}_>;\mathbb{C})$ satisfies*

$$\langle \mathrm{D}w, v\rangle = \int_{\mathbb{R}_>} \mathrm{D}_+^{\frac{1}{2}} w\, \mathrm{D}_-^{\frac{1}{2}} v \,\mathrm{d}t \quad \forall v \in \mathcal{F}(\mathbb{R}_>;\mathbb{C}) \ . \tag{2.13}$$

Proof. By definition we have

$$\langle \mathrm{D}w, v\rangle = \int_{\mathbb{R}_>} w\,(-\mathrm{D}v)\,\mathrm{d}t = (w, -\mathrm{D}v) \quad \forall v \in \mathcal{F}(\mathbb{R}_>;\mathbb{C}) \ .$$

Since $\mathcal{F}_0(\mathbb{R}_>;\mathbb{C}) \subset H^{\frac{1}{2}}_{00,\{0\}}(\mathbb{R}_>;\mathbb{C})$ is dense, it suffices to show

$$(w, -\mathrm{D}v) = (\mathrm{D}_+^{\frac{1}{2}} w, \mathrm{D}_-^{\frac{1}{2}} v) \quad \forall w \in \mathcal{F}_0(\mathbb{R}_>;\mathbb{C}) \ , \ v \in \mathcal{F}(\mathbb{R}_>;\mathbb{C}) \ .$$

If $w \in \mathcal{F}_0(\mathbb{R}_>;\mathbb{C})$, then $w(0) = 0$, so that $w = \mathrm{I}_+^1 \psi = \mathrm{I}_+^{\frac{1}{2}} \mathrm{I}_+^{\frac{1}{2}} \psi$, where $\psi = \mathrm{D}w$. Similarly, If $v \in \mathcal{F}(\mathbb{R}_>;\mathbb{C})$, then $v(\infty) = 0$, so that $v = \mathrm{I}_-^1 \phi = \mathrm{I}_-^{\frac{1}{2}} \mathrm{I}_-^{\frac{1}{2}} \phi$, where $\phi = -\mathrm{D}v$. Therefore, by integration by parts (2.9),

$$(w, -\mathrm{D}v) = (\mathrm{I}_+^{\frac{1}{2}} \mathrm{I}_+^{\frac{1}{2}} \psi, \phi) = (\mathrm{I}_+^{\frac{1}{2}} \psi, \mathrm{I}_-^{\frac{1}{2}} \phi) = (\mathrm{D}_+^{\frac{1}{2}} w, \mathrm{D}_-^{\frac{1}{2}} v) \ .$$

□

2.6. **Extension by zero.** In the previous Subsection 2.4 the space $H^{\frac{1}{2}}_{00,\{0\}}(\mathbb{R}_>;H)$ is characterized by means of an intrinsic norm. Here we give an alternative characterization in terms of extension by zero, denoted E_0.

Proposition 2.8. *The space $H^{\frac{1}{2}}_{00,\{0\}}(\mathbb{R}_>;H)$ is equals $\{w \in H^{\frac{1}{2}}(\mathbb{R}_>;H) : E_0 w \in H^{\frac{1}{2}}(\mathbb{R};H)\}$. A norm on $H^{\frac{1}{2}}_{00,\{0\}}(\mathbb{R}_>;H)$, which is equivalent to (2.8), is given by $\|E_0 \cdot\|_{H^{\frac{1}{2}}(\mathbb{R};H)}$; more precisely, for every $w \in H^{\frac{1}{2}}_{00,\{0\}}(\mathbb{R}_>;H)$ there holds*

$$\|w\|^2_{H^{\frac{1}{2}}_{00,\{0\}}(\mathbb{R}_>;H)} \le \|E_0 w\|^2_{H^{\frac{1}{2}}(\mathbb{R};H)} \le 2\|w\|^2_{H^{\frac{1}{2}}_{00,\{0\}}(\mathbb{R}_>;H)} . \tag{2.14}$$

Proof. For a proof of the first part, we refer to [12, Lemma 3.5]. It remains to show (2.14). Consider an arbitrary $w \in H^{\frac{1}{2}}_{00,\{0\}}(\mathbb{R}_>;H)$. Then $\tilde{w} = E_0 w \in H^{\frac{1}{2}}(\mathbb{R};H)$ and $\|w\|_{L^2(\mathbb{R}_>;H)} = \|\tilde{w}\|_{L^2(\mathbb{R};H)} = \|E_0 w\|_{L^2(\mathbb{R};H)}$. We next compute with the seminorm defined in (2.18) below:

$$\begin{aligned}
|E_0 w|^2_{H^{\frac{1}{2}}(\mathbb{R};H)} &= |\tilde{w}|^2_{H^{\frac{1}{2}}(\mathbb{R};H)} := \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\|\tilde{w}(t)-\tilde{w}(t')\|_H^2}{|t-t'|^2}\,\mathrm{d}t\,\mathrm{d}t' \\
&= \int_0^{\infty}\int_0^{\infty} \frac{\|\tilde{w}(t)-\tilde{w}(t')\|_H^2}{|t-t'|^2}\,\mathrm{d}t\,\mathrm{d}t' + 2\int_{t'=-\infty}^{0}\int_{t=0}^{\infty} \frac{\|\tilde{w}(t)-\tilde{w}(t')\|_H^2}{|t-t'|^2}\,\mathrm{d}t\,\mathrm{d}t' \\
&= \int_0^{\infty}\int_0^{\infty} \frac{\|w(t)-w(t')\|_H^2}{|t-t'|^2}\,\mathrm{d}t\,\mathrm{d}t' + 2\int_0^{\infty} \frac{\|w(t)\|_H^2}{t}\,\mathrm{d}t .
\end{aligned}$$

By comparison with the seminorm part of the norm (2.8), we conclude that

$$|w|^2_{H^{\frac{1}{2}}_{00,\{0\}}(\mathbb{R}_>;H)} \le |E_0 w|^2_{H^{\frac{1}{2}}(\mathbb{R};H)} \le 2|w|^2_{H^{\frac{1}{2}}_{00,\{0\}}(\mathbb{R}_>;H)}$$

and the proof is complete. □

2.7. **Further characterizations.** The preceding function spaces are intimately connected to the fractional derivatives $\mathrm{D}^{\frac{1}{2}}_+$ and $\mathrm{D}^{\frac{1}{2}}_-$ on $\mathcal{F}(\mathbb{R}_>;\mathbb{C})$. As these derivatives are essential in the proposed space-time formulation, we discuss their properties in detail. By continuity, the operators $\mathrm{D}^{\frac{1}{2}}_\pm$ extend to bounded operators from $H^{\frac{1}{2}}(\mathbb{R}_>;\mathbb{C})$ to $L^2(\mathbb{R}_>;\mathbb{C})$. The following proposition collects several properties of $\mathrm{D}^{\frac{1}{2}}_\pm$.

Proposition 2.9. *Let H denote an arbitrary Hilbert space over $\mathbb{R}$. Then there holds*

(1) *A function $u \in L^2(\mathbb{R}_>;H)$ belongs to $H^{\frac{1}{2}}_{00,\{0\}}(\mathbb{R}_>;H)$ if and only if its $\mathcal{F}'_0(\mathbb{R}_>;H)$-derivative $\mathrm{D}^{\frac{1}{2}}_+ u \in L^2(\mathbb{R}_>;H)$.*

(2) *A function $u \in L^2(\mathbb{R}_>;H)$ belongs to $H^{\frac{1}{2}}(\mathbb{R}_>;H)$ if and only if its $\mathcal{F}'(\mathbb{R}_>;H)$-derivative $\mathrm{D}^{\frac{1}{2}}_- u \in L^2(\mathbb{R}_>;H)$.*

(3) *A norm on $H^{\frac{1}{2}}_{00,\{0\}}(\mathbb{R}_>;H)$, equivalent to the norm (2.8), is given by*

$$\|u\|^2_{H^{\frac{1}{2}}_+(\mathbb{R}_>;H)} := \|u\|^2_{L^2(\mathbb{R}_>;H)} + \|\mathrm{D}^{\frac{1}{2}}_+ u\|^2_{L^2(\mathbb{R}_>;H)} . \tag{2.15}$$

A norm on $H^{\frac12}(\mathbb{R}_>;H)$*, equivalent to the norm* (2.7)*, is given by*

$$\|u\|^2_{H^{\frac12}_-(\mathbb{R}_>;H)} := \|u\|^2_{L^2(\mathbb{R}_>;H)} + \|\mathrm{D}_-^{\frac12}u\|^2_{L^2(\mathbb{R}_>;H)} \,. \tag{2.16}$$

Moreover, for every $u \in H^{\frac12}(\mathbb{R};H)$ *there holds*

$$\|u\|^2_{H^{\frac12}(\mathbb{R};H)} = \|u\|^2_{L^2(\mathbb{R};H)} + \|\mathrm{D}_-^{\frac12}u\|^2_{L^2(\mathbb{R};H)} = \|u\|^2_{L^2(\mathbb{R};H)} + \|\mathrm{D}_+^{\frac12}u\|^2_{L^2(\mathbb{R};H)} \,, \tag{2.17}$$

$$\|\mathrm{D}_+^{\frac12}u\|^2_{L^2(\mathbb{R},H)} \simeq |u|^2_{H^{\frac12}(\mathbb{R};H)} := \int_{s\in\mathbb{R}}\int_{t\in\mathbb{R}} \frac{\|u(s)-u(t)\|_H^2}{|s-t|^2}\,\mathrm{d}s\,\mathrm{d}t \,. \tag{2.18}$$

For the proof of (2.15), (2.16) we refer to [12, Lemmas 3.5, 3.8] and to [12, Lemmas 3.6, 3.9], respectively. The identity (2.17) is immediate from the Fourier characterizations of $\mathrm{D}_\pm^{\frac12}$ in [11, Sect. 3]. For (2.18), we refer to [11, (4.13)]. We remark that the expression $|\cdot|_{H^{\frac12}(\mathbb{R};H)}$ introduced in (2.18) is indeed a seminorm, as it vanishes on all functions $u \in H$ independent of t.

In view of Lemma 2.7 and Proposition 2.9, it is now clear that the bilinear form $\langle \mathrm{D}w, v\rangle$ is bounded on $H^{\frac12}_{00,\{0\}}(\mathbb{R}_>;H) \times H^{\frac12}(\mathbb{R}_>;H)$.

2.8. **Coercivity over** $\mathbb{R}$**.** A key ingredient in the theory of Fontes is that the time derivative is coercive in the sense of Corollary 2.2 for functions defined on $\mathbb{R}$. We demonstrate this here by considering the operator, with A as in Subsection 2.2,

$$Bv = \mathrm{D}v + Av \,, \quad v \in \mathcal{F}(\mathbb{R};V) \,.$$

By fractional integration by parts (immediate from the Fourier characterizations of $\mathrm{D}_\pm^{\frac12}$), we find

$$\langle Bw, v\rangle = \int_{\mathbb{R}} \Big((\mathrm{D}_+^{\frac12}w, \mathrm{D}_-^{\frac12}v)_H + a(w,v)\Big)\,\mathrm{d}t \,, \quad w,v \in \mathcal{F}(\mathbb{R};V) \,. \tag{2.19}$$

We also define the operator

$$\mathrm{H}^\alpha := \cos(\pi\alpha)\mathrm{I} + \sin(\pi\alpha)\mathrm{H} \,, \quad \alpha \in \mathbb{R} \,,$$

where H is the Hilbert transform acting with respect to the t-variable. By using (2.6), we then obtain the fundamental *coercivity inequality*: for any $w \in \mathcal{F}(\mathbb{R};V)$

$$\begin{aligned}
\langle Bw, \mathrm{H}^{-\alpha}w\rangle &= \langle \mathrm{D}w + Aw, \cos(\pi\alpha)w - \sin(\pi\alpha)\mathrm{H}w\rangle \\
&= \cos(\pi\alpha)\langle \mathrm{D}w, w\rangle - \sin(\pi\alpha)\langle \mathrm{D}_+^{\frac12}w, \mathrm{D}_-^{\frac12}\mathrm{H}w\rangle \\
&\quad + \int_{\mathbb{R}} \big(\cos(\pi\alpha)a(w,w) - \sin(\pi\alpha)a(w,\mathrm{H}w)\big)\,\mathrm{d}t \\
&\geq \sin(\pi\alpha)\|\mathrm{D}_+^{\frac12}w\|^2_{L^2(\mathbb{R};H)} + \big(\lambda_-\cos(\pi\alpha) - \lambda_+\sin(\pi\alpha)\big)\|w\|^2_{L^2(\mathbb{R};V)} \,,
\end{aligned}$$

because $\langle \mathrm{D}w, w\rangle = 0$, $\|\mathrm{H}w\|_{L^2(\mathbb{R};V)} \leq \|w\|_{L^2(\mathbb{R};V)}$, and $\mathrm{D}_-^{\frac12}\mathrm{H} = -\mathrm{D}_-^{\frac12}\mathrm{H}^{-\frac12} = -\mathrm{D}_+^{\frac12}$, see [12]. Fixing the parameter $\alpha > 0$ sufficiently small, by density of $\mathcal{F}(\mathbb{R};V)$ in $H^{\frac12}(\mathbb{R};H)\cap L^2(\mathbb{R};V)$, and (2.17), we find the coercivity inequality (cp. Corollary 2.2): there exists $c > 0$ such that

$$\langle Bw, \mathrm{H}^{-\alpha}w\rangle \geq c\big(\|w\|^2_{H^{\frac12}(\mathbb{R};H)} + \|w\|^2_{L^2(\mathbb{R};V)}\big) \quad \forall w \in H^{\frac12}(\mathbb{R};H)\cap L^2(\mathbb{R};V) \,.$$

Hence, by Corollary 2.2 and Proposition 2.1 we conclude that the bilinear form in (2.19) satisfies the inf-sup conditions (2.3), (2.4) with $X = Y = H^{\frac12}(\mathbb{R};H)\cap L^2(\mathbb{R};V)$.

2.9. **Coercivity over $\mathbb{R}_>$.** In order to prove the inf-sup condition (2.3) for functions on $\mathbb{R}_>$, we take an arbitrary $w \in H^{\frac12}_{00,\{0\}}(\mathbb{R}_>;H)\cap L^2(\mathbb{R}_>;V)$. Then its extension by zero, $\tilde w = E_0 w$, belongs to $H^{\frac12}(\mathbb{R};H)\cap L^2(\mathbb{R};V)$ according to Proposition 2.8. Similarly, if $\tilde v \in H^{\frac12}(\mathbb{R};H)\cap L^2(\mathbb{R};V)$, then its restriction to $\mathbb{R}_>$, $v = R_>\tilde v$, belongs to $H^{\frac12}(\mathbb{R}_>;H)\cap L^2(\mathbb{R}_>;V)$ according to Proposition 2.6 (1). We have the bounds

$$\|w\|_{H^{\frac12}_{00,\{0\}}(\mathbb{R}_>;H)\cap L^2(\mathbb{R}_>;V)} \le \|E_0 w\|_{H^{\frac12}(\mathbb{R};H)\cap L^2(\mathbb{R};V)}\,, \tag{2.20}$$

$$\|R_>\tilde v\|_{H^{\frac12}(\mathbb{R}_>;H)\cap L^2(\mathbb{R}_>;V)} \le \|\tilde v\|_{H^{\frac12}(\mathbb{R};H)\cap L^2(\mathbb{R};V)}\,. \tag{2.21}$$

Moreover,

$$(\mathrm{D}_+^{\frac12}\tilde w)(t) = \frac{1}{\Gamma(\frac12)}\mathrm{D}\int_{-\infty}^t (t-s)^{-\frac12}\tilde w(s)\,\mathrm{d}s = \begin{cases} \frac{1}{\Gamma(\frac12)}\mathrm{D}\int_0^t (t-s)^{-\frac12} w(s)\,\mathrm{d}s\,, & t>0\,,\\ 0\,, & t<0\,,\end{cases}$$

that is, $\mathrm{D}_+^{\frac12}E_0 w = E_0\mathrm{D}_+^{\frac12}w$. Similarly,

$$(\mathrm{D}_-^{\frac12}\tilde v)(t) = -\frac{1}{\Gamma(\frac12)}\mathrm{D}\int_t^\infty (s-t)^{-\frac12}\tilde v(s)\,\mathrm{d}s = -\frac{1}{\Gamma(\frac12)}\mathrm{D}\int_t^\infty (s-t)^{-\frac12} v(s)\,\mathrm{d}s\,,\quad t>0\,,$$

that is, $R_>\mathrm{D}_-^{\frac12}\tilde v = \mathrm{D}_-^{\frac12}R_>\tilde v$. Hence,

$$\begin{aligned}\int_{\mathbb{R}} (\mathrm{D}_+^{\frac12}E_0 w, \mathrm{D}_-^{\frac12}\tilde v)_H\,\mathrm{d}t &= \int_{\mathbb{R}} (E_0\mathrm{D}_+^{\frac12}w, \mathrm{D}_-^{\frac12}\tilde v)_H\,\mathrm{d}t = \int_{\mathbb{R}_>} (\mathrm{D}_+^{\frac12}w, R_>\mathrm{D}_-^{\frac12}\tilde v)_H\,\mathrm{d}t\\ &= \int_{\mathbb{R}_>} (\mathrm{D}_+^{\frac12}w, \mathrm{D}_-^{\frac12}R_>\tilde v)_H\,\mathrm{d}t\,.\end{aligned}$$

If we denote by $\mathcal{B}_{\mathbb{R}}(\cdot,\cdot)$ and $\mathcal{B}_{\mathbb{R}_>}(\cdot,\cdot)$ bilinear forms as in (2.19) computed over $\mathbb{R}$ and $\mathbb{R}_>$, respectively, then we conclude that

$$\mathcal{B}_{\mathbb{R}_>}(w, R_>\tilde v) = \mathcal{B}_{\mathbb{R}}(E_0 w, \tilde v)\,. \tag{2.22}$$

The inf-sup condition proved in the previous subsection means that for each $\tilde w \in H^{\frac12}(\mathbb{R};H)\cap L^2(\mathbb{R};V)$ there is a $\tilde v \in H^{\frac12}(\mathbb{R};H)\cap L^2(\mathbb{R};V)$ (namely, $\tilde v = \mathrm{H}^{-\alpha}\tilde w$) such that

$$\frac{\mathcal{B}_{\mathbb{R}}(\tilde w,\tilde v)}{\|\tilde v\|_{H^{\frac12}(\mathbb{R};H)\cap L^2(\mathbb{R};V)}} \ge c\|\tilde w\|_{H^{\frac12}(\mathbb{R};H)\cap L^2(\mathbb{R};V)}\,. \tag{2.23}$$

For arbitrary $w \in H^{\frac12}_{00,\{0\}}(\mathbb{R}_>;H)\cap L^2(\mathbb{R}_>;V)$, we let $\tilde w = E_0 w$ and take $\tilde v$ as above and set $v = R_>\tilde v$, that is, $v = R_>\mathrm{H}^{-\alpha}E_0 w$. Then, by (2.20), (2.21), (2.22), and (2.23), we obtain

$$\begin{aligned}\frac{\mathcal{B}_{\mathbb{R}_>}(w,v)}{\|v\|_{H^{\frac12}(\mathbb{R}_>;H)\cap L^2(\mathbb{R}_>;V)}} &\ge \frac{\mathcal{B}_{\mathbb{R}}(\tilde w,\tilde v)}{\|\tilde v\|_{H^{\frac12}(\mathbb{R};H)\cap L^2(\mathbb{R};V)}}\\ &\ge c\|\tilde w\|_{H^{\frac12}(\mathbb{R};H)\cap L^2(\mathbb{R};V)} \ge c\|w\|_{H^{\frac12}_{00,\{0\}}(\mathbb{R}_>;H)\cap L^2(\mathbb{R}_>;V)}\,.\end{aligned}$$

This is the desired inf-sup condition.

3. Linear parabolic evolution equations

We present a space-time variational formulation of the initial boundary value problem for the abstract, linear parabolic evolution equation (1.1) with homogeneous initial condition (1.2). For the operator $A \in \mathcal{L}(V, V^*)$, we assume (2.6). In what follows, all Hilbert spaces are taken over the coefficient field $\mathbb{R}$. Using the function spaces developed in Section 2, we now state the weak form of the linear parabolic initial-value problem (1.1), (1.2): it is based on the Bochner spaces

$$\begin{aligned} X &= H^{\frac{1}{2}}_{00,\{0\}}(\mathbb{R}_>; H) \cap L^2(\mathbb{R}_>; V) \simeq \left(H^{\frac{1}{2}}_{00,\{0\}}(\mathbb{R}_>) \otimes H\right) \cap \left(L^2(\mathbb{R}_>) \otimes V\right) , \\ Y &= H^{\frac{1}{2}}(\mathbb{R}_>; H) \cap L^2(\mathbb{R}_>; V) \simeq \left(H^{\frac{1}{2}}(\mathbb{R}_>) \otimes H\right) \cap \left(L^2(\mathbb{R}_>) \otimes V\right) . \end{aligned} \tag{3.1}$$

Here, $\otimes$ signifies the Hilbert tensor product space endowed with the (unique) cross norm. The parabolic operator takes the form $B = \mathrm{D} + A$ with the $\mathcal{F}_0'$-distributional derivative D introduced in Section 2.5, Lemma 2.7.

Besides the spaces X and Y in (3.1), we will also need the space

$$Z = H^{\frac{1}{2}}(\mathbb{R}; H) \cap L^2(\mathbb{R}; V) . \tag{3.2}$$

We shall make use of the following continuity properties of extensions and restrictions which follow from Proposition 2.6 and Proposition 2.8.

Proposition 3.1. *For X, Y, and Z as in* (3.1), (3.2) *there holds:*
(1) *$X \subset Z$ with continuous embedding given by the zero extension E_0.*
(2) *$Y = R_>(Z)$ with $R_>$ denoting the operator of restriction of elements of $L^2(\mathbb{R}; H)$ to $\mathbb{R}_>$.*
(3) *$Z^* \simeq (H^{\frac{1}{2}}(\mathbb{R}; H))^* + L^2(\mathbb{R}; V)^* \simeq H^{-\frac{1}{2}}(\mathbb{R}; H) + L^2(\mathbb{R}; V^*)$.*
(4) *Y^* is isomorphic to $\{g \in Z^* : \mathrm{supp}(g) \subseteq \mathbb{R}_>\}$.*
(5) *X is a dense subset of Y, that is, $\overline{X}^{\|\cdot\|_Y} = Y$.*

From $A \in \mathcal{L}(V, V^*)$ it follows that $B := \mathrm{D} + A \in \mathcal{L}(X, Y^*)$. More precisely, there holds for every $v \in X$,

$$\begin{aligned} Bv = (\mathrm{D} + A)v = \mathrm{D}v + Av &\in (H^{\frac{1}{2}}(\mathbb{R}_>; H))^* + L^2(\mathbb{R}_>; V^*) \\ &\simeq (H^{\frac{1}{2}}(\mathbb{R}_>; H))^* + L^2(\mathbb{R}_>; V)^* \simeq (H^{\frac{1}{2}}(\mathbb{R}_>; H) \cap L^2(\mathbb{R}_>; V))^* = Y^* . \end{aligned}$$

For any source term $f \in Y^*$, we consider the *space-time weak formulation* of (1.1), (1.2): find

$$u \in X : \quad \mathcal{B}_{\mathrm{D}+A}(u, v) = F(v) \quad \forall v \in Y . \tag{3.3}$$

Here, the linear functional $F(\cdot)$ is defined by

$$F(v) = \langle f, v \rangle \quad \forall v \in Y$$

with $\langle \cdot, \cdot \rangle$ denoting the $Y^* \times Y$ duality pairing. The bilinear form is given by, cp. Lemma 2.7,

$$\mathcal{B}_{\mathrm{D}+A}(w, v) := \int_{\mathbb{R}_>} \left\{ (\mathrm{D}_+^{\frac{1}{2}} w, \mathrm{D}_-^{\frac{1}{2}} v)_H + a(w, v) \right\} \mathrm{d}t , \quad w \in X, \ v \in Y , \tag{3.4}$$

where X and Y are as in (3.1). The form $\mathcal{B}_{\mathrm{D}+A}(\cdot, \cdot)$ in (3.4) is continuous by Proposition 2.9 (1) and (2), stating that for every $w \in H^{\frac{1}{2}}_{00,\{0\}}(\mathbb{R}_>; H)$ we have $\mathrm{D}_+^{\frac{1}{2}} w \in L^2(\mathbb{R}_>; H)$ and that for every $v \in H^{\frac{1}{2}}(\mathbb{R}_>; H)$ we have $\mathrm{D}_-^{\frac{1}{2}} v \in L^2(\mathbb{R}_>; H)$.

The unique solvability of (3.3) was proved in [12, Sect. 4.1] by extension to a problem over $\mathbb{R}$, where coercivity in the sense of Corollary 2.2 can be proved, see Subsection 2.8. As a

result of the unique solvability of (3.3) we conclude that the inf-sup conditions (2.3), (2.4) hold. We formulate this in the following proposition.

Proposition 3.2. *Suppose that assumption* (2.6) *holds. Then, for the choice* (3.1) *of spaces, the bilinear form* (3.4) *satisfies the continuity condition* (2.1) *and the inf-sup conditions* (2.3), (2.4). *In particular, for every $f \in Y^*$ there exists a unique solution $u \in X$ of* (3.3).

Proof. We observe that $Y \simeq B^{1,\frac{1}{2}}_{0,\cdot}(Q_+)$ and that $X \simeq B^{1,\frac{1}{2}}_{0,0}(Q_+)$ in the notation of [12, Thm. 4.3, Sect. 4.1] with $p = 2$. It is shown there that the operator $B = \mathrm{D} + A \in \mathcal{L}(X, Y^*)$ is bijective. Therefore, Proposition 2.1 implies the inf-sup conditions (2.3), (2.4) for the bilinear form (3.4) on the spaces $X \times Y$ in (3.1). □

4. Sparse tensor Galerkin discretization

Having established the well-posedness and the unique solvability of (1.1), (1.2) we now turn to Galerkin approximations. Rather than considering time-stepping (as studied, e.g., in [23]), we are interested in *compressive space-time Galerkin* discretizations, as analyzed for the first time in [20]. We present and analyze adaptive, compressive, space-time schemes which are based on the weak space-time formulation (3.3). The adaptive, and space-time compressive schemes inherit, being instances of the general theory in [5, 6], stability from the well-posedness of the infinite-dimensional problem shown in Proposition 3.2 and from the stability of the Riesz bases. As in [20], they are based on tensor product constructions of Riesz bases of X and Y; however, the variational formulation (3.3) obviates the need for stability of Riesz bases in negative order Sobolev spaces. We present classes of spline wavelets in the time domain and also in the spatial domain $D \subset \mathbb{R}^n$, which we assume to be a polygon or polyhedron. Rather than focusing on a particular family of wavelets, we specify several axioms from [20] to be satisfied by the tensorized multiresolution bases in the spatial and temporal domains in order for our analysis to apply. We assume that V and H are modeled on Sobolev spaces on the bounded Lipschitz polyhedron $D \subset \mathbb{R}^n$, $n \geq 1$. As in [20], our analysis accommodates two cases: case (A): $n = 2, 3$ and D is a bounded polyhedron with plane faces; and the high-dimensional case (B): $n \geq 1$ and $D = (0,1)^n$. In D we consider general elliptic operators A of order $2m$, $m \geq 1$. The generic example is $A = -\Delta$, $V = H^1_0(D)$, and $H = L^2(D)$, in which case $m = 1$. The domain for the parabolic initial-boundary value problem is the space-time cylinder $Q_> := \mathbb{R}_> \times D$.

4.1. **Space-time wavelet Galerkin discretization.** The Galerkin discretization of the space-time variational formulation (3.3) will be based on two dense, nested families $\{X^\ell\}_{\ell\in\mathbb{N}_0}$, $\{Y^\ell\}_{\ell\in\mathbb{N}_0}$ of subspaces of X and Y as in (3.1). The inf-sup condition (2.3) makes it necessary to allow $X^\ell \neq Y^\ell$ (leading in effect to Petrov-Galerkin discretizations), so that Proposition 2.1 is used in full generality. As indicated above, we choose $\{X^\ell\}_{\ell\in\mathbb{N}_0}$ as tensor-products of spaces of continuous, piecewise polynomial functions of $t \in \mathbb{R}_>$ and $x \in D$, in order to obtain good (space-time compressive) approximation of solutions, whereas Y^ℓ will be selected to ensure good stability. Multiresolution bases will be required to ensure: (a) multilevel preconditioning, i.e., all stiffness matrices have (generalized) condition numbers, which are bounded independently of ℓ; and (b) matrix and (space-time) solution compression.

Thus, we consider the Galerkin discretization: to find, for $\ell \in \mathbb{N}_0$,

$$u^\ell \in X^\ell : \quad \mathcal{B}_{\mathrm{D}+A}(u^\ell, v^\ell) = F(v^\ell) \quad \forall v^\ell \in Y^\ell \,. \tag{4.1}$$

We assume that

$$N_\ell = \dim(X^\ell) = \dim(Y^\ell) < \infty \,,$$

such that $X^\ell \subset X = H^{\frac{1}{2}}_{00,\{0\}}(\mathbb{R}_>; H) \cap L^2(\mathbb{R}_>; V)$ and $Y^\ell \subset Y = H^{\frac{1}{2}}(\mathbb{R}_>; H) \cap L^2(\mathbb{R}_>; V)$ are closed and $\cup_{\ell\in\mathbb{N}} X^\ell$ and $\cup_{\ell\in\mathbb{N}} Y^\ell$ are dense in X, respectively in Y. Proposition 2.1 implies

Proposition 4.1. *Assume that the Galerkin discretization* (4.1) *of* (3.3) *is stable, in the sense that there exists* $\bar\gamma$ *such that, for all* $\ell \in \mathbb{N}_0$,

$$\inf_{0\neq w\in X^\ell} \sup_{0\neq v\in Y^\ell} \frac{\mathcal{B}_{\mathrm{D}+A}(w,v)}{\|w\|_X \|v\|_Y} \geq \bar\gamma > 0 \,. \tag{4.2}$$

Then, for every $F \in Y^*$ *and for every* $\ell \in \mathbb{N}$, *the Galerkin approximation* (4.1) *admits a unique solution* $u^\ell \in X^\ell$. *In particular, the (in general, non-symmetric) stiffness matrix corresponding to* (4.1) *is nonsingular. Let* $u \in X$ *be the corresponding unique solution to* (3.3) *and* C *be the constant in* (2.1). *Then there holds the quasi-optimality estimate*

$$\|u - u^\ell\|_X \leq \frac{C}{\bar\gamma} \inf_{v^\ell\in X^\ell} \|u - v^\ell\|_X \,. \tag{4.3}$$

The proof of Proposition 4.1 is straightforward: existence and uniqueness of u^ℓ in (4.1) and the invertibility of the $N_\ell \times N_\ell$ matrix follows from (4.2) with Proposition 2.1. The error estimate (4.3) follows from the Galerkin orthogonality

$$\mathcal{B}_{\mathrm{D}+A}(u - u^\ell, v^\ell) = 0 \qquad \forall v^\ell \in Y^\ell \,,$$

by noting that the error is $u - u^\ell = (I - R^\ell)(u - v^\ell)$, where R^ℓ is the Ritz projector that maps $u \mapsto u^\ell$. Therefore, (4.3) holds with constant $\|I - R^\ell\|_{\mathcal{L}(X,X)} = \|R^\ell\|_{\mathcal{L}(X,X)} \leq C/\bar\gamma$, [27].

For preconditioning and efficient computation, as well for *adaptive space-time Galerkin discretizations* with optimality properties, the concept of *Riesz basis* takes a central role.

4.2. **Riesz bases and bi-infinite matrix vector equations.** We assume at hand a Riesz basis $\Psi^X = \{\psi^X_\lambda : \lambda \in \nabla^X\}$ for X. The Riesz basis property amounts to saying that the *synthesis operator*

$$s_{\Psi^X} : \ell_2(\nabla^X) \to X : \mathbf{c} \mapsto \mathbf{c}^\top \Psi^X := \sum_{\lambda\in\nabla^X} c_\lambda \psi^X_\lambda$$

is boundedly invertible. Its adjoint, known as the *analysis operator*, reads

$$s'_{\Psi^X} : X^* \to \ell_2(\nabla^X) : g \mapsto [g(\psi^X_\lambda)]_{\lambda\in\nabla^X} \,.$$

Similarly, let $\Psi^Y = \{\psi^Y_\lambda : \lambda \in \nabla^Y\}$ denote a Riesz basis for Y, with synthesis operator s_{Ψ^Y} and adjoint s'_{Ψ^Y}. Ahead, we construct Riesz bases Ψ^X and Ψ^Y by tensorization of wavelet bases in $\mathbb{R}_>$ and in D.

By Proposition 3.2, $B = \mathrm{D} + A \in \mathcal{L}(X, Y^*)$ is boundedly invertible with the choice of spaces in (3.1). We may write (3.3) equivalently as operator equation: given $f \in Y^*$, find

$$u \in X : \quad Bu = f \ \text{ in } \ Y^* \,. \tag{4.4}$$

Writing $u = s_{\Psi^X}\mathbf{u}$, (3.3) and (4.4) are equivalent to the bi-infinite matrix vector problem

$$\mathbf{B}\mathbf{u} = \mathbf{f} \,, \tag{4.5}$$

where $\mathbf{f} = s'_{\Psi^Y} f = [f(\psi^Y_\lambda)]_{\lambda\in\nabla^Y} \in \ell_2(\nabla^Y)$, and where the *"stiffness"* or *system matrix*

$$\mathbf{B} = s'_{\Psi^Y} B s_{\Psi^X} = [(B\psi^X_\mu)(\psi^Y_\lambda)]_{\lambda\in\nabla^Y,\mu\in\nabla^X} \in \mathcal{L}(\ell_2(\nabla^X), \ell_2(\nabla^Y))$$

is boundedly invertible. We may write

$$\mathcal{B}_{\mathrm{D}+A} : X \times Y \to \mathbb{R} : (w, v) \mapsto (Bw)(v) \ ,$$

and we also use the notations

$$\mathbf{B} = \mathcal{B}_{\mathrm{D}+A}(\Psi^X, \Psi^Y) \quad \text{and} \quad \mathbf{f} = f(\Psi^Y) \ .$$

With the Riesz constants

$$\Lambda^X_{\Psi^X} := \|s_{\Psi^X}\|_{\mathcal{L}(\ell_2(\nabla^X),X)} = \sup_{0\neq \mathbf{c}\in\ell_2(\nabla^X)} \frac{\|\mathbf{c}^\top \Psi^X\|_X}{\|\mathbf{c}\|_{\ell_2(\nabla^X)}} \ ,$$

$$\lambda^X_{\Psi^X} := \|s^{-1}_{\Psi^X}\|^{-1}_{\mathcal{L}(X,\ell_2(\nabla^X))} = \inf_{0\neq \mathbf{c}\in\ell_2(\nabla^X)} \frac{\|\mathbf{c}^\top \Psi^X\|_X}{\|\mathbf{c}\|_{\ell_2(\nabla^X)}} \ ,$$

and analogous constants $\Lambda^Y_{\Psi^Y}$ and $\lambda^Y_{\Psi^Y}$, the bounded invertibility of $B \in \mathcal{L}(X, Y^*)$ implies that the condition number of $\mathbf{B}$ is finite, i.e.,

$$\|\mathbf{B}\|_{\mathcal{L}(\ell_2(\nabla^X),\ell_2(\nabla^Y))} \le \|B\|_{\mathcal{L}(X,Y^*)} \Lambda^X_{\Psi^X} \Lambda^Y_{\Psi^Y} \ ,$$

$$\|\mathbf{B}^{-1}\|_{\mathcal{L}(\ell_2(\nabla^Y),\ell_2(\nabla^X))} \le \frac{\|B^{-1}\|_{\mathcal{L}(Y^*,X)}}{\lambda^X_{\Psi^X} \lambda^Y_{\Psi^Y}} \ .$$

We next construct *Riesz bases* of the spaces X and Y in (3.1).

4.3. **Riesz bases in $H^{\frac12}_{00,\{0\}}(\mathbb{R}_>)$ and $H^{\frac12}(\mathbb{R}_>)$.** We assume at our disposal two countable collections $\Theta^X, \Theta^Y \subset H^1(\mathbb{R}_>)$ of functions such that

$$\Theta^X = \{\theta^X_\lambda : \lambda \in \nabla^X_t\} \subset H^1_{0,\{0\}}(\mathbb{R}_>)$$

is a *normalized Riesz basis for $L^2(\mathbb{R}_>)$ which, when renormalized in $H^1(\mathbb{R}_>)$, is a Riesz basis for $H^1_{0,\{0\}}(\mathbb{R}_>)$*. Analogously, we assume available $\Theta^Y = \{\theta^Y_\lambda : \lambda \in \nabla^Y_t\} \subset H^1(\mathbb{R}_>)$, a Riesz basis of $L^2(\mathbb{R}_>)$ which, when renormalized in $H^1(\mathbb{R}_>)$, is a Riesz basis for $H^1(\mathbb{R}_>)$.

From Proposition 2.6 we obtain the following result.

Proposition 4.2. *Assume given two collections Θ^X and Θ^Y with the above properties. Then, for $0 \le s \le 1$, the collections $[\Theta^X]_s$ and $[\Theta^Y]_s$, which are obtained by rescaling Θ^X and Θ^Y by $\{2^{s|\lambda|} : \lambda \in \nabla_t\}$, (e.g., $[\Theta^X]_s = \{2^{s|\lambda|}\theta^X_\lambda : \lambda \in \nabla^X_t\}$) are Riesz bases of $[L^2(\mathbb{R}_>), H^1_{0,\{0\}}(\mathbb{R}_>)]_s$ and of $[L^2(\mathbb{R}_>), H^1(\mathbb{R}_>)]_s$, respectively. In particular, for $s = \frac12$, $[\Theta^X]_{\frac12}$ is a Riesz basis for $H^{\frac12}_{00,\{0\}}(\mathbb{R}_>)$ and $[\Theta^Y]_{\frac12}$ is a Riesz basis for $H^{\frac12}(\mathbb{R}_>)$.*

We denote by θ^X_λ elements of the collection Θ^X and, likewise, by θ^Y_λ elements of Θ^Y. Further assumptions on the bases Θ^X, Θ^Y are as in [20]: denoting by θ_λ a generic element in either of the collections Θ^X and Θ^Y, we require the θ_λ to be

- (t1) *local*: that is, $\sup_{t\in\mathbb{R}_>,\ell\in\mathbb{N}_0} \#\{\lambda : |\lambda| = \ell,\ t \in \operatorname{supp}\theta_\lambda\} < \infty$ and $|\operatorname{supp}\theta_\lambda| \lesssim 2^{-|\lambda|}$,
- (t2) *piecewise polynomial of order d_t*: here, "piecewise" means that the singular support consists of a finite number of points whose number is uniformly bounded with respect to $|\lambda|$,
- (t3) *globally continuous*: specifically, $\|\theta_\lambda\|_{W^k_\infty(\mathbb{R}_>)} \lesssim 2^{|\lambda|(\frac12+k)}$ for $k \in \{0, 1\}$,
- (t4) *vanishing moments*: for $|\lambda| > 0$, the θ_λ have $\tilde{d}_t \ge d_t$ vanishing moments.

Properties (t1)–(t4) are assumed to hold for both Θ^X and Θ^Y. We remark that property (t3), global continuity, is necessary to ensure $H^{\frac{1}{2}}(\mathbb{R}_>)$-conformity, even though $H^{\frac{1}{2}}(\mathbb{R}_>)$ is not embedded into $C^0(\overline{\mathbb{R}_>})$.

Properties (t1)–(t4) can be satisfied by collections Θ^X, Θ^Y that are continuous, piecewise polynomial wavelet bases on dyadic refinements of $\mathbb{R}_>$, which are of *order* $d_t > 1$. For $k \in \mathbb{N}_0$ we denote by $\nabla_t^{(k)}$ the set of $\lambda \in \nabla_t$ with *refinement level* $|\lambda| \le k$. It holds that $\#\nabla_t^{(k)} \eqsim 2^k$. Setting also $\nabla_t^{(-1)} := \emptyset$, we define the biorthogonal projector $Q_{k,t}^X := Q^X_{\nabla_t^{(k)}}$ by

$$Q_{k,t}^X v = \sum_{\lambda \in \nabla_t^{(k)X}} \langle v, \theta'^X_\lambda \rangle \theta^X_\lambda \ ,$$

where Θ'^X denotes the dual basis, and analogously for $Q_{k,t}^Y := Q^Y_{\nabla_t^{(k)}}$. We have

$$\|\mathrm{Id} - Q_{k,t}^X\|_{\mathcal{L}(H^{d_t}(\mathbb{R}_>),L^2(\mathbb{R}_>))} \lesssim 2^{-kd_t} \ , \quad \|\mathrm{Id} - Q_{k,t}^X\|_{\mathcal{L}(H^{d_t}(\mathbb{R}_>),H^{\frac{1}{2}}_{00,\{0\}}(\mathbb{R}_>))} \lesssim 2^{-k(d_t-\frac{1}{2})}$$

and analogously for $Q_{k,t}^Y$ with $H^{\frac{1}{2}}(\mathbb{R}_>)$ in place of $H^{\frac{1}{2}}_{00,\{0\}}(\mathbb{R}_>)$.

Constructions of compactly supported spline wavelet systems Θ on $(-1,1)$ and on $\mathbb{R}$, as well as *direct constructions* (i.e., not based on antisymmetry) of Riesz bases Θ^X and Θ^Y on $\mathbb{R}_>$ satisfying properties (t1)–(t4) with $\theta_\lambda^X(t)|_{t=0} = 0$ are available, for example, in [4, 8, 9, 26] and the references there.

4.4. **Riesz bases in H and V.** With $H = L^2(D)$ and the assumption that V coincides with a closed subspace (supporting homogeneous essential boundary conditions) of the Sobolev space $H^m(D)$ for some $m > 0$, we assume at our disposal a Riesz basis

$$\Sigma = \{\sigma_\lambda : \lambda \in \nabla_x\} \subset V \ .$$

Specifically, Σ is a collection of functions that is a *normalized Riesz basis for H which, upon renormalization in V, is a Riesz basis denoted $[\Sigma]_V$ for V*. Riesz bases of divergence-free functions in the context of Example 2.4 are constructed in [22, 26] and the references there. For the spatial wavelet basis Σ, we consider as in [20], two cases:

(A) it is a wavelet basis of order $d_x > m$ with isotropic supports constructed from a dyadic multiresolution analysis in $L^2(D)$,
(B) $D = (0,1)^n$ and Σ is the tensor product of (possibly different) univariate wavelet bases Σ_i as in (A) in each of the coordinate spaces.

In case (A), for some sufficiently large K depending on m, where $2m$ is the order of A, and for some $r_x \in \mathbb{N}_0$ such that $m - 1 \le r_x \le d_x - 2$ and $\tilde{d}_x \in \mathbb{N}_0$, we will assume that the σ_λ are

(s1) *local* and *piecewise smooth*: for any $\ell \in \mathbb{N}_0$ there exist collections $\{D_{\ell,v} : v \in \mathcal{O}_\ell\}$ of disjoint, uniformly shape regular, open subdomains such that $\overline{D} = \cup_{v\in\mathcal{O}_\ell} \overline{D_{\ell,v}}$, $\overline{D_{\ell,v}}$ is the union of some $\overline{D_{\ell+1,\tilde{v}}}$, $\mathrm{diam}(D_{\ell,v}) \eqsim 2^{-\ell}$, $\mathrm{supp}\,\sigma_\lambda$ is connected and is the union of a uniformly bounded number of $\overline{D_{|\lambda|,v}}$, each $\overline{D_{\ell,v}}$ has non-empty intersection with the supports of a uniformly bounded number of σ_λ with $|\lambda| = \ell$, and, for $k \in \{0, K\}$,

$$\|\sigma_\lambda\|_{W^k_\infty(D_{|\lambda|,v})} \lesssim 2^{|\lambda|(\frac{n}{2}+k)} \ ,$$

(s2) *globally C^{r_x}*: specifically, $\|\sigma_\lambda\|_{W^k_\infty(D)} \lesssim 2^{|\lambda|(\frac{n}{2}+k)}$ for $k \in \{0, r_x + 1\}$,

(s3) for $|\lambda| > 0$, have *cancellation properties of order* $\tilde{d}_x$:

$$\Big| \int_D w\sigma_\lambda \Big| \lesssim 2^{-|\lambda|(\frac{n}{2}+k)} \|w\|_{W^k_\infty(D)} \quad \text{for } k \in \{0, \tilde{d}_x\},\ w \in W^k_\infty(D) \cap V\ .$$

(s4) In addition to (s1), we assume that for any ℓ and $v \in \mathcal{O}_\ell$, there exists a sufficiently smooth transformation of coordinates κ, with derivatives bounded uniformly in ℓ and v, such that for all $|\lambda| = \ell$, $(\sigma_\lambda \circ \kappa)|_{\kappa^{-1}(D_{\ell,v})}$ is a polynomial of some fixed degree.

For case (B), we assume that each of the Σ_i satisfies the above conditions with $(D, n) = ((0,1),1)$. In this case, we assume that the wavelets are *piecewise polynomials of order* d_x, with those on positive levels being orthogonal to all polynomials of order $\tilde{d}_x$ that are in V.

Assumption 4.3. *The bi-infinite matrices* $\mathbf{M} = (\Sigma, \Sigma)_H$ *and* $\mathbf{A} = a([\Sigma]_V, [\Sigma]_V)$ *for the spatial operators in* (5.1) *are* s^* *computable, in the sense that for each* $N \in \mathbb{N}$*, there exist approximate matrices* $\mathbf{M}_N$ *and* $\mathbf{A}_N$ *with at most* N *non-zero entries in each column and such that, for every* $0 \le \bar{s} < s^*$*, the expressions*

$$\sup_{N\in\mathbb{N}} N\|\mathbf{M} - \mathbf{M}_N\|^{1/\bar{s}}\ , \quad \sup_{N\in\mathbb{N}} N\|\mathbf{A} - \mathbf{A}_N\|^{1/\bar{s}}$$

are finite. Here, $\|\cdot\|$ *denotes the spectral norm.*

A number of practically viable constructions of Riesz bases Σ, which satisfy Assumption 4.3 for several classes of operators $A \in \mathcal{L}(V, V^*)$ have become available in recent years: for example, for second order, elliptic divergence form differential operators A, and also for self-adjoint, integro-differential operators A of fractional order (in which case V coincides with the domain of $A^{\frac{1}{2}}$); also tensorized Σ for diffusions on $D = (0,1)^n$ have become available, which satisfy Assumption 4.3. We refer to [20, Sect. 8.3] for this. For $0 \le s \le 1$, we denote by $[\Sigma]_s$ the Riesz basis Σ rescaled to $[H, V]_s$.

4.5. **Riesz bases in X and Y.** We assume that we have at our disposal Riesz bases $\Theta^X = \{\theta^X_\lambda : \lambda \in \nabla^X_t\}$, $\Theta^Y = \{\theta^Y_\lambda : \lambda \in \nabla^Y_t\}$ of $L^2(\mathbb{R}_>)$ for which rescaling renders Θ^X a Riesz basis of $H^1_{0,\{0\}}(\mathbb{R}_>)$ and Θ^Y a Riesz basis of $H^1(\mathbb{R}_>)$. The bases $[\Theta^X]_{\frac{1}{2}}$ and $[\Theta^Y]_{\frac{1}{2}}$ are then defined as in Proposition 4.2. In the spatial domain D, we assume available a Riesz basis $\Sigma = \{\sigma_\lambda : \lambda \in \nabla_x\}$ of H which, when rescaled to V, becomes a Riesz basis $[\Sigma]_V$ for V: $[\Sigma]_V = \{\sigma_\lambda/\|\sigma_\lambda\|_V : \lambda \in \nabla_x\}$.

Proposition 4.4. *Given Riesz bases* Θ^X, Θ^Y *and* Σ *of* $L^2(\mathbb{R}_>)$ *and* H*, respectively, as above, the collections* $\Psi^X := \Theta^X \otimes \Sigma$, $\Psi^Y := \Theta^Y \otimes \Sigma$ *, are Riesz bases of* $L^2(\mathbb{R}_>; H) \simeq L^2(\mathbb{R}_>) \otimes H$*. Moreover, the collection*

$$\Psi^X := \left\{ (t,x) \mapsto \frac{\theta^X_\lambda(t)\sigma_\mu(x)}{\sqrt{\|\sigma_\mu\|^2_V + \|\theta^X_\lambda\|^2_{H^{\frac{1}{2}}_{00,\{0\}}(\mathbb{R}_>)}}} : (\lambda,\mu) \in \nabla^X := \nabla^X_t \times \nabla_x \right\}$$

is a Riesz basis for $X = H^{\frac{1}{2}}_{00,\{0\}}(\mathbb{R}_>; H) \cap L^2(\mathbb{R}_>; V)$*, and the collection*

$$\Psi^Y := \left\{ (t,x) \mapsto \frac{\theta^Y_\lambda(t)\sigma_\mu(x)}{\sqrt{\|\sigma_\mu\|^2_V + \|\theta^Y_\lambda\|^2_{H^{\frac{1}{2}}(\mathbb{R}_>)}}} : (\lambda,\mu) \in \nabla^Y := \nabla^Y_t \times \nabla_x \right\}$$

is a Riesz basis for $Y = H^{\frac{1}{2}}(\mathbb{R}_>; H) \cap L^2(\mathbb{R}_>; V)$.

The Riesz constants for Ψ^X *and* Ψ^Y *depend only on the respective Riesz constants for* Θ^X, $[\Theta^X]_{\frac{1}{2}}$, Θ^Y, $[\Theta^Y]_{\frac{1}{2}}$ *and for* Σ, $[\Sigma]_V$.

Proof. The Riesz basis property for Ψ follows from our assumptions on Θ and Σ, the result [15, Prop. 1, Prop. 2] on tensor products of Riesz bases and from Proposition 4.2. □

4.6. **Space-time compressible approximation rates of smooth solutions.** Using the tensor product Riesz bases Ψ^X of X in Proposition 4.4 in a Petrov-Galerkin discretization (4.1) of the space-time variational formulation (3.3) allows for space-time compressive approximations of smooth solutions, *provided* test function spaces Y^ℓ are available which are *stable*, i.e., which satisfy (4.2). The approximate solutions thus obtained will be quasi-optimal. Such stable test spaces can be constructed on the basis of the coercivity property in Subsection 2.9. However, we shall not develop this here but refer to [10, Chapt. 5]. Likewise, in the adaptive setting, sequences of approximate solutions are produced, which converge at best possible rates, when compared to best N-term approximations of the solution. We therefore exemplify the best possible approximation rates in X which can be achieved in terms of the parameters d_t and d_x.

4.6.1. *Best rate in case* (A). For any $\Lambda \subset \nabla_x$, let $Q_\Lambda : L^2(D) \to \mathrm{span}(\theta_\lambda : \lambda \in \Lambda)$ denote the $L^2(D)$-biorthogonal projector associated to Σ and Λ. The assumption of Σ being of order d_x means that, with $\nabla_x^{(k)}$ being the set of $\lambda \in \nabla_x$ with refinement level $|\lambda| \le k \in \mathbb{N}_0$, it holds that $\#\nabla_x^{(k)} \eqsim 2^{kn}$. Setting $\nabla_x^{(-1)} := \emptyset$, we obtain for the projector $Q_{k,x} := Q_{\nabla_x^{(k)}}$ that

$$\|\mathrm{Id} - Q_{k,x}\|_{\mathcal{L}(H^{d_x}(D)\cap V, V)} \lesssim 2^{-k(d_x-m)}, \quad \|\mathrm{Id} - Q_{k,x}\|_{\mathcal{L}(H^{d_x}(D)\cap V, H)} \lesssim 2^{-kd_x} .$$

In case $d_t < \frac{d_x-m}{n}$, with $\ell/k \in [\frac{d_t}{d_x-m} + \varepsilon, \frac{1}{n} - \varepsilon]$ for (small) $\varepsilon > 0$, we have

$$\Big\|\mathrm{Id} - \sum_{p=0}^{k}\sum_{q=0}^{\ell}(Q_{p,t} - Q_{p-1,t}) \otimes (Q_{q,x} - Q_{q-1,x})\Big\|_{\mathcal{L}(H^{d_t}(\mathbb{R}_>)\otimes(H^{d_x}(D)\cap V), L^2(\mathbb{R}_>)\otimes V)} \lesssim 2^{-kd_t} .$$

Here $\sum_{p=0}^{k}\sum_{q=0}^{\ell}(Q_{p,t} - Q_{p-1,t}) \otimes (Q_{q,x} - Q_{q-1,x})$ is the $L^2(D)$-biorthogonal projector associated to the tensor product basis $\Psi = \Theta \otimes \Sigma$ and the "sparse" tensor-product index set

$$\Lambda_{\mathrm{A}} := \cup_{p=0}^{k} \cup_{q=0}^{\ell} (\nabla_t^{(p)} \setminus \nabla_t^{(p-1)}) \times (\nabla_x^{(q)} \setminus \nabla_x^{(q-1)}) ,$$

which satisfies $\#(\Lambda_{\mathrm{A}}) \lesssim 2^k$, see [14].

In view of the approximation orders of the bases being applied, and the tensor product structure of $X = H^{\frac{1}{2}}_{00,\{0\}}(\mathbb{R}_>; H) \cap L^2(\mathbb{R}_>; V)$, by interpolation we obtain the rate

$$2^{-k[\min(d_t - \frac{1}{2}, \frac{d_x-m}{n}) - \varepsilon]} . \tag{4.6}$$

with $\varepsilon > 0$ arbitrarily small due to the appearance of logarithmic factors. This rate is best possible for functions which are smooth with respect to x and t, and for Riesz bases Σ with isotropic supports in D as are admitted in case (A).

4.6.2. *Best rate in case* (B). Throughout the discussion of case (B), we assume $n \geq 2$ (the case $n = 1$ being a particular instance of (A)). For $1 \leq i \leq n$, let V_i be either $H^m(0,1)$ or a closed subspace incorporating essential boundary conditions. Let $\Sigma_i = \{\sigma_{i,\lambda_i} : \lambda_i \in \nabla_i\}$ be a normalized Riesz basis for $H_i := L^2(0,1)$, that renormalized in V_i is a Riesz basis for V_i. For any $\Lambda_i \subset \nabla_i$ we denote by $Q_{\Lambda_i} : L^2(0,1) \to \operatorname{span}(\theta_\lambda : \lambda \in \nabla_i)$ the $L^2(0,1)$-biorthogonal projectors associated to Σ_i and Λ_i. The assumption of Σ_i consisting of continuous, piecewise polynomial functions of order d_x means that, with $\nabla_i^{(k)} = \{\lambda \in \nabla_i : |\lambda| \leq k \in \mathbb{N}_0\}$, on any finite subinterval $(0,T) \subset \mathbb{R}_>$ it holds that $\#\nabla_i^{(k)} \eqsim 2^k$ (with the constant implied in $\eqsim$ being $O(T)$). With the convention $\nabla_i^{(-1)} := \emptyset$, and $Q_{-1,i} \equiv 0$, we have for $Q_{k,i} := Q_{\nabla_i^{(k)}}$ that

$$\|\mathrm{Id} - Q_{k,i}\|_{\mathcal{L}(H^{d_x}(0,1)\cap V_i, V_i)} \lesssim 2^{-k(d_x-m)} , \quad \|\mathrm{Id} - Q_{k,i}\|_{\mathcal{L}(H^{d_x}(0,1)\cap V_i, H_i)} \lesssim 2^{-kd_x} .$$

The collection $\Sigma := \otimes_{i=1}^n \Sigma_i = \{\sigma_\lambda := \otimes_{i=1}^n \sigma_{i,\lambda_i} : \lambda \in \nabla_x := \prod_{i=1}^n \nabla_i\}$ is a normalized Riesz basis for $L^2(D)$. Rescaling this basis in

$$V := \cap_{i=1}^n \otimes_{j=1}^n W_{ij} , \text{ where } W_{ij} := \begin{cases} H_j , & \text{when } j \neq i , \\ V_i , & \text{when } j = i , \end{cases}$$

it is a Riesz basis for V as well.

Recall that for any $\Lambda \subset \nabla_x$, Q_Λ denotes the $L^2(D)$-biorthogonal projector associated to Σ and Λ. As shown in [14, 24], there exist "optimized" sparse product sets $\nabla_x^{(1)} \subset \nabla_x^{(1)} \subset \cdots \subset \nabla_x$ and $\hat\nabla_x^{(1)} \subset \hat\nabla_x^{(1)} \subset \cdots \subset \nabla_x$ with $\#\nabla_x^{(1)} \eqsim 2^k \eqsim \hat\nabla_x^{(k)}$, such that with $Q_{k,x} := Q_{\nabla_x^{(k)}}$ and with $\hat Q_{k,x} := Q_{\hat\nabla_x^{(k)}}$, and

$$\mathcal{H}^{d_x}(D) := \cap_{i=1}^n \otimes_{j=1}^n Z_{ij} , \text{ where } Z_{ij} := \begin{cases} H_j , & \text{when } j \neq i , \\ H^{d_x}(0,1) \cap V_i , & \text{when } j = i , \end{cases}$$

it holds that

$$\|\mathrm{Id} - Q_{k,x}\|_{\mathcal{L}(\mathcal{H}^{d_x}(D),V)} \lesssim 2^{-k(d_x-m)} , \quad \|\mathrm{Id} - \hat Q_{k,x}\|_{\mathcal{L}(\mathcal{H}^{d_x}(D),H)} \lesssim 2^{-kd_x} .$$

Choosing as index set Λ_{B} the union of sparse products of the index sets $(\nabla_t^{(p)})_{0\leq p\leq k}$ with $(\nabla_x^{(q)})_{0\leq q\leq \ell}$ or $(\hat\nabla_x^{(q)})_{0\leq q\leq \ell}$ for suitable k and ℓ, we obtain $L^2(\mathbb{R}_> \times D)$-biorthogonal projectors associated to $X_{\Lambda_{\mathrm{B}}} \subset X = \operatorname{clos}_X(\Theta \otimes \Sigma)$ that, for $u \in (H^{d_t} \cap H^1_{0,\{0\}})(\mathbb{R}_>) \otimes \mathcal{H}^{d_x}(D)$, with a set of at most N basis functions give rise to an error in $H^s_{0,\{0\}}(\mathbb{R}_>; V)$ of order $2^{-k\min(d_t-s,d_x-m)}$, for $s = 0, 1$. Interpolation between $L^2(\mathbb{R}_>)$ and $H^1_{0,\{0\}}(\mathbb{R}_>)$ results, by Proposition 2.6, in the norm of the Bochner space $X \simeq H^{\frac12}_{00,\{0\}}(\mathbb{R}_>) \otimes V$ in the (best possible, for smooth functions) rate

$$\min(d_t - \frac12, d_x - m) . \tag{4.7}$$

Summarizing (4.6) and (4.7), for solutions which are smooth functions of space and time, the rate

$$s_{\max} := \begin{cases} \min(d_t - \frac12, \frac{d_x-m}{n}) - \varepsilon & \text{in case (A)} , \\ \min(d_t - \frac12, d_x - m) & \text{in case (B)} . \end{cases} \tag{4.8}$$

is realized with the index sets $\Lambda_{\mathrm{A}}, \Lambda_{\mathrm{B}} \subset \nabla^X$.

5. Adaptivity

The sparse tensor space-time Galerkin discretization (4.1) based on the a priori choices X_{Λ_A}, X_{Λ_B} of sparse tensor product trial spaces and the corresponding testfunction spaces Y_{Λ_A}, Y_{Λ_B} lead to quasi-optimal approximations; the quality of the Galerkin approximation thus being determined by the best approximation property. Alternatively, following [6, 13], (sequences of) subspaces $X^\ell = X_{\Lambda_\ell^X} \subset X$ and $Y^\ell = Y_{\Lambda_\ell^Y} \subset Y$ may be selected *adaptively*, with sequences $\{\Lambda_k\}_{k\geq 0} \subset \nabla_t \times \nabla_x$ of sets of "active" basis elements $\theta_\lambda \otimes \sigma_\mu \in \Psi = \Theta \otimes \Sigma$ determined so as to ensure optimality properties of the corresponding Galerkin approximations $u_{\Lambda_k^X}$ *for the given set of data*. In doing this, a key role is played by the (approximate) computability of (finite sections of) the bi-infinite matrix $\mathbf{B}$ defined by

$$\mathbf{B} = \big(\big(\mathrm{D}_+^{\frac{1}{2}}[\Theta^X]_{\frac{1}{2}}\big), \mathrm{D}_-^{\frac{1}{2}}[\Theta^Y]_{\frac{1}{2}}\big)\big)_{L^2(\mathbb{R}_>)} \otimes \big(\Sigma, \Sigma\big)_H + \big(\Theta^X, \Theta^Y\big)_{L^2(\mathbb{R}_>)} \otimes a\big([\Sigma]_V, [\Sigma]_V\big)\,. \tag{5.1}$$

We recapitulate basic properties of adaptive wavelet-Galerkin methods, in particular, the notions of admissibility and computability of the corresponding discretized operators; our presentation will be synoptic, and we refer readers who are unfamiliar with these to [21, 20]. We will, in particular, review the notions of s-admissibility, s-computability and s-compressibility of Galerkin matrices of operators. Finally, we obtain an optimality result for the adaptive wavelet Galerkin discretization of the space-time variational formulation (3.3): the sequence of Galerkin solutions produced by the adaptive scheme is optimal in the norm of X with respect to the best N-term approximation of the solution in space-time tensor product wavelet bases; thereby offering the first result on optimality for a nonlinear and compressive algorithm for long-time parabolic evolution problems. This is distinct from [4, 20], where the constants in the error and complexity estimates depend on the length of the time interval.

5.1. **Nonlinear approximation.** Nonlinear approximations to $u \in X$ are obtained from its coefficient vector $\mathbf{u}$ by best N-term approximations $\mathbf{u}_N$. These vectors, with supports of size $N \in \mathbb{N}_0$, encode the N largest coefficients in modulus of $\mathbf{u}$. For $s > 0$, the approximation class $\mathcal{A}^s_\infty(\ell_2(\nabla^X)) := \big\{\mathbf{v} \in \ell_2(\nabla^X) : \|\mathbf{v}\|_{\mathcal{A}^s_\infty(\ell_2(\nabla^X))} < \infty\big\}$, where

$$\|\mathbf{v}\|_{\mathcal{A}^s_\infty(\ell_2(\nabla^X))} := \sup_{\delta>0} \delta \times [\min\{N \in \mathbb{N}_0 : \|\mathbf{v} - \mathbf{v}_N\|_{\ell_2(\nabla^X)} \leq \delta\}]^s$$

contains all $\mathbf{v}$ whose best N-term approximations converge to $\mathbf{v}$ with rate s.

Since best N-term approximations involve searching the entire vector $\mathbf{v}$, they cannot be realized in practice. In addition, for a solution $u \in X$ of the PDE (1.1), the vector $\mathbf{u}$ to be approximated is not explicitly available. It is only given implicitly via (1.1), (1.2) through the (equivalent) bi-infinite matrix vector problem (4.5) with respect to some Riesz basis Ψ^X. Our aim is to construct a practical method that produces approximations to $\mathbf{u}$ which, whenever $\mathbf{u} \in \mathcal{A}^s_\infty(\ell_2(\nabla^X))$ for some $s > 0$, converge with this rate s in linear computational complexity.

5.2. **Adaptive Galerkin methods.** Let $s > 0$ be such that $\mathbf{u} \in \mathcal{A}^s_\infty(\ell_2(\nabla^X))$. In [6] and the references there, *adaptive wavelet Galerkin methods* for solving (4.5) were introduced. These methods are iterative methods which address the non-elliptic nature of the operator (1.1) by iterating, instead of (5.1), on the associated normal equations, i.e., on the linear system

$$\mathbf{B}^*\mathbf{B}\mathbf{u} = \mathbf{B}^*\mathbf{f}\,. \tag{5.2}$$

Key ingredients in the estimates of their complexity are asymptotic cost bounds for approximate matrix-vector products in terms of the prescribed tolerance ε.

Definition 5.1. (s^*-admissibility) $\mathbf{B} \in \mathcal{L}(\ell_2(\nabla^X), \ell_2(\nabla^Y))$ *is s^*-admissible if there exists a routine*

$$\mathbf{APPLY}_{\mathbf{B}}[\mathbf{w}, \varepsilon] \to \mathbf{z}$$

which yields, for any $\varepsilon > 0$ and any finitely supported $\mathbf{w} \in \ell_2(\nabla^X)$, a finitely supported $\mathbf{z} \in \ell_2(\nabla^Y)$ with $\|\mathbf{B}\mathbf{w} - \mathbf{z}\|_{\ell_2(\nabla^Y)} \leq \varepsilon$ and for which, for any $\bar{s} \in (0, s^*)$, there exists an admissibility constant $a_{\mathbf{B},\bar{s}}$ *such that $\#\operatorname{supp}\mathbf{z} \leq a_{\mathbf{B},\bar{s}}\varepsilon^{-1/\bar{s}}\|\mathbf{w}\|^{1/\bar{s}}_{\mathcal{A}^{\bar{s}}_\infty(\ell_2(\nabla^X))}$, and the number of arithmetic operations and storage locations used by the call $\mathbf{APPLY}_{\mathbf{B}}[\mathbf{w}, \varepsilon]$ is bounded by some absolute multiple of*

$$a_{\mathbf{B},\bar{s}}\varepsilon^{-1/\bar{s}}\|\mathbf{w}\|^{1/\bar{s}}_{\mathcal{A}^{\bar{s}}_\infty(\ell_2(\nabla^X))} + \#\operatorname{supp}\mathbf{w} + 1.$$

One key step in adaptive wavelet methods for (4.5) is thus the construction of a valid routine $\mathbf{APPLY}_{\mathbf{B}}[\mathbf{w}, \varepsilon]$ for the bi-infinite matrices $\mathbf{B}$ defined in (5.1).

In order to approximate $\mathbf{u}$ one should be able to approximate $\mathbf{f}$. *Throughout what follows, we therefore assume availability of the following routine.*

$\mathbf{RHS}_{\mathbf{f}}[\varepsilon] \to \mathbf{f}_\varepsilon$: *For given $\varepsilon > 0$, it yields a finitely supported $\mathbf{f}_\varepsilon \in \ell_2(\nabla^Y)$ with*

$$\|\mathbf{f} - \mathbf{f}_\varepsilon\|_{\ell_2(\nabla^Y)} \leq \varepsilon \quad \text{and} \quad \#\operatorname{supp}\mathbf{f}_\varepsilon \lesssim \min\{N : \|\mathbf{f} - \mathbf{f}_N\| \leq \varepsilon\},$$

with the number of arithmetic operations and storage locations used by the call $\mathbf{RHS}_{\mathbf{f}}[\varepsilon]$ bounded by some absolute multiple of $\#\operatorname{supp}\mathbf{f}_\varepsilon + 1$.

The availability of $\mathbf{APPLY}_{\mathbf{B}}$ and $\mathbf{RHS}_{\mathbf{f}}$ implies the following result.

Proposition 5.2. *Let $\mathbf{B}$ in (4.5) be s^*-admissible. Then for any $\bar{s} \in (0, s^*)$, we have $\|\mathbf{B}\|_{\mathcal{L}(\mathcal{A}^{\bar{s}}_\infty(\ell_2(\nabla^X)),\mathcal{A}^{\bar{s}}_\infty(\ell_2(\nabla^Y)))} \leq a^{\bar{s}}_{\mathbf{B},\bar{s}}$. For $\mathbf{z}_\varepsilon := \mathbf{APPLY}_{\mathbf{B}}[\mathbf{w}, \varepsilon]$, there holds $\|\mathbf{z}_\varepsilon\|_{\mathcal{A}^{\bar{s}}_\infty(\ell_2(\nabla^Y))} \leq a^{\bar{s}}_{\mathbf{B},\bar{s}}\|\mathbf{w}\|_{\mathcal{A}^{\bar{s}}_\infty(\ell_2(\nabla^X))}$.*

For proofs, we refer to [6] or [7, Prop. 3.3]. Using the definition of $\mathcal{A}^s_\infty(\ell_2(\nabla^Y))$ and the properties of $\mathbf{RHS}_{\mathbf{f}}$, we have

Corollary 5.3. *If, in (4.5), $\mathbf{B}$ is s^*-admissible and $\mathbf{u} \in \mathcal{A}^s_\infty(\ell_2(\nabla^X))$ for $s < s^*$, then for $\mathbf{f}_\varepsilon = \mathbf{RHS}_{\mathbf{f}}[\varepsilon]$, $\#\operatorname{supp}\mathbf{f}_\varepsilon \lesssim a_{\mathbf{B},s}\varepsilon^{-1/s}\|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_\infty(\ell_2(\nabla^X))}$ with the number of arithmetic operations and storage locations used by the call $\mathbf{RHS}_{\mathbf{f}}[\varepsilon]$ being bounded by some absolute multiple of*

$$a_{\mathbf{B},s}\varepsilon^{-1/s}\|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_\infty(\ell_2(\nabla^X))} + 1.$$

Remark 5.4. *Besides $\|\mathbf{f} - \mathbf{f}_\varepsilon\|_{\ell_2(\nabla^Y)} \leq \varepsilon$, the complexity bounds in Corollary 5.3 with $a_{\mathbf{B},s} > 0$ being independent of ε are essential for the use of $\mathbf{RHS}_{\mathbf{f}}$ in the adaptive wavelet methods.*

The following corollary of Proposition 5.2 can be used for example for the construction of valid **APPLY** and **RHS** routines in case the adaptive wavelet algorithms are applied to a preconditioned system.

Corollary 5.5. *If $\mathbf{B} \in \mathcal{L}(\ell_2(\nabla^X), \ell_2(\nabla^Y))$, $\mathbf{C} \in \mathcal{L}(\ell_2(\nabla^Y), \ell_2(\nabla^Z))$ are both s^*-admissible, then so is $\mathbf{CB} \in \mathcal{L}(\ell_2(\nabla^X), \ell_2(\nabla^Z))$. A valid routine $\mathbf{APPLY}_{\mathbf{CB}}$ is*

$$[\mathbf{w}, \varepsilon] \mapsto \mathbf{APPLY}_{\mathbf{C}}\big[\mathbf{APPLY}_{\mathbf{B}}[\mathbf{w}, \varepsilon/(2\|\mathbf{C}\|)], \varepsilon/2\big] \,, \tag{5.3}$$

with admissibility constant $a_{\mathbf{CB},\bar{s}} \lesssim a_{\mathbf{B},\bar{s}}(\|\mathbf{C}\|^{1/\bar{s}} + a_{\mathbf{C},\bar{s}})$ for $\bar{s} \in (0, s^*)$.

For some $s^* > s$, let $\mathbf{C} \in \mathcal{L}(\ell_2(\nabla^Y), \ell_2(\nabla^Z))$ be s^*-admissible. Then for

$$\mathbf{RHS}_{\mathbf{Cf}}[\varepsilon] := \mathbf{APPLY}_{\mathbf{C}}[\mathbf{RHS}_{\mathbf{f}}[\varepsilon/(2\|\mathbf{C}\|)], \varepsilon/2] \,, \tag{5.4}$$

there holds

$$\#\operatorname{supp}\mathbf{RHS}_{\mathbf{Cf}}[\varepsilon] \lesssim a_{\mathbf{B},s}(\|\mathbf{C}\|^{1/s} + a_{\mathbf{C},s})\varepsilon^{-1/s}\|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_\infty(\ell_2(\nabla^X))} ,$$
$$\|\mathbf{Cf} - \mathbf{RHS}_{\mathbf{Cf}}[\varepsilon]\|_{\ell_2(\nabla^Z)} \leq \varepsilon ,$$

with the number of arithmetic operations and storage locations used by the call $\mathbf{RHS}_{\mathbf{Cf}}[\varepsilon]$ *bounded by a multiple of*

$$a_{\mathbf{B},s}(\|\mathbf{C}\|^{1/s} + a_{\mathbf{C},s})\varepsilon^{-1/s}\|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_\infty(\ell_2(\nabla^X))} + 1 .$$

Remark 5.6. $\mathbf{RHS}_{\mathbf{Cf}}$ *allows to approximate* $\mathbf{Cf}$ *in the sense of Remark 5.4.*

Consider first the case that $\mathbf{B}$ is *self-adjoint positive definite*, i.e., $\nabla^X = \nabla^Y$ and $\mathbf{B} = \mathbf{B}^* > 0$. In this case the adaptive wavelet methods from [6] are *optimal* in the following sense.

Theorem 5.7. *([6, 13]) If in (4.5)* $\mathbf{B}$ *is self-adjoint positive definite and* s^**-admissible, then for any* $\varepsilon > 0$*, the adaptive wavelet method from* [6] *produces an approximation* $\mathbf{u}_\varepsilon$ *to* $\mathbf{u}$ *with* $\|\mathbf{u} - \mathbf{u}_\varepsilon\|_{\ell_2(\nabla^X)} \leq \varepsilon$*. If in* (4.5) *for some* $s > 0$ *it holds* $\mathbf{u} \in \mathcal{A}^s_\infty(\ell_2(\nabla^X))$*, then* $\#\operatorname{supp}\mathbf{u}_\varepsilon \lesssim \varepsilon^{-1/s}\|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_\infty(\ell_2(\nabla^X))}$ *and if, in addition,* $s < s^*$*, the number of arithmetic operations and storage locations required by one call of either of these adaptive wavelet solvers with tolerance* ε *is bounded by a multiple of*

$$\varepsilon^{-1/s}(1 + a_{\mathbf{B},s})\|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_\infty(\ell_2(\nabla^X))} + 1.$$

The factor depends only on s *when it tends to* 0 *or* ∞*, and on* $\|\mathbf{B}\|$ *and* $\|\mathbf{B}^{-1}\|$*.*

The adaptive Galerkin discretization method from [5] for self-adjoint operators $\mathbf{B}$ consists of the application of a damped Richardson iteration to $\mathbf{Bu} = \mathbf{f}$, where the required residual computations are approximated using calls of $\mathbf{APPLY}_{\mathbf{B}}$ and $\mathbf{RHS}_{\mathbf{f}}$ within tolerances that decrease linearly with the iteration counter.

With the method from [5], a sequence $\Xi_0 \subset \Xi_1 \subset \cdots \subset \nabla^X$ is produced, together with corresponding (approximate) Galerkin solutions $\mathbf{u}_i \in \ell_2(\Xi_i)$. The coefficients of approximate residuals $\mathbf{f} - \mathbf{Bu}_i$ are used as indicators how to expand Ξ_i to Ξ_{i+1} such that it gives rise to an improved Galerkin approximation.

The method of [5] relies on a recurrent coarsening of the approximation vectors, where small coefficients are removed to maintain optimal balance between accuracy and support length. We have s^*-admissibility of $\mathbf{B}$ once the stiffness matrix with respect to suitable wavelet bases is close to a computable sparse matrix. The next definition makes this precise.

Definition 5.8. (s^*-computability) $\mathbf{B} \in \mathcal{L}(\ell_2(\nabla^X), \ell_2(\nabla^Y))$ *is* s^**-computable if, for each* $N \in \mathbb{N}$*, there exists a* $\mathbf{B}_N \in \mathcal{L}(\ell_2(\nabla^X), \ell_2(\nabla^Y))$ *having in each column at most* N *non-zero entries whose joint computation takes an absolute multiple of* N *operations, such that the* computability constants

$$c_{\mathbf{B},\bar{s}} := \sup_{N\in\mathbb{N}} N\|\mathbf{B} - \mathbf{B}_N\|^{1/\bar{s}}_{\ell_2(\nabla^X)\to\ell_2(\nabla^Y)}$$

are finite for any $\bar{s} \in (0, s^*)$*.*

Theorem 5.9. *An* s^**-computable* $\mathbf{B}$ *is* s^**-admissible. Moreover, for* $\bar{s} < s^*$*,* $a_{\mathbf{B},\bar{s}} \lesssim c_{\mathbf{B},\bar{s}}$ *where the constant in this estimate depends only on* $\bar{s} \downarrow 0$*,* $\bar{s} \uparrow s^*$*, and on* $\|\mathbf{B}\| \to \infty$*.*

This theorem is proven by the construction of a suitable $\mathbf{APPLY}_{\mathbf{B}}$ routine as was done in [5, §6.4], see also [21] and the references there.

The non-elliptic nature of $\mathbf{B}$ was addressed in [6] by applying the adaptive schemes to the *normal equations* (5.2): From Subsection 4.2 we deduce that $\mathbf{B}^*\mathbf{B} \in \mathcal{L}(\ell_2(\nabla^X), \ell_2(\nabla^X))$ is boundedly invertible, self-adjoint positive definite, with

$$\|\mathbf{B}^*\mathbf{B}\|_{\mathcal{L}(\ell_2(\nabla^X),\ell_2(\nabla^X))} \le \|\mathbf{B}\|^2_{\mathcal{L}(\ell_2(\nabla^X),\ell_2(\nabla^Y))}\,,$$
$$\|(\mathbf{B}^*\mathbf{B})^{-1}\|_{\mathcal{L}(\ell_2(\nabla^X),\ell_2(\nabla^X))} \le \|\mathbf{B}^{-1}\|^2_{\mathcal{L}(\ell_2(\nabla^Y),\ell_2(\nabla^X))}\,.$$

Now let $\mathbf{u} \in \mathcal{A}^s_\infty(\ell_2(\nabla^X))$, and assume that for some $s^* > s$, both $\mathbf{B}$ and $\mathbf{B}^*$ are s^*-admissible. By Corollary 5.5, with $\mathbf{B}^*$ in place of $\mathbf{C}$, a valid $\mathbf{RHS}_{\mathbf{B}^*\mathbf{f}}$ routine is given by (5.4), and $\mathbf{B}^*\mathbf{B}$ is s^*-admissible with a valid $\mathbf{APPLY}_{\mathbf{B}^*\mathbf{B}}$ routine given by (5.3). A combination of Theorem 5.7 and Corollary 5.5 yields the following result.

Theorem 5.10. *For any $\varepsilon > 0$, the adaptive wavelet methods from* [6] *applied to the normal equations* (5.2) *using above $\mathbf{APPLY}_{\mathbf{B}^*\mathbf{B}}$ and $\mathbf{RHS}_{\mathbf{B}^*\mathbf{f}}$ routines produce approximations $\mathbf{u}_\varepsilon$ to $\mathbf{u}$ which satisfy $\|\mathbf{u}-\mathbf{u}_\varepsilon\|_{\ell_2(\nabla^X)} \le \varepsilon$. If for some $s > 0$, $\mathbf{u} \in \mathcal{A}^s_\infty(\ell_2(\nabla^X))$, then $\#\operatorname{supp}\mathbf{u}_\varepsilon \lesssim \varepsilon^{-1/s}\|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_\infty(\ell_2(\nabla^X))}$, with constant only dependent on s when it tends to 0 or ∞, and on $\|\mathbf{B}\|$ and $\|\mathbf{B}^{-1}\|$ when they tend to infinity.*

If $s < s^*$, then the number of arithmetic operations and storage locations required by a call of either of these adaptive wavelet methods with tolerance $\varepsilon > 0$ is bounded by some multiple of

$$1 + \varepsilon^{-1/s}(1 + a_{\mathbf{B},s}(1 + a_{\mathbf{B}^*,s}))\|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_\infty(\ell_2(\nabla^X))}$$

where this multiple only depends on s when it tends to 0 or ∞, and on $\|\mathbf{B}\|$ and $\|\mathbf{B}^{-1}\|$ when they tend to infinity.

5.3. **s^*-computability of $\mathbf{B}$ in** (5.1)**.** We apply the general concepts to the space-time variational formulation (3.3) and the space-time tensor-product wavelet bases $\Psi^X = \Theta^X \otimes \Sigma$ and $\Psi^Y = \Theta^Y \otimes \Sigma$ in Proposition 4.4.

Due to the discussion in Section 4.6, it suffices to show s^*-admissibility of both, $\mathbf{B}$ and $\mathbf{B}^*$, for $s^* > s_{\max}$ with $s_{\max}$ as defined in (4.8). The bi-infinite matrix $\mathbf{B}$ defined in (5.1) comprises of a sum of tensor products of bi-infinite matrices, each factor matrix corresponding to either the Gram matrices $\langle\Theta_>, \Theta_>\rangle_{L^2(\mathbb{R}_>)}$ or $\langle\Sigma, \Sigma\rangle_{L^2(D)}$ or of the "stiffness" matrices $a([\Sigma]_V, [\Sigma]_V)$ with respect to the Riesz bases $\Theta_>$ and Σ (cp. Section 4.5).

To apply the general theory of adaptive wavelet discretizations of [5, 6, 21], the key step is the verification of s^*-compressibility and of s^*-computability of the matrix $\mathbf{B}$ in (5.1).

We verify s^*-computability of $\mathbf{B}$ in (5.1) with the following result [20, Prop. 8.1].

Proposition 5.11. *Let for some $s^* > 0$, $\mathbf{D}$, $\mathbf{E}$ be s^*-computable. Then*

(a) *$\mathbf{D} \otimes \mathbf{E}$ is s^*-computable with computability constant satisfying, for $0 < \bar{s} < \tilde{s} < s^*$, $c_{\mathbf{D}\otimes\mathbf{E},\bar{s}} \lesssim (c_{\mathbf{D},\tilde{s}} c_{\mathbf{E},\tilde{s}})^{\tilde{s}/\bar{s}}$ and*

(b) *for any $\varepsilon \in (0, s^*)$, $\mathbf{D} \otimes \mathbf{E}$ is $(s^* - \varepsilon)$-computable, with computability constant $c_{\mathbf{D}\otimes\mathbf{E},\bar{s}}$ satisfying, for $0 < \bar{s} < s^* - \varepsilon < \tilde{s} < s^*$, $c_{\mathbf{D}\otimes\mathbf{E},\bar{s}} \lesssim \max(c_{\mathbf{D},\tilde{s}}, 1)\max(c_{\mathbf{E},\tilde{s}}, 1)$.*

The constants implicit by $\lesssim$ in the bounds on the computability constants in (a) *and* (b) *depend only on $\tilde{s}$, $\tilde{s} \to \infty$ and on $\tilde{s} - \bar{s} \downarrow 0$.*

We recall that we work under Assumption 4.3, so that the bi-infinite mass matrix $\mathbf{M} = \langle \Sigma, \Sigma \rangle_{L^2(D)}$ and the bi-infinite stiffness matrix $\mathbf{A} = a([\Sigma]_V, [\Sigma]_V)$ are both s^* computable and compressible under our assumptions (s1)–(s4).

5.4. **s^*-computability of the fractional time derivatives.** Proposition 5.11 and Assumption 4.3 reduce the analysis of s^*-compressibility of $\mathbf{B}$ in (5.1) to the verification of the s^*-compressibility of the temporal "stiffness" and "mass" matrices

$$\mathbf{D} := \left\langle \mathrm{D}_+^{\frac{1}{2}}[\Theta^X]_{\frac{1}{2}}, \mathrm{D}_-^{\frac{1}{2}}[\Theta^Y]_{\frac{1}{2}} \right\rangle_{L^2(\mathbb{R}_>)} , \qquad \mathbf{G} := \left\langle \Theta^X, \Theta^Y \right\rangle_{L^2(\mathbb{R}_>)} , \tag{5.5}$$

i.e., on the compressibility of the "stiffness" matrix $\mathbf{D}$ and of the "mass"-matrix $\mathbf{G}$ of the fractional time derivative in (3.4).

We discuss s^*-computability of $\mathbf{D}$ and $\mathbf{G}$ in the sense of Definition 5.8. We assume at our disposal Riesz bases Θ^X of $H^{\frac{1}{2}}_{00,\{0\}}(\mathbb{R}_>)$ and Θ^Y of $H^{\frac{1}{2}}(\mathbb{R}_>)$ as in Section 4.3 and, in particular, that properties (t1)–(t4) of that section hold for elements of either of these bases.

The s^*-computability of $\mathbf{G}$ follows as in [20, Sect. 8.2] from the properties (t1)–(t4) of Θ^X and Θ^Y. It remains to address s^*-computability of $\mathbf{D}$ in (5.5).

To this end, we observe that by a density argument, Lemma 2.7 and, in particular, the fractional integration by parts identity (2.13) remain valid for $w \in H^{\frac{1}{2}}_{00,\{0\}}(\mathbb{R}_>)$ and for $v \in H^{\frac{1}{2}}(\mathbb{R}_>)$. Since Θ^X is a Riesz basis of $H^{\frac{1}{2}}_{00,\{0\}}(\mathbb{R}_>)$ and Θ^Y of $H^{\frac{1}{2}}(\mathbb{R}_>)$, we obtain from (2.13) that

$$\mathbf{D} = \left\langle \mathrm{D}_+^{\frac{1}{2}}[\Theta^X]_{\frac{1}{2}}, \mathrm{D}_-^{\frac{1}{2}}[\Theta^Y]_{\frac{1}{2}} \right\rangle_{L^2(\mathbb{R}_>)} = \left\langle \mathrm{D}[\Theta^X]_1, \Theta^Y \right\rangle_{L^2(\mathbb{R}_>)} .$$

Now using properties (t1)–(t4) of the temporal wavelet bases Θ^X and Θ^Y, we establish s^*-computability of $\mathbf{D}$ as in [20, Sect. 8.2].

5.5. **Optimality.** The preceding considerations can be combined into

Theorem 5.12. *Consider the parabolic problem* (1.1), (1.2) *in the weak form* (3.3) *with spatial bilinear form as in Section 2.2. Consider its representation* $\mathbf{Bu} = \mathbf{f}$ *using temporal and spatial wavelet bases* Θ *and* Σ *as above.*

Then for any $\varepsilon > 0$*, the adaptive wavelet methods from* [6] *applied to the normal equations* (5.2) *produce an approximation* $\mathbf{u}_\varepsilon$ *with*

$$\|u - \mathbf{u}_\varepsilon^\top [\Theta \otimes \Sigma]\|_X \eqsim \|\mathbf{u} - \mathbf{u}_\varepsilon\| \leq \varepsilon .$$

If for some $s > 0$, $\mathbf{u} \in \mathcal{A}^s_\infty(\ell_2(\nabla^X))$*, then* $\operatorname{supp} \mathbf{u}_\varepsilon \lesssim \varepsilon^{-1/s} \|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_\infty(\ell_2(\nabla^X))}$*, with the implied constant only dependent on* s *when it tends to* 0 *or* ∞*.*

If, for arbitrary $s^* > 0$*, it holds* $s < s^*$*, then the number of operations and storage locations required by one call of the space-time adaptive algorithm with tolerance* $\varepsilon > 0$ *is bounded by some absolute multiple of*

$$\varepsilon^{-1/s} n^2 \|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_\infty(\ell_2(\nabla^X))} + 1 .$$

Here, the implied constant depends only on the Riesz and the admissibility constants of the spatial wavelet bases Σ*.*

6. Acknowledgement

This work was initiated during a visit of S.L. at the research institute for mathematics (FIM) of ETH, and continued during a visit of C.S. at Chalmers and during workshops at the Mathematical Research Institute Oberwolfach.

References

[1] Roman Andreev. Stability of sparse space-time finite element discretizations of linear parabolic evolution equations. *IMA J. Numer. Anal.*, 33(1):242–260, 2013.

[2] Claudio Baiocchi and Franco Brezzi. Optimal error estimates for linear parabolic problems under minimal regularity assumptions. *Calcolo*, 20(2):143–176 (1984), 1983.

[3] Jöran Bergh and Jörgen Löfström. *Interpolation Spaces. An Introduction*. Springer-Verlag, Berlin, 1976. Grundlehren der Mathematischen Wissenschaften, No. 223.

[4] Nabi Chegini and Rob Stevenson. Adaptive wavelet schemes for parabolic problems: sparse matrices and numerical results. *SIAM J. Numer. Anal.*, 49(1):182–212, 2011.

[5] Albert Cohen, Wolfgang Dahmen, and Ronald DeVore. Adaptive wavelet methods for elliptic operator equations: convergence rates. *Math. Comp.*, 70(233):27–75, 2001.

[6] Albert Cohen, Wolfgang Dahmen, and Ronald DeVore. Adaptive wavelet methods. II. Beyond the elliptic case. *Found. Comput. Math.*, 2(3):203–245, 2002.

[7] Stephan Dahlke, Thorsten Raasch, Manuel Werner, Massimo Fornasier, and Rob Stevenson. Adaptive frame methods for elliptic operator equations: the steepest descent approach. *IMA J. Numer. Anal.*, 27(4):717–740, 2007.

[8] Wolfgang Dahmen, Angela Kunoth, and Karsten Urban. Biorthogonal spline wavelets on the interval—stability and moment conditions. *Appl. Comput. Harmon. Anal.*, 6(2):132–196, 1999.

[9] Tammo Jan Dijkema, Christoph Schwab, and Rob Stevenson. An adaptive wavelet method for solving high-dimensional elliptic PDEs. *Constr. Approx.*, 30(3):423–455, 2009.

[10] Magnus Fontes. Parabolic Equations with Low Regularity. Doctoral Dissertation, Department of Mathematics, University of Lund, Sweden, 1996.

[11] Magnus Fontes. A monotone operator method for elliptic-parabolic equations. *Comm. Partial Differential Equations*, 25(3-4):681–702, 2000.

[12] Magnus Fontes. Initial-boundary value problems for parabolic equations. *Ann. Acad. Sci. Fenn. Math.*, 34(2):583–605, 2009.

[13] T. Gantumur, H. Harbrecht, and R. Stevenson. An optimal adaptive wavelet method without coarsening of the iterands. *Math. Comp.*, 76(258):615–629, 2007.

[14] Michael Griebel and Helmut Harbrecht. On the construction of sparse tensor product spaces. *Math. Comp.*, 82(282):975–994, 2013.

[15] Michael Griebel and Peter Oswald. On the abstract theory of additive and multiplicative Schwarz algorithms. *Numer. Math.*, 70(2):163–180, 1995.

[16] Pierre Grisvard. *Elliptic Problems in Nonsmooth Domains*, volume 24 of *Monographs and Studies in Mathematics*. Pitman (Advanced Publishing Program), Boston, MA, 1985.

[17] U. Langer and M. Wolfmayr. Multiharmonic finite element analysis of a time-periodic parabolic optimal control problem. *J. Numer. Math.*, 21(4):265–300, 2013.

[18] J.-L. Lions and E. Magenes. *Problèmes aux limites non homogènes et applications. Vol. 1.* Travaux et Recherches Mathématiques, No. 17. Dunod, Paris, 1968.

[19] Stefan G. Samko, Anatoly A. Kilbas, and Oleg I. Marichev. *Fractional Integrals and Derivatives*. Gordon and Breach Science Publishers, Yverdon, 1993. Theory and applications, Edited and with a foreword by S. M. Nikol′skiĭ, Translated from the 1987 Russian original, Revised by the authors.

[20] Christoph Schwab and Rob Stevenson. Space-time adaptive wavelet methods for parabolic evolution problems. *Math. Comp.*, 78(267):1293–1318, 2009.

[21] Rob Stevenson. Adaptive wavelet methods for solving operator equations: an overview. In *Multiscale, Nonlinear and Adaptive Approximation*, pages 543–597. Springer, Berlin, 2009.

[22] Rob Stevenson. Divergence-free wavelet bases on the hypercube: general boundary conditions. *Preprint Erwin Schrödinger Inst.*, 13, 2013.

[23] Vidar Thomée. *Galerkin Finite Element Methods for Parabolic Problems*, volume 25 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 1997.
[24] Radu-Alexandru Todor. A new approach to energy-based sparse finite-element spaces. *IMA J. Numer. Anal.*, 29(1):72–85, 2009.
[25] Hans Triebel. *Interpolation Theory, Function Spaces, Differential Operators*. Johann Ambrosius Barth, Heidelberg, second edition, 1995.
[26] Karsten Urban. *Wavelet Methods for Elliptic Partial Differential Equations*. Numerical Mathematics and Scientific Computation. Oxford University Press, Oxford, 2009.
[27] Jinchao Xu and Ludmil Zikatanov. Some observations on Babuška and Brezzi theories. *Numer. Math.*, 94(1):195–202, 2003.

(Stig Larsson)
Department of Mathematical Sciences
Chalmers University of Technology
and University of Gothenburg
SE–412 96 Gothenburg, Sweden.
E-mail address: stig@chalmers.se

(Christoph Schwab)
Seminar für Angewandte Mathematik
ETH Zürich
Rämistrasse 101, CH–8092 Zürich, Switzerland.
E-mail address: schwab@math.ethz.ch

Recent Research Reports

Nr.	Authors/Title

2014-34 S. Lanthaler and S. Mishra
Computation of measure valued solutions for the incompressible Euler equations.

2014-35 P. Grohs and A. Obermeier
Ridgelet Methods for Linear Transport Equations

2014-36 P. Chen and Ch. Schwab
Sparse-Grid, Reduced-Basis Bayesian Inversion

2014-37 R. Kaeppeli and S. Mishra
Well-balanced schemes for gravitationally stratified media

2014-38 D. Schoetzau and Ch. Schwab
Exponential Convergence for hp-Version and Spectral Finite Element Methods for Elliptic Problems in Polyhedra

2014-39 P. Grohs and M. Sprecher
Total Variation Regularization by Iteratively Reweighted Least Squares on Hadamard Spaces and the Sphere

2014-40 R. Casagrande and R. Hiptmair
An A Priori Error Estimate for Interior Penalty Discretizations of the Curl-Curl Operator on Non-Conforming Meshes

2015-01 X. Claeys and R. Hiptmair
Integral Equations for Electromagnetic Scattering at Multi-Screens

2015-02 R. Hiptmair and S. Sargheini
Scatterers on the substrate: Far field formulas

2015-03 P. Chen and A. Quarteroni and G. Rozza
Reduced order methods for uncertainty quantification problems