Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Seminar for Applied Mathematics

Wavelet compressed, modified Hilbert transform in the space-time discretization of the heat equation

H. Harbrecht and Ch. Schwab and M. Zank

Research Report No. 2024-08
February 2024

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

AMS Subject Classification: primary 65N30

WAVELET COMPRESSED, MODIFIED HILBERT TRANSFORM IN THE SPACE-TIME DISCRETIZATION OF THE HEAT EQUATION

HELMUT HARBRECHT, CHRISTOPH SCHWAB, AND MARCO ZANK

Abstract. On a finite time interval $(0,T)$, we consider the multiresolution Galerkin discretization of a modified Hilbert transform $\mathcal{H}_T$ which arises in the space-time Galerkin discretization of the linear diffusion equation. To this end, we design spline-wavelet systems in $(0,T)$ consisting of piecewise polynomials of degree ≥ 1 with sufficiently many vanishing moments which constitute Riesz bases in the Sobolev spaces $H^s_{0,}(0,T)$ and $H^s_{,0}(0,T)$. These bases provide multilevel splittings of the temporal discretization spaces into "increment" or "detail" spaces of direct sum type. Via algebraic tensor-products of these temporal multilevel discretizations with standard, hierarchic finite element spaces in the spatial domain (with standard Lagrangian FE bases), *sparse space-time tensor-product spaces* are obtained, which afford a substantial reduction in the number of the degrees of freedom as compared to time-marching discretizations. In addition, temporal spline-wavelet bases allow to compress certain nonlocal integrodifferential operators which appear in stable space-time variational formulations of initial-boundary value problems, such as the heat equation and the acoustic wave equation. An efficient preconditioner is proposed that affords linear complexity solves of the linear system of equations which results from the sparse space-time Galerkin discretization.

1. Introduction

1.1. Motivation and background.

The efficient numerical solution of initial-boundary value problems (IBVPs) is central in computational science and engineering. Accordingly, numerical methods have been developed to a high degree of sophistication and maturity. Foremost among these are time-stepping schemes, which are motivated by the causality of the physical phenomena modeled by the equations. They discretize the evolution equation via sequential numerical solution of a sequence of spatial problems [36]. In recent years, however, principally motivated by applications from numerical optimal control, so-called *space-time methods* have emerged: these methods aim at the "one-shot" solution of the initial-boundary value problem as a well-posed operator equation on a space-time cylinder. The present article develops an efficient space-time method for linear, parabolic initial boundary value problems.

For linear parabolic evolution equations such as the heat equation, the *analytic semigroup* property will imply exponential convergence rates with respect to the number of temporal degrees of freedom if spectral or so-called *hp* Petrov–Galerkin discretizations are employed, see, e.g. [26, 30], *provided* that the initial boundary

2010 *Mathematics Subject Classification.* Primary 65N30, 65J15.

Key words and phrases. Wavelets, Modified Hilbert transform, Space-time variational formulation of parabolic PDEs, Generalized Hilbert transform, Sparse space-time approximation of evolution .

value problem is subject to a forcing function which depends analytically on the temporal variable.

In *space-time Galerkin discretizations* developed in [26], the exponential convergence rate of the corresponding *hp* time discretizations from [30] implies correspondingly small temporal stiffness and mass matrices, which can be efficiently treated by standard, dense linear algebra.

For *linear hyperbolic evolution equations* such as acoustic waves or time-domain Maxwell equations (e.g. [20]), temporal analyticity of solutions is not to be taken for granted. Similarly, for the aforementioned parabolic IBVPs subject to body forces with low temporal regularity (as, e.g., arise in pathwise solutions of linear, parabolic stochastic PDEs driven by rough noises, such as cylindrical Wiener processes), the solution has in turn low temporal regularity. In these settings, *fixed, low-order temporal discretization* are usually employed. Due to the possibly dense singular support in the time variable of solutions, adaptive time-stepping will in such settings likewise not afford significant order improvements. This implies, in the presently considered temporal Petrov–Galerkin formulation in a duality pairing which is realized by the temporal Hilbert transform, *large, dense* stiffness and mass matrices of size $O(N_t \times N_t)$ with N_t denoting the number of temporal degrees of freedom.

A prototypical parabolic initial-boundary value problem considered in the present article is to find the function $u(x,t)$ such that

$$\left\{ \begin{array}{rcll} \partial_t u + \mathcal{L}_x u & = & f & \text{in } Q = \Omega \times (0,1), \\ \gamma_{x,0}(u) & = & 0 & \text{on } \Gamma_{\mathrm{D}} \times [0,1], \\ \gamma_{x,1}(u) & = & 0 & \text{on } \Gamma_{\mathrm{N}} \times [0,1], \\ u(\cdot,0) & = & 0 & \text{in } \Omega \end{array} \right. \tag{1.1}$$

with given right-hand side f and, for simplicity at this stage, homogeneous Dirichlet boundary conditions on $\Gamma_{\mathrm{D}} \subset \partial\Omega$, homogeneous Neumann boundary conditions on $\Gamma_{\mathrm{N}} \subset \partial\Omega$ and homogeneous initial conditions. Here, $\mathcal{L}_x$ is a spatial, linear and strongly elliptic differential operator, $\gamma_{x,0}$ denotes the spatial Dirichlet trace map, and $\gamma_{x,1}$ is the spatial conormal trace operator on $\partial\Omega$, see Subsection 2.3 for further details.

1.2. **Previous results.** As already mentioned, space-time discretizations are motivated by applications from optimal control, and also in the context of space-time a-posteriori discretization error estimation. These applications require access to the entire approximate solution over the entire time horizon $(0,T)$ of the IBVP. Space-time discretizations of discontinuous Galerkin (dG) type also fall in this rubric of time-marching schemes ([26, 30, 36]). These *hp*-methods leverage *exponential convergence rates* of *hp*-time discretizations typically afforded by the *analytic semigroup* property of the solution operator of the parabolic IBVP.

Although linear parabolic evolution equations such as (1.1) are long known to be well-posed as operator equations in suitable Bochnerian function spaces on the space-time cylinder Q (e.g. [12, 31, 32]), the nonsymmetry and anisotropy of the parabolic operator $\partial_t + \mathcal{L}_x$, where $\mathcal{L}_x$ denotes a positive, second-order linear elliptic divergence form differential operator such as $-\Delta_x$ have obstructed development of stable space-time variational discretizations. Progress towards stable space-time discretizations has been made in [31], where the well-posedness of linear, parabolic initial-boundary value problems in Q as a well-posed operator-equation in suitable (in general fractional-order) Bochner–Sobolev spaces in Q has been leveraged in connection with adaptive wavelet methods in Q. The stability of the continuous parabolic operator $\partial_t + \mathcal{L}_x$ and Riesz bases in x and in t in the corresponding function spaces were shown in [31] to afford stable, adaptive Galerkin algorithms, with certain optimality properties. These properties imply that these algorithms produce finite-parametric approximations of the solution u on Q which converge at (essentially optimal) best n-term rates provided by suitable Besov-scale smoothness of solutions u in Q. In addition, the Riesz basis properties of the multiresolution analyses in Q enable optimal preconditioning of the resulting finite-parametric approximations of the IBVP. A wide range of linear and nonlinear parabolic IVBPs has been shown to admit corresponding well-posedness results (e.g. [31, 32] and the references there). One obstruction to the wide applicability of such space-time adaptive wavelet schemes is the need to build spline-wavelet bases in possibly complicated, polytopal spatial domains Ω with Riesz basis properties in a scale of Sobolev spaces in Ω. Recent effort has therefore been spent to develop efficient numerical methods for compressive or sparse discretizations of space-time solvers of IBVPs. We mention [35] for linear parabolic IBVPs and [1] for the acoustic wave equation in polygons.

1.3. **Contributions.** We consider sparse space-time discretizations of IBVPs for the heat equation, with low-order discretizations in the temporal variable t and the spatial variable x, with N_t temporal degrees of freedom in t and N_x degrees of freedom in the spatial domain. We adopt a nonlocal, space-time variational formulation of the parabolic IBVP which was recently proposed and investigated in [33, 34]. As in [35], we employ spline-wavelet bases for the temporal discretization and one-scale Galerkin finite elements in the spatial variable. The fractional-order operator, comprising a modified Hilbert transform $\mathcal{H}_T$ which appears in the discretization of the temporal variable, implies dense matrices with N_t^2 entries resulting from the temporal (Petrov–)Galerkin discretization with N_t many degrees of freedom. Leveraging the (by now classical) theory of *wavelet compression of pseudodifferential operators* developed in the 90ies in e.g. [7, 29] and the references there, we prove here that the matrices of the discrete, nonlocal evolution operators can, in suitable, so-called *spline-wavelet bases* resp. multiresolution analyses be *compressed to $O(N_t)$ nonvanishing entries without loss of consistency orders.*

While $\mathcal{H}_T$ is not a classical pseudodifferential operator, we prove that, nevertheless, in suitable temporal spline-wavelet bases subject to causal boundary conditions which we construct of any fixed, given polynomial order, with a corresponding number of vanishing moments, will imply *optimal compressibility of the temporal stiffness and mass matrices* of Petrov–Galerkin discretizations in these bases. To this end, we verify here that derivatives of the distributional kernel $K(s,t)$ of $\mathcal{H}_T$ satisfy so-called Calderón–Zygmund estimates, which are at the heart of wavelet compression analysis of classical pseudodifferential operators in, e.g., [7] and the references there. Combined with local analyticity estimates of $K(s,t)$ and the corresponding (exponentially convergent) quadrature techniques from [39], we develop here efficient, fully discrete algorithms for the $O(N_t)$ computation, storage and application of compressed Petrov–Galerkin discretizations of $\mathcal{H}_T$.

The computational methodology proposed here is based on multiresolution Galerkin discretizations of $\mathcal{H}_T$ in the time interval $I = (0, T)$ by biorthogonal spline-wavelets. We construct concrete multiresolution analyses (MRAs) for which most matrix entries in the resulting (dense) Galerkin matrix are numerically negligible. I.e., they can be replaced by zero without compromising the overall accuracy of the Galerkin discretization. The number of relevant, nonzero matrix coefficients (which are located in a-priori known matrix entries) scales linearly with N_t, the number of degrees of freedom, while discretization error accuracy of the underlying Galerkin scheme is retained. Due to the piecewise polynomial structure of the MRA, the analyticity of the kernel function of $\mathcal{H}_T$ can be leveraged in exponentially convergent numerical quadrature proposed in [38].

Based on the sparsity pattern of the wavelet compressed system matrix, a fill-in reducing reordering of the matrix entries by means of nested dissection is employed, see [13, 24]. As firstly proposed and demonstrated in [18], this reordering in turn allows for the rapid inversion of the system matrix by the Cholesky decomposition or more generally by a nested-dissection version of the LU decomposition. As no additional approximation errors are introduced, this is a major difference to other approaches for the discretization and the arithmetics of nonlocal operators, e.g. by means of hierarchical matrices. As the hierarchical matrix format is not closed under arithmetic operations, a recompression step after each arithmetic (block) operation has to be performed, which results in accumulating and hardly controllable consistency errors for matrix factorizations, see [16, 17].

The efficient iterative solution of the (large) linear system of equations resulting from sparse tensor-product space-time Galerkin discretizations requires a suitable preconditioner. We apply here the standard BPX scheme in the spatial domain combined with sparse, direct inversion of the wavelet-compressed stiffness and mass matrices for the first-order, temporal derivative. That way, we arrive at an iterative solver which essentially requires a fixed number of iterations to achieve a prescribed accuracy.

Having available a hierarchical basis like wavelets in the temporal variable, we can simply build *sparse tensor-product spaces* which use standard finite elements in space. Hence, we can apply sparse tensor-product spaces to approximate the heat equation. From an asymptotic point of view, the time is then for free. We show how to modify the current implementation for the discretization with respect to full tensor-product space, using only ingredients that have already been used for the full space-time approximation. Our algorithm admits sparse space-time Galerkin approximation for linear parabolic evolution equations. Unlike earlier works (e.g. [26, 32] and the references there), the presently proposed discretization handles low spatial and temporal regularity, is non-adaptive, with the number of degrees of freedom, work and memory scaling essentially as those for a multilevel solve of one stationary, elliptic boundary value problem. It allows a numerical solution of essentially optimal complexity in terms of the number of degrees of freedom. The fractional temporal order variational formulation is facilitated by a nonlocal, $H^{1/2}(0, T)$-type duality pairing with forward/backward causality, which is realized numerically by an optimal, $O(N_t)$ compression and preconditioning based on spline-wavelets in the time-domain.

We develop theory and compression estimates for $T = 1$. All results in the present article generalize to arbitrary, finite time horizon $0 < T < \infty$ by scaling.

1.4. **Notation.** We denote by $\mathbb{N} = \{1, 2, \dots\}$ the natural numbers, and set $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. With the Lipschitz domain Ω, we shall denote a bounded, polytopal subset of the Euclidean space $\mathbb{R}^n$ where, mostly, $n = 2$, and with $I = (0, 1)$ the time interval. With $Q = \Omega \times I$, we shall denote the space-time cylinder. We shall use in various places tensor-products $\otimes$ of spaces. For finite-dimensional spaces, $\otimes$ shall always denote the algebraic tensor-product. In the infinite-dimensional case, the symbol $\otimes$ shall denote the Hilbertian ("weak", or w_2) tensor-product of separable, real Hilbert spaces (see [27]). For a countable set $\mathcal{S}$, $|\mathcal{S}|$ shall denote the number of elements in $\mathcal{S}$ whenever this number is finite. All function spaces are real-valued, and dualities are with respect to $\mathbb{R}$ throughout. We shall denote duality between Sobolev spaces in the temporal domain I and the spatial domain Ω by a $'$. We identify $L^2(I) \simeq L^2(I)'$ and $L^2(\Omega) \simeq L^2(\Omega)'$. Further, we denote by $\langle \cdot, \cdot \rangle_I$ the duality pairing in $[H^{1/2}_{,0}(I)]'$ and $H^{1/2}_{,0}(I)$ as extension of the inner product in $L^2(I)$. Lastly, for $s \in \mathbb{R}$, $s > 0$, the Hilbert spaces $H^s(I)$ are the usual Sobolev spaces, endowed with their usual norms $\|\cdot\|_{H^s(I)}$.

1.5. **Layout.** The outline of this article is as follows. In Section 2, we introduce the space-time variational formulation of the heat equation. The Galerkin discretization of this variational formulation in full tensor-product spaces is presented in Section 3. In order to arrive at a data-sparse representation of the temporal system matrices, biorthogonal spline-wavelets are defined in Section 4, enabling the wavelet matrix compression of the temporal matrices defined in Section 5. Respective numerical experiments are performed in Section 6. In Section 7, we show that a sparse tensor-product discretization is much more efficient than the full tensor-product approximation introduced before. Finally, concluding remarks are stated in Section 8 while the coefficients of the wavelets used are given in the Appendix A.

2. Space-time variational formulation

We present the space-time variational formulations upon which our Galerkin discretizations will be based. They are based on fractional-order Sobolev spaces on the finite time interval, which we first introduce. Without loss of generality, we assume that $T = 1$.

2.1. **Sobolev spaces on $I = (0, 1)$.** In this subsection, we recall the Sobolev spaces on intervals. Without loss of generality, we stay with the unit interval $I = (0, 1)$. The Lebesgue space $L^2(I)$ is endowed with the usual norm $\|\cdot\|_{L^2(I)}$, whereas the Sobolev space $H^1(I)$ is equipped with the norm $\left(\|\cdot\|^2_{L^2(I)} + \|\partial_t \cdot\|^2_{L^2(I)}\right)^{1/2}$. Due to Poincaré inequalities, the subspaces

$$H^1_{0,}(I) = \{z \in H^1(I) |\ z(0) = 0\}$$

and

$$H^1_{,0}(I) = \{z \in H^1(I) |\ z(1) = 0\}$$

are endowed with the norm $\|\partial_t \cdot\|_{L^2(I)}$. We also define fractional-order Sobolev spaces by interpolation [23, Chapter 1], i.e., with the notation of [23]

$$H^s_{0,}(I) := [H^1_{0,}(I), L^2(I)]_s$$

and

$$H^s_{,0}(I) := [H^1_{,0}(I), L^2(I)]_s$$

for $s \in [0,1]$. Here, for the Sobolev space $H^s_{0,}(I)$, $s \in [0,1]$, we consider the interpolation norm

$$\|z\|_{H^s_{0,}(I)} = \left(\sum_{\ell=0}^{\infty} \lambda_\ell^s \, |z_\ell|^2 \right)^{1/2}, \quad z \in H^s_{0,}(I),$$

with the expansion coefficients $z_\ell = \int_0^1 z(t) V_\ell(t) \mathrm{d}t$. This norm coincides with the norm induced on $H^s_{0,}(I) := [H^1_{0,}(I), L^2(I)]_s$ through the real method of interpolation (see [23, Theorem 15.1 on page 98]).

For $\ell \in \mathbb{N}_0$, we designate eigenfunctions $V_\ell(t) = \sqrt{2} \sin\left(\left(\frac{\pi}{2} + \ell\pi\right) t\right)$ and eigenvalues $\lambda_\ell = \frac{\pi^2}{4}(2\ell+1)^2$ of the eigenvalue problem

$$-\partial_{tt} V_\ell = \lambda_\ell V_\ell \quad \text{in } I, \quad V_\ell(0) = \partial_t V_\ell(1) = 0, \quad \|V_\ell\|_{L^2(I)} = 1,$$

see [37, Subsection 3.4.1] for details. Analogously, for $H^s_{,0}(I)$, $s \in [0,1]$, we consider the interpolation norm

$$\|w\|_{H^s_{,0}(I)} = \left(\sum_{\ell=0}^{\infty} \hat{\lambda}_\ell^s \, |w_\ell|^2 \right)^{1/2}, \quad w \in H^s_{,0}(I),$$

with the coefficients $w_\ell = \int_0^1 w(t) W_\ell(t) \mathrm{d}t$. Here, $W_\ell(t) = \sqrt{2} \cos\left(\left(\frac{\pi}{2} + \ell\pi\right) t\right)$ and $\hat{\lambda}_\ell = \frac{\pi^2}{4}(2\ell+1)^2$ are eigenpairs of

$$-\partial_{tt} W_\ell = \hat{\lambda}_\ell W_\ell \quad \text{in } I, \quad \partial_t W_\ell(0) = W_\ell(1) = 0, \quad \|W_\ell\|_{L^2(I)} = 1.$$

Note that $\lambda_\ell = \hat{\lambda}_\ell$ for all $\ell \in \mathbb{N}_0$.

2.2. **Bochner–Sobolev spaces on the space-time cylinder Q.** Let $\Omega \subset \mathbb{R}^n$, $n \in \{1,2,3\}$, be a bounded, polytopal Lipschitz domain. We equip the Lebesgue space $L^2(\Omega)$ with the usual norm $\|\cdot\|_{L^2(\Omega)}$ and the Sobolev space $H^1_0(\Omega)$, fulfilling homogeneous Dirichlet conditions on $\partial\Omega$, with the norm $\|\cdot\|_{H^1_0(\Omega)} = |\cdot|_{H^1(\Omega)} = \|\nabla_x \cdot\|_{L^2(\Omega)^n}$. More generally, we divide the boundary $\partial\Omega$ into a Dirichlet boundary Γ_D and a Neumann boundary Γ_N, where each boundary part consists of a union of $n_\mathrm{D} \in \mathbb{N}$ resp. $n_\mathrm{N} \in \mathbb{N}_0$ faces of the spatial domain Ω. Note that, by assumption, we have $n_\mathrm{D} > 0$, $n_\mathrm{N} \geq 0$. Thus, $|\Gamma_\mathrm{D}| > 0$ if $n = 2,3$ or $\Gamma_\mathrm{D} \neq \emptyset$ if $n = 1$, i.e. there is a nonempty Dirichlet part. With this notation, the space $H^1_{\Gamma_\mathrm{D}}(\Omega)$ consists of all functions in the Sobolev space $H^1(\Omega)$, fulfilling homogeneous Dirichlet conditions on Γ_D, and is equipped with the norm $\|\cdot\|_{H^1_{\Gamma_\mathrm{D}}(\Omega)} = |\cdot|_{H^1(\Omega)} = \|\nabla_x \cdot\|_{L^2(\Omega)^n}$. Further, for $s \in \mathbb{R}$, $s \geq 0$, the Hilbert spaces $H^s(\Omega)$ are the usual Sobolev spaces endowed with their usual norms $\|\cdot\|_{H^s(\Omega)}$.

For the space-time cylinder $Q = \Omega \times (0,1)$, we extend the Sobolev spaces on intervals of Subsection 2.1 to vector-valued Bochner–Sobolev spaces, e.g.,

$$\begin{aligned}
L^2((0,1); H^1_{\Gamma_\mathrm{D}}(\Omega)) &\simeq H^1_{\Gamma_\mathrm{D}}(\Omega) \otimes L^2(0,1), \\
H^s_{0,}((0,1); L^2(\Omega)) &\simeq L^2(\Omega) \otimes H^s_{0,}(0,1) \quad \text{for } s \in [0,1], \\
H^s_{,0}((0,1); L^2(\Omega)) &\simeq L^2(\Omega) \otimes H^s_{,0}(0,1) \quad \text{for } s \in [0,1],
\end{aligned}$$

where $\otimes$ signifies the Hilbertian tensor-product of separable Hilbert spaces, see e.g. [37, Section 2.4] for details. Last, we introduce the intersection spaces

$$H^{1,1/2}_{\Gamma_{\mathrm{D}};0,}(Q) := \big(H^1_{\Gamma_{\mathrm{D}}}(\Omega)\otimes L^2(0,1)\big) \cap \big(L^2(\Omega)\otimes H^{1/2}_{0,}(0,1)\big),$$
$$H^{1,1/2}_{\Gamma_{\mathrm{D}};,0}(Q) := \big(H^1_{\Gamma_{\mathrm{D}}}(\Omega)\otimes L^2(0,1)\big) \cap \big(L^2(\Omega)\otimes H^{1/2}_{,0}(0,1)\big),$$

equipped with the sum norm, and the duality pairing

$$\langle\cdot,\cdot\rangle_Q : \big[H^{1,1/2}_{\Gamma_{\mathrm{D}};,0}(Q)\big]' \times H^{1,1/2}_{\Gamma_{\mathrm{D}};,0}(Q) \to \mathbb{R} \tag{2.1}$$

as continuous extension of the $L^2(Q)$ inner product.

2.3. **Well-posedness and isomorphism results.** In this subsection, we state the space-time variational setting for the IBVP (1.1). First, we introduce the spatial operators $\mathcal{L}_x$, $\gamma_{x,0}$, $\gamma_{x,1}$, occurring in IBVP (1.1). We assume that the spatial differential operator $\mathcal{L}_x$ is linear, self-adjoint, and in divergence form, i.e.

$$\mathcal{L}_x = -\nabla_x \cdot (A(\cdot)\nabla_x). \tag{2.2}$$

Here, the diffusion coefficient $A \in L^\infty(\Omega;\mathbb{R}^{n\times n}_{\mathrm{sym}})$ is a symmetric and positive definite matrix function $A(x)$ of $x\in\Omega$, which does not depend on the temporal variable t. Further, we assume uniform positive definiteness of A:

$$\operatorname*{ess\,inf}_{x\in\Omega} \inf_{0\neq\xi\in\mathbb{R}^n} \frac{\xi^\intercal A(x)\xi}{\xi^\intercal\xi} > 0\ . \tag{2.3}$$

Next, $\gamma_{x,0}$ is the spatial Dirichlet trace operator, which is given by $\gamma_{x,0}(v) = v|_{\partial\Omega\times I}$ for a continuous function v on $\overline{Q}$. Moreover, $\gamma_{x,1}$ denotes the spatial conormal trace map, given for a sufficiently smooth function v by $\gamma_{x,1}(v) = n_x\cdot(A(\cdot)\nabla_x v)|_{\partial\Omega\times I}$ with the exterior unit normal vector $n_x \in L^\infty(\partial\Omega;\mathbb{R}^n)$ of Ω.

The space-time variational formulation of IBVP (1.1) is to find $u \in H^{1,1/2}_{\Gamma_{\mathrm{D}};0,}(Q)$ such that

$$\forall w \in H^{1,1/2}_{\Gamma_{\mathrm{D}};,0}(Q): \quad b(u,w) = \langle f, w\rangle_Q, \tag{2.4}$$

where $f \in \big[H^{1,1/2}_{\Gamma_{\mathrm{D}};,0}(Q)\big]'$ is given. Here, $b(\cdot,\cdot)\colon H^{1,1/2}_{\Gamma_{\mathrm{D}};0,}(Q)\times H^{1,1/2}_{\Gamma_{\mathrm{D}};,0}(Q)\to\mathbb{R}$ is the continuous bilinear form given by

$$b(u,w) := \langle \partial_t u, w\rangle_Q + \langle A\nabla_x u, \nabla_x w\rangle_{L^2(Q)^n}. \tag{2.5}$$

Proceeding as in [33, Theorem 3.2] or [37, Theorem 3.4.19], the space-time variational formulation (2.4) has a unique solution. We summarize this result in the following theorem.

Theorem 2.1. *Assume that* $|\Gamma_{\mathrm{D}}| > 0$ *if* $n = 2,3$ *or* $\Gamma_{\mathrm{D}} \neq \emptyset$ *if* $n = 1$, *and that the spatial differential operator* $\mathcal{L}_x$ *in* (2.2) *with the diffusion coefficient* $A \in L^\infty(\Omega;\mathbb{R}^{n\times n}_{\mathrm{sym}})$ *satisfies* (2.3).

Then, the space-time variational formulation (2.4) *of IBVP* (1.1) *is uniquely solvable, and induces an isomorphism* $\partial_t + \mathcal{L}_x \in \mathcal{L}_{\mathrm{iso}}\big(H^{1,1/2}_{\Gamma_{\mathrm{D}};0,}(Q), \big[H^{1,1/2}_{\Gamma_{\mathrm{D}};,0}(Q)\big]'\big)$.

Remark 2.2. *Inhomogeneous Neumann boundary conditions on* Γ_{N} *can be incorporated in a weak sense, resulting in an additional term on the right-hand side of the variational formulation* (2.4).

Inhomogeneous Dirichlet boundary conditions on Γ_{D} *and initial conditions can be treated by using inverse trace theorems and by superposition since the IBVP* (1.1) *is linear.*

3. Disrectization

We present finite-dimensional subspaces of the Bochnerian function spaces on Q which shall be used in the ensuing space-time Galerkin discretizations at the end of this section. As all subspaces are algebraic tensor-products of subspaces on $I = (0, T)$ and on Ω, we introduce the factor spaces separately. Without loss of generality, we assume $T = 1$, i.e. $I = (0, 1)$.

3.1. **Tensor-product trial spaces on $Q = \Omega \times I$.** In this subsection, we introduce the temporal and spatial finite element spaces as well as their tensor-product.

3.1.1. *Temporal trial space.* For a refinement level $j \in \mathbb{N}_0$ and given $N_j^t \in \mathbb{N}$, consider in $I = (0, 1)$ the temporal uniform mesh

$$\mathcal{T}_j^t := \{\tau_{j,\ell}^t \subset I : \ \tau_{j,\ell}^t = (t_{j,\ell-1}, t_{j,\ell}) \text{ for } \ell = 1, \ldots, N_j^t\}$$

given by $t_{j,\ell} = \ell / N_j^t$ for $\ell = 0, \ldots, N_j^t$ with uniform mesh size $h_j^t = 1/N_j^t$. With the temporal mesh $\mathcal{T}_j^t$, we associate the trial space

$$S_j^t = \{v_j^t \in C(\overline{I}) : \forall \tau \in \mathcal{T}_t^j : v_j^t|_{\overline{\tau}} \in \mathbb{P}^{d-1}(\overline{\tau}) \text{ and } v_j^t(0) = 0\}$$

of globally continuous, piecewise polynomial functions, fulfilling the homogeneous initial condition. Here, $\mathbb{P}^{d-1}(A)$ denotes the space of polynomials of order d on a set $A \subset \mathbb{R}^m$, $m \in \mathbb{N}$. For $d = 2$, we obtain the usual hat functions $\phi_{j,k}^t : \overline{I} \to \mathbb{R}$, $k = 1, \ldots, N_j^t$, fulfilling $\phi_{j,k}^t(t_{j,\ell}) = \delta_{k,\ell}$ for $\ell = 0, \ldots, N_j^t$ with the Kronecker delta $\delta_{k,\ell}$. For $d > 2$, the present setting includes finite elements of higher order and B-splines. Note that in case of B-splines appropriate wavelet bases used later for matrix compression on are available.

3.1.2. *Spatial trial space.* To define the spatial trial space in Ω, we assume $\Omega \subset \mathbb{R}^n$, $n \in \{1, 2, 3\}$, to be a given bounded Lipschitz domain, which is an interval $\Omega = (0, L)$ with $L > 0$ for $n = 1$, polygonal for $n = 2$, or polyhedral for $n = 3$. For a refinement level $j \in \mathbb{N}_0$, we decompose Ω into an admissible decomposition

$$\mathcal{T}_j^x := \{\tau_{j,i}^x \subset \mathbb{R}^n : i = 1, \ldots, n_j^x\},$$

where n_j^x is the number of spatial elements $\tau_{j,i}^x$. We denote by $h_{j,i}^x$, $i = 1, \ldots, n_j^x$, the local mesh size and $h_j^x = \max_i h_{j,i}^x$ is the maximal mesh size. To a shape-regular sequence $(\mathcal{T}_j^x)_{j \in \mathbb{N}_0}$, we relate the spatial finite element spaces

$$S_j^x = \left\{w_j^x \in C(\overline{\Omega}) : \forall \tau \in \mathcal{T}_j^x : w_j^x|_{\overline{\tau}} \in \mathbb{P}^{p_x}(\overline{\tau}) \text{ and } w_j^x|_{\Gamma_{\mathrm{D}}} = 0\right\}$$

of functions, which are globally continuous, piecewise polynomial functions of degree $p_x \geq 1$ and fulfill the homogeneous Dirichlet condition on Γ_{D}. We denote by $\phi_{j,k}^x : \overline{\Omega} \to \mathbb{R}$, $k = 1, \ldots, N_j^x$, basis functions of S_j^x, i.e., we have that $S_j^x = \operatorname{span}\{\phi_{j,k}^x : k = 1, \ldots, N_j^x\}$.

3.1.3. *Space-time tensor-product trial space.* The space-time trial space is obtained via the tensor-product of the trial spaces in space and time. To this end, fixing a level $j \in \mathbb{N}_0$, we define the space-time partition of Q

$$\mathcal{T}_j^{x,t} = \left\{\tau^x \times \tau^t \subset \mathbb{R}^{n+1} : \tau^x \in \mathcal{T}_j^x, \, \tau^t \in \mathcal{T}_j^t\right\}$$

with space-time mesh size $h_j = \max\{h_j^x, h_j^t\}$. The respective space-time trial space is given by

$$S_j^{x,t} = S_j^x \otimes S_j^t \subset H^{1,1/2}_{\Gamma_{\mathrm{D}};0,}(Q),$$

fulfilling the homogeneous Dirichlet and initial conditions. A basis is via dyads of the basis functions in space and time (being here equivalent to pointwise products), that is

$$S_j^{x,t} = \operatorname{span}\left\{\phi_{j,k}^x \otimes \phi_{j,\ell}^t : k = 1, \ldots, N_j^x,\ \ell = 1, \ldots, N_j^t\right\}.$$

3.2. **Modified Hilbert transform.** We detail next the modified Hilbert transform $\mathcal{H}_T$, which was introduced in [33, 37], and recall its main properties. It is an essential ingredient in a stable Petrov–Galerkin discretization of the variational formulation (2.4) for the time derivative ∂_t. We only recap definitions and analytic properties. For details and proofs, we refer to [25, 33, 37, 38, 39].

For a given function $z \in L^2(I)$ with Fourier coefficients

$$z_k = \sqrt{2} \int_0^1 z(t) \sin\left(\left(\frac{\pi}{2} + k\pi\right) t\right) \mathrm{d}t = \int_0^1 z(t) V_k(t) \mathrm{d}t$$

and the series representation

$$z(t) = \sum_{k=0}^{\infty} z_k \sqrt{2} \sin\left(\left(\frac{\pi}{2} + k\pi\right) t\right) = \sum_{k=0}^{\infty} z_k V_k(t), \quad t \in (0,1),$$

the modified Hilbert transform is defined by the series

$$(\mathcal{H}_T z)(t) = \sum_{k=0}^{\infty} z_k \sqrt{2} \cos\left(\left(\frac{\pi}{2} + k\pi\right) t\right) = \sum_{k=0}^{\infty} z_k W_k(t), \quad t \in (0,1). \tag{3.1}$$

Here, V_k and W_k are eigenfunctions of corresponding eigenvalue problems given in Section 2.1. For $s \in [0,1]$, the mapping

$$\mathcal{H}_T \colon H_{0,}^s(I) \to H_{,0}^s(I)$$

is an isometry. This follows from

$$\|z\|_{H_{0,}^s(I)}^2 = \sum_{\ell=0}^{\infty} \lambda_\ell^s \left|z_\ell\right|^2 = \sum_{\ell=0}^{\infty} \hat{\lambda}_\ell^s \left|z_\ell\right|^2 = \|\mathcal{H}_T z\|_{H_{,0}^s(I)}^2, \quad z \in H_{0,}^s(I).$$

We also have

$$\forall v, w \in H_{0,}^{1/2}(I) : \langle \partial_t v, \mathcal{H}_T w \rangle_I \le \|v\|_{H_{0,}^{1/2}(I)} \|w\|_{H_{0,}^{1/2}(I)}, \tag{3.2}$$

see [37, Eq. (3.40)]. Further, the properties

$$\forall v \in H_{0,}^{1/2}(I) : \langle \partial_t v, \mathcal{H}_T v \rangle_I = \|v\|_{H_{0,}^{1/2}(I)}^2 \tag{3.3}$$

and

$$\forall v \in L^2(I) : \langle v, \mathcal{H}_T v \rangle_{L^2(I)} \ge 0 \tag{3.4}$$

proven in [33, Corollary 2.5] and [33, Lemma 2.6] hold true. Adapting the proof of [33, Lemma 2.6], we even have

$$\forall s \in (0,1] : \forall v \in H_{0,}^s(I) : \langle v, \mathcal{H}_T v \rangle_{L^2(I)} > 0,$$

which is proven in [25, Lemma 2.2].

In addition, we have that

$$\forall u, w \in L^2(I) : \langle \mathcal{H}_T u, w \rangle_{L^2(I)} = \left\langle u, \mathcal{H}_T^{-1} w \right\rangle_{L^2(I)}$$

by [37, Lemma 3.4.7] and

$$\forall u \in H_{0,}^1(I) : \forall v \in L^2(I) : \langle \partial_t \mathcal{H}_T u, v \rangle_{L^2(I)} = -\left\langle \mathcal{H}_T^{-1} \partial_t u, v \right\rangle_{L^2(I)}$$

by applying the same arguments as in the proof of [33, Lemma 2.3].

The modified Hilbert transform $\mathcal{H}_T$ allows for integral representations, see [34, 39], which is the motivation for the wavelet compression, originally introduced in the context of boundary integral equations. To present these, we recall that we considered without loss of generality $T = 1$ (upon a dilation $t \mapsto t/T$).

Lemma 3.1 (Lemma 2.1 in [34])**.** *For $v \in L^2(I)$, the modified Hilbert transform $\mathcal{H}_T$, defined in* (3.1)*, allows the integral representation*

$$(\mathcal{H}_T v)(t) = \text{v.p.} \int_0^1 K(s,t)\, v(x) \mathrm{d}s, \quad t \in (0,1),$$

as a Cauchy principal value integral, where the kernel function is given as

$$K(s,t) := \frac{1}{2} \left[\frac{1}{\sin \frac{\pi(s+t)}{2}} + \frac{1}{\sin \frac{\pi(s-t)}{2}} \right]. \tag{3.5}$$

For $v \in H^1(I)$, the operator $\mathcal{H}_T$ allows the integral representation

$$(\mathcal{H}_T v)(t) = -\frac{2}{\pi} v(0) \ln \tan \frac{\pi t}{4} + \int_0^1 K_{-1}(s,t)\, \partial_s v(s) \mathrm{d}s, \quad t \in (0,1),$$

as a weakly singular integral, where

$$K_{-1}(s,t) := -\frac{1}{\pi} \ln \left[\tan \frac{\pi(s+t)}{4} \tan \frac{\pi\, |t-s|}{4} \right]. \tag{3.6}$$

Remark 3.2. *We adopt the convention (with slight abuse of notation) that $\mathcal{H}_T$ applied to functions in Bochner spaces $H^s_{0,}(I;H)$ taking values in a separable Hilbert space H, i.e. $\mathcal{H}_T \otimes \mathrm{Id}_H$, is denoted again by $\mathcal{H}_T$.*

3.3. **Conforming space-time Galerkin discretization.** We shall next introduce the conforming space-time Galerkin discretization of the parabolic evolution equation (1.1) by employing the modified Hilbert transform $\mathcal{H}_T$, which leads to the $H^{1/2}_{0,}(I;L^2(\Omega))$-ellipticity of the bilinear form $b(\cdot,\cdot)$ in (2.5). More precisely, tensorizing (3.3) and (3.4) with $L^2(\Omega)$ we have that (cp. (2.1))

$$\begin{aligned} \|v\|^2_{H^{1/2}_{0,}(I;L^2(\Omega))} &= \langle \partial_t v, \mathcal{H}_T v \rangle_Q \\ &\le \langle \partial_t v, \mathcal{H}_T v \rangle_Q + \langle A \nabla_x v, \nabla_x \mathcal{H}_T v \rangle_{L^2(Q)^n} = b(v, \mathcal{H}_T v) \end{aligned} \tag{3.7}$$

for all $v \in H^{1,1/2}_{\Gamma_\mathrm{D};0,}(Q)$.

Then, for any refinement level $j \in \mathbb{N}_0$, the full tensor-product space-time Galerkin finite element method to find $u_j \in S_j^{x,t} \subset H^{1,1/2}_{\Gamma_\mathrm{D};0,}(Q)$ such that

$$\forall v_j \in S_j^{x,t} : \quad b(u_j, \mathcal{H}_T v_j) = \langle f, \mathcal{H}_T v_j \rangle_Q \tag{3.8}$$

is well-defined, with $b(\cdot,\cdot)$ as in (2.5). The tensor-product space-time Galerkin projection $u_j \in S_j^{x,t}$ in (3.8) is unconditionally stable with the space-time stability estimate

$$\|u_j\|_{H^{1/2}_{0,}(I;L^2(\Omega))} \le \|f\|_{[H^{1/2}_{,0}(I;L^2(\Omega))]'},$$

provided that $f \in [H^{1/2}_{,0}(I;L^2(\Omega))]'$, see [37, Theorem 3.4.20].

Theorem 3.3. *Let $j \in \mathbb{N}_0$ be a given refinement level. In addition, let $u \in H^{1,1/2}_{\Gamma_D;0,}(Q)$ be the unique solution of the space-time variational formulation (2.4). Suppose the regularity assumption $\mathcal{L}_x u \in L^2(Q)$ holds. Let $u_j \in S^{x,t}_j$ be the space-time approximation given by (3.8).*

Then, the space-time error estimate

(3.9)
$$\|u-u_j\|_{H^{1/2}_{0,}(I;L^2(\Omega))} \leq 2\|u-P_ju\|_{H^{1/2}_{0,}(I;L^2(\Omega))}+\left\|(\mathrm{Id}_{xt}-P^t_j\otimes \mathrm{Id}_x)\mathcal{L}_x u\right\|_{[H^{1/2}_{,0}(I;L^2(\Omega))]'}$$

holds true with $P_j u := (P^t_j \otimes P^x_j)u \in S^{x,t}_j$, where $P^t_j \colon L^2(I) \to S^t_j$ is any temporal projection and P^x_j denotes the spatial Ritz projection $P^x_j \colon H^1_{\Gamma_D}(\Omega) \to S^x_j$. It is defined as follows: given $z \in H^1_{\Gamma_D}(\Omega)$, find $P^x_j z \in S^x_j$ such that

$$\forall w_j \in S^x_j : \quad \langle A\nabla_x P^x_j z, \nabla_x w_j\rangle_{L^2(\Omega)^n} = \langle A\nabla_x z, \nabla_x w_j\rangle_{L^2(\Omega)^n}. \tag{3.10}$$

Proof. The proof is a generalization of the proof of [33, Theorem 3.4] or [37, Theorem 3.4.24]. The triangle inequality yields

$$\|u-u_j\|_{H^{1/2}_{0,}(I;L^2(\Omega))} \leq \|u-P_ju\|_{H^{1/2}_{0,}(I;L^2(\Omega))} + \|u_j-P_ju\|_{H^{1/2}_{0,}(I;L^2(\Omega))}.$$

The last term is estimated as follows.

Using the $H^{1/2}_{0,}(I;L^2(\Omega))$-ellipticity of the bilinear form $b(\cdot,\cdot)$ in (3.3) and (3.4) with the Galerkin orthogonality corresponding to (3.8) give

$$\begin{aligned}
\|u_j-P_ju\|^2_{H^{1/2}_{0,}(I;L^2(\Omega))} &\leq b\big(u_j-P_ju, \mathcal{H}_T(u_j-P_ju)\big)\\
&= b\big(u-P_ju, \mathcal{H}_T(u_j-P_ju)\big)\\
&= \langle \partial_t(u-P_ju), \mathcal{H}_T(u_j-P_ju)\rangle_Q\\
&\quad + \langle A\nabla_x u, \nabla_x \mathcal{H}_T(u_j-P_ju)\rangle_{L^2(Q)^n}\\
&\quad - \langle A\nabla_x (P^t_j\otimes P^x_j)u, \nabla_x \mathcal{H}_T(u_j-P_ju)\rangle_{L^2(Q)^n}\ .
\end{aligned}$$

From property (3.2) of $\mathcal{H}_T$ and the Ritz projection property (3.10) of P^x_j, it follows from the preceding bound with integration by parts in Ω that

$$\begin{aligned}
\|u_j-P_ju\|^2_{H^{1/2}_{0,}(I;L^2(\Omega))} &\leq \|u-P_ju\|_{H^{1/2}_{0,}(I;L^2(\Omega))}\|u_j-P_ju\|_{H^{1/2}_{0,}(I;L^2(\Omega))}\\
&\quad + \langle A\nabla_x u, \nabla_x \mathcal{H}_T(u_j-P_ju)\rangle_{L^2(Q)^n}\\
&\quad - \langle A\nabla_x (P^t_j\otimes \mathrm{Id}_x)u, \nabla_x \mathcal{H}_T(u_j-P_ju)\rangle_{L^2(Q)^n}\\
&= \|u-P_ju\|_{H^{1/2}_{0,}(I;L^2(\Omega))}\|u_j-P_ju\|_{H^{1/2}_{0,}(I;L^2(\Omega))}\\
&\quad - \langle \mathcal{L}_x u, \mathcal{H}_T(u_j-P_ju)\rangle_{L^2(Q)}\\
&\quad + \langle \big(P^t_j\otimes \mathrm{Id}_x\big)\mathcal{L}_x u, \mathcal{H}_T(u_j-P_ju)\rangle_{L^2(Q)}.
\end{aligned}$$

Duality and the isometry property of $\mathcal{H}_T$ finally yield

$$\begin{aligned}
\|u_j-P_ju\|^2_{H^{1/2}_{0,}(I;L^2(\Omega))} &\leq \|u-P_ju\|_{H^{1/2}_{0,}(I;L^2(\Omega))}\|u_j-P_ju\|_{H^{1/2}_{0,}(I;L^2(\Omega))}\\
&\quad + \left\|\big(\mathrm{Id}_{xt}-P^t_j\otimes \mathrm{Id}_x\big)\mathcal{L}_x u\right\|_{[H^{1/2}_{,0}(I;L^2(\Omega))]'}\|u_j-P_ju\|_{H^{1/2}_{0,}(I;L^2(\Omega))},
\end{aligned}$$

which implies the assertion. □

To prove convergence rates given in the forthcoming Corollary 3.4, we define the $H^{1/2}_{0,}(I)$-projection $P^{1/2,t}_j \colon H^{1/2}_{0,}(I) \to S^t_j$ by

$$\forall z_j \in S^t_j : \quad \langle \partial_t P^{1/2,t}_j v, \mathcal{H}_T z_j\rangle_{L^2(I)} = \langle \partial_t v, \mathcal{H}_T z_j\rangle_I \tag{3.11}$$

for a given function $v \in H^{1/2}_{0,}(I)$, fulfilling the error estimate

$$\|v - P^{1/2,t}_j v\|_{H^{1/2}_{0,}(I)} \le c(h^t_j)^{d-1/2}\|v\|_{H^d(I)}$$

provided that $v \in H^{1/2}_{0,}(I) \cap H^d(I)$ with a constant $c > 0$. Further, setting the function $w = \int_0^{\cdot} \mathcal{H}_T(v - P^{1/2,t}_j v) \in H^1_{0,}(I)$, i.e. an Aubin–Nitsche-type argument, yields

$$\begin{aligned}
\|v - P^{1/2,t}_j v\|^2_{L^2(I)} &= \langle \mathcal{H}_T(v - P^{1/2,t}_j v), \mathcal{H}_T(v - P^{1/2,t}_j v)\rangle_{L^2(I)} \\
&= \langle \partial_t w, \mathcal{H}_T(v - P^{1/2,t}_j v)\rangle_{L^2(I)} \\
&= \langle \partial_t (w - P^{1/2,t}_j w), \mathcal{H}_T(v - P^{1/2,t}_j v)\rangle_{L^2(I)} \\
&\le \|w - P^{1/2,t}_j w\|_{H^{1/2}_{0,}(I)} \|v - P^{1/2,t}_j v\|_{H^{1/2}_{0,}(I)} \\
&\le c(h^t_j)^{1/2}\|\partial_t w\|_{L^2(I)} \|v - P^{1/2,t}_j v\|_{H^{1/2}_{0,}(I)} \\
&= c(h^t_j)^{1/2}\|v - P^{1/2,t}_j v\|_{L^2(I)} \|v - P^{1/2,t}_j v\|_{H^{1/2}_{0,}(I)},
\end{aligned}$$

where the isometry property of $\mathcal{H}_T$, the fundamental theorem of calculus, the boundedness (3.2) and an $H^{1/2}_{0,}(I)$-error estimate of the $H^{1/2}_{0,}(I)$-projection are used. Thus, we arrive at the $L^2(I)$-error estimate

$$\|v - P^{1/2,t}_j v\|_{L^2(I)} \le c(h^t_j)^d \|v\|_{H^d(I)} \tag{3.12}$$

whenever $v \in H^{1/2}_{0,}(I) \cap H^d(I)$.

Corollary 3.4. *Let the assumptions of Theorem 3.3 be satisfied. Consider temporal biorthogonal spline-wavelets S^t_j of order $d \ge 2$ in I with uniform mesh size h^t_j and Lagrangian finite elements S^x_j of degree $p_x \ge 1$ on a regular, simplicial partition of Ω with maximal mesh size h^x_j.*

Assume also that the exact solution u of the variational IBVP (2.4) *admits the regularity $u \in H^{1/2}_{0,}(I; H^{p_x+1/2}(\Omega)) \cap H^d(I; L^2(\Omega))$ and $\mathcal{L}_x u \in H^{d-1}(I; L^2(\Omega))$. Assume finally that the spatial Ritz projection $P^x_j \colon H^1_{\Gamma_D}(\Omega) \to S^x_j$ satisfies*

$$\|u - (\mathrm{Id}_t \otimes P^x_j)u\|_{H^{1/2}_{0,}(I;L^2(\Omega))} \le c(h^x_j)^{p_x+1/2}\|u\|_{H^{1/2}_{0,}(I;H^{p_x+1/2}(\Omega))} \tag{3.13}$$

with a constant $c > 0$.

Then, the space-time error estimate

$$\|u - u_j\|_{H^{1/2}_{0,}(I;L^2(\Omega))} \le C[(h^t_j)^{d-1/2} + (h^x_j)^{p_x+1/2}]$$

holds true with a constant $C > 0$.

Proof. Let the temporal projection P^t_j be the $L^2(I)$-projection on S^t_j. Using the triangle inequality for the first part on the right-hand side of the estimate (3.9) yields

$$\begin{aligned}
\|u - P_j u\|_{H^{1/2}_{0,}(I;L^2(\Omega))} \le &\|(\mathrm{Id}_{xt} - \mathrm{Id}_t \otimes P^x_j)u\|_{H^{1/2}_{0,}(I;L^2(\Omega))} \\
&+ \|(\mathrm{Id}_{xt} - P^t_j \otimes \mathrm{Id}_x)u\|_{H^{1/2}_{0,}(I;L^2(\Omega))} \\
&+ \|(\mathrm{Id}_{xt} - P^t_j \otimes \mathrm{Id}_x)(\mathrm{Id}_{xt} - \mathrm{Id}_t \otimes P^x_j)u\|_{H^{1/2}_{0,}(I;L^2(\Omega))}.
\end{aligned}$$

For $\|(\mathrm{Id}_{xt} - \mathrm{Id}_t \otimes P_j^x)u\|_{H^{1/2}_{0,}(I;L^2(\Omega))}$, we apply estimate (3.13). For $\|(\mathrm{Id}_{xt} - P_j^t \otimes \mathrm{Id}_x)u\|_{H^{1/2}_{0,}(I;L^2(\Omega))}$, the triangle inequality upon inserting the temporal $H^{1/2}_{0,}(I)$-projection $P_j^{1/2,t}$ given in (3.11) and then using the temporal inverse inequality

$$\forall z_j \in S_j^t : \|z_j\|_{H^{1/2}_{0,}(I)} \le c_{\mathrm{I}}(h_j^t)^{-1/2}\|z_j\|_{L^2(I)}$$

give the error bound

$$\begin{aligned}
\|(\mathrm{Id}_{xt} - P_j^t \otimes \mathrm{Id}_x)u\|_{H^{1/2}_{0,}(I;L^2(\Omega))} \le{} & \|(\mathrm{Id}_{xt} - P_j^{1/2,t} \otimes \mathrm{Id}_x)u\|_{H^{1/2}_{0,}(I;L^2(\Omega))} \\
& + \|(P_j^{1/2,t} \otimes \mathrm{Id}_x - P_j^t \otimes \mathrm{Id}_x)u\|_{H^{1/2}_{0,}(I;L^2(\Omega))} \\
\le{} & c_1(h_j^t)^{d-1/2} + c_{\mathrm{I}}(h_j^t)^{-1/2}\|(P_j^{1/2,t} \otimes \mathrm{Id}_x - \mathrm{Id}_{xt})u\|_{L^2(Q)} \\
& + c_{\mathrm{I}}(h_j^t)^{-1/2}\|(\mathrm{Id}_{xt} - P_j^t \otimes \mathrm{Id}_x)u\|_{L^2(Q)} \\
\le{} & c_2(h_j^t)^{d-1/2}
\end{aligned}$$

with constants $c_1, c_2 > 0$, when $L^2(I)$-error estimates for the $L^2(I)$- and $H^{1/2}_{0,}(I)$-projection are applied, where the latter follows from the Aubin–Nitsche argument (3.12) for the $H^{1/2}_{0,}(I)$-projection. For the term $\|(\mathrm{Id}_{xt} - P_j^t \otimes \mathrm{Id}_x)(\mathrm{Id}_{xt} - \mathrm{Id}_t \otimes P_j^x)u\|_{H^{1/2}_{0,}(I;L^2(\Omega))}$, we employ the triangle inequality to get

$$\begin{aligned}
\|(\mathrm{Id}_{xt}-P_j^t\otimes\mathrm{Id}_x)(\mathrm{Id}_{xt}-\mathrm{Id}_t\otimes P_j^x)u\|_{H^{1/2}_{0,}(I;L^2(\Omega))} \le{} & \|(\mathrm{Id}_{xt}-\mathrm{Id}_t\otimes P_j^x)u\|_{H^{1/2}_{0,}(I;L^2(\Omega))} \\
& + \|(P_j^t \otimes \mathrm{Id}_x)(\mathrm{Id}_{xt} - \mathrm{Id}_t \otimes P_j^x)u\|_{H^{1/2}_{0,}(I;L^2(\Omega))}.
\end{aligned}$$

For the first term, we use error estimate (3.13). For the second term, we apply the stability of the $L^2(I)$-projection in $H^{1/2}_{0,}(I)$ and again, error estimate (3.13). With this, we derive the error bound $\|(\mathrm{Id}_{xt} - P_j^t \otimes \mathrm{Id}_x)(\mathrm{Id}_{xt} - \mathrm{Id}_t \otimes P_j^x)u\|_{H^{1/2}_{0,}(I;L^2(\Omega))} \le c(h_j^x)^{p_x+1/2}\|u\|_{H^{1/2}_{0,}(I;H^{p_x+1/2}(\Omega))}$ with a constant $c > 0$. Thus, we conclude that

$$\begin{aligned}
\|u - P_j u\|_{H^{1/2}_{0,}(I;L^2(\Omega))} \le{} & c(h_j^x)^{p_x+1/2}\|u\|_{H^{1/2}_{0,}(I;H^{p_x+1/2}(\Omega))} \\
& + c(h_j^t)^{d-1/2}\|u\|_{H^d(I;L^2(\Omega))}
\end{aligned}$$

with a constant $c > 0$.

For the second part on the right-hand side of the estimate (3.9), we use the standard error estimate of the temporal $L^2(I)$-projection with respect to $\|\cdot\|_{[H^{1/2}_{,0}(I)]'}$ to conclude that

$$\big\|(\mathrm{Id}_{xt} - P_j^t \otimes \mathrm{Id}_x)\mathcal{L}_x u\big\|_{[H^{1/2}_{,0}(I;L^2(\Omega))]'} \le c(h_j^t)^{d-1/2}\|\mathcal{L}_x u\|_{H^{d-1}(I;L^2(\Omega))}$$

with a constant $c > 0$. □

Remark 3.5. *The error estimate* (3.13) *can be proven by an Aubin–Nitsche argument and function space interpolation, provided that the unique solution $w_g \in H^1_{\Gamma_{\mathrm{D}}}(\Omega)$ of the spatial variational formulation*

$$\forall z \in H^1_{\Gamma_{\mathrm{D}}}(\Omega) : \quad \langle A\nabla_x z, \nabla_x w_g\rangle_{L^2(\Omega)^n} = \langle g, z\rangle_{L^2(\Omega)}$$

for a given $g \in L^2(\Omega)$ satisfies the regularity assumption $w_g \in H^2(\Omega)$ with estimate $\|w_g\|_{H^2(\Omega)} \le C\|g\|_{L^2(\Omega)}$ with a constant $C > 0$.

The discrete variational formulation (3.8) is equivalent to the global linear system

$$\big(\mathbf{A}_j^t \otimes \mathbf{M}_j^x + \mathbf{M}_j^t \otimes \mathbf{A}_j^x\big)\mathbf{u}_j = \mathbf{f}_j$$

with the Kronecker matrix product $\otimes$. Here, the temporal matrices given by

$$[\mathbf{A}_j^t]_{k,k'} = \big\langle \partial_t \phi_{j,k'}^t, \mathcal{H}_T \phi_{j,k}^t \big\rangle_{L^2(I)}, \quad [\mathbf{M}_j^t]_{k,k'} = \big\langle \phi_{j,k'}^t, \mathcal{H}_T \phi_{j,k}^t \big\rangle_{L^2(I)}$$

for $k, k' = 1, \ldots, N_j^t$ and the spatial matrices are

$$[\mathbf{M}_j^x]_{k,k'} = \big\langle \phi_{j,k'}^x, \phi_{j,k}^x \big\rangle_{L^2(\Omega)}, \quad [\mathbf{A}_j^x]_{k,k'} = \big\langle \nabla_x \phi_{j,k'}^x, \nabla_x \phi_{j,k}^x \big\rangle_{L^2(\Omega)^n}$$

for $k, k' = 1, \ldots, N_j^x$, where $\mathbf{f}_j$ is the related right-hand side.

The matrices $\mathbf{A}_j^t$, $\mathbf{M}_j^t$, $\mathbf{A}_j^x$, $\mathbf{M}_j^x$ are positive definite, see the properties of $\mathcal{H}_T$ for the temporal matrices. Moreover, the matrices $\mathbf{A}_j^t$, $\mathbf{A}_j^x$, $\mathbf{M}_j^x$ are symmetric, whereas $\mathbf{M}_j^t$ is nonsymmetric.

4. Wavelets and multiresolution analysis

Instead of using the single-scale basis in S_j^t for the temporal discretization, we shall switch to a wavelet basis in order to apply wavelet matrix compression for the data sparse representation of the dense matrices $\mathbf{A}_j^t$ and $\mathbf{M}_j^t$. To this end, we assume without loss of generality that $T = 1$ and consider the interval $I := (0, 1)$. Moreover, for the sake of simplicity in representation, we omit the suffix t in this section.

A *multiresolution analysis* (MRA for short) consists of a nested family of finite dimensional approximation spaces

$$\{0\} = S_{-1} \subset S_0 \subset S_1 \subset \cdots \subset S_j \subset \cdots \subset L^2(I), \tag{4.1}$$

such that

$$\overline{\bigcup_{j \geq 0} S_j} = L^2(I) \quad \text{and} \quad \dim S_j \sim 2^j.$$

We will refer to j as the *level* of S_j in the multiresolution analysis. Each space S_j is endowed with a *single-scale basis*

$$\Phi_j = \{\phi_{j,k} : k \in \Delta_j\},$$

i.e. $S_j = \operatorname{span} \Phi_j$, where Δ_j denotes a suitable index set with cardinality $|\Delta_j| = \dim S_j$. For convenience, we write in the sequel bases in the form of row vectors, such that, for $\mathbf{v} = [v_k]_{k \in \Delta_j} \in \ell^2(\Delta_j)$, the corresponding function can simply be written as a dot product according to

$$v_j = \Phi_j \mathbf{v} = \sum_{k \in \Delta_j} v_k \phi_{j,k}.$$

In addition, we shall assume that the single-scale bases Φ_j are *uniformly stable*, this means that

$$\|\mathbf{v}\|_{\ell^2(\Delta_j)} \sim \|\Phi_j \mathbf{v}\|_{L^2(I)} \quad \text{for all } \mathbf{v} \in \ell^2(\Delta_j)$$

uniformly in j, and that they satisfy the locality condition

$$\operatorname{diam}(\operatorname{supp} \phi_{j,k}) \sim 2^{-j}.$$

Additional properties of the spaces S_j are required for using them as trial spaces in a Galerkin scheme. The temporal approximation spaces shall have the *regularity*

$$\gamma := \sup\{s \in \mathbb{R} : S_j \subset H^s(I)\}$$

and the *approximation order* $d \in \mathbb{N}$, that is

$$d = \sup\Big\{s \in \mathbb{R} : \inf_{v_j \in S_j} \|v - v_j\|_{L^2(I)} \lesssim 2^{-js}\|v\|_{H^s(I)}\Big\}.$$

Rather than using the multiresolution analysis corresponding to the hierarchy in (4.1), the pivotal idea of wavelets is to keep track of the increment of information between two consecutive levels $j-1$ and j. Since we have $S_{j-1} \subset S_j$, we may decompose

$$S_j = S_{j-1} \oplus W_j, \text{ i.e. } S_{j-1} \cup W_j = S_j \text{ and } S_{j-1} \cap W_j = \{0\},$$

with an appropriate *detail space* W_j. Of practical interest is the particular choice of the basis of the detail space W_j in S_j. This basis will be denoted by

$$\Psi_j = \{\psi_{j,k} : k \in \nabla_j\}$$

with the index set $\nabla_j := \Delta_j \setminus \Delta_{j-1}$. In particular, we shall assume that the collections $\Phi_{j-1} \cup \Psi_j$ form uniformly stable bases of S_j, as well. If $\Psi = \bigcup_{j\geq 0} \Psi_j$, where $\Psi_0 := \Phi_0$, is even a Riesz-basis of $L^2(I)$, then it is called a *wavelet basis*. We require the functions $\psi_{j,k}$ to be localized with respect to the corresponding level j, i.e.

$$\operatorname{diam}(\operatorname{supp}\psi_{j,k}) \sim 2^{-j},$$

and we normalize them such that

$$\|\psi_{j,k}\|_{L^2(I)} \sim 1.$$

At first glance, it would be very convenient to deal with a single orthonormal system of wavelets. However, it has been shown in [7, 10, 29] that orthogonal wavelets are not optimal for the efficient approximation nonlocal operator equations. For this reason, we rather use *biorthogonal wavelet bases*. In this case, we also have a dual, multiresolution analysis, i.e. dual single-scale bases and wavelets

$$\widetilde{\Phi}_j = \{\widetilde{\phi}_{j,k} : k \in \Delta_j\}, \quad \widetilde{\Psi}_j = \{\widetilde{\psi}_{j,k} : k \in \nabla_j\},$$

respectively. They are biorthogonal to the primal ones: there hold the orthogonality conditions

$$\langle \Phi_j, \widetilde{\Phi}_j\rangle_{L^2(I)} = \mathbf{I}, \quad \langle \Psi_j, \widetilde{\Psi}_j\rangle_{L^2(I)} = \mathbf{I}.$$

The corresponding spaces $\widetilde{S}_j := \operatorname{span}\widetilde{\Phi}_j$ and $\widetilde{W}_j := \operatorname{span}\widetilde{\Psi}_j$ satisfy

$$S_{j-1} \perp \widetilde{W}_j, \quad \widetilde{S}_{j-1} \perp W_j. \tag{4.2}$$

Moreover, the dual spaces are supposed to exhibit some approximation order $\widetilde{d} \in \mathbb{N}$ and regularity $\widetilde{\gamma} > 0$.

Denoting in complete analogy to the primal basis $\widetilde{\Psi} = \bigcup_{j\geq 0} \widetilde{\Psi}_j$, where $\widetilde{\Psi}_0 := \widetilde{\Phi}_0$, then every $v \in L^2(I)$ has unique representations

$$v = \sum_{j\geq 0}\sum_{k\in\nabla_j} \langle v, \widetilde{\psi}_{j,k}\rangle_{L^2(I)}\psi_{j,k} = \sum_{j\geq 0}\sum_{k\in\nabla_j} \langle v, \psi_{j,k}\rangle_{L^2(I)}\widetilde{\psi}_{j,k}$$

such that

$$\|v\|_{L^2(I)}^2 \sim \sum_{j\geq 0}\sum_{k\in\nabla_j} \big\|\langle v, \widetilde{\psi}_{j,k}\rangle_{L^2(I)}\big\|_{\ell^2(\nabla_j)}^2 \sim \sum_{j\geq 0}\sum_{k\in\nabla_j} \big\|\langle v, \psi_{j,k}\rangle_{L^2(I)}\big\|_{\ell^2(\nabla_j)}^2.$$

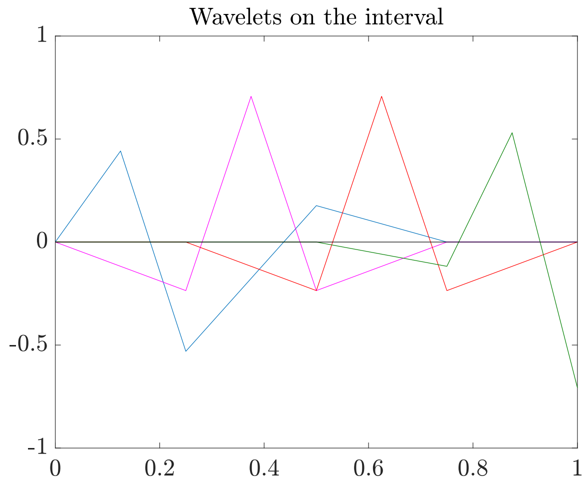

FIGURE 1. Continuous, piecewise linear wavelets on level $j = 2$ with two vanishing moments and a homogeneous boundary condition in $t = 0$. Blue and green: boundary wavelets, accounting for homogeneous Dirichlet BCs at $x = 0$. Orange and magenta: interior wavelets.

In particular, relation (4.2) implies that the wavelets exhibit *vanishing moments* of order $\widetilde{d}$, i.e.

$$\left|\langle v, \psi_{j,k}\rangle_{L^2(I)}\right| \lesssim 2^{-j(1/2+\widetilde{d})}|v|_{W^{\widetilde{d},\infty}(\operatorname{supp}\psi_{j,k})}$$

for all $v \in W^{\widetilde{d},\infty}(I)$. Herein, the quantity $|v|_{W^{\widetilde{d},\infty}(I)} := \|\partial^{\widetilde{d}} v\|_{L^\infty(I)}$ is the seminorm in $W^{\widetilde{d},\infty}(I)$. We refer to [6] for further details.

For the compression of the mass- and stiffness matrices of the operator $\mathcal{H}_T$ presently considered, the $\Phi_j, \widetilde{\Phi}_j$ are generated by constructing first dual pairs of single-scale bases on the interval $[0, 1]$, using B-splines for the primal bases and the dual components from [5] adapted to the interval [9]. Then, we follow the construction of [11] to incorporate the homogeneous boundary condition in $t = 0$. In case of continuous, piecewise linear spline-wavelets, i.e., $d = 2$, we obtain the wavelets shown in Figure 1, providing $\widetilde{d} = 2$ vanishing moments. The detailed refinement relation of these wavelets and also of piecewise linear spline-wavelets with $\widetilde{d} = 4$ vanishing moments are found in the appendix.

5. WAVELET MATRIX COMPRESSION

The basic ingredients in the analysis of the wavelet compression are decay estimates for the matrix entries. They are based on Calderón–Zygmund-like estimates of the kernel function in (3.5) which we shall provide first. They enable to apply results from [7] to define optimally compressed versions of the matrices $\mathbf{A}_j^t$ and $\mathbf{M}_j^t$.

5.1. **Calderón–Zygmund estimates.** For the analysis of wavelet Galerkin discretizations, the splitting

$$\mathcal{H}_T = \mathcal{H}_T^+ + \mathcal{H}_T^-, \tag{5.1}$$

motivated by the kernel function in (3.5), will be important. In (5.1), the operators $\mathcal{H}_T^+$ and $\mathcal{H}_T^-$ are given by

$$(\mathcal{H}_T^{\pm} v)(t) = \text{v.p.} \int_0^1 v(s)\, K^{\pm}(s,t)\, \mathrm{d}s, \quad t \in (0,1),\ v \in L^2(0,1)\ .$$

Here, v.p. indicates the Cauchy principal value. The kernel functions $K^{\pm}$ comprising K in (3.5) are given as

$$K^+(s,t) := \frac{1}{2\sin\frac{\pi(s+t)}{2}} \quad \text{and} \quad K^-(s,t) := \frac{1}{2\sin\frac{\pi|s-t|}{2}}\ , \quad (s,t) \in (0,1)^2 \backslash \{s=t\}.$$

The property $\ln(ab) = \ln(a) + \ln(b)$ of the natural logarithm implies in (3.6) a decomposition $K_{-1} = K_{-1}^+ + K_{-1}^-$ which corresponds to (5.1).

The following properties of the kernels $K^{\pm}$ are the basis for the ensuing wavelet compression results.

Proposition 5.1. *There exist constants $C, d > 0$ such that for every $(s,t) \in (0,1)^2 \backslash \{s=t\}$, and for every $p, p' \in \mathbb{N}$ hold*

$$\left| D_s^p D_t^{p'} K^{\pm}(s,t) \right| \leq C \frac{d^{p+p'}(p+p')!}{|s-t|^{p+p'+1}}\ . \tag{5.2}$$

Proof. The argument for K^- is elementary, due to the explicit relation

$$D_x^p(1/x) = (-1)^p p! x^{-p-1}.$$

The argument for K^+ follows from this and the elementary bound

$$\forall (s,t) \in (0,1)^2 : \quad |s-t| \leq s+t\ ,$$

since then for all $q \in \mathbb{N}$ holds $|s+t|^{-q} \leq |s-t|^{-q}$. □

Remark 5.2. *The Calderón–Zygmund estimate in* (5.2) *is actually stronger than what is needed in e.g.* [7, 29] *and the references there to justify the ensuing matrix-compressibility analysis. There, only weaker bounds of the form*

$$\left| D_s^p D_t^{p'} K^{\pm}(s,t) \right| \leq C_{pp'} \frac{1}{|s-t|^{p+p'+1}}$$

with $0 \leq p, p' \leq P(d)$ and some (implicit) dependence of the constant $C_{pp'}$ on p, p' are sufficient for optimal operator compression rates.

The stronger bounds (5.2) *correspond to what is called in $\mathcal{H}$-matrix analysis in* [16, 17] *"asymptotic smoothness" of kernels, implying in particular that also $\mathcal{H}$-matrix formats can be used for approximate representations of $\mathcal{H}_T^{\pm}$ of log-linear complexity, see* [34, Section 3.3]. *The analytic derivative bounds* (5.2) *are also essential for exponentially convergent numerical quadrature methods for the numerical evaluation of the nonzero matrix entries.*

5.2. **Matrix compression.** The basic ingredients in the analysis of the compression procedure are estimates for the matrix entries $\langle B\psi_{j',k'}, \psi_{j,k}\rangle_{L^2(0,1)}$ with $k \in \nabla_j$, $k' \in \nabla_{j'}$ and $j, j' > 0$, where $B\colon H_{0,}^q(0,1) \to [H_{0,}^q(0,1)]'$ is an integral operator of order $2q$ with $q \in [0,1]$. In the present application, we are interested in the operators $\langle \mathcal{H}_T \psi_{j',k'}, \partial_t \psi_{j,k}\rangle_{L^2(0,1)}$, which is of order $q = 1/2$, and $\langle \mathcal{H}_T \psi_{j',k'}, \psi_{j,k}\rangle_{L^2(0,1)}$, which is of order $q = 0$. However, our results apply also to the operator $\langle \mathcal{H}_T \partial_t \psi_{j',k'}, \partial_t \psi_{j,k}\rangle_{L^2(0,1)}$, which is of order $q = 1$ and which would appear in the discretization of the wave equation.

The supports of the wavelets will be denoted by

$$\Xi_{j,k} := \operatorname{supp}(\psi_{j,k}) \subset \overline{I}. \tag{5.3}$$

Moreover, the so-called "second compression" involves the *s*ingular support of a wavelet $\psi_{j,k}$,

$$\Xi^s_{j,k} := \operatorname{sing\,supp} \psi_{j,k} \subset \Xi_{j,k}. \tag{5.4}$$

It consists of all points of the wavelets' support, where $\psi_{j,k}$ is not smooth. A proof of the following estimates for the matrix entries is based on the Calderón–Zygmund estimates (5.2) and can be found e.g. in [29].

Theorem 5.3. *Suppose $j, j' \geq 0$ and let $k \in \nabla_j$, $k' \in \nabla_{j'}$ be such that*

$$\operatorname{dist}(\Xi_{j,k}, \Xi_{j',k'}) > 0.$$

Then, one has

$$|\langle B\psi_{j',k'}, \psi_{j,k}\rangle_{L^2(0,1)}| \lesssim \frac{2^{-(j+j')(\widetilde{d}+1/2)}}{\operatorname{dist}(\Xi_{j,k}, \Xi_{j',k'})^{1+2q+2\widetilde{d}}}$$

uniformly with respect to j and j'.

Theorem 5.4. *Suppose that $j' > j \geq 0$. Then, for $k \in \nabla_j$, $k' \in \nabla_{j'}$, the coefficients $\langle B\psi_{j',k'}, \psi_{j,k}\rangle_{L^2(0,1)}$ and $\langle B\psi_{j,k}, \psi_{j',k'}\rangle_{L^2(0,1)}$ satisfy*

$$|\langle B\psi_{j',k'}, \psi_{j,k}\rangle_{L^2(0,1)}|,\ |\langle B\psi_{j,k}, \psi_{j',k'}\rangle_{L^2(0,1)}| \lesssim \frac{2^{j/2}2^{-j'(\widetilde{d}+1/2)}}{\operatorname{dist}(\Xi^s_{j,k}, \Xi_{j',k'})^{2q+\widetilde{d}}}$$

uniformly with respect to j and j', provided that $\operatorname{dist}(\Xi^s_{j,k}, \Xi_{j',k'}) \gtrsim 2^{-j'}$.

Based on Theorems 5.3 and 5.4, we arrive at the following theorem, which defines a compressed version of the matrix $\mathbf{B}_j = [\langle B\phi_{j,k'}, \phi_{j,k}\rangle_{L^2(0,1)}]_{k,k'\in\Delta_j}$ when expressed in wavelet coordinates.

Theorem 5.5 (A-priori compression [7])**.** *Let the supports $\Xi_{j,k}$ and $\Xi^s_{j,k}$ be given as in* (5.3) *and* (5.4) *and define the compressed system matrix $\mathbf{B}^c_j$, corresponding to the discretization of the integral operator B with respect to the trial space S_j, by*

$$[\mathbf{B}^c_j]_{(\ell,k),(\ell,',k')} := \begin{cases} 0, & \operatorname{dist}(\Xi_{\ell,k}, \Xi_{\ell',k'}) > \mathcal{B}_{\ell,\ell'} \text{ and } \ell,\ell' \geq 0,\\ 0, & \operatorname{dist}(\Xi_{\ell,k}, \Xi_{\ell',k'}) \leq 2^{-\min\{\ell,\ell'\}} \text{ and}\\ & \operatorname{dist}(\Xi^s_{\ell,k}, \Xi_{\ell',k'}) > \mathcal{B}^s_{\ell,\ell'} \text{ if } \ell' > \ell \geq 0,\\ & \operatorname{dist}(\Xi_{\ell,k}, \Xi^s_{\ell',k'}) > \mathcal{B}^s_{\ell,\ell'} \text{ if } \ell > \ell' \geq 0,\\ \langle B\psi_{\ell',k'}, \psi_{\ell,k}\rangle_{L^2(0,1)}, & \text{otherwise.} \end{cases} \tag{5.5}$$

For given, fixed parameters

$$a > 1, \qquad d < \delta < \widetilde{d} + 2q,$$

for all $\ell, \ell' \geq 0$ the matrix compression parameters $\mathcal{B}_{\ell,\ell'}$ and $\mathcal{B}^s_{\ell,\ell'}$ are set according to

$$\begin{aligned} \mathcal{B}_{\ell,\ell'} &:= a\ \max\left\{2^{-\min\{\ell,\ell'\}}, 2^{\frac{2j(\delta-q)-(\ell+\ell')(\delta+\widetilde{d})}{2(\widetilde{d}+q)}}\right\},\\ \mathcal{B}^s_{\ell,\ell'} &:= a\max\left\{2^{-\max\{\ell,\ell'\}}, 2^{\frac{2j(\delta-q)-(\ell+\ell')\delta-\max\{\ell,\ell'\}\widetilde{d}}{\widetilde{d}+2q}}\right\}. \end{aligned} \tag{5.6}$$

With the compression according to (5.5)–(5.6), *the compressed system matrices* $\mathbf{B}_j^c$ *have only* $\mathcal{O}(N_j)$ *nonzero entries. In addition, the error estimates*

$$\|u - u_j\|_{H^{q-t}(I)} \lesssim 2^{j(d-q+t)} \|u\|_{H^d(I)}, \quad \textit{for all } 0 \leq t \leq d - q$$

hold for the solution u_j *of the compressed Galerkin system* $\mathbf{B}_j^c \mathbf{u}_j = \mathbf{f}_j$ *provided that* $u = B^{-1} f \in H^d(I)$.

6. Numerical experiments

6.1. **1D test example.** In order to demonstrate the wavelet method and to validate its accuracy, we consider the ordinary differential equation

$$\partial_t u(t) + \mu u(t) = f(t) \quad \text{for } t \in I := (0, T), \quad u(0) = 0 \tag{6.1}$$

and choose $T = 2$, $\mu = 10$, and the right-hand side $f(t)$ such that

$$u(t) = -2 \sin\left(\frac{3\pi}{4} t\right) + \sin\left(\frac{9\pi}{4} t\right)$$

becomes the known solution to be determined.

j	N_j^t	%	$L^2(I)$-error		$H^1(I)$-error		$H^{1/2}(I)$-error	
4	16	98.62	$3.28 \cdot 10^{-2}$		1.88		$2.48 \cdot 10^{-1}$	
5	32	84.57	$7.64 \cdot 10^{-3}$	(2.10)	$9.28 \cdot 10^{-1}$	(1.02)	$8.42 \cdot 10^{-2}$	(1.56)
6	64	62.32	$1.87 \cdot 10^{-3}$	(2.03)	$4.62 \cdot 10^{-1}$	(1.00)	$2.94 \cdot 10^{-2}$	(1.52)
7	128	42.09	$4.67 \cdot 10^{-4}$	(2.01)	$2.31 \cdot 10^{-1}$	(1.00)	$1.04 \cdot 10^{-2}$	(1.50)
8	256	26.76	$1.17 \cdot 10^{-4}$	(2.00)	$1.15 \cdot 10^{-1}$	(1.00)	$3.67 \cdot 10^{-3}$	(1.50)
9	512	16.27	$2.91 \cdot 10^{-5}$	(2.00)	$5.77 \cdot 10^{-2}$	(1.00)	$1.30 \cdot 10^{-3}$	(1.50)
10	1024	9.58	$7.28 \cdot 10^{-6}$	(2.00)	$2.89 \cdot 10^{-2}$	(1.00)	$4.58 \cdot 10^{-4}$	(1.50)
11	2048	5.50	$1.82 \cdot 10^{-6}$	(2.00)	$1.44 \cdot 10^{-2}$	(1.00)	$1.62 \cdot 10^{-4}$	(1.50)
12	4096	3.09	$4.55 \cdot 10^{-7}$	(2.00)	$7.21 \cdot 10^{-3}$	(1.00)	$5.73 \cdot 10^{-5}$	(1.50)
13	8192	1.71	$1.14 \cdot 10^{-7}$	(2.00)	$3.61 \cdot 10^{-3}$	(1.00)	$2.03 \cdot 10^{-5}$	(1.50)

Table 1. Results for the analytical example (6.1) on the interval for piecewise linear wavelets with $\tilde{d} = 2$ vanishing moments.

j	N_j^t	%	$L^2(I)$-error		$H^1(I)$-error		$H^{1/2}(I)$-error	
5	32	96.51	$7.63 \cdot 10^{-3}$		$9.28 \cdot 10^{-1}$		$8.42 \cdot 10^{-2}$	
6	64	78.37	$1.87 \cdot 10^{-3}$	(2.03)	$4.62 \cdot 10^{-1}$	(1.00)	$2.94 \cdot 10^{-2}$	(1.52)
7	128	56.14	$4.66 \cdot 10^{-4}$	(2.01)	$2.31 \cdot 10^{-1}$	(1.00)	$1.04 \cdot 10^{-2}$	(1.50)
8	256	37.19	$1.16 \cdot 10^{-4}$	(2.00)	$1.15 \cdot 10^{-1}$	(1.00)	$3.67 \cdot 10^{-3}$	(1.50)
9	512	23.22	$2.91 \cdot 10^{-5}$	(2.00)	$5.77 \cdot 10^{-2}$	(1.00)	$1.30 \cdot 10^{-3}$	(1.50)
10	1024	13.90	$7.28 \cdot 10^{-6}$	(2.00)	$2.89 \cdot 10^{-2}$	(1.00)	$4.58 \cdot 10^{-4}$	(1.50)
11	2048	8.06	$1.82 \cdot 10^{-6}$	(2.00)	$1.44 \cdot 10^{-2}$	(1.00)	$1.62 \cdot 10^{-4}$	(1.50)
12	4096	4.57	$4.55 \cdot 10^{-7}$	(2.00)	$7.21 \cdot 10^{-3}$	(1.00)	$5.73 \cdot 10^{-5}$	(1.50)
13	8192	2.54	$1.14 \cdot 10^{-7}$	(2.00)	$3.61 \cdot 10^{-3}$	(1.00)	$2.03 \cdot 10^{-5}$	(1.50)

Table 2. Results for the analytical example (6.1) on the interval for piecewise linear wavelets with $\tilde{d} = 4$ vanishing moments.

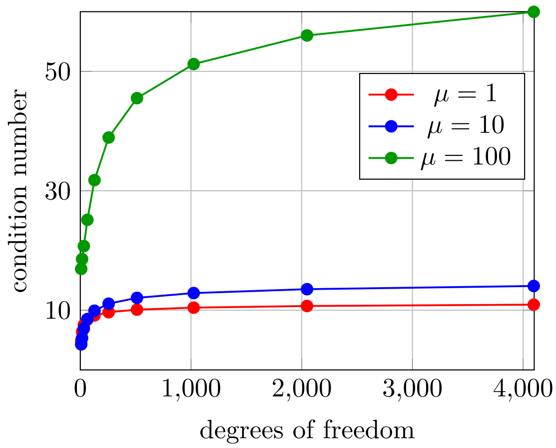

FIGURE 2. Condition numbers of the diagonally scaled system matrix $\mathbf{A}_j^t + \mu\mathbf{M}_j^t$ in wavelet coordinates for $\mu = 1, 10, 100$.

In Table 1 and Table 2, we find the results when discretizing (6.1) by a Galerkin method using wavelet matrix compression with $\tilde{d} = 2$ and $\tilde{d} = 4$ vanishing moments, respectively. The respective compression rates for the associated system matrices $(\mathrm{nnz}(\mathbf{A}_j^t + \mu\mathbf{M}_j^t)/N^2$ in percent) are found in the column entitled "%". Although the system matrices are quite sparse, we achieve the optimal rate of convergence of order 2 in $L^2(I)$ and of order 1 in $H^1(I)$. In particular, the accuracy of the compressed scheme is the same as that of the original scheme.

6.2. **Preconditioning of space-time methods.** It is well-known that proper scaling of wavelet bases leads to norm equivalences [6, 8, 21] for a whole scale of Sobolev spaces. Indeed, a diagonal scaling of matrices in wavelet coordinates yields uniformly bounded condition numbers provided that the underlying operators are elliptic and continuous with respect to the given Sobolev space provided that the wavelets and their duals are regular enough.

The condition numbers of the diagonally scaled system matrix $\mathbf{A}_j^t + \mu\mathbf{M}_j^t$ from the previous example are shown in Figure 2 for $\widetilde{d} = 2$ and different values of the coupling parameter μ. The condition numbers appear indeed uniformly bounded with respect to h as the matrix $\mathbf{A}_j^t$ is equivalent to the $H_{0,}^{1/2}(I)$-norm and $\mathbf{M}_j^t$ is positive and bounded by the L^2-norm. However, we observe a strong dependence on the parameter μ. In fact, the diagonally scaled version of $\mathbf{A}_j^t$ is also uniformly bounded, but the condition number of the matrix $\mathbf{M}_j^t$ grows linearly in $1/h_j^t$. Diagonal scaling makes the situation even worse. This rules out wavelet preconditioning for space-time formulations, where we have to precondition – in case of the heat equation – the system matrix

$$\mathbf{A}_j^t \otimes \mathbf{M}_j^x + \mathbf{M}_j^t \otimes \mathbf{A}_j^x \tag{6.2}$$

with $\mathbf{M}_j^x$ being the spatial mass matrix and $\mathbf{A}_j^x$ corresponding to the spatial Laplacian.

In [22], it was proposed to use the Bartels–Stewart method for the efficient solution of (6.2). This, however, requires the Schur decomposition of the matrix

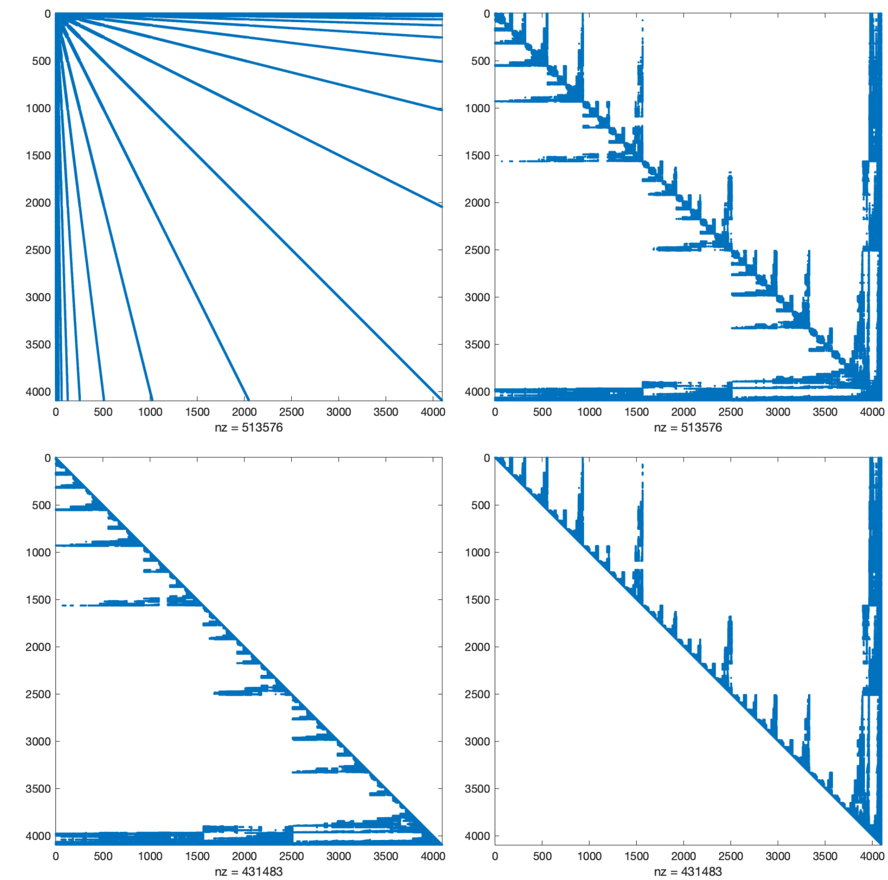

FIGURE 3. Sparsity patterns of the compressed system matrix (top left), its reordered version (top right) and of the associated LU factors (bottom).

$\left(\mathbf{A}_j^t\right)^{-1}\mathbf{M}_j^t$ which is computationally expensive and hence time-consuming. We shall therefore make use of the observation from [18] that the LU factorization of the compressed wavelet matrices can be computed very efficiently by means of nested dissection. Figure 3 depicts the sparsity pattern of the compressed system matrix $\mathbf{A}_j^t + \mu\mathbf{M}_j^t$ in wavelet coordinates. The fill-in of the LU factorization is minimal as both, the left factor as well as the right factor, have fewer nonzero entries than the original matrix. As a consequence, we can rapidly apply a direct solver in the temporal coordinate.

In order to exploit this fact, we assume that the sequence

$$S_0^x \subset S_1^x \subset S_2^x \subset \cdots, \quad \text{where } S_j^x = \operatorname{span}\{\phi_{j,k}^x : k \in \Delta_j^x\}$$

consists of Lagrangian finite element spaces of continuous, piecewise polynomials of total degree $p_x \geq 1$ on a nested sequence of regular, simplicial quasiuniform partitions of the spatial domain Ω that is obtained by a systematic refinement, for example by the newest edge bisection or dyadic subdivision, which we will use

unknowns		full tensor-product space			sparse tensor-product space		
N_j^t	N_j^x	iter	$L^2(Q)$-error		iter	$L^2(Q)$-error	
16	289	19	$3.0029 \cdot 10^{-3}$		22	$3.0022 \cdot 10^{-3}$	
32	1089	20	$7.5041 \cdot 10^{-4}$	(2.00)	37	$7.5041 \cdot 10^{-4}$	(2.00)
64	4225	20	$1.8752 \cdot 10^{-4}$	(2.00)	45	$1.8752 \cdot 10^{-4}$	(2.00)
128	16641	20	$4.6868 \cdot 10^{-5}$	(2.00)	40	$4.6868 \cdot 10^{-5}$	(2.00)
256	66049	20	$1.1715 \cdot 10^{-5}$	(2.00)	37	$1.1715 \cdot 10^{-5}$	(2.00)
512	263169	21	$2.9284 \cdot 10^{-6}$	(2.00)	39	$2.9285 \cdot 10^{-6}$	(2.00)
1024	1050625	21	$7.3195 \cdot 10^{-7}$	(2.00)	35	$7.3226 \cdot 10^{-7}$	(2.00)

TABLE 3. Approximation error and number of iterations needed to solve the linear system of equations up to a relative error of 10^{-8} in $2+1$ dimensions. The table contains the results of both, the full and the sparse tensor-product discretization.

here. Here, $\Delta_j^x = \{1, \ldots, N_j^x\}$ denotes the set of indices for (one-scale) nodal basis functions of S_j^x. For preconditioning the system matrix (6.2), we switch to multilevel frame coordinates of BPX-type in $S_j^x(\Omega)$, see [2, 15, 19] for details. Then, on each grid-scale in the spatial domain Ω, we have to perform a diagonal scaling of the system matrix (6.2) for the current spatial meshwidth h_j^x. Since the entries of the finite element mass matrix $\operatorname{diag}(\mathbf{M}_j^x)$ scale asymptotically like 1 and those of the stiffness matrix $\operatorname{diag}(\mathbf{A}_j^x)$ like $(h_j^x)^{-2}$ provided that the finite elements in space are L^2-normalized, we need to apply the diagonal scaling to

$$\mathbf{A}_j^t \otimes \operatorname{diag}(\mathbf{M}_j^x) + \mathbf{M}_j^t \otimes \operatorname{diag}(\mathbf{A}_j^x) \sim \big(\mathbf{A}_j^t + (h_j^x)^{-2}\mathbf{M}_j^t\big) \otimes \mathbf{I}_j^x,$$

where $\mathbf{I}_j^x$ denotes the identity matrix in $\mathbb{R}^{N_j^x \times N_j^x}$. Since the inverse of $\mathbf{A}_j^t + (h_j^x)^{-2}\mathbf{M}_j^t$ can be cheaply computed, this diagonal scaling is feasible. The implementation of this preconditioner is along the lines of [2]. Denoting the (spatial) prolongation from level ℓ to level $j > \ell$ by $\mathbf{I}_\ell^j$ and the (spatial) restriction from level j to level $\ell < j$ by $\mathbf{I}_j^\ell$, one can implement the multilevel preconditioner in accordance with

$$\sum_{\ell=0}^{j} \big(\mathbf{I}_j^t \otimes \mathbf{I}_\ell^j\big)\Big(\big(\mathbf{A}_j^t + (h_\ell^x)^{-2}\mathbf{M}_j^t\big)^{-1} \otimes \mathbf{I}_\ell^x\Big)\big(\mathbf{I}_j^t \otimes \mathbf{I}_j^\ell\big). \tag{6.3}$$

Here, in analogy to $\mathbf{I}_j^x$ introduced above, $\mathbf{I}_j^t$ denotes the identity matrix in $\mathbb{R}^{N_j^t \times N_j^t}$. Note that the preconditioner (6.3) is optimal *except for a logarithmic factor which appears since the multilevel frame is not stable in* $L^2(\Omega)$, compare [19] for example.

6.3. **Heat equation in two spatial dimensions.** Let Ω be the unit square $(0,1)^2$ and consider the heat equation (1.1) with right-hand side

$$f(x,t) := \sin(2\pi x_1)\sin(2\pi x_2)\big(\cos(t) + 2\pi^2 \sin(t)\big).$$

This yields the known solution $u(x,t) = \sin(2\pi x_1)\sin(2\pi x_2)\sin(t)$ of the heat equation.

On the level j, we subdivide the unit square into $2^j \times 2^j$ squares and the discretize in the spatial coordinate by piecewise bilinear finite elements. In the temporal coordinate, we apply piecewise linear wavelets with four vanishing moments. For the solution of the linear system of equations, we use the iterative Krylov subspace solver GMRES (generalized minimal residual method, see [28]), preconditioned by the multilevel preconditioner proposed in the previous subsection. We started the iteration with the initial guess $\mathbf{0}$ and stopped it when the relative norm of the residual is smaller than 10^{-8}.

The observed numerical results for the levels $j = 4, \dots, 10$ are reported in Table 3 in the columns entitled "full tensor-product space". Therein, the number of iterations needed by the preconditioned GMRES solver appears to be essentially bounded. The expected rate 4^{-j} of convergence is realized.

7. Sparse space-time tensor-product approximation

In this section, we introduce a sparse space-time tensor-product Galerkin approximation to the solution of the parabolic evolution equation (1.1), including space-time error estimates.

7.1. **Sparse tensor-product spaces.** For a given refinement level $j \in \mathbb{N}_0$, instead of using the full tensor-product space $S_j^{x,t} = S_j^x \otimes S_j^t \subset H^{1,1/2}_{\Gamma_{\mathrm{D}};0,}(Q)$, we can also consider the *sparse tensor-product space* $\widehat{S}_j^{x,t}$ for the discretization of the heat equation. In view of the multilevel decomposition $S_j^t = \bigoplus_{\ell=0}^j W_\ell^t$ in the temporal variable, the sparse tensor-product space is defined by

$$\widehat{S}_j^{x,t} := \bigoplus_{\ell=0}^{j} S_\ell^x \otimes W_{j-\ell}^t \subset H^{1,1/2}_{\Gamma_{\mathrm{D}};0,}(Q). \tag{7.1}$$

With the representations $S_j^x = \operatorname{span}\{\phi_{j,k}^x : k \in \Delta_j^x\}$ and $W_j^t = \operatorname{span}\{\psi_{j,k}^t : k \in \nabla_j^t\}$, a basis in the sparse tensor-product space is given by

$$\widehat{S}_j^{x,t} = \operatorname{span}\{\phi_{\ell,k}^x \otimes \psi_{j-\ell,k'}^t : k \in \Delta_\ell^x,\ k' \in \nabla_{j-\ell}^t,\ \ell = 0, \dots, j\}. \tag{7.2}$$

We refer to [14, Section 6.1] for all the details.

7.2. **Sparse tensor space-time Galerkin projection.** The dimension of the sparse tensor-product space $\widehat{S}_j^{x,t}$ scales in general linearly in the maximum of the number of degrees of freedom in space and time, i.e., it scales as $\mathcal{O}(\max\{N_j^x, N_j^t\})$, with an additional logarithm appearing if $N_j^x = N_j^t$, compare [14, Theorem 4.1]. In [14] also more general settings of the sparse tensor-product compared to the setting (7.1) have been studied.

The *sparse space-time tensor-product Galerkin approximation* is to find $\widehat{u}_j \in \widehat{S}_j^{x,t} \subset H^{1,1/2}_{\Gamma_{\mathrm{D}};0,}(Q)$ such that

$$\forall \widehat{v}_j \in \widehat{S}_j^{x,t} : \quad b(\widehat{u}_j, \mathcal{H}_T \widehat{v}_j) = \langle f, \mathcal{H}_T \widehat{v}_j \rangle_Q. \tag{7.3}$$

The sparse space-time tensor-product approximation scheme (7.3) is well-defined and unconditionally stable with the space-time stability estimate

$$\|\widehat{u}_j\|_{H^{1/2}_{0,}(I;L^2(\Omega))} \le \|f\|_{[H^{1/2}_{,0}(I;L^2(\Omega))]'},$$

provided that $f \in [H^{1/2}_{,0}(I; L^2(\Omega))]'$, see [37, Theorem 3.4.20]. This is essential to ensure the stability of the projection defined by (7.3), as in the sparse tensor-product (7.1) with basis (7.2) all possible combinations of spatial and temporal meshwidths appear simultaneously.

7.3. **Convergence rates.** In this subsection, we prove convergence of the sparse tensor-product approximation $\widehat{u}_j$ given by (7.3).

Theorem 7.1. *Let $j \in \mathbb{N}_0$ be a given refinement level and let $u \in H^{1,1/2}_{\Gamma_{\mathrm{D}};0,}(Q)$ be the unique solution of the space-time variational formulation (2.4). Let $\widehat{u}_j \in \widehat{S}^{x,t}_j$ be the sparse tensor-product approximation of (2.4) given by (7.3), and assume that $\mathcal{L}_x u \in L^2(Q)$ holds.*

Then, the space-time error estimate

$$\begin{aligned}\|u - \widehat{u}_j\|_{H^{1/2}_{0,}(I;L^2(\Omega))} &\le 2\|u - \widehat{P}_j u\|_{H^{1/2}_{0,}(I;L^2(\Omega))} \\ &\quad + \left\|(\mathrm{Id}_{xt} - P^t_j \otimes \mathrm{Id}_x)\mathcal{L}_x u\right\|_{[H^{1/2}_{,0}(I;L^2(\Omega))]'} \\ &\quad + C\sum_{\ell=0}^{j} (h^x_\ell)^{-1} \left\|\big((P^t_j - P^t_{j-\ell}) \otimes (P^x_\ell - P^x_{\ell-1})\big)u\right\|_{H^1(\Omega)\otimes[H^{1/2}_{0,}(I)]'}.\end{aligned} \tag{7.4}$$

holds true, with a constant $C > 0$ and the sparse space-time tensor-product projection defined by

$$\widehat{P}_j u := \sum_{\ell=0}^{j} \big(P^t_{j-\ell} \otimes (P^x_\ell - P^x_{\ell-1})\big)u \in \widehat{S}^{x,t}_j.$$

Here, for $\ell \in \mathbb{N}_0$, $P^t_\ell\colon L^2(I) \to S^t_\ell$ is any temporal projection and $P^x_\ell\colon H^1_{\Gamma_{\mathrm{D}}}(\Omega) \to S^x_\ell$ denotes the spatial Ritz projection as defined in Theorem 3.3 Further, we set $P^x_{-1} = 0$.

Proof. The triangle inequality yields

$$\|u - \widehat{u}_j\|_{H^{1/2}_{0,}(I;L^2(\Omega))} \le \|u - \widehat{P}_j u\|_{H^{1/2}_{0,}(I;L^2(\Omega))} + \|\widehat{u}_j - \widehat{P}_j u\|_{H^{1/2}_{0,}(I;L^2(\Omega))}.$$

The second term in the upper bound is estimated as follows.

In view of the $H^{1/2}_{0,}(I;L^2(\Omega))$-ellipticity of the bilinear form $b(\cdot,\cdot)$ in (3.7) and the Galerkin orthogonality corresponding to (7.3), we calculate

$$\begin{aligned}\|\widehat{u}_j - \widehat{P}_j u\|^2_{H^{1/2}_{0,}(I;L^2(\Omega))} &\le b(\widehat{u}_j - \widehat{P}_j u, \mathcal{H}_T(\widehat{u}_j - \widehat{P}_j u)) \\ &= b(u - \widehat{P}_j u, \mathcal{H}_T(\widehat{u}_j - \widehat{P}_j u)) \\ &= \langle \partial_t(u - \widehat{P}_j u), \mathcal{H}_T(\widehat{u}_j - \widehat{P}_j u)\rangle_Q \\ &\quad + \langle A\nabla_x(\mathrm{Id}_{xt} - P_j)u, \nabla_x \mathcal{H}_T(\widehat{u}_j - \widehat{P}_j u)\rangle_{L^2(Q)^n} \\ &\quad + \langle A\nabla_x(P_j - \widehat{P}_j)u, \nabla_x \mathcal{H}_T(\widehat{u}_j - \widehat{P}_j u)\rangle_{L^2(Q)^n}\end{aligned}$$

with $P_j u = (P^t_j \otimes P^x_j)u \in S^{x,t}_j$. For the first term, we have with (3.2) that

$$\langle \partial_t(u - \widehat{P}_j u), \mathcal{H}_T(\widehat{u}_j - \widehat{P}_j u)\rangle_Q \le \|u - \widehat{P}_j u\|_{H^{1/2}_{0,}(I;L^2(\Omega))} \|\widehat{u}_j - \widehat{P}_j u\|_{H^{1/2}_{0,}(I;L^2(\Omega))},$$

while we find for the second term by the definition of the Ritz projection P^x_j and by integration by parts

$$\begin{aligned}&\langle A\nabla_x(\mathrm{Id}_{xt} - P_j)u, \nabla_x \mathcal{H}_T(\widehat{u}_j - \widehat{P}_j u)\rangle_{L^2(Q)^n} \\ &\quad = \langle A\nabla_x(\mathrm{Id}_{xt} - P^t_j \otimes \mathrm{Id}_x)u, \nabla_x \mathcal{H}_T(\widehat{u}_j - \widehat{P}_j u)\rangle_{L^2(Q)^n} \\ &\quad \le \left\|(\mathrm{Id}_{xt} - P^t_j \otimes \mathrm{Id}_x)\mathcal{L}_x u\right\|_{[H^{1/2}_{,0}(I;L^2(\Omega))]'} \|\widehat{u}_j - \widehat{P}_j u\|_{H^{1/2}_{0,}(I;L^2(\Omega))}.\end{aligned}$$

The last term is treated as follows. We set $\widehat{v}_j := \widehat{u}_j - \widehat{P}_j u \in \widehat{S}_j^{x,t}$ and insert the definition of the sparse grid projection and the representation $P_j = \sum_{\ell=0}^{j} P_j^t \otimes (P_\ell^x - P_{\ell-1}^x)$:

$$
\begin{aligned}
&\langle A\nabla_x (P_j - \widehat{P}_j)u, \nabla_x \mathcal{H}_T \widehat{v}_j\rangle_{L^2(Q)^n}\\
&\quad= \sum_{\ell=0}^{j} \langle A\nabla_x \big((P_j^t - P_{j-\ell}^t)\otimes(P_\ell^x - P_{\ell-1}^x)\big)u, \nabla_x \mathcal{H}_T \widehat{v}_j\rangle_{L^2(Q)^n}\\
&\quad= \sum_{\ell=0}^{j} \langle A\nabla_x \big((P_j^t - P_{j-\ell}^t)\otimes(P_\ell^x - P_{\ell-1}^x)\big)u, \nabla_x \mathcal{H}_T \big(\mathrm{Id}_t \otimes (P_\ell^x - P_{\ell-1}^x)\big)\widehat{v}_j\rangle_{L^2(Q)^n}\\
&\quad\le \sum_{\ell=0}^{j} \big\| \big((P_j^t - P_{j-\ell}^t)\otimes(P_\ell^x - P_{\ell-1}^x)\big)u\big\|_{H^1(\Omega)\otimes[H_{0,}^{1/2}(I)]'}\\
&\qquad\qquad\qquad\qquad \times \big\|\big(\mathrm{Id}_t \otimes (P_\ell^x - P_{\ell-1}^x)\big)\widehat{v}_j\big\|_{H_{0,}^{1/2}(I;H^1(\Omega))}.
\end{aligned}
$$

Using finally the inverse inequality and the L^2-stability of the Ritz projections P_ℓ^x, we derive

$$
\big\|\big(\mathrm{Id}_t \otimes (P_\ell^x - P_{\ell-1}^x)\big)\widehat{v}_j\big\|_{H_{0,}^{1/2}(I;H^1(\Omega))} \le C(h_\ell^x)^{-1}\big\|\widehat{v}_j\big\|_{H_{0,}^{1/2}(I;L^2(\Omega))},
$$

with a constant $C > 0$, which proves the claim. □

The error bound in Theorem 7.1 implies convergence rate bounds for solutions of sufficient spatio-temporal regularity. However, compared to the convergence rate in the full tensor-product space, we only obtain a reduced convergence rate in the spatial variable, even for regular solutions when we choose $P_j^t = P_j^{1/2,t}$ as defined in (3.11). Namely, for the single terms in (7.4) we have

$$
\begin{aligned}
\|u - \widehat{P}_j u\|_{H_{0,}^{1/2}(I;L^2(\Omega))} &\le C \sum_{\ell=0}^{j} (h_\ell^x)^{p_x+1}(h_{j-\ell}^t)^{d-1/2}\|u\|_{H^d(I;H^{p_x+1}(\Omega))},\\
\big\|(\mathrm{Id}_{xt} - P_j^t \otimes \mathrm{Id}_x)\mathcal{L}_x u\big\|_{[H_{,0}^{1/2}(I;L^2(\Omega))]'} &\le C(h_j^t)^{d+1/2}\|u\|_{H^d(I;H^2(\Omega))},
\end{aligned}
$$

and

$$
\begin{aligned}
&\sum_{\ell=0}^{j} (h_\ell^x)^{-1}\big\|\big((P_j^t - P_{j-\ell}^t)\otimes(P_\ell^x - P_{\ell-1}^x)\big)u\big\|_{H^1(\Omega)\otimes[H_{0,}^{1/2}(I)]'}\\
&\qquad\qquad \le C\sum_{\ell=0}^{j} (h_\ell^x)^{p_x-1}(h_{j-\ell}^t)^{d+1/2}\|u\|_{H^d(I;H^{p_x+1}(\Omega))}.
\end{aligned}
$$

Altogether, this implies the estimate

$$
\|u - \widehat{u}_j\|_{H_{0,}^{1/2}(I;L^2(\Omega))} \le C\max\big\{(h_j^x)^{p_x-1}, (h_j^t)^{d-1/2}\big\}\|u\|_{H^d(I;H^{p_x+1}(\Omega))}. \tag{7.5}
$$

This rate, however, is not observed in our numerical experiments: for smooth solutions, essentially the same rate as obtained with the full tensor-product approximation space is observed.

Remark 7.2. *A possible explanation for this observation could be that the expression $\big(\mathrm{Id}_t \otimes (P_\ell^x - P_{\ell-1}^x)\big)\widehat{v}_j$ contains for $\widehat{v}_j \in \widehat{S}_j^{x,t}$ only contributions from the tensor-product spaces $S_k^x \otimes S_{j-k}^t$ for $k < \ell$. Hence, if P_ℓ^t is chosen to be the projection*

$P_\ell^t\colon\ L^2(I) \to S_\ell^t$ defined by

$$\forall z_\ell \in S_\ell^t: \quad \langle P_\ell^t v, \mathcal{H}_T z_\ell\rangle_{L^2(I)} = \langle v, \mathcal{H}_T z_\ell\rangle_{L^2(I)} \tag{7.6}$$

for a given function $v \in L^2(I)$, then the term $\langle A\nabla_x(P_j - \widehat{P}_j)u, \nabla_x\mathcal{H}_T\widehat{v}_j\rangle_{L^2(Q)^n}$ vanishes in the above proof. This would lead to the error estimate

$$\begin{aligned}\|u-\widehat{u}_j\|_{H^{1/2}_{0,}(I;L^2(\Omega))} &\le 2\|u-\widehat{P}_j u\|_{H^{1/2}_{0,}(I;L^2(\Omega))} \\ &\quad + \big\|(\mathrm{Id}_{xt} - P_j^t\otimes \mathrm{Id}_x)\mathcal{L}_x u\big\|_{[H^{1/2}_{,0}(I;L^2(\Omega))]'}.\end{aligned}$$

Stability and consistency bounds for the projection (7.6) *are delicate and topic of current investigation* [25]*.*

Corollary 7.3. *Assume the sparse space-time Galerkin solution $\widehat{u}_j$ in* (7.3) *has been obtained with temporal biorthogonal spline-wavelets of order $d \ge 2$ in I and with Lagrangian finite elements of degree $p_x \ge 1$ in Ω. Assume also that the exact solution u of the variational IBVP* (2.4) *admits the regularity $u \in H^d(I;H^{1+p_x}(\Omega)) \simeq H^{1+p_x}(\Omega)\otimes H^d(I)$.*

Then, there holds the asymptotic error bound (7.5)*. Further, subject to the wavelet matrix compression of the temporal stiffness and mass-matrices in Section 4 according to the specifications in Section 5.2, the numerical solution $\widehat{u}_j$ can be computed in essentially*[1] *$O(2^{nj})$ operations and memory.*

7.4. **Iterative solution.** We show the claim on solver complexity in Corollary 7.3 and detail the sparse space-time matrix-vector multiplication, and the corresponding preconditioner. For the solution of the linear system of equations which arises from the discretization (7.3) of the heat equation with respect to the sparse space-time tensor-product space $\widehat{S}_j^{x,t}$, we use preconditioned GMRES. To this end, we require matrix-vector products of the form

$$\mathbf{v}_{j-\ell',\ell'} = \sum_{\ell=0}^{j}\big(\mathbf{A}^t_{j-\ell,j-\ell'}\otimes\mathbf{M}^x_{\ell,\ell'} + \mathbf{M}^t_{j-\ell,j-\ell'}\otimes\mathbf{A}^x_{\ell,\ell'}\big)\mathbf{u}_{j-\ell,\ell} \tag{7.7}$$

for all $0 \le \ell' \le j$. Here, the matrices $\mathbf{A}^t_{j-\ell,j-\ell'}$ and $\mathbf{M}^t_{j-\ell,j-\ell'}$ are specific blocks of the wavelet representations of the temporal matrices and therefore available. In contrast, the matrices $\mathbf{A}^x_{\ell,\ell'}$ and $\mathbf{M}^x_{\ell,\ell'}$ correspond to spatial finite element mass and stiffness matrices with trial and test functions from different discretization levels when $\ell \neq \ell'$. Such matrices are usually not available in practice. Therefore, we replace them with square finite element matrices and apply inter-grid restrictions to derive the desired ones. This means we employ the identities

$$\mathbf{A}^x_{\ell,\ell'} = \mathbf{I}^\ell_{\ell'}\mathbf{A}^x_{\ell'}, \quad \mathbf{M}^x_{\ell,\ell'} = \mathbf{I}^\ell_{\ell'}\mathbf{M}^x_{\ell'}, \quad \text{if } \ell < \ell', \tag{7.8}$$

and

$$\mathbf{A}^x_{\ell,\ell'} = \mathbf{A}^x_{\ell}\mathbf{I}^\ell_{\ell'}, \quad \mathbf{M}^x_{\ell,\ell'} = \mathbf{M}^x_{\ell}\mathbf{I}^\ell_{\ell'}, \quad \text{if } \ell > \ell'. \tag{7.9}$$

In both cases, the application of $\mathbf{A}^x_{\ell,\ell'}$ and $\mathbf{M}^x_{\ell,\ell'}$, respectively, scales linearly in the number $\max\{N^x_\ell, N^x_{\ell'}\}$ of degrees of freedom.

[1]Up to an $O(j)$ factor.

Note that the multilevel preconditioner in (6.3) needs also to be adapted to the present setting. Namely, restricting (6.3) to the sparse tensor-product space yields the multilevel preconditioner

$$\sum_{\ell=0}^{j} \big(\mathbf{I}_{j-\ell}^{t} \otimes \mathbf{I}_{\ell}^{j}\big)\Big(\big(\mathbf{A}_{j-\ell}^{t} + (h_\ell^x)^{-2}\mathbf{M}_{j-\ell}^{t}\big)^{-1} \otimes \mathbf{I}_{\ell}^{x}\Big)\big(\mathbf{I}_{j-\ell}^{t} \otimes \mathbf{I}_{j}^{\ell}\big).$$

7.5. **Fast matrix-vector products.** We shall explain how to compute matrix-vector products in (7.7) in an efficient way. To this end, we will exploit that, given matrices $\mathbf{B} \in \mathbb{R}^{k\times\ell}$, $\mathbf{X} \in \mathbb{R}^{\ell\times m}$, $\mathbf{A} \in \mathbb{R}^{n\times m}$, and $\mathbf{Y} \in \mathbb{R}^{k\times n}$, there holds the identity

$$\operatorname{vec}(\mathbf{Y}) = (\mathbf{A} \otimes \mathbf{B}) \operatorname{vec}(\mathbf{X}) \quad \Longleftrightarrow \quad \mathbf{B}\mathbf{X}\mathbf{A}^{\intercal} = \mathbf{Y}, \tag{7.10}$$

where the vectorization $\operatorname{vec}(\cdot)$ converts a matrix into a column vector. We will use this equivalence to realize a fast matrix-vector multiplication. For the sake of simplicity in presentation, we assume that the vector $\widehat{\mathbf{u}}_j = [\mathbf{u}_{\ell,j-\ell}]_{0\le\ell\le j}$ is partitioned into blocks of coefficients associated with the sparse tensor-product basis in $\widehat{S}_j^{x,t}$ (compare (7.2)) and is blockwise stored in matrix form, i.e. $\mathbf{u}_{\ell,j-\ell} \in \mathbb{R}^{|\Delta_\ell^x|\times|\nabla_{j-\ell}^t|}$. Recall that we have (cp. Sec. 1.4) $|\Delta_\ell^x| = N_\ell^x$ and $|\nabla_{j-\ell}^t| \sim N_{j-\ell}^t$.

Computation of one matrix-vector multiplication (7.10) requires to compute products of the form

$$\operatorname{vec}(\mathbf{z}) = \big(\mathbf{A}_{j-\ell',j-\ell}^{t} \otimes \mathbf{B}_{\ell',\ell}^{x}\big) \operatorname{vec}(\mathbf{u}_{\ell,j-\ell}), \quad 0 \le \ell, \ell' \le j.$$

In view of (7.10), this means that

$$\mathbf{z} = \mathbf{B}_{\ell',\ell}^{x}\mathbf{u}_{\ell,j-\ell}\big(\mathbf{A}_{j-\ell',j-\ell}^{t}\big)^{\intercal}, \quad 0 \le \ell, \ell' \le j.$$

To optimize the complexity bound, care has to be taken with respect to the order in which one performs the multiplications. If the matrix-vector multiplication is performed in accordance with

$$\mathbf{y} = \mathbf{u}_{\ell,j-\ell}\big(\mathbf{A}_{j-\ell',j-\ell}^{t}\big)^{\intercal}, \quad \mathbf{z} = \mathbf{B}_{\ell',\ell}^{x}\mathbf{y}, \quad \text{if } |\nabla_{j-\ell'}^t| \cdot |\Delta_\ell^x| \le |\nabla_{j-\ell}^t| \cdot |\Delta_{\ell'}^x|,$$

and

$$\mathbf{y} = \mathbf{B}_{\ell',\ell}^{x}\mathbf{u}_{\ell,j-\ell}, \quad \mathbf{z} = \mathbf{y}\big(\mathbf{A}_{j-\ell',j-\ell}^{t}\big)^{\intercal}, \quad \text{if } |\nabla_{j-\ell'}^t| \cdot |\Delta_\ell^x| > |\nabla_{j-\ell}^t| \cdot |\Delta_{\ell'}^x|,$$

then we require at most $\mathcal{O}(\max\{N_j^x, N_j^t\})$ operations since the matrices $\mathbf{B}_{\ell',\ell}^{x}$ and $\mathbf{A}_{j-\ell',j-\ell}^{t}$ have only $\mathcal{O}(\max\{|\Delta_\ell^x|, |\Delta_{\ell'}^x|\})$ and $\mathcal{O}(|\nabla_{j-\ell}^t|, |\nabla_{j-\ell'}^t|\})$ nonzero coefficients, respectively. Note that the application of the identities (7.8) and (7.9) do not change this complexity bound but avoid the non-quadratic finite element matrices $\mathbf{B}_{\ell',\ell}^{x}$ when $\ell \ne \ell'$. In all, we thus arrive at an algorithm that computes all matrix-vector product of the form (7.7) in essentially linear complexity. This is achieved via the so-called "unidirectional principle", specifically Algorithm UNIDIRML in [40], see also [3].

7.6. **Numerical results.** We recompute the example in Subsection 6.3 with respect to the sparse tensor-product space (7.3) instead of the full tensor-product space (3.8). The results are displayed in Table 3 in the columns labeled "sparse tensor-product space". Using the sparse tensor-product space results in an approximation accuracy that is basically the same as furnished with the full tensor-product space. Whereas, the number of iterations of the iterative solver is about a factor two larger in comparison with the full tensor-product space, compare the 3rd and 5th column in Table 3. For example, for the finest resolution, where $N_j^x = 1050625$

and $N_j^t = 1024$, we have only about 12.5 million degrees of freedom in the sparse tensor-product space instead of about one billion degrees of freedom in the full tensor-product space. Since a single iteration of the GMRES solver is therefore drastically cheaper due to the much smaller amount of degrees of freedom, the computing time and memory requirement for the iterative solution is considerably smaller.

8. Conclusion and further directions

We proposed and analyzed a class of variational space-time discretizations for initial-boundary value problems of linear, parabolic evolution equations, with finite time horizon $0 < T < \infty$. As in [26, 33], the space-time Galerkin algorithm is based on combining the classical variational formulation in the spatial variable, combined with a new, $H_{0,}^{1/2}(0,T)$-coercive variational formulation of the first-order evolution operator ∂_t. Whereas in [26], we leveraged the *analytic semigroup property* of the parabolic solution operator by exponentially convergent temporal Petrov–Galerkin discretizations, we focus in the present work on settings with solutions of low temporal regularity. These typically arise in pathwise discretizations of linear, parabolic stochastic PDEs which are driven by noises of low pathwise temporal regularity, such as cylindrical Wiener processes. Here, low time-regularity of pathwise solutions obstructs the exponential convergence of *hp*-time discretizations of [26]. Time-adaptive schemes are likewise not offering substantial savings, as typically the temporal singular support of pathwise solutions is dense in $(0,T)$. Accordingly, low-order, non-adaptive discretizations in $(0,T)$ afford optimal convergence rates. Furthermore, *multilevel space-time discretization* is desirable in connection with so-called multilevel Monte-Carlo discretizations of such stochastic PDEs.

Stability of the continuous variational formulation in the finite time case and of its Galerkin discretization is based on an explicit, linear and nonlocal operator, a modified Hilbert transform, introduced first in [33]. Upon (Petrov–)Galerkin discretization with finite-dimensional spaces of Lagrangian finite element functions of degree $p_t \geq 1$ and of dimension N_t in $(0,T)$, the evolution operator induces fully populated matrices, i.e. a complexity $O(N_t^2)$ of memory and work for the time-derivative. Similar to [35], where stability was ensured by an adaptive wavelet algorithm, we use suitable biorthogonal spline-wavelet bases in $(0,T)$ and standard Lagrangian finite element spaces in the physical domain Ω.

We proved that the densely populated matrices resulting from non-adaptive (Petrov–)Galerkin discretization are optimally compressible, i.e. to $O(N_t)$ many nonvanishing entries, and, with the aid of the exponential convergence quadrature schemes from [39], can be approximated computationally in $O(N_t)$ work and memory, while maintaining the asymptotic discretization error bounds. Location of these $O(N_t)$ many nonvanishing matrix entries are known in advance and the compressed temporal Galerkin matrix corresponding to $\mathcal{H}_T$ can be precomputed in $O(N_t)$ work and memory. We also obtained an $O(N_t)$ direct factorization by applying a nested dissection strategy to the compressed wavelet Petrov–Galerkin matrix resulting from the time-discretization.

We combined this multiresolution temporal (Petrov–)Galerkin discretization with a multilevel finite element Galerkin discretization in the spatial domain Ω. It is based on standard, first-order Lagrangian finite element spaces on a sequence of nested, regular partitions of the physical domain Ω, with $O(N_x)$ many degrees of freedom.

We showed that the nested nature of both, the spatial and the temporal subspaces used in the Galerkin discretizations allows for straightforward realization of *sparse space-time tensor-product* subspaces for the discretization of the evolution problem. The resulting sparse space-time tensor-product Galerkin discretizations afford optimal approximation rates in $Q = \Omega \times (0,T)$ while only using $O(N_x)$ work and memory, which are required for one spatial elliptic problem.

We proposed a new, iterative solver for the very large, linear systems resulting from the sparse tensor-product discretizations. It is based on standard, BPX-type multilevel iterative solvers combined with the nested dissection LU factorizations of the evolution operator, and we found that this results in $O(N_x)$ complexity solution of the fully discrete, sparse space-time (Petrov–)Galerkin approximation.

Numerical experiments for first-order discretizations in space and time confirmed and illustrated the above findings.

We address several directions for further work: in our consistency and discretization error bounds, we assumed maximal spatial and temporal regularity of solutions. While this regularity is available under conditions (such as compatible initial- and boundary data, sufficient smooth forcing, convex domains), the sparse space-time discretization and the multilevel solution algorithm adapt straightforwardly to more general situations: for non-compatible initial data, the solutions are known to exhibit an *algebraic temporal singularity as* $t \downarrow 0$, and for nonconvex spatial domains Ω solutions will exhibit corner singularities in space dimension $d = 2$ and corner-edge singularities in addition in polyhedral $\Omega \subset \mathbb{R}^3$, causing reduced convergence rates of Galerkin discretizations on (quasi-)uniform meshes. Remedies restoring the full convergence rates afforded by Lagrangian finite element methods in Ω consist of graded or bisection-tree meshes with corner (resp. corner-edge) refinement (see, e.g., [26, Section 5.3]), and by local addition of wavelet-coefficients in $(0,T)$ near $t = 0$. In both cases, extensions of the presently proposed, fast solution algorithms are available (e.g. [4]) which would yield a performance comparable to the algorithms proposed here.

Appendix A. Piecewise linear wavelets on the interval

Let $\{\phi_{j,k}\}_{k=0}^{2^j+1}$ denote the standard piecewise linear hat functions on the interval $[0,1]$ which is supposed to be subdivided into 2^j equidistant subintervals of length 2^{-j}. Then, the piecewise linear wavelets on level j with zero Dirichlet boundary conditions at $x = 0$ are given as follows.

A.1. Two vanishing moments.

- Left boundary wavelet:

$$\psi_{j,1} = \frac{5}{8}\phi_{j,1} - \frac{3}{4}\phi_{j,2} - \frac{1}{4}\phi_{j,3} + \frac{1}{4}\phi_{j,4} + \frac{1}{8}\phi_{j,5}$$

- Stationary wavelets ($k = 2, \dots, 2^{j-1} - 1$):

$$\psi_{j,k} = -\frac{1}{8}\phi_{j,2k-3} - \frac{1}{4}\phi_{j,2k-2} + \frac{3}{4}\phi_{j,2k-1} - \frac{1}{4}\phi_{j,2k} - \frac{1}{8}\phi_{j,2k+1}$$

- Right boundary wavelet:

$$\psi_{j,2^{j-1}} = -\frac{1}{16}\phi_{j,2^j-2} - \frac{1}{8}\phi_{j,2^j-1} + \frac{9}{16}\phi_{j,2^j} - \frac{3}{4}\phi_{j,2^j+1}$$

A.2. **Four vanishing moments.**

- Left boundary wavelets:

$$\begin{aligned}\psi_{j,1} &= \frac{63}{128}\phi_{j,1} - \frac{65}{64}\phi_{j,2} - \frac{1}{16}\phi_{j,3} + \frac{57}{64}\phi_{j,4} + \frac{13}{64}\phi_{j,5}\\ &\quad - \frac{31}{64}\phi_{j,6} - \frac{3}{16}\phi_{j,7} + \frac{7}{64}\phi_{j,8} + \frac{7}{128}\phi_{j,9}\\ \psi_{j,2} &= -\frac{7}{128}\phi_{j,1} - \frac{7}{64}\phi_{j,2} + \frac{21}{32}\phi_{j,3} - \frac{37}{64}\phi_{j,4} - \frac{11}{64}\phi_{j,5}\\ &\quad + \frac{15}{64}\phi_{j,6} + \frac{3}{32}\phi_{j,7} - \frac{3}{64}\phi_{j,8} - \frac{3}{128}\phi_{j,9}\end{aligned}$$

Stationary wavelets ($k = 3, \ldots, 2^{j-1} - 2$):

$$\begin{aligned}\psi_{j,k} &= \frac{3}{128}\phi_{j,2k-5} + \frac{3}{64}\phi_{j,2k-4} - \frac{1}{8}\phi_{j,2k-3} - \frac{19}{64}\phi_{j,2k-2} + \frac{45}{64}\phi_{j,2k-1}\\ &\quad - \frac{19}{64}\phi_{j,2k} - \frac{1}{8}\phi_{j,2k+1} + \frac{3}{64}\phi_{j,2k+2} + \frac{3}{128}\phi_{j,2k+3}\end{aligned}$$

- Right boundary wavelets:

$$\begin{aligned}\psi_{j,2^{j-1}-1} &= \frac{9}{512}\phi_{j,2^j-6} + \frac{9}{256}\phi_{j,2^j-5} - \frac{53}{512}\phi_{j,2^j-4} - \frac{31}{128}\phi_{j,2^j-3}\\ &\quad + \frac{345}{512}\phi_{j,2^j-2} - \frac{105}{256}\phi_{j,2^j-1} - \frac{45}{512}\phi_{j,2^j} + \frac{15}{64}\phi_{j,2^j+1}\\ \psi_{j,2^{j-1}} &= -\frac{5}{512}\phi_{j,2^j-6} - \frac{5}{256}\phi_{j,2^j-5} + \frac{67}{1536}\phi_{j,2^j-4} + \frac{41}{384}\phi_{j,2^j-3}\\ &\quad - \frac{53}{512}\phi_{j,2^j-2} - \frac{241}{768}\phi_{j,2^j-1} + \frac{875}{1536}\phi_{j,2^j} - \frac{35}{64}\phi_{j,2^j+1}\end{aligned}$$

Acknowledgments

M. Zank has been partially funded by the Austrian Science Fund (FWF) through project P 33477. Part of the work was done when M. Zank was a NAWI Graz PostDoc Fellow at the Institute of Applied Mathematics, TU Graz. M. Zank acknowledges NAWI Graz for the financial support.

References

1. P. Bansal, A. Moiola, I. Perugia, and C. Schwab, *Space-time discontinuous Galerkin approximation of acoustic waves with point singularities*, IMA J. Numer. Anal. **41** (2021), no. 3, 2056–2109.
2. J. Bramble, J. Pasciak, and J. Xu, *Parallel multilevel preconditioners*, Math. Comput. **55** (1990), 1–22.
3. H.-J. Bungartz and M. Griebel, *Sparse grids*, Acta Numer. **13** (2004), 1–123.
4. L. Chen, R.H. Nochetto, and J. Xu, *Optimal multilevel methods for graded bisection grids*, Numer. Math. **120** (2012), no. 1, 1–34.
5. A. Cohen, I. Daubechies, and J.-C. Feauveau, *Biorthogonal bases of compactly supported wavelets*, Pure Appl. Math. **45** (1992), 485–560.
6. W. Dahmen, *Wavelet and multiscale methods for operator equations*, Acta Numer. **6** (1997), 55–228.
7. W. Dahmen, H. Harbrecht, and R. Schneider, *Compression techniques for boundary integral equations — asymptotically optimal complexity estimates*, SIAM J. Numer. Anal. **43** (2006), no. 6, 2251–2271.
8. W. Dahmen and A. Kunoth, *Multilevel preconditioning*, Numer. Math. **63** (1992), no. 3, 315–344.

9. W. Dahmen, A. Kunoth, and K. Urban, *Biorthogonal spline-wavelets on the interval – stability and moment conditions*, Appl. Comp. Harm. Anal. **6** (1999), 259–302.
10. W. Dahmen, S. Prößdorf, and R. Schneider, *Multiscale methods for pseudodifferential equations on smooth manifolds*, Proceedings of the International Conference on Wavelets: Theory, Algorithms, and Applications (C.K. Chui, L. Montefusco, and L. Puccio, eds.), 1995, pp. 385–424.
11. W. Dahmen and R. Schneider, *Wavelets with complementary boundary conditions. function spaces on the cube*, Results Math. **34** (1998), 255–293.
12. R. Dautray and J.-L. Lions, *Mathematical analysis and numerical methods for science and technology. Vol. 5*, Springer-Verlag, Berlin, 1992.
13. A. George, *Nested dissection of a regular finite element mesh*, SIAM J. Numer. Anal. **2** (1973), no. 10, 345–363.
14. M. Griebel and H. Harbrecht, *On the construction of sparse tensor product spaces*, Math. Comput. **82** (2013), no. 282, 975–994.
15. M. Griebel and P. Oswald, *On the abstract theory of additive and multiplicative Schwarz algorithms*, Numer. Math. **70** (1995), no. 2, 163–180.
16. W. Hackbusch, *A sparse matrix arithmetic based on $\mathcal{H}$-matrices. Part I: Introduction to $\mathcal{H}$-matrices*, Computing **62** (1999), no. 2, 89–108.
17. W. Hackbusch and B.N. Khoromskij, *A sparse $\mathcal{H}$-matrix arithmetic. General complexity estimates*, J. Comput. Appl. Math. **125** (2000), no. 1–2, 479–501.
18. H. Harbrecht and M.D. Multerer, *A fast direct solver for nonlocal operators in wavelet coordinates*, J. Comput. Phys. **428** (2021), 110056.
19. H. Harbrecht, R. Schneider, and C. Schwab, *Multilevel frames for sparse tensor product spaces*, Numer. Math. **110** (2008), 199–220.
20. J.I.M. Hauser and M. Zank, *Numerical study of conforming space-time methods for Maxwell's equations*, Numer. Methods Partial Differ. Equ. **40** (2024), no. 2, e23070.
21. S. Jaffard, *Wavelet methods for fast resolution of elliptic equations*, SIAM J. Numer. Anal. **29** (1992), 965–986.
22. U. Langer and M. Zank, *Efficient direct space-time finite element solvers for parabolic initial-boundary value problems in anisotropic sobolev spaces*, SIAM J. Sci. Comput. **4** (2021), no. 43, A2714–A2736.
23. J.-L. Lions and E. Magenes, *Non-homogeneous boundary value problems and applications. Volume I.*, Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen, vol. 181, Springer, New York-Heidelberg, 1972.
24. R.J. Lipton, D.J. Rose, and R.E. Tarjan, *Generalized nested dissection*, SIAM J. Numer. Anal. **2** (1979), no. 16, 346–358.
25. R. Löscher, O. Steinbach, and M. Zank, *On a modified Hilbert transformation, the discrete inf-sup condition, and error estimates*, [math.NA] arXiv:2402.08291, arXiv.org, 2024.
26. I. Perugia, C. Schwab, and M. Zank, *Exponential convergence of hp-time-stepping in space-time discretizations of parabolic PDEs*, ESAIM Math. Model. Numer. Anal. **57** (2023), no. 1, 29–67.
27. R.A. Ryan, *Introduction to tensor products of Banach spaces*, Springer Monographs in Mathematics, Springer-Verlag London, Ltd., London, 2002.
28. Y. Saad and M.H. Schultz, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput. **7** (1986), no. 3, 856–869.
29. R. Schneider, *Multiskalen- und Wavelet-Matrixkompression*, Advances in Numerical Mathematics, B.G. Teubner, Stuttgart, 1998.
30. D. Schötzau and C. Schwab, *Time discretization of parabolic problems by the hp-version of the discontinuous Galerkin finite element method*, SIAM J. Numer. Anal. **38** (2000), no. 3, 837–875.
31. C. Schwab and R. Stevenson, *Space-time adaptive wavelet methods for parabolic evolution problems*, Math. Comput. **78** (2009), no. 267, 1293–1318.
32. ———, *Fractional space-time variational formulations of (Navier-) Stokes equations*, SIAM J. Math. Anal. **49** (2017), no. 4, 2442–2467.
33. O. Steinbach and M. Zank, *Coercive space-time finite element methods for initial boundary value problems*, Electron. Trans. Numer. Anal. **52** (2020), 154–194.
34. ———, *A note on the efficient evaluation of a modified Hilbert transformation*, J. Numer. Math. **29** (2021), no. 1, 47–61.

35. R. Stevenson, R. van Venetië, and J. Westerdiep, *A wavelet-in-time, finite element-in-space adaptive method for parabolic evolution equations*, Adv. Comput. Math. **48** (2022), no. 3, Paper No. 17, 43.
36. V. Thomée, *Galerkin finite element methods for parabolic problems*, 2nd ed., Springer Series in Computational Mathematics, vol. 25, Springer, Berlin, 2006.
37. M. Zank, *Inf-sup stable space-time methods for time-dependent partial differential equations*, Monographic Series TU Graz: Computation in Engineering and Science, vol. 36, TU Graz, Austria, 2020.
38. ———, *An exact realization of a modified Hilbert transformation for space-time methods for parabolic evolution equations*, Comput. Methods Appl. Math. **21** (2021), no. 2, 479–496.
39. ———, *Integral representations and quadrature schemes for the modified Hilbert transformation*, Comput. Methods Appl. Math. **23** (2023), no. 2, 473–489.
40. A. Zeiser, *Fast matrix-vector multiplication in the sparse-grid Galerkin method*, J. Sci. Comput. **47** (2011), no. 3, 328–346.

H. Harbrecht, Department of Mathematics and Computer Science, University of Basel, Basel, Switzerland
Email address: helmut.harbrecht@unibas.ch

Ch. Schwab, Seminar for Applied Mathematics, ETH Zürich, Zürich, Switzerland
Email address: schwab@math.ethz.ch

M. Zank, Institute of Applied Mathematics, TU Graz, Graz, Austria
Email address: zank@math.tugraz.at