Exponential convergence of hp-Time-stepping in space-Time discretizations of parabolic PDES

Journal Article

Author(s):
Perugia, Ilaria; Schwab, Christoph; Zank, Marco

Publication date:
2023-01

Permanent link:
https://doi.org/10.3929/ethz-b-000594935

Rights / license:
Creative Commons Attribution 4.0 International

Originally published in:
ESAIM: Mathematical Modelling and Numerical Analysis 57(1), https://doi.org/10.1051/m2an/2022081
EXPONENTIAL CONVERGENCE OF \( h/p \)-TIME-STEPPING IN SPACE-TIME DISCRETIZATIONS OF PARABOLIC PDES*

ILARIA PERUGIA\(^1\), CHRISTOPH SCHWAB\(^2\) AND MARCO ZANK\(^1, **\)

Abstract. For linear parabolic initial-boundary value problems with self-adjoint, time-homogeneous elliptic spatial operator in divergence form with Lipschitz-continuous coefficients, and for incompatible, time-analytic forcing term in polygonal/polyhedral domains \( D \), we prove time-analyticity of solutions. Temporal analyticity is quantified in terms of weighted, analytic function classes, for data with finite, low spatial regularity and without boundary compatibility. Leveraging this result, we prove exponential convergence of a conforming, semi-discrete \( h/p \)-time-stepping approach. We combine this semi-discretization in time with first-order, so-called “\( h \)-version” Lagrangian Finite Elements with corner-refinements in space into a tensor-product, conforming discretization of a space-time formulation. We prove that, under appropriate corner- and corner-edge mesh-refinement of \( D \), error vs. number of degrees of freedom in space-time behaves essentially (up to logarithmic terms), to what standard FEM provide for one elliptic boundary value problem solve in \( D \). We focus on two-dimensional spatial domains and comment on the one- and the three-dimensional case.

Mathematics Subject Classification. 65N30, 65J15.

Received March 22, 2022. Accepted September 27, 2022.

1. Introduction

Efficient numerical solution of parabolic evolution problems is required in many applications. In addition to the plain numerical solution of associated initial-boundary value problems, in recent years the efficient numerical treatment of optimal control problems and of uncertain input data has been considered. Here, often a large number of cases needs to be treated, and the (numerical) solution must be stored in a data-compressed format. Rather than the (trivial) option of a posteriori compressing a numerical solution obtained by a standard scheme, novel algorithms have emerged featuring some form of space-time compressibility in the numerical solution process. I.e., the numerical scheme will obtain directly, at runtime, a numerical solution in a compressed format. As examples, we mention only sparse-grid and wavelet-based methods (e.g., \cite{19}), and wavelet-based compressive schemes (e.g., \cite{20, 31} and the references there). Key to successful compressive space-time discretizations is an appropriate variational formulation of the evolution problem under consideration. Accordingly, recent years have seen the development of a variety of, in general nonequivalent, space-time variational formulations of parabolic initial-boundary value problems. Departing from the classical, Bochner-space perspective used to establish

---

Keywords and phrases. Parabolic IBVP, space-time methods, \( h/p \)-FEM, exponential convergence.

---

1 Faculty of Mathematics, University of Vienna, Vienna, Austria.

2 Seminar for Applied Mathematics, ETH Zürich, 8092 Zürich, Switzerland.

* Dedicated to the memory of Dominik Schötzau

** Corresponding author: marco.zank@univie.ac.at

© The authors. Published by EDP Sciences, SMAI 2023

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
I. PERUGIA

well-posedness, the novel formulations adopt the perspective of treating the parabolic evolution problem as an operator equation between appropriate function spaces, the primary motivation being accommodation of efficient, compressive space-time numerical schemes. We mention only \cite{3, 9, 15, 16, 22, 26, 31, 33, 35, 37} and the references there. A comprehensive account of the numerical analysis of fixed order time-discretizations is provided in \cite{38} and the references there. In the results given in that volume, the semigroup perspective is adopted, and the mathematical setting is based on homogeneous Sobolev spaces $\dot{H}^s(D)$, which impose implicit boundary compatibilities of regular data, see Chapter 19 of \cite{38}.

The presently investigated time-discretization approach is based on the space-time variational formulation in \cite{35}. It is of Petrov–Galerkin type, and is based on a fractional order Sobolev space in the temporal direction. It has been proposed and developed in a series of papers \cite{21, 34–36, 40}. We briefly recapitulate it here, and refer to \cite{35} for full development of details. The compressive aspect is here realized by the $hp$-time discretization for this formulation. One motivation for considering this formulation is that it was used for the analysis of the resulting boundary integral operators; see \cite{10}. Thus, conforming discretizations of this variational formulation in fractional order Sobolev spaces are a natural choice for the interior problems of FEM-BEM couplings for transmission problems. Moreover, when combined with a Hilbert-type transformation in time of the test functions, this formulation is unconditionally stable for any conforming space-time finite element space, see \cite{35}.

Throughout, we denote by $D \subset \mathbb{R}^d$ a bounded interval (if $d = 1$), or a bounded polygonal (if $d = 2$) or polyhedral (if $d = 3$) domain, with a Lipschitz boundary $\Gamma = \partial D$ consisting of a finite number of plane faces, and by $T > 0$ a finite time horizon. In the space-time cylinder $Q = (0, T) \times D$, we consider the parabolic initial-boundary value problem (IBVP for short) governed by the partial differential equation

$$Bu := \partial_t u + A(\partial_x)u = g \quad \text{in} \quad (0, T) \times D. \quad (1)$$

Here, the forcing function $g : Q \to \mathbb{R}$ is assumed to belong to $A([0, T]; L^2(D))$, i.e., it is analytic as a map from $[0, T]$ into $L^2(D)$. The spatial differential operator $A(\partial_x)$ is assumed linear, self-adjoint, in divergence form, i.e.,

$$A(\partial_x) = -\nabla_x \cdot (A(x) \nabla_x)$$

with $A \in L^\infty(D; \mathbb{R}^{d \times d})$ being a symmetric, positive definite matrix function of $x \in D$ which does not depend on the temporal variable $t$. The PDE (1) is completed by initial condition

$$u|_{t=0} = u_0, \quad (2)$$

and by mixed boundary conditions

$$\gamma_0(u) = u_D \quad \text{on} \quad \Gamma_D, \quad \gamma_1(u) = u_N \quad \text{on} \quad \Gamma_N. \quad (3)$$

Here, $\Gamma_D$ and $\Gamma_N$ denote a partitioning of $\Gamma = \partial D$ into a Dirichlet and a Neumann part, $\gamma_0$ denotes the Dirichlet trace map, and $\gamma_1$ denotes the conormal trace operator, given (in strong form) by $\gamma_1(v) = n_x \cdot (A(x)\nabla_x v)|_{\Gamma}$, with $\Gamma = \partial D$ denoting the boundary of $D$, and $n_x \in L^\infty(\Gamma; \mathbb{R}^d)$ the exterior unit normal vector field on $\Gamma$.

**Remark 1.1.** In the rest of this paper, the results are formulated for $u_0 = 0$, $u_D = 0$, and $u_N = 0$. Since the IBVP (1)–(3) is linear, superposition for a sufficiently regular function $U(x, t)$ in $Q$, which satisfies (2) and (3), will imply that the function $u - U$ will solve (1)–(3) with $g - BU$ in place of $g$ in (1), and with homogeneous initial and boundary data in (2) and (3). All regularity hypotheses which we will impose below on the source term $g$ in (1) (in particular, time-analyticity (39)) entail via $U$ corresponding assumptions on $u_0$, $u_D$, and $u_N$.

Exploiting the analytic semigroup property of the parabolic evolution operator, we provide in Section 3.1 sufficient conditions for the time-analyticity of solutions when considered as maps from the time interval $[0, T]$ into a suitable Sobolev space $W \subset L^2(D)$ on the bounded spatial domain $D \subset \mathbb{R}^d$.

**Contributions of the present paper** are a weighted analytic, temporal regularity analysis based on the analytic semigroup theory for linear, parabolic evolution equations, for source terms and coefficients of finite spatial
regularity, and the proof of exponential convergence of a temporal $hp$-discretization. For polygonal spatial
domain $D \subset \mathbb{R}^2$, and for data without boundary compatibility, we establish a priori convergence rate bounds
for fully discrete, space-time approximations which are based on a fractional order space-time formulation, on
$hp$-time-stepping and on $h$-FEM with corner-refined, regular graded triangulations in $D$. The diffusion coefficient
$A(x)$ is assumed to be independent of $t$, and to belong to $W^{1,\infty}(D; \mathbb{R}^{2\times 2})$. We comment on the cases $d = 1$
(when $D$ is a bounded interval) and $d = 3$ (when $D$ is a polyhedron).

The layout of this paper is as follows: In Section 2, we introduce notation and function spaces of tensor product
and of Bochner type, which will be used in the following. We also provide the space-time variational formulation
in fractional order spaces and the subspaces used in discretization. Section 3 addresses the solution regularity,
with particular attention to temporal analytic regularity in weighted, analytic Bochner spaces of functions taking
values in corner-weighted, Kondrat’ev type spaces on the domain $D$. Section 4 then introduces the Galerkin
approximations in space and time that will be used, and their approximation properties. Section 5 contains the
main results on the convergence rate of the discretization. Section 6 describes the numerical realization of the
nonlocal temporal bilinear form, and reports numerical results which are in full agreement with the convergence
rate analysis.

We use standard notation: $\mathbb{N} = \{1, 2, \ldots \}$ shall denote the natural numbers, and $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. For Banach
spaces $X$ and $Y$, $\mathcal{L}(X,Y)$ denotes the space of bounded linear operators from $X$ to $Y$, and $X' := \mathcal{L}(X, \mathbb{R})$
denotes the dual of $X$. For $q \in [1, \infty]$, the usual notation $L^q(D)$ is adopted for Lebesgue spaces of $q$-integrable
functions $u: D \to \mathbb{R}$ over some (bounded) domain $D$ in the Euclidean space $\mathbb{R}$. For nonnegative integers $k$,
Hilbertian Sobolev spaces (where $q = 2$) on such domains $D$ are denoted by $H^k(D)$. For $k = 0$, as usual,
$H^0(D) = L^2(D)$. Hilbertian Sobolev spaces of noninteger order $s = k + \theta$ for $k \in \mathbb{N}_0$ and $0 < \theta < 1$ are defined
by interpolation (real method, with fine index 2).

2. FUNCTION SPACES AND SPACE-TIME VARIATIONAL FORMULATION

We introduce several Bochner-type Sobolev spaces in the space-time cylinder $Q := J \times D$, with the finite
time interval $J := (0, T)$ and the bounded spatial domain $D \subset \mathbb{R}^d$.

2.1. Function spaces

Bochner-type function spaces defined on the space-time cylinder $Q = J \times D$ are spaces of strongly measurable
maps $u: J \to H^l(D)$, such that $u \in H^k(J; H^l(D))$ for nonnegative integers $k,l$. Due to the Hilbertian structure
of $H^k$, these separable Hilbert spaces admit tensor product structure, i.e.,

$$H^k(J; H^l(D)) \simeq H^k(J) \otimes H^l(D) \simeq H^l(D; H^k(J)),$$

where $\simeq$ denotes (isometric) isomorphism and $\otimes$ the Hilbertian tensor product.

For any integer $k \geq 1$, we denote by $H^k_0(D)$ the closed subspace of $H^k$ of functions with homogeneous boundary
values in the sense of closure of $C_0^\infty$ with respect to the norm of $H^k$. For instance, $H^1_0$ denotes the closed
nullspace of the Dirichlet trace operator $\gamma_0$.

To consider mixed boundary value problems on $D$, we partition $\Gamma = \partial D$ into two disjoint pieces $\Gamma_D$ and $\Gamma_N$.
Assuming positive $(d-1)$-dimensional measure of $\Gamma_D$ if $d = 2, 3$, or that $\Gamma_D$ contains at least one endpoint of
$D$ if $d = 1$, we set

$$H^1_D(D) := \left\{v \in H^1(D) \mid \gamma_0(v)|_{\Gamma_D} = 0 \right\}.$$

Evidently, for $\Gamma_D \subset \Gamma$, $H^0_\Gamma(D) = H^1_\Gamma(D) \subset H^1_D(D) \subset H^1(D)$.

In the following, we introduce Sobolev spaces for functions defined on an interval $(a, b) \subset \mathbb{R}$ with $a < b$.
For simplicity, we consider real-valued functions $v: (a, b) \to \mathbb{R}$. All results and proofs can be generalized
straightforwardly to $X$-valued functions $v: (a, b) \to X$ for a Hilbert space $X$, i.e., Bochner–Sobolev spaces. We
write

\[ H^1_0(a, b) = H^1_{\{a\}}(a, b) = \{ v \in H^1(a, b) \mid v(a) = 0 \}, \]
\[ H^1_{\partial b}(a, b) = H^1_{\{b\}}(a, b) = \{ v \in H^1(a, b) \mid v(b) = 0 \}. \]

In either of these two spaces, the seminorm \( | \circ |_{H^1(a, b)} = \| \partial_t \circ \|_{L^2(a, b)} \) is a norm. Thus, \( | \circ |_{H^1(a, b)} \) is considered as the norm in \( H^1_0(a, b) \) and \( H^1_{\partial b}(a, b) \), whereas the space \( H^1(a, b) \) is endowed with the norm \( \| \circ \|_{H^1(a, b)} = (\| \circ \|_{L^2(a, b)} + \| \partial_t \circ \|_{L^2(a, b)})^{1/2} \).

Fractional order spaces shall be defined by interpolation, via the real method of interpolation (see, e.g., [39], Chap. 1). We use the fine index \( q = 2 \) to preserve the Hilbertian structure. Of particular interest will be the space

\[ H^{1/2}_0(a, b) := (H^1_0(a, b), L^2(a, b))_{1/2, 2}, \]

where \( | \circ |_{H^1(a, b)} = \| \partial_t \circ \|_{L^2(a, b)} \) is the norm of the space \( H^1_0(a, b) \). The Sobolev space \( H^{1/2}_0(a, b) \) is a Hilbert space endowed with the interpolation norm (see [35], Sect. 2.3 for \((a, b) = (0, T)\)) defined by

\[ \| v \|_{H^{1/2}_0(a, b)} := \left( \sum_{k=0}^{\infty} \frac{\pi(2k+1)}{2(b-a)} |v_k|^2 \right)^{1/2}, \quad v \in H^{1/2}_0(a, b), \]  

where the Fourier coefficients \( v_k \) are given by \( v_k = \int_a^b v(s)V_k(s) \, ds \). Here, we use that any \( z \in L^2(a, b) \) admits a representation as a Fourier series

\[ z(t) = \sum_{k=0}^{\infty} z_k V_k(t), \quad z_k = \int_a^b z(s)V_k(s) \, ds, \quad k \in \mathbb{N}_0, \]  

where \( V_k \) denotes an eigenfunction corresponding to eigenvalue \( \lambda_k = \frac{x^2(2k+1)^2}{4(b-a)^2} \) of

\[ -\partial_t V_k(t) = \lambda_k V_k(t) \quad \text{for} \ t \in (a, b), \quad V_k(a) = \partial_t V_k(b) = 0, \quad \| V_k \|_{L^2(a, b)} = 1. \]  

In particular for \( J = (0, T) = (a, b) \), we have

\[ \| v \|_{H^{1/2}_0(J)} = \left( \frac{\pi}{2T} \sum_{k=0}^{\infty} (2k+1)|v_k|^2 \right)^{1/2}, \quad v \in H^{1/2}_0(J), \]

with the Fourier representation

\[ v(t) = \sum_{k=0}^{\infty} v_k \sqrt{\frac{2}{T}} \sin \left( \left( \frac{\pi}{2} + k\pi \right) \frac{t}{T} \right), \quad v_k = \int_0^T v(s) \sqrt{\frac{2}{T}} \sin \left( \left( \frac{\pi}{2} + k\pi \right) \frac{s}{T} \right) \, ds. \]  

Analogous to \( H^{1/2}_0(J) \), the Hilbert space \( H^{1/2}_0(J) := (H^1_0(J), L^2(J))_{1/2, 2} \) is endowed with the Hilbertian norm (see [35], Sect. 2.3) defined by

\[ \| w \|_{H^{1/2}_0(J)} := \left( \frac{\pi}{2T} \sum_{k=0}^{\infty} (2k+1)|w_k|^2 \right)^{1/2}, \quad w \in H^{1/2}_0(J), \]

where the Fourier coefficients are given by \( w_k = \int_0^T w(s) \sqrt{\frac{2}{T}} \cos \left( \left( \frac{\pi}{2} + k\pi \right) \frac{s}{T} \right) \, ds. \)
To prove exponential convergence of a temporal $hp$-discretization, we need further investigations of the Sobolev space $H^{1/2}_{0,1}(a,b)$ and its norm $\| \circ \|_{H^{1/2}_{0,1}(a,b)}$. For this purpose, let the classical Sobolev space $H^{1/2}(a,b)$ be endowed with the Slobodetskii norm ([25], p. 74)

$$\| v \|_{H^{1/2}(a,b)} := \left( \| v \|_{L^2(a,b)}^2 + |v|_{H^{1/2}(a,b)}^2 \right)^{1/2}$$

for $v \in H^{1/2}(a,b)$ with

$$|v|_{H^{1/2}(a,b)} := \left( \int_a^b \int_a^b \frac{|v(s) - v(t)|^2}{|s-t|^2} \, ds \, dt \right)^{1/2}.$$  

With the Slobodetskii norm (8), we endow $H^{1/2}_{0,1}(a,b)$ with the norm

$$\| v \|_{H^{1/2}_{0,1}(a,b)} := \left( \| v \|_{L^2(a,b)}^2 + |v|_{H^{1/2}(a,b)}^2 + \int_a^b \frac{|v(t)|^2}{t-a} \, dt \right)^{1/2}$$

for $v \in H^{1/2}_{0,1}(a,b)$. We have the following equivalence result for the norms defined in (4) and (10), which is proven, e.g., in [25] (see the proof in Appendix A for the characterization of the equivalence constants).

**Lemma 2.1.** There are constants $C_{\text{Int,1}}, C_{\text{Int,2}} > 0$, which are independent of $a,b$, such that

$$C_{\text{Int,1}} \| v \|_{H^{1/2}_{0,1}(a,b)} \leq \| v \|_{H^{1/2}_{0,1}(a,b)} \leq C_{\text{Int,2}} \sqrt{1 + \frac{4(b-a)^2}{\pi^2}} \| v \|_{H^{1/2}_{0,1}(a,b)}$$

for all $v \in H^{1/2}_{0,1}(a,b)$.

The next result is used for the proof of the temporal $hp$-error estimate in Section 5. It localizes the $H^{1/2}(a,b)$ norm in a certain sense. We report its proof in Appendix A, and refer to [14] for a more general localization result.

**Lemma 2.2.** For a number $\tau \in (a,b)$, the estimate

$$|v|^2_{H^{1/2}(a,b)} \leq |v|^2_{H^{1/2}(a,\tau)} + 4 \int_a^\tau \frac{|v(t)|^2}{\tau-t} \, dt + 4 \int_\tau^b \frac{|v(s)|^2}{s-\tau} \, ds + |v|^2_{H^{1/2}(\tau,b)}$$

holds true for $v \in H^{1/2}(a,b)$, if all occurring integrals on the right side exist.

### 2.2. Hilbert transformation $\mathcal{H}_T$

A key role in the space-time variational formulation of IBVP (1) is taken by the nonlocal operator $\mathcal{H}_T \in \mathcal{L}(L^2(J), L^2(J))$, which is defined by

$$(\mathcal{H}_Tv)(t) := \sum_{k=0}^\infty v_k \sqrt{\frac{2}{T}} \cos \left( \frac{\pi}{2} + k\pi \right) \frac{t}{T}, \quad t \in J.$$  

(11)

Here, $v \in L^2(J)$ and its Fourier coefficients $v_k = \int_0^T v(s) \sqrt{\frac{2}{T}} \sin \left( \frac{\pi}{2} + k\pi \right) \frac{s}{T} \, ds$ are represented as in (7). We collect some properties of $\mathcal{H}_T$. 


Proposition 2.3 ([35], Sect. 2.4, [36, 40]). The modified Hilbert transformation \( \mathcal{H}_T \) defined in (11) is a linear isometry as mapping
\[
\mathcal{H}_T : H^\nu_0(J) \to H^\nu_0(J) \quad \text{for } \nu \in \{0, 1/2, 1\}
\]
and is \( H^{1/2}_0(J) \)-elliptic, satisfying
\[
\forall v \in H^{1/2}_0(J) : \langle \partial_t v, \mathcal{H}_T v \rangle_{L^2(J)} = \|v\|_{H^{1/2}_0(J)}^2.
\]

Additionally, \( \mathcal{H}_T \) fulfills the following properties:
\[
\begin{align*}
\forall v, w \in H^{1/2}_0(J) : & \quad \langle \partial_t w, \mathcal{H}_T v \rangle_{L^2(J)} = \langle \mathcal{H}_T w, \partial_t v \rangle_{L^2(J)} = \langle w, v \rangle_{H^{1/2}_0(J)}, \\
\forall w \in H^1_0(J), \forall v \in L^2(J) : & \quad \langle \partial_t \mathcal{H}_T w, v \rangle_{L^2(J)} = -\langle \mathcal{H}_T^{-1} \partial_t w, v \rangle_{L^2(J)}, \\
\forall v, w \in L^2(J) : & \quad \langle \mathcal{H}_T v, w \rangle_{L^2(J)} = \langle v, \mathcal{H}_T^{-1} w \rangle_{L^2(J)}, \\
\forall v \in L^2(J) : & \quad \langle v, \mathcal{H}_T v \rangle_{L^2(J)} \geq 0, \\
\forall \nu \in \{1/2, 1\}, \forall v \in H^\nu_0(J), v \neq 0 : & \quad \langle v, \mathcal{H}_T v \rangle_{L^2(J)} > 0.
\end{align*}
\]

Remark 2.4. We remark that (12)–(18) are valid for all \( T > 0 \). In particular, these identities remain stable under passage to the limit \( T \to \infty \), with appropriate modifications of spaces. We refer to [12] for a space-time variational formulation and a discussion of a Petrov–Galerkin discretization for the resulting limiting problems.

2.3. Model scalar initial value problem

In \( J = (0, T) \), for a given right-hand side \( f \), consider the scalar IVP to find a function \( u : J \to \mathbb{R} \) such that
\[
\partial_t u = f \quad \text{in } J, \quad u(0) = 0.
\]

A weak formulation relevant for treatment of IBVP (1) is to find \( u \in H^{1/2}_0(J) \) such that
\[
\forall w \in H^{1/2}_0(J) : \langle \partial_t u, w \rangle_{L^2(J)} = \langle f, w \rangle_{L^2(J)}
\]
for given \( f \in [H^{1/2}_0(J)]' \). Here, \( \langle \cdot, \cdot \rangle_{L^2(J)} \) denotes the inner product in \( L^2(J) \) and as continuous extension of it, also the duality pairing with respect to \( [H^{1/2}_0(J)]' \) and \( H^{1/2}_0(J) \). The continuous bilinear form on the left side of (19) is inf-sup stable:
\[
\inf_{0 \neq u \in H^{1/2}_0(J)} \sup_{0 \neq w \in H^{1/2}_0(J)} \frac{\langle \partial_t u, w \rangle_{L^2(J)}}{\|u\|_{H^{1/2}_0(J)} \|w\|_{H^{1/2}_0(J)}} \geq 1.
\]

This is shown in Remark 2.10 of [35] by observing that, for every \( u \in H^{1/2}_0(J) \),
\[
\|u\|_{H^{1/2}_0(J)} = \frac{\langle \partial_t u, \mathcal{H}_T u \rangle_{L^2(J)}}{\|\mathcal{H}_T u\|_{H^{1/2}_0(J)}} \leq \sup_{0 \neq w \in H^{1/2}_0(J)} \frac{\langle \partial_t u, w \rangle_{L^2(J)}}{\|w\|_{H^{1/2}_0(J)}}.
\]

For every \( f \in [H^{1/2}_0(J)]' \), IVP (19) then admits a unique solution \( u \in H^{1/2}_0(J) \).

For the derivation of a space-time variational formulation of (1), it is useful to consider a parametric IVP: for a given parameter \( \mu \geq 0 \) (eventually in the spectrum of the spatial operator of (1)) and for \( f \in [H^{1/2}_0(J)]' \), find \( u \in H^{1/2}_0(J) \) such that \( \partial_t u + \mu u = f \) in \([H^{1/2}_0(J)]' \). A Petrov–Galerkin variational form of this problem is to find \( u \in H^{1/2}_0(J) \) such that
\[
\forall w \in H^{1/2}_0(J) : \langle \partial_t u, w \rangle_{L^2(J)} + \langle \mu u, w \rangle_{L^2(J)} = \langle f, w \rangle_{L^2(J)}.
\]
A Bubnov–Galerkin variational form with equal trial and test function spaces is to find \( u \in H^{1/2}_0(J) \) such that

\[
\forall v \in H^{1/2}_0(J) : \langle \partial_t u, \mathcal{H}Tv \rangle_{L^2(J)} + \mu \langle u, \mathcal{H}Tv \rangle_{L^2(J)} = \langle f, \mathcal{H}Tv \rangle_{L^2(J)}.
\]  

(22)

Both formulations (21) and (22) admit unique solutions due to (20) and (17).

2.4. Temporal hp-discretization

To discretize (22), we use some space \( V^M_t \subset H^{1/2}_0(J) \) of finite dimension \( M = \dim(V^M_t) \). In the hp-time discretization, we build \( V^M_t \) as follows: on a partition \( \mathcal{G} = \{ I_j \}_{j=1}^m \) of \( J \) into \( m \) time intervals \( I_j := (t_{j-1}, t_j) \), where \( 0 =: t_0 < t_1 < \cdots < t_m := T \), we choose \( V^M_t \) as a space of continuous, piecewise polynomials of degrees \( p_j \geq 1 \), which we collect in the degree vector \( p := (p_j)_{j=1}^m \in \mathbb{N}^m \). We define

\[
V^M_t = S^{p,1}_{0,0}(J; \mathcal{G}) := \left\{ v \in H^1_0(J) : v|_{I_j} \in P^{p_j}, I_j \in \mathcal{G} \right\}.
\]  

(23)

Here, continuity between adjacent time-intervals is required to ensure \( S^{p,1}_{0,0}(J; \mathcal{G}) \subset H^{1/2}_0(J) \). Then \( M = \dim(S^{p,1}_{0,0}(J; \mathcal{G})) = \left( \sum_{j=1}^m (p_j + 1) \right) - m = \sum_{j=1}^m p_j \).

We restrict (22) to \( V^M_t \) to obtain the temporal hp-approximation: find \( u^M_t \in V^M_t \) such that

\[
\forall v \in V^M_t : \langle \partial_t u^M_t, \mathcal{H}Tv \rangle_{L^2(J)} + \mu \langle u^M_t, \mathcal{H}Tv \rangle_{L^2(J)} = \langle f, \mathcal{H}Tv \rangle_{L^2(J)}.
\]  

(24)

Due to the inf-sup stability (20) and \( \mu \geq 0 \), the discretization (24) is well-posed with inf-sup constant independent of \( \mathcal{G} \) and \( p \). Its numerical implementation will require, similar to \([12,35]\), the efficient evaluation of \( \mathcal{H}Tv \) for \( v \in V^M_t \). We shall address this in Section 6.1 below.

2.5. Space-time variational formulation

We consider the source problem corresponding to the spatial part of (1). Its variational form reads: given a source term \( f \in L^2(D) \), find

\[
w \in H^1_D(D) \text{ such that } \forall v \in H^1_D(D) : a(w, v) = \langle f, v \rangle_{L^2(D)}.
\]  

(25)

Here, \( a(w, v) = \int_D A(x) \nabla w(x) \cdot \nabla x v(x) \, dx \). We assume uniform positive definiteness of \( A \):

\[
a_{\text{min}} := \text{ess inf}_{x \in D} \inf_{0 \neq \xi \in \mathbb{R}^d} \frac{\xi^\top A(x) \xi}{\xi^\top \xi} > 0.
\]  

(26)

With assumption (26), we have

\[
\forall w \in H^1_D(D) : a(w, w) \geq a_{\text{min}} \| \nabla x w \|^2_{L^2(D)} \geq a_{\text{min}} c \| w \|^2_{H^1(D)}
\]

due to \( |\Gamma_D| > 0 \) if \( d = 2,3 \) or \( \Gamma_D \neq \emptyset \) if \( d = 1 \), and the Poincaré inequality.

The spectral theorem and the symmetry \( a(w, v) = a(v, w) \) for all \( v, w \in H^1(D) \) ensure that the corresponding eigenvalue problem to find

\[
0 \neq \phi \in H^1_D(D), \mu \in \mathbb{R} : \forall v \in H^1_D(D) : a(\phi, v) = \mu \langle \phi, v \rangle_{L^2(D)}
\]  

(27)

admits a sequence of eigenpairs \( \{(\mu_k, \phi_k)\}_{k \geq 1} \) enumerated in increasing order of the real eigenvalues \( \mu_k > 0 \), repeated according to multiplicity, with the eigenfunctions \( \phi_k \) orthonormal in \( L^2(D) \) and orthogonal in \( H^1_D(D) \), and with \( \mu_k \) accumulating only at \( \infty \). In view of the forthcoming analysis, \textit{in what follows, we endow} \( H^1_D(D) \) \textit{with the “energy” norm} \( a(\cdot, \cdot)^{1/2} \). We remark that, for \( v \in H^1_D(D) \), \( a(v, v) = \sum_{i=1}^\infty \mu_i |v_i|^2 \), where \( v_i = \langle v, \phi_i \rangle_{L^2(D)} \),
The space-time variational formulation of (1) is based on the intersection space
\[ H_{Γ,D,0}^{1/2}(Q) := \left( L^2(J) ⊗ H^{1}_Γ(D) \right) \cap \left( H^{1/2}_{Γ,D}(J) ⊗ L^2(D) \right), \]
which we equip with the corresponding sum norm. The space \( H_{Γ,D,0}^{1/2}(Q) \) is defined analogously. Proceeding as in Theorem 3.2 of [35], the initial-boundary value problem (1)–(3) is set as a well-posed operator equation.

**Theorem 2.5.** Consider (1)–(3) with homogeneous data \( u_0 = 0 \) in (2) and \( u_D, u_N = 0 \) in (3). Assume \( |Γ_D| > 0 \) if \( d = 2, 3 \) or \( Γ_D ≠ \emptyset \) if \( d = 1 \), and that the coefficient \( A ∈ L^∞(D; \mathbb{R}^{d×d}) \) satisfies (26).

Then, the space-time variational formulation of (1) to find \( u ∈ H_{Γ,D,0}^{1/2}(Q) \) such that
\[ ∀ v ∈ H_{Γ,D,0}^{1/2}(Q) : \langle ∂_t u, v \rangle_{L^2(Q)} + \langle A\nabla_x u, \nabla_x v \rangle_{L^2(Q)} = \langle g, v \rangle_{L^2(Q)} \tag{28} \]
induces an isomorphism
\[ B := ∂_t + A(∂_x) ∈ \mathcal{L}_{iso}\left( H_{Γ,D,0}^{1/2}(Q), \left[ H_{Γ,D,0}^{1/2}(Q) \right]' \right). \]
In particular, for every \( g ∈ [H_{Γ,D,0}^{1/2}(Q)]' \), IBVP \( Bu = g \) in (1) admits a unique solution \( u ∈ H_{Γ,D,0}^{1/2}(Q) \).

We remark that \( ⟨\circ, \circ⟩_{L^2(Q)} \) denotes the inner product in \( L^2(Q) \) and as continuous extension of it, also the duality pairing with respect to \([H_{Γ,D,0}^{1/2}(Q)]' \) and \( H_{Γ,D,0}^{1/2}(Q) \). The space-time discretization of (28) is straightforward: for any conforming, spatial finite element subspace \( V^N_x ⊂ H^{1}_Γ(D) \) of finite dimension \( N \), and for the temporal \( hp \)-subspace \( V^M_t ⊂ H^{1/2}_D(J) \) introduced in (23), we restrict (28) to the space-time approximation space
\[ V^M_t ⊗ V^N_x ⊂ H_{Γ,D,0}^{1/2}(Q). \tag{29} \]
That is, we seek an approximate solution \( u^{MN} ∈ V^M_t ⊗ V^N_x \) such that
\[ \langle ∂_t u^{MN}, v \rangle_{L^2(Q)} + \langle A\nabla_x u^{MN}, \nabla_x v \rangle_{L^2(Q)} = \langle g, v \rangle_{L^2(Q)} \tag{30} \]
holds true for all \( v ∈ (\mathcal{H}_T V^M_t) ⊗ V^N_x ⊂ H_{Γ,D,0}^{1/2}(Q) \).

For these choices of test function spaces and for \( any \) subspace \( V^N_x ⊂ V \) of finite dimension \( N \), as in Section 3 of [35], existence and uniqueness of the discrete solution \( u^{MN} ∈ V^M_t ⊗ V^N_x ⊂ H_{Γ,D,0}^{1/2}(Q) \) of (30) follow from the continuous inf-sup condition
\[ \inf_{0 ≠ u ∈ H_{Γ,D,0}^{1/2}(Q)} \sup_{0 ≠ w ∈ H_{Γ,D,0}^{1/2}(Q)} \frac{⟨∂_t u, w⟩_{L^2(Q)} + ⟨A\nabla_x u, \nabla_x w⟩_{L^2(Q)}}{∥u∥_{H_{Γ,D,0}^{1/2}(Q)}∥w∥_{H_{Γ,D,0}^{1/2}(Q)}} ≥ \frac{1}{2} \]
With \( H^{1}_Γ(D) \) endowed with the \( a(\circ, \circ)^{1/2} \) norm, the proof of this condition with constant independent of \( A \) follows verbatim that of Theorem 3.2, Corollary 3.3 of [35] for the case \( A = \mathbb{I} \). Evidently, the stability of the discrete problem is a consequence of the choice of the test function space \( \mathcal{H}_T V^M_t \), whose efficient numerical realization will be discussed in Section 6.

### 3. Regularity

To obtain convergence rate bounds, we address the regularity of the solution \( u ∈ H_{Γ,D,0}^{1/2}(Q) \). We consider separately the temporal and spatial regularity. The solution operator to the parabolic equation (1) being an analytic semigroup, for time-analytic forcing \( g \) in (1) we expect time-analyticity of \( u \). This, in turn, is well-known to imply exponential convergence of \( hp \)-time-stepping as shown, e.g., in [12, 28] and the references there. We shall verify this in Sections 4 and 5 below.
3.1. Time-analyticity

We quantify the temporal analyticity of the solution $u : t \mapsto u(t)$ with $u(t) := u(t, \circ) \in L^2(D)$. To this end, we recall the eigenvalue problem (27). Setting $H := L^2(D)$ and thus denoting by $\langle \circ, \circ \rangle_H$ the $L^2(D)$ inner product, the solution $u(t)$ of (1) at time $t > 0$ for $g = 0$, $u_D = u_N = 0$, and for initial data $u_0 \in H$ may be written as

$$u(t) = E(t)u_0 := \sum_{i=1}^{\infty} \exp(-\mu_i t) \langle u_0, \phi_i \rangle_H \phi_i$$

(31)

with convergence of the series in $H$. The operators $\{E(t)\}_{t \geq 0}$ satisfy the semigroup property in $H$, i.e.,

$$\forall s, t > 0 : \ E(s + t) = E(s)E(t), \ E(0) = \text{Id}.$$

For $r \geq 0$, we define the scale of spaces $X_r \subset H = X_0$

$$X_r := \left\{ v \in H : \|v\|_{X_r}^2 := \sum_{i=1}^{\infty} \mu_i^r |v_i|^2 < \infty \right\}. \quad (32)$$

Here, $v_i = \langle v, \phi_i \rangle_H$ denotes the $i$-th coefficient in the eigenfunction expansion of $v$ (recall from (27) that the sequence $\{\phi_i\}_{i \geq 1}$ was assumed to be an orthonormal basis of $H = X_0$). We remark that the norm $\circ \| \cdot \|_{X_i}$ is the energy-norm on the space $V = \mathcal{H}_D^1(D)$, due to

$$\forall v \in V : \|v\|_{X_i}^2 = a(v, v) = \sum_{i=1}^{\infty} \mu_i |v_i|^2.$$

For $|\Gamma_D| > 0$ if $d = 2, 3$ or $\Gamma_D \neq \emptyset$ if $d = 1$, $\| \circ \|_{X_1}$ is equivalent to the $H^1(D)$ norm on $V$ and the norm bounds $\|v\|_{X_r} \leq c\|v\|_{X_{r'}}$ for $r' \geq r$ follow from (32) and the assumed enumeration of the real eigenvalues $\mu_i > 0$ with $\mu_i \uparrow \infty$ as $i \uparrow \infty$:

$$\|v\|_{X_r}^2 = \sum_{i=1}^{\infty} \mu_i^r |v_i|^2 \leq \left( \sup_{m \in \mathbb{N}} \mu_m^{r'} \right) \sum_{i=1}^{\infty} \mu_i^{r'} |v_i|^2 \leq \mu_1^{r-r'} \|v\|_{X_{r'}}^2. \quad (33)$$

For $\theta, r \geq 0$ and for any $t > 0$, $E(t)$ in (31) belongs to $\mathcal{L}(X_\theta, X_r)$. In fact, for any $t > 0$ and $v \in X_\theta$, we have

$$\|E(t)v\|_{X_r}^2 = \sum_{i=1}^{\infty} \mu_i^{r\theta} \exp(-2\mu_i t) |v_i|^2 = \sum_{i=1}^{\infty} \mu_i^{r\theta} \exp(-2\mu_i t) \mu_i^\theta |v_i|^2. \quad (34)$$

For $\theta \geq r \geq 0$, identity (34) implies

$$\|E(t)v\|_{X_r}^2 \leq \mu_1^{-(\theta-r)} \exp(-2\mu_1 t) \sum_{i=1}^{\infty} \mu_i^\theta |v_i|^2 = \mu_1^{-(\theta-r)} \exp(-2\mu_1 t) \|v\|_{X_\theta}^2$$

for all $v \in X_\theta$, i.e., $E(t) \in \mathcal{L}(X_\theta, X_r)$ for any $t > 0$ with

$$\forall \theta \geq r \geq 0, \forall t > 0 : \|E(t)v\|_{X_r}^2 \leq \mu_1^{-(\theta-r)} \exp(-2\mu_1 t). \quad (35)$$

For $r \geq \theta \geq 0$, for any $t > 0$ and $v \in X_\theta$, identity (34) implies

$$\|E(t)v\|_{X_r}^2 \leq \sup_{i \in \mathbb{N}} \{ \mu_i^{r\theta} \exp(-2\mu_i t) \} \sum_{i=1}^{\infty} \mu_i^\theta |v_i|^2 =: G_{r\theta}(t) \|v\|_{X_\theta}^2. \quad (36)$$
To provide an upper bound for \( G_{r-\theta}(t) \), we observe that, for fixed \( t, \sigma > 0 \), the function \( 0 < \mu \mapsto \mu^{2\sigma} \exp(-2\mu t) \) takes its maximum at \( \mu_* := \sigma/t \) whence

\[
\forall t > 0 : \quad G_{2\sigma}(t) \leq G_{\max}(\sigma, t) := [\mu_*^\sigma \exp(-\mu_* t)]^2 = \left( \frac{\sigma}{te} \right)^{2\sigma}. \tag{37}
\]

Inserting (37) with \( \sigma = (r - \theta)/2 > 0 \) into (36), we arrive at

\[
\forall r \geq \theta \geq 0, \forall t > 0 : \quad \| E(t) \|_{L(X_\theta, X_r)}^2 \leq \left( \frac{r - \theta}{2te} \right)^{r - \theta}.
\]

The exponential decay of the Fourier coefficients for \( t > 0 \) implied by the exponential weighting \( \exp(-\mu_* t) \) entails time-analyticity of the solution \( t \mapsto u(t) \) for \( t > 0 \). To prove exponential convergence rates of \( hp \)-approximation in \( J = (0, T) \), we quantify the time-regularity of the solution \( u \) of (1) for \( u_0 = 0 \) and \( u_D = u_N = 0 \) with the Duhamel representation (see, e.g., [27])

\[
u(t) = \int_0^t E(t-s)g(s) \, ds, \quad 0 < t \leq T.
\tag{38}
\]

We work under the following time-analyticity assumption on the forcing \( g \) in (1): There exist constants \( C > 0 \) and \( \delta \geq 1 \) such that, for some \( \varepsilon \in (0, 1) \), we have

\[
\forall l \in \mathbb{N}_0 : \quad \sup_{0 \leq t \leq T} \left\| g^{(l)}(t) \right\|_{X_\varepsilon} \leq C \delta^l \Gamma(l + 1), \tag{39}
\]

where \( \Gamma(o) \) denotes the gamma function fulfilling \( \Gamma(l) = (l - 1)! \) for all \( l \in \mathbb{N} \). Formally differentiating (38) \( l \)-times with respect to \( t \), upon writing it equivalently as \( u(t) = \int_0^t E(s)g(t-s) \, ds \), gives

\[
\frac{d^l}{dt^l} u(t) = \sum_{i=0}^{l-1} E^{(i)}(t)g^{(l-1-i)}(0) + \int_0^t E(s)g^{(l)}(t-s) \, ds, \quad l \in \mathbb{N}, t > 0.
\tag{40}
\]

The right limits at \( t = 0 \) of the time-derivatives of the forcing \( g \) in (1) contribute to the time-regularity. We estimate the norm of the operators \( E^{(l)}(t) \) in \( L(X_\theta, X_r) \).

**Lemma 3.1.** For \( r \geq \theta \geq 0 \), we have

\[
\forall l \in \mathbb{N}_0, \forall t > 0 : \quad \left\| E^{(l)}(t) \right\|_{L(X_\theta, X_r)}^2 \leq \frac{1}{\sqrt{2\pi}} \left( \frac{1}{2} \right)^{2l+r-\theta} \Gamma(2l + 1 + r - \theta) t^{-2l-(r-\theta)}. \tag{41}
\]

**Proof.** For \( v \in H = X_\theta \), with \( v_i = \langle v, \phi_i \rangle_H \), the time-derivative of order \( l \in \mathbb{N} \) applied to \( v(t) = E(t)v \) represented as in (31) yields (with formal, term-by-term differentiation)

\[
\frac{d^l}{dt^l} v(t) = \sum_{i=1}^\infty (-\mu_i)^l \exp(-\mu_i t) v_i \phi_i
\]

with convergence in \( H \) for arbitrary, fixed \( t > 0 \). Therefore

\[
\forall t > 0 : \quad \left\| v^{(l)}(t) \right\|_{X_r}^2 = \sum_{i=1}^\infty \mu_i^{2l+r-\theta} \exp(-2\mu_i t) \mu_i^\theta |v_i|^2.
\]

It follows from (37) that for every \( t > 0 \)

\[
\left\| v^{(l)}(t) \right\|_{X_r}^2 \leq G_{2l+[r-\theta]/2}(t) \left\| v \right\|_{X_\theta}^2 \leq G_{\max}(l+[r-\theta]/2, t) \left\| v \right\|_{X_\theta}^2.
\]
Therefore, for every $v \in X_\theta$ and every $r \geq \theta \geq 0$, we have
\[
\forall l \in \mathbb{N}, t > 0 : \quad \left\|v^{(l)}(t)\right\|_{X_r}^2 \leq \left(\frac{2l + r - \theta}{2te}\right)^{2l+r-\theta} \left\|v\right\|_{X_\theta}^2 = \left(\frac{1}{2}\right)^{2l+r-\theta} \left(\frac{2l + r - \theta}{e}\right)^{2l+r-\theta} t^{-2l-(r-\theta)} \left\|v\right\|_{X_\theta}^2.
\]
For $x \in \mathbb{R}_+$, the Stirling’s formula (C.1) states $\sqrt{2\pi} x^{x-1/2} e^{-x} \leq \Gamma(x)$, which implies $(x/e)^x \leq \frac{1}{\sqrt{2\pi}} x^{-1/2} \Gamma(x+1)$. With $x = 2l + r - \theta$, this gives the claimed bound, as $(2l + r - \theta)^{-1/2} \leq 1$.

**Lemma 3.2.** Assume (39) with some $\varepsilon \in (0, 1)$ and some $\delta \geq 1$. For $r \in [0, 2]$, there exists a constant $C > 0$ (independent of $\delta, l, t$) such that, for every $l \in \mathbb{N}_0$ and $t > 0$, we have
\[
\left\|u^{(l)}(t)\right\|_{X_r} \leq C\delta l \Gamma(l+1) \left(t^{(2-r+\min(\varepsilon, \varepsilon))}/2 + \sum_{i=0}^{l-1} t^{-i-r/2+\varepsilon/2}\right).
\]
For $l = 0$, this bound is valid without the sum.

**Proof.** From (40), we estimate for every $0 < t \leq T$
\[
\left\|u^{(l)}(t)\right\|_{X_r} \leq \sum_{i=0}^{l-1} \left\|E^{(i)}(t)\right\|_{\ell(L(X_r, X_r))} \left\|g^{(l-i-1)}(0)\right\|_{X_r} + \int_0^t \left\|E(s)\right\|_{\ell(L(X_r, X_r))} \left\|g^{(l)}(t-s)\right\|_{X_r} ds.
\]
To estimate the sum, we use (41) with $\theta = \varepsilon$ and assumption (39) and obtain
\[
\sum_{i=0}^{l-1} \left\|E^{(i)}(t)\right\|_{\ell(L(X_r, X_r))} \left\|g^{(l-i-1)}(0)\right\|_{X_r} \leq \sum_{i=0}^{l-1} C \left(\frac{1}{2}\right)^{i+r/2-\varepsilon/2} \Gamma(2i+1+r-\varepsilon)^{1/2} t^{-i-r/2+\varepsilon/2} \delta^{l-i} \Gamma(l-i)
\]
\[
\leq C\delta^{l-1} \sum_{i=0}^{l-1} \left(\frac{1}{2}\right)^{i+r/2-\varepsilon/2} \Gamma(2i+1+r-\varepsilon)^{1/2} \Gamma(l-i) t^{-i-r/2+\varepsilon/2}
\]
\[
\leq C\delta^{l-1} \sum_{i=0}^{l-1} \Gamma(i+1+r/2-\varepsilon/2) \Gamma(l-i) t^{-i-r/2+\varepsilon/2}
\]
\[
\leq C\delta^{l-1} \Gamma(l+1) \sum_{i=0}^{l-1} t^{-i-r/2+\varepsilon/2},
\]
where in the third inequality we have used the duplication formula $\Gamma(z)\Gamma(z+1/2) = 2^{1-2z}\sqrt{\pi}\Gamma(2z)$ with $z = (2i+1+r-\varepsilon)/2$, and the fourth inequality follows from $\max_{0 \leq i \leq l-1} \Gamma(i+1+r/2-\varepsilon/2) \Gamma(l-i) \leq \max_{0 \leq i \leq l-1} \Gamma(i+2) \Gamma(l-i) \leq \Gamma(l+1)$.

To estimate the integral term for $\varepsilon \leq r \leq 2$, we use assumption (39) with $\varepsilon \in (0, 1)$ and (41) with $l = 0$, $\theta = \varepsilon > 0$ and $\varepsilon \leq r \leq 2$, and obtain, for every $l \in \mathbb{N}_0$,
\[
\int_0^t \left\|E(s)\right\|_{\ell(L(X_r, X_r))} \left\|g^{(l)}(t-s)\right\|_{X_r} ds \leq C\delta \Gamma(l+1) \int_0^t s^{-(r-\varepsilon)/2} ds = CC_\varepsilon \Gamma(l+1) t^{(2-r+\varepsilon)/2}
\]
with $C_\varepsilon = 2/(2-r+\varepsilon)$. It remains to estimate the integral term for $0 \leq r \leq \varepsilon$. In this case, for every $l \in \mathbb{N}_0$, we have
\[
\int_0^t \left\|E(s)\right\|_{\ell(L(X_r, X_r))} \left\|g^{(l)}(t-s)\right\|_{X_r} ds \leq C\delta \Gamma(l+1) \int_0^t \mu_1^{-(r-\varepsilon)/2} \exp(-\mu_1 s) ds \leq C\tilde{C}_\varepsilon \Gamma(l+1) t,
\]
where the bound (35) is used ($\tilde{C}_\varepsilon = \mu_1^{-(r-\varepsilon+2)/2}$). This completes the proof of the assertion. \qed
Remark 3.3. For $0 \leq r < 2$, the preceding result is valid under hypothesis (39) with $\varepsilon = 0$, as used, e.g., in [28], but with $C(r) \uparrow \infty$ as $r \uparrow 2$.

Lemma 3.4. Assume (39) with some $\varepsilon \in (0, 1)$ and some $\delta \geq 1$. Let $u$ be the solution of (28). For $T \geq b > a \geq 1$, the estimate

$$\forall l \in \mathbb{N}_{0} : \left( \int_{a}^{b} \left\| u^{(l)}(t) \right\|_{X_{r}}^{2} \right)^{1/2} \leq \delta^{l} \Gamma(l + 2) C(\varepsilon, a, b)$$

(43)

holds true with a constant $C(\varepsilon, a, b) > 0$ independent of $l$ and $\delta$. Furthermore, for $J = (0, T)$, we have $u \in H^{1}_{0}(J; H)$ with

$$\left( \int_{0}^{T} \left\| u(t) \right\|_{H}^{2} \, dt \right)^{1/2} \leq C T^{3/2}, \quad \left( \int_{0}^{T} \left\| u'(t) \right\|_{H}^{2} \, dt \right)^{1/2} \leq C \delta (T^{1+\varepsilon} + T^{3})^{1/2}$$

(44)

and $u \in L^{2}(J; X_{2})$ with

$$\left\| u \right\|_{L^{2}(J; X_{2})} \leq C T^{\varepsilon/2+1/2},$$

(45)

where the constant $C > 0$ is independent of $\varepsilon$, $\delta$ and $T$.

Proof. The bound (43) follows from (42). For the estimates (44) and (45), we use (42) for $r = 0$ with $l = 0$ or $l = 1$, and $r = 2$ with $l = 0$, respectively. \qed

Proposition 3.5. Assume (39) with some $\varepsilon \in (0, 1)$ and some $\delta \geq 1$. For $r \in [0, 2]$, there exists a constant $C > 0$ (independent of $l$, $\delta$, $a$, $b$, $t$, $q$) such that the solution $u$ of (28) satisfies

$$\forall l \in \mathbb{N}_{0}, \forall t \in (0, \min\{1, T\}] : \left\| u^{(l)}(t) \right\|_{X_{r}} \leq C \delta^{l} \Gamma(l + 2) t^{-l+1-r/2+\varepsilon/2},$$

(46)

and, for $0 < a < b \leq \min\{1, T\}$,

$$\forall l \in \mathbb{N}, l \geq 2 : \left( \int_{a}^{b} \left\| u^{(l)}(t) \right\|_{X_{r}}^{2} \, dt \right)^{1/2} \leq C \delta^{l} \Gamma(l + 2) a^{-l+3/2-r/2+\varepsilon/2},$$

(47)

and, for arbitrary $q \geq 2$,

$$\left\| u \right\|_{H^{q}(a, b; X_{r})} \leq C \delta^{q} \Gamma(q + 3) a^{-q+3/2-r/2+\varepsilon/2}.$$ (48)

Proof. The bound (46) follows from (42). Estimate (47) is obtained by integrating the pointwise bound (46). The Sobolev bound (48) follows by interpolation. \qed

For the proof of the exponential convergence rate of the space-time discretization proposed in this work, we need the following regularity result, which is proven in Appendix B.

Lemma 3.6. Assume (39) with some $\varepsilon \in (0, 1)$ and some $\delta \geq 1$ for $l = 0, 1$. Then, for $b \in (0, T]$, the solution $u$ of (28) belongs to $H^{1/2}_{0}((0, b); X_{2})$ and the estimate

$$\left\| u \right\|_{H^{1/2}_{0}((0, b); X_{2})} \leq \frac{4}{\sqrt{\pi}} \frac{1}{\varepsilon} b^{\varepsilon/2} \left( \frac{b}{1+\varepsilon} + \frac{3}{\varepsilon} + \frac{4b^{2}}{(\varepsilon+1)(\varepsilon+2)} \right)^{1/2} C_{g}$$

holds true, with

$$C_{g} := \left\| g \right\|_{W^{1, \infty}((0, b); X_{r})} = \max \left\{ \sup_{0 \leq t \leq b} \left\| g(t) \right\|_{X_{r}}, \sup_{0 \leq t \leq b} \left\| g'(t) \right\|_{X_{r}} \right\}.$$ (49)
Then, eigenfunction expansions of $f_{\theta} \in \mathcal{D}_n$ established also for transmission problems. Assume $D$ to be partitioned into scalar $d$ subintervals $\mathcal{D}_n$. Furthermore, the solution operator is bijective, since from (32) and (50) it follows that

$$
\|u\|_{H^\theta_0((a,b);X_2)} \leq C(b,\varepsilon,\theta,g)
$$

for $\theta \in [1/2, 1/2 + \varepsilon) \cap [1/2, 1]$ with a constant $C(b,\varepsilon,\theta,g) > 0$.

### 3.2. Spatial regularity

We elaborate here on the regularity of the solution with respect to the spatial variable $x \in D$. For (1), this regularity is, of course, dependent on the temporal variable $t$, and the spaces $X_r$ defined in (32) via eigensystems, which are intrinsic to the spatial operator (25) with (26), play a prominent role. In order to leverage spatial approximation results, we relate these spaces to standard $(d = 1)$ or corner-weighted $(d \geq 2)$ Sobolev spaces. As we shall consider in detail only $P^1$-Lagrangian FEM approximation in $D$, for the ensuing convergence rate analysis in Section 4 we are mainly interested in the spaces $X_r$ for $r = 0, 1, 2$ as defined in (32). The cases $0 \leq r \leq 1$ coincide with standard Sobolev spaces endowed with equivalent norms.

**Proposition 3.8.** For space dimension $d \geq 2$, assume that $D \subset \mathbb{R}^d$ is a bounded Lipschitz domain. Assume further that $A \in L^\infty(D;\mathbb{R}^{d\times d})$ is uniformly positive definite in the sense that (26) is satisfied. Then, $X_0 = L^2(D)$ and $X_1 \simeq H^1_0(D)$ and for $0 < r < 1$, $X_r \simeq (L^2(D), H^1_0(D))^r.2$.

Consider next $1 < r \leq 2$. Once we characterize $X_2$, for $1 < r < 2$, $X_r$ is characterized by real interpolation. To characterize $X_2$, we consider the source diffusion problem (25), with assumption (26) in place. In addition, we assume

$$
f \in L^2(D), \quad A \in W^{1,\infty}(D;\mathbb{R}^{d\times d}).
$$

Then, eigenfunction expansions of $f \in L^2(D)$ imply that the unique solution $u \in X_1$ of (25) belongs to $X_2$. Furthermore, the solution operator is bijective, since from (32) and (50) it follows that

$$
\|u\|_{X_2}^2 = \sum_{k=1}^\infty \mu_k^2 |u_k|^2 = \|A(\partial_x)u\|_H^2 = \sum_{k=1}^\infty |f_k|^2 = \|f\|_{L^2(D)}^2.
$$

It remains to relate the space $X_2$, which is defined in terms of the spatial operator $A(\partial_x)$, to an intrinsic function space in $D$. Due to (51), $X_2 = (A(\partial_x))^{-1}L^2(D)$. To characterize elements in $X_2$, we use the elliptic regularity of the BVP (25) with time-independent data $f \in L^2(D)$ in standard (if $d = 1$) or corner-weighted (if $d \geq 2$) Sobolev spaces in $D \subset \mathbb{R}^d$.

#### 3.2.1. Case $d = 1$

The spatial domain $D$ is an open, bounded and connected interval, and, by (50), the diffusion coefficient is a scalar $a \in W^{1,\infty}(D)$ such that (26) is satisfied.

Standard elliptic regularity results imply that there exists a constant $c > 0$ such that, for every $f \in L^2(D)$, the solution $u = A(\partial_x)^{-1}f$ belongs to $H^2(D)$ and satisfies $\|v\|_{H^2(D)} \leq c\|f\|_{L^2(D)}$. This, combined with (51), gives that $X_2 \subset H^2(D)$ and

$$
\forall v \in X_2 : \quad \|v\|_{H^2(D)} \leq c\|v\|_{X_2}.
$$

**Remark 3.9.** For $d = 1$, a continuous embedding of $X_2$ into a nonintrinsic function space can be easily established also for transmission problems. Assume $D$ to be partitioned into $n_{\text{sub}}$ disjoint, open and connected subintervals $D = \{D_i\}_{i=1}^{n_{\text{sub}}}$ and denote the corresponding broken Sobolev spaces $W^{1,\infty}(D) = \{a \in L^\infty(D) : a|_{D_i} \in W^{1,\infty}(D_i), \ i = 1,\ldots,n_{\text{sub}}\}$ and $H^2(D) := \{v \in H^1(D) : v|_{D_i} \in H^2(D_i), \ i = 1,\ldots,n_{\text{sub}}\}$. We set
\[ \|v\|^2_{H^2(D)} := \|v\|_{H^1(D)}^2 + \sum_{i=1}^{n_{\text{sub}}} |v|_{H^2(D_i)}^2. \]

We assume that the diffusion coefficient \( a \) belongs to \( W^{1,\infty}(D) \) and satisfies (26). In this case, standard elliptic regularity results imply that there exists a constant \( c > 0 \) such that, for every \( \begin{array}{c} \frac{c}{v} \\ \end{array} \) \( v \in H^2(D) \), \( u = A(\partial_x)^{-1} f \in H^2(D) \) and \( \|v\|_{H^2(D)} \leq c \|f\|_{H^2(D)}. \) This, combined with (51), gives \( X_2 \subset H^2(D) \) and (52) is valid with \( \|v\|_{H^2(D)} \) on the left side.

3.2.2. Case \( d = 2 \)

Under (50), for polygonal domains \( D \subset \mathbb{R}^2 \), weak solutions of the source problem (25) are known to belong to a weighted Sobolev space of Kondrat’ev type which is defined as follows.

**Definition 3.10 (Kondrat’ev Spaces in dimension \( d = 2 \)).** Assume that \( D \subset \mathbb{R}^2 \) is a bounded polygonal domain with \( g \geq 3 \) corners and straight sides, whose boundary \( \partial D \) is Lipschitz.

Denote by \( r_D : D \rightarrow \mathbb{R}_{\geq 0} \) a smooth function that locally, in a (sufficiently small) open neighborhood of each corner of \( D \), coincides with the Euclidean distance to that corner. Then, for \( m \in \mathbb{N}_0 \) and for some constant \( a \in \mathbb{R} \), the Kondrat’ev corner-weighted Sobolev space \( \mathcal{K}_a^m(D) \) is defined as

\[
\mathcal{K}_a^m(D) := \left\{ v : D \rightarrow \mathbb{R} : \forall |\alpha| \leq m : r_D^{|\alpha|-a} \partial^\alpha v \in L^2(D) \right\},
\]

with \( \|v\|_{\mathcal{K}_a^m(D)} := \left( \sum_{|\alpha| \leq m} \|r_D^{|\alpha|-a} \partial^\alpha v\|_{L^2(D)}^2 \right)^{1/2}. \)

The regularity result in question is a special case of Corollary 4.5 from [8], which we state here for definiteness in the form required by us.

**Proposition 3.11.** Assume that \( D \subset \mathbb{R}^2 \) is a bounded polygon with boundary \( \partial D \) consisting of a finite number of straight sides. Consider the elliptic source problem (25) with assumptions (26) and (50) in place.

Then, there exist \( c > 0 \) and a constant \( a > 0 \) such that, for every \( f \in L^2(D) \), the weak solution \( u \in X_1 = H^1_{\partial D}(D) \) of (25) belongs to \( \mathcal{K}_a^2(D) \) and satisfies the a priori estimate

\[ \|u\|_{\mathcal{K}_a^2(D)} \leq c \|f\|_{L^2(D)}. \]

In particular, therefore, \( X_2 \subset \mathcal{K}_a^2(D) \) and there exists \( c > 0 \) such that

\[ \forall v \in X_2 : \|v\|_{\mathcal{K}_a^2(D)} \leq c \|v\|_{X_2}. \]

**Proof.** Assumption (50) implies that \( A \in W^{1,\infty}(D) \) as defined in equation (5) of [8], and that \( \|A\|_{W^{1,\infty}(D)} \leq C(D)\|A\|_{W^{1,\infty}(D)}. \) We may then use Corollary 4.5 of [8] with \( b_1 = c = 0, m = 1, \) to conclude the a priori estimate

\[ \|u\|_{\mathcal{K}_a^2(D)} \leq c \|f\|_{\mathcal{K}_a^1(D)}. \]

for all \( |a| < \eta \) for some (sufficiently small) \( \eta > 0 \). We assume, without loss of generality, that \( 0 < \eta < 1 \). Then, definition (53) states that \( f \in \mathcal{K}_a^0(D) \) means \( r_D^{-(a-1)} f \in L^2(D) \). As \(-(a-1) > 0, r_D^{-(a-1)} \in L^\infty(D), \) so that \( \|f\|_{\mathcal{K}_a^0(D)} \leq c(a,D)\|f\|_{L^2(D)}. \) The a priori estimate implies then (54). Since \( \|f\|_{L^2(D)} = \|u\|_{X_2} \) (see (51)), the a priori estimate also implies (55).

**Remark 3.12.** For transmission problems in a polygonal domain \( D \), with piecewise constant, isotropic coefficients in materials occupying a finite number \( n_{\text{sub}} \) of polygonal subdomains \( D_i \subset D \), regularity in the weighted spaces \( \mathcal{K}_a^2(D) \) with radial weights also at multi-material intersection points in \( D \) are stated in Theorem 4.7 of [23]. The assumptions in [23] on \( A \) are more restrictive than just (26) and \( A \in W^{1,\infty}(D; \mathbb{R}^{d \times d}) \) with \( D = \{D_i\}_{i=1}^{n_{\text{sub}}} \). The regularity result in Theorem 4.7 of [23] with \( m = 1 \) will imply for \( u \in X_2 \) a splitting \( u = u_{\text{reg}} + u_s \), with the bound (54) for \( u_{\text{reg}}\|D_i \) on each subdomain \( D_i \), and with \( u_s \) in a finite-dimensional space \( W_s \), see Section 4.2 of [23].
3.2.3. Case d = 3

Proposition 3.11 remains valid in space dimension d = 3. To detail a precise statement, we still assume (50). Then, Theorem 1.1 of [2] implies (54) and (55) in bounded, polyhedral domains D ⊂ R³ with Lipschitz boundary ∂D consisting of a finite number of plane faces. Similar results are shown in [24] and, for the Poisson equation with Γ = Γ_D, in Theorem 1.2 of [6] (with μ = 1 in the statement of that theorem).

4. Approximation

We introduce the spatial and temporal (quasi-) interpolation operators that shall allow us to deduce convergence rates of the space-time variational approximation of formulation (28). In order to use the tensor product construction of subspaces in (29), we specify the choice of temporal subspaces \( \mathcal{V} \) and specify the choice of temporal subspaces \( \mathcal{V} \) in (29), we fix the geometric subdivision parameter \( \sigma \in (0, 1) \) and the number of elements \( m := m_1 + m_2 \in \mathbb{N} \) with given \( 2 < m_1 \in \mathbb{N}, m_2 \in \mathbb{N}_0 \). We set \( T_1 := \min \{1, T\} \). Then, we define the time steps by

\[
t_j := \begin{cases} 
0, & j = 0, \\
T_1 \sigma^{m_1-j}, & j \in \{1, \ldots, m_1\}, \\
\frac{T - T_1}{m_2} (j - m_1) + T_1, & j \in \{m_1 + 1, \ldots, m_1 + m_2\}, \quad \text{if } m_2 > 0,
\end{cases}
\]

where the last line is omitted in the case \( T_1 = T \), i.e., we assume \( m_2 = 0 \) whenever \( T_1 = T \). Furthermore, we denote by \( I_j = (t_{j-1}, t_j) \subset J \) the corresponding time intervals of lengths \( k_j := |I_j| = t_j - t_{j-1} \), fulfilling

\[
k_j = \begin{cases} 
T_1 \sigma^{m_1-1}, & j = 1, \\
T_1 \sigma^{m_1-j}(1 - \sigma), & j \in \{2, \ldots, m_1\}, \\
k_T := \frac{T - T_1}{m_2}, & j \in \{m_1 + 1, \ldots, m_1 + m_2\}, \quad \text{if } m_2 > 0.
\end{cases}
\]

Note that the splitting of \( J = [0, T] \) into the parts \( [0, T_1] \) and \( [T_1, T] \) is necessary for the proofs of the \( hp \)-error estimate in Section 5, since Proposition 3.5 states estimates for \( b \leq T_1 = \min \{1, T\} \) only. In other words, we apply the temporal \( hp \)-FEM in \([0, T_1]\), whereas in \([T_1, T]\) we use a temporal \( p \)-FEM in the case \( T > 1 \). With this notation, we define a geometric partition \( \mathcal{G}_\sigma^m = \{I_j\}_{j=1}^m \) of \( J = (0, T) \). On \( \mathcal{G}_\sigma^m \), we introduce the distribution \( p = (p_1, \ldots, p_m) \in \mathbb{N}^m \) of polynomial degrees as follows: For a given slope parameter \( \mu_{hp} \in \mathbb{R}, \mu_{hp} \geq 1 \), we set

\[
p_j := \begin{cases} 
1, & j = 1, \\
\lfloor \mu_{hp} j \rfloor, & j \in \{2, \ldots, m_1\}, \\
n_T := \lfloor \mu_{hp} m_1 \rfloor, & j \in \{m_1 + 1, \ldots, m_1 + m_2\}, \quad \text{if } m_2 > 0,
\end{cases}
\]

where \( \lfloor \cdot \rfloor \) denotes the floor function. Again, in the case \( m_2 = 0 \), the last line is omitted. Thus, we set \( \mathcal{S}_{0}^{1}(J; \mathcal{G}_\sigma^m) := \{v \in C^0(\mathcal{J}) : v_{ij} \in \mathbb{P}^{p_j}\} \), and the temporal subspace \( \mathcal{V}_t^M \) in (29) is defined as

\[
\mathcal{S}_{0}^{1}(J; \mathcal{G}_\sigma^m) := \{v \in \mathcal{S}_{0}^{1}(J; \mathcal{G}_\sigma^m) : v(0) = 0\} \subset H_{0,\sigma}^{1/2}(J).
\]

Due to the continuity requirement at \( t_j \) for \( j = 1, \ldots, m - 1 \), which is mandated by the \( H^{1/2} \)-conformity, and the zero trace at \( t = 0 \), it holds that

\[
M = \dim\left( \mathcal{S}_{0}^{1}(J; \mathcal{G}_\sigma^m) \right) = \sum_{j=1}^{m} p_j.
\]
We introduce the temporal quasi-interpolant $\Pi_{G^p_\omega}^{p,1} v$ for a sufficiently smooth function $v : [0, T] \to \mathbb{R}$ by

$$
(\Pi_{G^p_\omega}^{p,1} v)(t) := \begin{cases} v(t_1), & t \in T_1 \\
v(t_j) + \frac{t - t_j}{t_{j+1} - t_j} (\Pi_{L^2(I_j)}^{p-1} v^j)(\xi), & t \in T_j, 
\end{cases}
$$

where $\Pi_{L^2(I_j)}^{p-1}$ denotes the $L^2(I_j)$ projection onto $\mathbb{P}_p$. As (60) uses point values of the interpolated function, $\Pi_{G^p_\omega}^{p,1}$ is only defined on a subspace of the continuous functions $C^0(\mathcal{T})$. Note that the nodal property

$$
\forall j \in \{0, \ldots, m\} : \quad (\Pi_{G^p_\omega}^{p,1} v)(t_j) = v(t_j)
$$

holds true for a sufficiently smooth function $v$ with $v(0) = 0$. Our approach to convergence rate bounds in the fractional Sobolev norms is to first obtain estimates in the additive integer order $L^2$ and $H^1$ norms in the usual fashion by scaling estimates on unit size reference domains, then to interpolate the global $L^2$ and $H^1$ norm error bounds. For $j \geq 2$, the error bounds in $I_j$ are standard $hp$-interpolation error estimates as can be found, e.g., in Chapter 3 of [30]. We recall the error bound on $\hat{I} = (-1, 1)$, with the estimates on $I_j$ following by scaling.

**Lemma 4.1.** On $\hat{I} = (-1, 1)$, for every $p \in \mathbb{N}$, a projector $\hat{\Pi}_1^p : H^1(\hat{I}) \to \mathbb{P}^p(\hat{I})$ exists such that, for all $v \in H^{r+1}(\hat{I})$ with some $r \in \mathbb{N}$,

$$
\left\| v' - (\hat{\Pi}_1^p v)' \right\|_{L^2(\hat{I})}^2 \leq \frac{(p - s)!}{(p + s)!} \left\| v^{(s+1)} \right\|_{L^2(\hat{I})}^2
$$

and

$$
\left\| v - \hat{\Pi}_1^p v \right\|_{L^2(\hat{I})}^2 \leq \frac{1}{p(p + 1)} \frac{(p - s)!}{(p + s)!} \left\| v^{(s+1)} \right\|_{L^2(\hat{I})}^2
$$

are valid for every integer $s$ with $0 \leq s \leq \min\{r, p\}$. Furthermore,

$$
(\hat{\Pi}_1^p v)(\pm 1) = v(\pm 1).
$$

We remark that the projectors $\hat{\Pi}_p^1$ for $p \geq 1$ are given by

$$
(\hat{\Pi}_p^1 v)(t) := v(-1) + \int_{-1}^t \hat{\Pi}_0^{p-1} (v')(\xi) \, d\xi, \quad t \in \hat{I},
$$

with $\hat{\Pi}_0^{p-1}$ denoting the $L^2(\hat{I})$ projection onto $\mathbb{P}^{p-1}$.

For $I_j \in G^m_\omega$ with $j \geq 2$, the global quasi-interpolation projectors $\Pi_{G^p_\omega}^{p,1}$ are obtained by transporting $\hat{\Pi}_1^p$ from $\hat{I}$ to $I_j \in G^m_\omega$ via affine transformations $T_j : \hat{I} \to I_j$, resulting in local projections $\Pi_{G^p_\omega}^{p,1}$. We scale the projection error bounds (62) and (63) to $I_j$, and apply them to strongly measurable maps $v : I_j \to X$ for separable Hilbert space $X$ by Hilbertian tensorization of Bochner spaces. We denote by $\mathbb{P}^p(I_j; X)$ the linear space of polynomial maps of degree $p$ with coefficients in $X$. We obtain the following result.

**Lemma 4.2.** For every $I_j \in G^m_\omega$ with $j \geq 2$ with time-step size $k_j = |I_j|$, and for every $p \in \mathbb{N}$, there exists a projector $\Pi_{I_j}^p : H^1(I_j; X) \to \mathbb{P}^p(I_j; X)$ such that, for every $v \in H^{r+1}(I_j; X)$ with some $r \in \mathbb{N}$, the error bounds

$$
\left\| \partial_t v - \partial_t \Pi_{I_j}^p v \right\|_{L^2(I_j; X)}^2 \leq \frac{(p - s)!}{(p + s)!} \left(\frac{k_j}{2}\right)^{2s} \left\| \partial_t^{s+1} v \right\|_{L^2(I_j; X)}^2
$$

and

$$
\left\| v - \Pi_{I_j}^p v \right\|_{L^2(I_j; X)}^2 \leq \frac{1}{p(p + 1)} \frac{(p - s)!}{(p + s)!} \left(\frac{k_j}{2}\right)^{2(s+1)} \left\| \partial_t^{s+1} v \right\|_{L^2(I_j; X)}^2
$$

are valid for every integer $s$ with $0 \leq s \leq \min\{r, p\}$. Furthermore,

$$
(\Pi_{I_j}^p v)(t) = v(t) \text{ in } X \text{ for } t \in \partial I_j = \{t_{j-1}, t_j\}.
$$
4.2. $P^1$-FEM approximation in $D$

We consider the choice of subspaces $V^N_x \subset H^1_{(D)}$ in (29) as standard, conforming $P^1$-Lagrangian finite elements on simplicial meshes $T$ of $D$. We denote by $S^1(D; T)$ the space of continuous, piecewise linear functions on $T$, and further, we define the closed subspace

$$S^1_{a_D}(D; T) := S^1(D; T) \cap H^1_{(D)} \subset H^1_{(D)},$$ (64)

4.2.1. Case $d = 1$

For any finite partition $T$ of the open, bounded and connected interval $D$ into $N$ open subintervals that is quasi-uniform with mesh width $h := \max\{|I_j| : I_j \in T\} > 0$, there exists a constant $c > 0$ independent of $N = O(h^{-1})$ such that the nodal interpolant $I^N : C^0(D) \to S^1(D; T)$ satisfies

$$\forall v \in X_2 : \quad \|v - I^N v\|_{L^2(D)} + N^{-1}\|v - I^N v\|_{H^1(D)} \leq cN^{-2}\|v\|_{H^2(D)}.$$ (65)

With (52), for any $f \in L^2(D)$, we also have that the solution $u = A(\partial_x)^{-1}f$ satisfies

$$\|u - I^N u\|_{L^2(D)} + N^{-1}\|u - I^N u\|_{H^1(D)} \leq cN^{-2}\|f\|_{L^2(D)}.$$ (66)

Remark 4.3. For transmission problems with diffusion coefficient $a \in W^{1,\infty}(D)$ as in Remark 3.9, assuming that $T$ is compatible with the partition $D$ (i.e., the set of nodes of $T$ includes all interfaces in $D$), the nodal interpolant $I^N : C^0(D) \to S^1(D; T)$ satisfies (65) with $\|v\|_{H^2(D)}$ instead of $\|v\|_{H^2(D)}$ on the right side. The subsequent estimate for $u = A(\partial_x)^{-1}f, f \in L^2(D)$, follows from (52) with $\|v\|_{H^2(D)}$ on the left side (see Remark 3.9).

4.2.2. Case $d = 2$

$D \subset \mathbb{R}^2$ is a polygon with a finite number of corners and straight sides. We assume furthermore that each entire side $\Gamma_j$ has either the Dirichlet or the Neumann boundary condition (this is possible by subdividing sides of $D$ with changing boundary conditions and by increasing $M$ appropriately; points where boundary conditions change become then “corner points”).

As it is well-known (e.g., [1, 4, 5] and the references there), functions $u \in K^2_{\alpha+1}(D)$ allow for rate-optimal approximation in $H^1(D)$ and $L^2(D)$ norms in terms of continuous, piecewise linear nodal Lagrangian $\text{FEM}$ in $D$, on regular, simplicial partitions $T^N_\beta$ (see, e.g., [1, 4, 5] and the references there for constructions) of $D$ with $O(N)$ triangles and algebraic corner-refinement towards the vertices of $D$. The subscript $\beta \in (0, 1]$ denotes the corner-refinement parameter, with $\beta = 1$ corresponding to quasi-uniform meshes. As $K^2_{\alpha+1}(D) \subset C(D)$ (see, e.g., [5]), the nodal interpolation operator $I^N_\beta$ is well-defined for $u \in K^2_{\alpha+1}(D)$. Also, for $u \in K^2_{\alpha+1}(D) \cap H^1_{(D)}$, the interpolants $I^N_\beta u$ satisfy exactly the homogeneous Dirichlet boundary conditions on $\Gamma_D$. Furthermore, for suitably strong mesh grading as expressed by the parameter $\beta$ (depending on $D$, and the corner angles at the vertices of $D$), the interpolants $I^N_\beta u$ of $u \in K^2_{\alpha+1}(D)$ converge at optimal rates under mesh refinement: there exists a constant $c > 0$ such that, for all $N = \dim(S^1_{a_D}(D; T^N_\beta)) \in \mathbb{N}$,

$$\|u - I^N_\beta u\|_{L^2(D)} + N^{-\frac{1}{2}}\|u - I^N_\beta u\|_{H^1(D)} \leq cN^{-1}\|u\|_{K^2_{\alpha+1}(D)} \leq cN^{-1}\|f\|_{L^2(D)}.$$ (66)

Here, we used (54) in the last step.

Remark 4.4. The interpolation error bound (66) is based on the graded mesh family $\{T^N_\beta\}_{\beta \geq 1}$. The bound (66) also holds on families of bisection tree meshes, as shown in Theorem 2.2 of [17]. Such families are typically generated by adaptive algorithms, and will also be used in the ensuing numerical experiments in Section 6 below.
Remark 4.5. For transmission problems in $D$, with $A$ as in (26), piecewise smooth on a fine partition $\{D_i\}_{i=1}^{n_{\text{sub}}}$ of $D$ in straight-sided polygons $D_i$, the results in Theorem 4.7 of [23] imply that with graded meshes in each $D_i$ with grading towards multmaterial intersection points, the interpolation error bound (66) is based on the graded mesh family $\{T^N_\beta\}_{N\geq 1}$ still remains true by approximating $u_{\text{reg}}$ and $w_s$ in the decomposition of Theorem 4.7 from [23] separately.

4.2.3. Case $d = 3$

Only partial extensions of (66) to space dimension $d = 3$ are available. We indicate the argument in one particular case. Specifically, we assume (26), (50) and, in addition, that $A(x) = a(x)$, with $a \in W^{1,\infty}(D)$. Furthermore, we consider that $\Gamma_D = \Gamma$, i.e., we consider homogeneous Dirichlet boundary conditions on the entire $\Gamma$. The temporal (analytic) regularity in Section 3.1 is then still valid and, as outlined in Section 3.2.3, the space $X_2$ is continuously embedded into a weighted Kondrat’ev space in $D$ with corner- and edge-weights. A convergence estimate analogous to the $H^1$ bound in (66) (with rate $N^{-1/3}$ instead of $N^{-1/2}$) is stated in Theorem 2.1 of [6] with $m = 1$, and proven in [7], for standard, first-order Langrangian FEM in $D$ on regular triangulations of $D$ into simplices, with anisotropic edge refinements.

5. Convergence Rate of the Space-Time Discretization

We are in a position to establish the convergence rate of the space-time Galerkin discretization (30) with $V_t^M = S^1_0(J; \mathcal{G}^m)$ as defined in (59) and with $V_x^N = S^1_{1,\beta}(D; T^N_\beta)$ as given in (64), where $\beta = 1$ in the case $d = 1$.

We will require the temporal $H^{1/2}_0(J)$ projector $Q^{1/2}_t$ onto $V_t^M$ and the spatial $H^{1/2}_0(D)$ “Ritz” projector $Q^1_x$ into $V_x^N$. Being orthogonal projections, they are stable, i.e.,

$$\left\|v - Q^{1/2}_t v\right\|_{H^{1/2}_0(J)} \leq \left\|v\right\|_{H^{1/2}_0(J)} \quad \text{and} \quad \left\|v - Q^1_x v\right\|_{X_1} \leq \left\|v\right\|_{X_1},$$

optimal in the respective spaces, i.e.,

$$\left\|v - Q^{1/2}_t v\right\|_{H^{1/2}_0(J)} = \min_{w \in V_t^M} \left\|v - w\right\|_{H^{1/2}_0(J)} \quad \text{and} \quad \left\|v - Q^1_x v\right\|_{X_1} = \min_{w \in V_x^N} \left\|v - w\right\|_{X_1}.$$
Lemma 5.1. Let $u$ and $u^{MN}$ be the solutions to (28) and (30), respectively. We have

$$
\|u - u^{MN}\|_{H^{1/2}_{0_1}(J;L^2(D))} \leq \|u - Q^1_{t/2} u\|_{H^{1/2}_{0_1}(J;L^2(D))} \\
+ \|u - Q^1_{x} u\|_{H^{1/2}_{0_1}(J;L^2(D))} + \|(I - Q^1_{t/2})(I - Q^1_{x}) u\|_{H^{1/2}_{0_1}(J;L^2(D))} \\
+ \|u - Q^1_{x} u\|_{H^{1/2}_{0_1}(J;L^2(D))} + \|A(\partial_x)(u - Q^1_{t/2} u)\|_{H^{1/2}_{0_1}(J;L^2(D))}.
$$

(70)

We combine (67)–(70) with the preceding regularity, proven in Section 3, and the approximation properties of the projections $Q^1_{t/2}, Q^1_{x}$ to obtain our main convergence rate bound. For this purpose, we address Term1 through Term5 in the upper bound (70). To this end, we use that the solution $u$ to (28) belongs to $H^{1/2}_{0_1}(J;X_2)$, which was proven in Lemma 3.6.

We start by deriving upper bounds for Term1 and Term5. We have $L^2(Q) \simeq [L^2(Q)]' \hookrightarrow [H^{1/2}_{0_1}(J;L^2(D))]'$ and $H^{1/2}_{0_1}(J;X_2) \hookrightarrow L^2(J;X_2)$ with continuous and dense injections. This, together with (51), gives the following bound for Term5:

$$
\left\| A(\partial_x)(u - Q^1_{t/2} u) \right\|_{[H^{1/2}_{0_1}(J;L^2(D))]'} \leq \tilde{c}(T) \left\| A(\partial_x)(u - Q^1_{t/2} u) \right\|_{L^2(Q)} = \tilde{c}(T) \left\| u - Q^1_{t/2} u \right\|_{L^2(J;X_2)} \\
\leq c(T) \left\| u - Q^1_{t/2} u \right\|_{H^{1/2}_{0_1}(J;X_2)}.
$$

Using estimate (33) yields that Term1 can be bounded by

$$
\left\| u - Q^1_{t/2} u \right\|_{H^{1/2}_{0_1}(J;L^2(D))} \leq c \left\| u - Q^1_{t/2} u \right\|_{H^{1/2}_{0_1}(J;X_2)}
$$

with a constant $c > 0$, i.e., for both Term1 and Term5, we need an estimate of the term $\left\| u - Q^1_{t/2} u \right\|_{H^{1/2}_{0_1}(J;X_2)}$.

For this purpose, we use the temporal quasi-interpolant $\Pi^{p,l}_{t,v}$ of Section 4.1 and the inequality (69). First, note that $\Pi^{p,l}_{t,v} u$ is well-defined since $u: [0,T] \rightarrow X_2$ is continuous, see estimate (46) for $l = 0, r = 2$, and since $u: [0,T] \rightarrow X_2$ is smooth for $t > 0$ due to Lemma 3.2. Second, we have $u \in H^{1/2}_{0_1}(J;X_2)$ because of Lemma 3.6, hence $u - \Pi^{p,l}_{t,v} u \in H^{1/2}_{0_1}(J;X_2)$. Thus, it remains to estimate $\left\| u - \Pi^{p,l}_{t,v} u \right\|_{H^{1/2}_{0_1}(J;X_2)}$, which is done in the following lemmas.

Lemma 5.2. Let $\alpha > 0$ and $m \in \mathbb{N}_0$ be given. For $\mu \geq 1$ with $\mu > \alpha$, there exist a constant $C_\Gamma > 0$, depending on $\alpha, \mu$, but independent of $m$ such that

$$
\sum_{j=0}^{m} \alpha^{2j} \Gamma(\frac{\lfloor \mu j \rfloor - j + 1}{\lfloor \mu j \rfloor + j + 1}) \Gamma(j + 3)^2 \leq C_\Gamma.
$$

Proof. The proof is based on Lemma 3.4 of [12], see Appendix C. \qed

Lemma 5.3. Assume (39) with some $\varepsilon \in (0,1)$ and some $\delta \geq 1$. Let the subdivision parameter $\sigma \in (0,1)$ be given. Choose the slope parameter $\mu_{hp} \geq 1$ such that

$$
\mu_{hp} > \frac{1 - \sigma}{2\sigma(3+\varepsilon)/2},
$$

and fix the number of elements $m_2 \in \mathbb{N}_0$ such that

$$
m_2 \begin{cases} = 0, & T \leq T_h, \\
> \frac{T - T_h}{T} \cdot \delta \sigma^{-\frac{1}{\mu_{hp}}}, & T > T_h,
\end{cases}
$$

(72)
where $T_1 = \min\{1, T\}$. Then, for every $m_1 \in \mathbb{N}$ with $m_1 \geq \max\{3, m_2\}$ and $m = m_1 + m_2$, the geometric partition $\mathcal{G}_\sigma^m$ of $J = (0, T)$, which is given by the time steps $t_j$ in (56) with time-step sizes $k_j$ in (57), and the temporal order distribution $\mathbf{p} \in \mathbb{N}^m$ defined by (58), lead to the error bound

$$\left\| u - \Pi_{\mathcal{G}_\sigma^m}^{1/2} u \right\|_{H_0^{1/2}(t_2, T); X_2}^2 \leq C\sigma^{m_1},$$

with $t_2 = T_1\sigma^{m_1-2}$ and a constant $C > 0$ independent of $m_1$.

Proof. Set $w = u - \Pi_{\mathcal{G}_\sigma^m}^{1/2} u$. Since $w \in H_0^1((t_2, T); X_2)$, see the nodal property (61), the interpolation estimate (Lem. A.2) yields

$$\left\| w \right\|_{H_0^{1/2}((t_2, T); X_2)}^2 \leq \left\| \partial_t w \right\|_{L^2((t_2, T); X_2)} \left\| w \right\|_{L^2((t_2, T); X_2)}.$$  

We estimate both factors on the right side using Proposition 3.5, which states estimates for $b \leq \min\{1, T\} = T_1$ only. Thus, we split $[0, T]$ into the two intervals $[0, T_1]$ and $[T_1, T]$ for the case $T > 1$. Without loss of generality, let us assume that $T > 1$, i.e., $T_1 = 1$ (otherwise we examine only $[0, T] \subset [0, 1]$ and omit the considerations for the second interval $[T_1, T]$). We investigate the intervals $[0, T_1]$ and $[T_1, T]$ separately.

Interval $[0, T_1]$ with $\lambda = \frac{1-\sigma}{2}$, the time-step size fulfills $k_j = t_j - t_{j-1} = t_j\lambda$ for $j = 2, \ldots, m_1$, Lemma 4.2 with $p_j = |\mu_{hp}^j|$, $s_j = j$ and estimate (47) in Proposition 3.5 yield

$$\left\| \partial_t w \right\|_{L^2((t_2, T_1); X_2)}^2 = \sum_{j=3}^{m_1} \left\| \partial_t w \right\|_{L^2((t_j; X_2)}^2 \leq C \sum_{j=3}^{m_1} \frac{\Gamma(|\mu_{hp}^j| - j + 1)}{\Gamma(|\mu_{hp}^j| + j + 1)} \frac{(\lambda^2)^{j-1}}{\delta^2} \frac{\lambda \delta}{2\sigma(-1+\varepsilon/2)^2} \Gamma(j + 3)^2 \leq C \sigma^{m_1(-1+\varepsilon)},$$

where, in the last step, Lemma 5.2 is applied for $\mu_{hp} > \alpha = \frac{(1-\sigma)\delta}{2\sigma(-1+\varepsilon/2)} = \frac{(1-\sigma)\delta}{2\sigma(-1+\varepsilon/2)}$ with (71) and the constant $C_1 > 0$ is independent of $m_1$. In the same way, we get from Lemma 5.2 for $\mu_{hp} > \alpha = \frac{(1-\sigma)\delta}{2\sigma(-1+\varepsilon/2)}$ with (71) that

$$\left\| w \right\|_{L^2((t_2, T_1); X_2)}^2 \leq \sum_{j=3}^{m_1} \left\| w \right\|_{L^2((t_j; X_2)}^2 \leq C \sigma^{m_1+1} \sum_{j=3}^{m_1} \frac{\Gamma(|\mu_{hp}^j| - j + 1)}{\Gamma(|\mu_{hp}^j| + j + 1)} \frac{\lambda \delta}{2\sigma(1+\varepsilon/2)^2} \Gamma(j + 3)^2 \leq C \sigma^{m_1(1+\varepsilon)}$$

with a constant $C_2 > 0$ independent of $m_1$.

Interval $[T_1, T]$ in the case $T > 1$: first, note that $T_1 = 1$. From Lemma 4.2 with the choices $p_j = s_j = p_T := |\mu_{hp}^m|$ and $k_j = k_T$, estimate (43) in Lemma 3.4, and the Stirling’s formula $\sqrt{2\pi n(\frac{n}{e})^n} < n! < \sqrt{2\pi n(\frac{n}{e})^n} e^{\frac{1}{2}}$ with $e^{\frac{1}{2}} < 2$, we get

$$\left\| \partial_t w \right\|_{L^2((1, T); X_2)}^2 \leq \sum_{j=m_1+1}^{m_2} \left\| \partial_t w \right\|_{L^2((t_j; X_2)}^2 \leq C \frac{1}{(2p_T)^2} \sum_{j=m_1+1}^{m_2} \left\| \partial_t^{(p_T+1)} u \right\|_{L^2((t_j; X_2)}^2 \leq C \sigma^2 \frac{1}{(2p_T)^2} \Gamma(p_T + 3)^2 C(\varepsilon, 1, T)^2.$$
Proof. Set \( a \) the triangle inequality, Lemmas 2.1, 3.6, the Poincaré inequality (Lem. A.1), definition (60), and First term: estimate the three terms on the right side.\( \sum \) with a constant \( C_3 > 0 \) independent of \( m_1 \). In the last step, due to (72), we use that a constant \( q \in (0, 1) \) exists such that
\[
k_T = \frac{T - T_1}{m_2} = 4 \delta m_1 \frac{1}{\mu_{hp} m_1} \leq 4 \delta m_1 \frac{1}{\mu_{hp} m_1} = 4 \delta m_1 \frac{1}{\mu_{hp} m_1}
\]
and therefore, \( (p_T + 2)^4 \sqrt{p_T} \to 0 \) as \( p_T \to \infty \). Analogously, we obtain
\[
\|w\|_{L^2((1,T);X_2)}^2 \leq C_4 \sigma^{m_1(1+\varepsilon)}
\]
with a constant \( C_4 > 0 \) independent of \( m_1 \).

With all estimates above, we conclude that
\[
\|w\|_{H_{0}^{1/2}(t_2,T);X_2}^2 \leq \|w\|_{L^2((t_2,T);X_2)}^2 + \|w\|_{H^{1/2}(0,t_2);X_2}^2 + 4 \int_{t_2}^{T} \|w(t)\|_{X_2}^2 \, dt
\]
and therefore, \( (p_T + 2)^4 \sqrt{p_T} \to 0 \) as \( p_T \to \infty \). Analogously, we obtain
\[
\|w\|_{L^2((1,T);X_2)}^2 \leq C_4 \sigma^{m_1(1+\varepsilon)}
\]
with a constant \( C_4 > 0 \) independent of \( m_1 \).

Lemma 5.4. Under the assumptions of Lemma 5.3, the estimate
\[
\|u - \Pi_{G_{m}^T}^{s} u\|_{H_{0}^{1/2}(J,X_2)} \leq C \exp(-b\sqrt{T})
\]
holds true with a constant \( C > 0 \) independent of \( b \) and \( M \), where \( b = -\varepsilon \ln \sigma/\sqrt{8\mu_{hp}} > 0 \) and \( M = \text{dim}(S_{0,1}^1(J;G_{m}^T)) \leq 2\mu_{hp} m_1^2 \leq 2\mu_{hp} m_2^2 \).

Proof. Set \( w = u - \Pi_{G_{m}^T}^{s} u \). Then, for \( X_2 \)-valued functions, the norm equivalence in Lemma 2.1 and the localization in Lemma 2.2 for \( s = 0, b = T, \tau = t_2 \) yield
\[
(C_{int,2})^2 \|w\|_{H_{0}^{1/2}(J,X_2)}^2 \leq \|w\|_{H_{0}^{1/2}(J,X_2)}^2 + \|w\|_{H^{1/2}(0,t_2);X_2}^2 + 4 \int_{t_2}^{T} \|w(t)\|_{X_2}^2 \, dt
\]
and therefore, \( (p_T + 2)^4 \sqrt{p_T} \to 0 \) as \( p_T \to \infty \). Analogously, we obtain
\[
\|w\|_{L^2((1,T);X_2)}^2 \leq C_4 \sigma^{m_1(1+\varepsilon)}
\]
with a constant \( C_4 > 0 \) independent of \( m_1 \).

First term: the triangle inequality, Lemmas 2.1, 3.6, the Poincaré inequality (Lem. A.1), definition (60), and estimates (46), (47) yield
\[
\|w\|_{H_{0}^{1/2}(0,t_2);X_2}^2 \leq 2 \|w\|_{H_{0}^{1/2}(0,t_2);X_2}^2 + 2 \|\Pi_{G_{m}^T}^{s} u\|_{H_{0}^{1/2}(0,t_2);X_2}^2
\]
and therefore, \( (p_T + 2)^4 \sqrt{p_T} \to 0 \) as \( p_T \to \infty \). Analogously, we obtain
\[
\|w\|_{L^2((1,T);X_2)}^2 \leq C_4 \sigma^{m_1(1+\varepsilon)}
\]
we find with the bound (46), the nodal property (61) and second term:

with a constant $C_1 > 0$ independent of $m_1$, where we used

$$t_2 t_1^{1+\varepsilon} = T_1 \sigma^{m_1-2} T_1^{1+\varepsilon} \sigma^{(m_1-1)} = T_1^{1-\varepsilon} \sigma^{m_1}.$$  

Second term: with the bound (46), the nodal property (61) and $k_1 = t_1 = T_1 \sigma^{m_1-1}$, $k_2 = T_1 \sigma^{m_1-2}(1-\sigma)$, we find

$$4 \int_0^{t_2} \frac{\|w(t)\|^2_{X_2}}{t_2-t^2} \, dt = 4 \int_0^{t_1} \frac{\|w(t)\|^2_{X_2}}{t_2-t} \, dt + 4 \int_{t_1}^{t_2} \frac{\|w(t)\|^2_{X_2}}{t_2-t} \, dt$$

$$\leq \frac{8}{k_2} \int_0^{t_1} \frac{\|w(t)\|^2_{X_2}}{t_2-t} \, dt + \frac{8}{k_2} \int_0^{t_1} \frac{\|w(t)\|^2_{X_2}}{k_1^2 t_2} \, dt + 4 \int_{t_1}^{t_2} \frac{\|\partial_t w(t)\|_{X_2}^2}{t_2-t} \, dt$$

$$\leq \frac{C}{k_2} \int_0^{t_1} t^\varepsilon \, dt + \frac{C T_1^{1-\varepsilon}}{k_2^2} \int_0^{t_1} t^2 \, dt + 4 \int_{t_1}^{t_2} \frac{\|\partial_t w(t)\|_{L^2((t_2):X_2)}^2}{t_2-t} \, dt$$

$$\leq 2C T_1^{1-\varepsilon} \frac{1-\varepsilon}{1-\sigma} \sigma^{m_1} + 4k_2 \|\partial_t w\|_{L^2((t_2):X_2)} \leq C_2 \sigma^{m_1},$$

with a constant $C_2 > 0$ independent of $m_1$, where in the last step we have used the estimate (47). This yields

$$4k_2 \|\partial_t w\|_{L^2((t_2):X_2)}^2 \leq CT_1 \sigma^{m_1-2}(1-\sigma)t_1^{1+\varepsilon} = CT_1^{1-\varepsilon} \frac{1-\varepsilon}{1+\varepsilon} \sigma^{m_1}.$$  

Third term: Lemmas 2.1 and 5.3 give

$$5 \|w\|^2_{H^{1/2}_{0}(t_2,T_1):X_2} \leq C(C_{\text{Int},2})^2 \sqrt{1+T^2} \|w\|^2_{H^{1/2}_{0}(t_2,T_1):X_2} \leq C_3 \sigma^{m_1},$$

with a constant $C_3 > 0$ independent of $m_1$.

Conclusion of the proof: as the temporal number of degrees of freedom $M$ fulfills

$$M \leq \sum_{j=1}^{m_1} |\mu_{hp,j}| + |\mu_{hp,m_1}| m_2 \leq \mu_{hp} \frac{m_1 (m_1 + 1)}{2} + \mu_{hp} m_1^2 \leq 2\mu_{hp} m_1^2$$  

(73)

with $m_2 \leq m_1$, using all the estimates above, we conclude

$$\|w\|^2_{H^{1/2}_{0}(J):X_2} \leq C_4 \sigma^{m_1} \leq C_4 \exp\left(-2b\sqrt{M}\right),$$

with a constant $C_4 > 0$ independent of $m_1$, $M$ and $b = -\varepsilon \ln \sigma / \sqrt{8\mu_{hp}} > 0$, i.e., the assertion follows. □

Remark 5.5. Regarding the possibility of optimizing the choice of $\sigma$ and $\mu_{hp}$ in order to improve the exponent $b > 0$ in the estimate of Lemma 5.4, assume that $\varepsilon \in (0,1)$ and $\delta \geq 1$ are given. Then, one has to maximize
Based on these spaces satisfies the error bound on triangulations the spatial FE space \( \mathcal{G} \) with the temporal \( h_p \) satisfies the temporal analytic regularity (39) \( \mu \) can choose \( \mu_h \geq 1 \) and maximizing \( b \) with respect to \( \sigma \in (0,1) \) lead to a violation of (71). Thus, one can choose \( \mu_h \approx (1-\sigma)\delta \over 2\sigma(\delta+\sigma)^{1/2} \) and maximize

\[
b \approx \frac{-\varepsilon \ln \sigma}{\sqrt{8(1-\sigma)\delta}}
\]

with respect to \( \sigma \in (0,1) \), which gives the optimal exponent \( b \).

As Lemma 5.4 implies exponential convergence bounds on Term1 and Term5, it remains to treat Terms2–4 in (70). Term2 and Term4 are identical. We focus on Term3. Using that \( Q_{t}^{1/2} \) is a projector in the Hilbert space \( H_{0}^{1/2}(J) \), the triangle inequality gives

\[
\left\| (I - Q_{t}^{1/2})(I - Q_{x}^{1})u \right\|_{H_{0}^{1/2}(J; L^{2}(D))} \leq 2 \left\| u - Q_{x}^{1}u \right\|_{H_{0}^{1/2}(J; L^{2}(D))}.
\]

Thus, Term3 can be estimated in the same way as Term2 and Term4. Using

\[
H_{0}^{1/2}(0,T; L^{2}(D)) \cong H_{0}^{1/2}(J) \otimes L^{2}(D) \cong L^{2}(D) \otimes H_{0}^{1/2}(J) \cong L^{2}(D; H_{0}^{1/2}(J))
\]

we may use the \( L^{2}(D) \) error bound (68) on the Ritz projection \( Q_{x}^{1} \) and the regularity result in Lemma 3.6 for \( b = T \), in connection with the norm equivalence in Lemma 2.1 for \( a = 0 \), \( b = T \), to arrive at

\[
\left\| u - Q_{x}^{1}u \right\|_{H_{0}^{1/2}(J; L^{2}(D))} \leq C N^{-2/d} \| u \|_{H_{0}^{1/2}(J; X_2)} \leq C N^{-2/d},
\]

with a constants \( c > 0 \), \( C > 0 \) independent of \( N \).

We combine the previous estimates to obtain the main result of this paper.

**Theorem 5.6.** Let the space dimension \( d \) be either \( d = 1 \) or \( d = 2 \). Assume that the diffusion coefficient \( A \in W^{1,\infty}(D; \mathbb{R}^{d \times d}) \) is uniformly positive definite, i.e., that (26) is satisfied, and that the forcing \( g \) in (1) satisfies the temporal analytic regularity (39). Furthermore, assume that the assumptions of Lemma 5.3 on the temporal mesh \( \mathcal{G}^{m} \) in (56) with \( \mu_h \geq 1 \) and \( m_{2} \in \mathbb{N}_{0} \) fulfilling (71) and (72), respectively, and the temporal order distribution \( p \in \mathbb{N}^{m} \) in (58) are satisfied.

Then the space-time Galerkin approximation (30) admits a unique solution \( u^{MN} \in S_{0}^{1}(J; \mathcal{G}^{m}) \otimes S_{1}^{1}(D; \mathcal{T}^{N}) \) with the temporal \( hp \)-FE space \( S_{0}^{1}(J; \mathcal{G}^{m}) \) of dimension \( M = \dim(S_{0}^{1}(J; \mathcal{G}^{m})) \) as defined in (59), and with the spatial FE space \( S_{1}^{1}(D; \mathcal{T}^{N}) \) of continuous, piecewise linear FEM on a sequence of suitably graded, regular triangulations \( \mathcal{T}_{\beta}^{N} \) in \( D \) (\( \beta = 1 \), i.e., quasi-uniform partitions, if \( d = 1 \)) of dimension \( N = \dim(S_{1}^{1}(D; \mathcal{T}^{N}_{\beta})) \).

Moreover, a constant \( C > 0 \) (independent of \( M \) and \( N \)) exists such that the space-time discretization (30) based on these spaces satisfies the error bound

\[
\left\| u - u^{MN} \right\|_{H_{0}^{1/2}(J; L^{2}(D))} \leq C \left( \exp \left( -b \sqrt{M} \right) + N^{-2/d} \right)
\]

with \( b = -\varepsilon \ln \sigma / \sqrt{8 \mu_{hp}} > 0 \).

**Proof.** Existence and uniqueness of the solution \( u^{MN} \) were established at the end of Section 2.5. Estimate (75) follows from Lemma 5.1, taking into account Lemma 5.4, and estimate (74).

Balancing the terms in the upper bound (75) results in

\[
M \approx O((\log N)^2) \quad \text{or} \quad m_{1} \approx O(\log N),
\]
where \( M \leq 2\mu_{hp}m_1^2 \leq 2\mu_{hp}m^2 \), see (73). Then, the number of degrees of freedom for the space-time discretization behaves, as \( N \to \infty \), as
\[
MN \simeq O\left(N(\log N)^2\right),
\]
i.e., it is essentially (up to the \((\log N)^2\) factor) equal to the number of degrees of freedom for the discretization of one spatial problem. Importantly, in the solution algorithms of [21], \( M \simeq O\left((\log N)^2\right) \) will reduce time and memory requirements.

**Remark 5.7.** Theorem 5.6 remains valid for solutions \( u(t, \circ) \) which depend analytically on \( t \in [0, T] \). Classical results on exponential rates of convergence for polynomial approximation of analytic functions in \([0, T]\) (e.g., [11], Chap. 12) imply that for any constant number of temporal elements \( m \in \mathbb{N} \), with \( p \in \mathbb{N} \), temporal exponential convergence follows when \( p \to \infty \) (\( p \)-method). Under the otherwise exact same assumptions as in Theorem 5.6, one obtains in place of (75) the error bound
\[
\| u - u^{MN} \|_{H_0^{1/2}(J;L^2(D))} \leq C \left( \exp(-bp) + N^{-2/d} \right)
\]
with \( M = mp \) and constants \( b > 0, C > 0 \) independent of \( p \) and \( N \). This allows to improve (76) to
\[
MN \simeq O\left(N \log N\right).
\]

6. Numerical experiments

In this section, we present numerical examples for the space-time Galerkin approximation (30) of the heat equation with homogeneous Dirichlet conditions
\[
\partial_t u - \Delta_x u = g \quad \text{in} \; Q, \quad u|_{t=0} = 0, \quad \gamma_0(u) = 0 \quad \text{on} \; \partial D,
\]
i.e., \( A(\partial_x) = -\Delta_x \) in (1), \( u_0 = 0 \) in (2) and \( u_D = 0 \) with \( \Gamma_D = \Gamma = \partial D \) in (3). We use globally continuous functions, which are piecewise linear in space and piecewise polynomials of higher-order in time, see Theorem 5.6. We start by deriving the algebraic linear system associated with (30), and by describing the realization of the operator \( \mathcal{H}_T \) for a temporal \( hp \)-FEM.

For (79), we solve the discrete space-time variational formulation to find \( u^{MN} \in S_{0,T}^{p,1}(J;G^m) \otimes S_{1,D}^{1,1}(D;\mathcal{T}_N) \) such that
\[
\langle \partial_t u^{MN}, v \rangle_{L^2(Q)} + \langle \nabla_x u^{MN}, \nabla_x v \rangle_{L^2(Q)} = \langle \Pi^{MN} g, v \rangle_{L^2(Q)}
\]
is satisfied for all \( v \in (\mathcal{H}_T S_{0,T}^{p,1}(J;G^m) \otimes S_{1,D}^{1,1}(D;\mathcal{T}_N)) \). Here, we use the notation of Section 4 with
\[
S_{0,T}^{p,1}(J;G^m) =: V_i^M := \text{span}\{\varphi_l\}_{l=1}^M, \quad (81)
\]
and
\[
S_{1,D}^{1,1}(D;\mathcal{T}_N) =: V_i^N := \text{span}\{\psi_i\}_{i=1}^N, \quad (82)
\]
where the functions \( \varphi_l \) are basis functions in time, and the functions \( \psi_i \) are the usual nodal basis functions in space. The total number of degrees of freedom is
\[
MN = \dim\left( S_{0,T}^{p,1}(J;G^m) \otimes S_{1,D}^{1,1}(D;\mathcal{T}_N) \right)
\]
In addition, for an easier implementation, we approximate the right-hand side \( g \in L^2(Q) \) by \( g \approx \Pi^{MN} g \), where \( \Pi^{MN} : L^2(Q) \to S^{p,1}(J;G^m) \otimes S^{1}(D;\mathcal{T}_N) \) is the space-time \( L^2 \) projection, namely \( \Pi^{MN} g \in S^{p,1}(J;G^m) \otimes S^{1}(D;\mathcal{T}_N) \) is such that
\[
\forall w \in S^{p,1}(J;G^m) \otimes S^{1}(D;\mathcal{T}_N) : \langle \Pi^{MN} g, w \rangle_{L^2(Q)} = \langle g, w \rangle_{L^2(Q)}.
\]
Note that the spaces \( S^1(D; T^N_\beta) \) and \( S^{n,1}(J; \mathcal{G}^m) \) do not necessarily satisfy the homogeneous Dirichlet and initial conditions, respectively; see beginning of Sections 4.1 and 4.2. We denote the temporal mesh width (i.e., the maximal time-step size) by \( k_{\max} = \max_{j=1,\ldots,m} k_j \), the spatial mesh width by \( h_x \), and the space-time mesh width by \( h_{x,t} = \max\{k_{\max}, h_x\} \). The space-time error \( \left\| u - u^{MN} \right\|_{H^{1/2}_0(J;L^2(D))} \) mandates the numerical evaluation of the fractional order norm \( \left\| \partial_t^{\alpha} \right\|_{H^{1/2}_0(J;L^2(D))} \). In order to overcome this problem, we introduce the quantity

\[
[u]_{H^{1/2}_0(J;L^2(D))} := \sqrt{\left\| u \right\|_{L^2(Q)} \cdot \left\| \partial_t u \right\|_{L^2(Q)}},
\]

which is defined for \( u \in H^1_0(J; L^2(D)) \), and observe that \( u \in H^1_0(J; L^2(D)) \),

\[
\left\| u - u^{MN} \right\|_{H^{1/2}_0(J;L^2(D))} \leq \left\| u - u^{MN} \right\|_{H^{1/2}_0(J;L^2(D))},
\]
due to the interpolation estimate (Lem. A.2). Therefore, in the experiments below, instead of the space-time error \( \left\| u - u^{MN} \right\|_{H^{1/2}_0(J;L^2(D))} \), we consider its upper bound \( \left\| u - u^{MN} \right\|_{H^{1/2}_0(J;L^2(D))} \) which can be numerically evaluated via local integration.

The fully discrete, space-time variational formulation (80) is equivalent to the global linear system

\[
B^{MN} u = G,
\]

with the system matrix

\[
B^{MN} = A^{HT} \otimes M_x + M^{HT} \otimes A_x \in \mathbb{R}^{M \cdot N \times M \cdot N},
\]

where \( \otimes \) is the Kronecker product, \( M_x \in \mathbb{R}^{N \times N} \) and \( A_x \in \mathbb{R}^{N \times N} \) denote the spatial mass and stiffness matrices given by

\[
M_x[i,j] = \langle \psi_j, \psi_i \rangle_{L^2(D)}, \quad A_x[i,j] = \langle \nabla_x \psi_j, \nabla_x \psi_i \rangle_{L^2(D)}
\]

for \( i, j = 1, \ldots, N \), and \( M^{HT} \in \mathbb{R}^{M \times M} \) and \( A^{HT} \in \mathbb{R}^{M \times M} \) are defined by

\[
M^{HT}[k,l] := \langle \varphi_l, \mathcal{H}_T \varphi_k \rangle_{L^2(J)}, \quad A^{HT}[k,l] := \langle \partial_t \varphi_l, \mathcal{H}_T \varphi_k \rangle_{L^2(J)}
\]

for \( k, l = 1, \ldots, M \). Note that, due to the nonlocality of \( \mathcal{H}_T \), the matrices \( M^{HT} \) and \( A^{HT} \) are densely populated. Furthermore, the temporal stiffness matrix \( A^{HT} \) is symmetric (due to (14)) and positive definite (due to (13)), whereas \( M^{HT} \) is nonsymmetric and positive definite (due to (18)). The assembling of the matrices \( M^{HT} \) and \( A^{HT} \) is described in Section 6.1 below. For the right-hand side \( G \), the integrals for computing the projection \( \Pi^{MN} g \) in (82) are calculated by using high-order quadrature rules. The global linear system (83) is solved in MATLAB by using the Bartels–Stewart method with real-Shur decomposition, see Algorithm 4.1 of [21]. All calculations presented in this section were performed on a PC with two Intel Xeon E5-2687W v4 CPUs 3.00 GHz, i.e., in sum 24 cores and 512 GB main memory.

6.1. Numerical implementation of \( \mathcal{H}_T \)

We describe the assembling of the matrices \( M^{HT} \) and \( A^{HT} \) in (84). The crucial point is the realization of the modified Hilbert transformation \( \mathcal{H}_T \), for which different possibilities exist, see [36, 40]. In particular, for a uniform degree vector \( p := (p, p, \ldots, p) \) with a fixed, low polynomial degree \( p \in \mathbb{N} \), e.g., \( p = 1 \) or \( p = 2 \), the matrices \( M^{HT} \) and \( A^{HT} \) in (84) can be calculated using a series expansion based on the Legendre chi function, which converges very fast, independently of the temporal mesh widths; see Section 2.2 of [40]. As for the temporal \( hp \)-FEM the degree vector \( p \) is not uniform, it is convenient to apply numerical quadrature rules to approximate the matrix entries.

From the integral representation of \( \mathcal{H}_T \),

\[
(\mathcal{H}_T v)(t) = -\frac{2}{\pi} v(0) \ln \tan \frac{\pi t}{4T} - \frac{1}{\pi} \int_0^T \ln \left[ \tan \frac{\pi(s + t)}{4T} \tan \frac{\pi(t - s)}{4T} \right] \partial_t v(s) \, ds,
\]
indices \( \boldsymbol{\zeta} \) for \( \boldsymbol{\alpha} \) time intervals \( t \). The functions and the affine transformation i.e. polynomials (or integrated Legendre polynomials) as hierarchical shape functions, compute the integrals in (88). We split these integrals into regular and singular parts, see Section 3.1 of [36].

In summary, the matrix entries of the matrices \( M_t^{\mathcal{H}T} \) in (86), since the matrix entries \( A_t^{\mathcal{H}T} \) in (85) can be computed in the same way.

The matrix entries \( M_t^{\mathcal{H}T} \) in (86) are computed element-wise for the partition \( \mathcal{G}_m = \{ I_j \}_{j=1}^m \) of \( J \) into time intervals \( I_j = (t_{j-1}, t_j) \subset J \), \( j = 1, \ldots, m \). Fix two time intervals \( I_i = (t_{i-1}, t_i) \), \( I_j = (t_{j-1}, t_j) \) with indices \( i, j \in \{ 0, 1, \ldots, M \} \) and related local polynomial degrees \( p_i, p_j \in \mathbb{N} \). We define the local matrix \( M_t^{\mathcal{H}T; i,j} \) by

\[
M_t^{\mathcal{H}T; i,j}[\kappa, \ell] = -\frac{1}{\pi} \int_{t_{j-1}}^{t_j} \int_{t_{i-1}}^{t_i} \ln \left( \frac{\pi(s+t)}{4T} \right) \frac{\pi|t-s|}{4T} \partial_t \varphi_{\kappa}(s) \, ds \, dt
\]

for \( \kappa = 1, \ldots, p_i + 1 \) and \( \ell = 1, \ldots, p_j + 1 \). Here, \( \alpha(\kappa, i) \in \{ 0, 1, \ldots, M \} \) is the global index related to the local index \( \kappa \) for the time interval \( I_i \); similarly for \( \alpha(\ell, j) \). Notice that the function \( \varphi_0 \), corresponding to the vertex \( t = 0 \), does not contribute to the global matrix \( M_t^{\mathcal{H}T} \). On the reference interval \((-1, 1)\), we use the Lobatto polynomials (or integrated Legendre polynomials) as hierarchical shape functions, i.e., we set

\[
N_1(\xi) = \frac{1 - \xi}{2}, \quad N_2(\xi) = \frac{1 + \xi}{2}, \quad N_\ell(\xi) = \int_{-1}^{\xi} L_{\ell-2}(\zeta) \, d\zeta \quad \text{for } \ell \geq 3,
\]

\( \xi \in [-1, 1] \), where \( L_\ell \) denotes the \( \ell \)-th Legendre polynomial on \([-1, 1]\), see Chapter 3 of [30]. With these shape functions and the affine transformation \( T_\iota \colon [-1, 1] \rightarrow [t_{\iota-1}, t_\iota] \) for \( \iota \in \{ 1, \ldots, m \} \), the entries (87) of the local matrix \( M_t^{\mathcal{H}T; i,j} \) are

\[
M_t^{\mathcal{H}T; i,j}[\kappa, \ell] = -\frac{k_j}{2\pi} \int_{-1}^{1} N_\ell(\eta) \int_{-1}^{1} \ln \left( \frac{\pi(T_\iota(\xi) + T_\iota(\eta))}{4T} \right) \frac{\pi|T_\iota(\eta) - T_\iota(\xi)|}{4T} \, N_\iota^*(\xi) \, d\xi \, d\eta
\]

for \( \kappa = 1, \ldots, p_i + 1 \) and \( \ell = 1, \ldots, p_j + 1 \), where \( k_j = |t_j - t_{j-1}| \) is the length of the time interval \( I_j \). To compute the integrals in (88), we split these integrals into regular and singular parts, see Section 3.1 of [36]. For the regular parts, a tensor Gauss quadrature is applied. In Section 3.1 of [36], it is proposed to calculate the singular parts analytically or with an adapted numerical integration. As the polynomial degrees \( p_i, p_j \) may be high, we use the latter. The singularity of the singular parts is of logarithmic type. Thus, we apply so-called classical and nonclassical Gauss–Jacobi quadrature rules of order adapted to \( p_i, p_j \), see equations (1.6), (1.7) of [18], to the singular parts. These adapted integration rules allow us to calculate the singular parts exactly. In summary, the matrix entries of the matrices \( M_t^{\mathcal{H}T} \) and \( A_t^{\mathcal{H}T} \) in (84) are computable to high float point accuracy efficiently.

6.2. Numerical examples in 1D

We present a numerical example in the one-dimensional spatial domain \( D = (0, 1) \subset \mathbb{R} \) with final time \( T = 2 \), i.e., \( Q = J \times D = (0,2) \times (0,1) \subset \mathbb{R}^2 \). We choose the constant right-hand side \( g_1 \equiv 1 \), for which the solution to problem (79) is given by the Fourier series

\[
u_1(t, x) = \sum_{\eta=1}^{\infty} \frac{4 - 4e^{-\pi^2(2\eta - 1)^2 t}}{\pi^3(2\eta - 1)^3} \sin(\pi(2\eta - 1)x), \quad (t,x) \in \overline{Q}.
\]
Table 1. Numerical results with the space-time Galerkin approximation (80) for the 1D example with the right-hand side $g_1 \equiv 1$ and solution $u_1$ in (89), for a uniform mesh refinement strategy and piecewise linear polynomials both in space and time.

| $MN$ | $h_x$ | $k_{\text{max}}$ | $[u_1 - u_1^{MN}]_{H^{1/2}(J; L^2(D))}$ | eoc |
|------|-------|-----------------|-------------------------------------|-----|
| 12   | 0.25000 | 0.50000        | 7.330e−02                            | −   |
| 56   | 0.12500 | 0.25000        | 3.423e−02                            | 0.99|
| 240  | 0.06250 | 0.12500        | 1.355e−02                            | 1.27|
| 992  | 0.03125 | 0.06250        | 5.396e−03                            | 1.30|
| 4032 | 0.01562 | 0.03125        | 2.267e−03                            | 1.24|
| 16256| 0.00781 | 0.01562        | 9.531e−04                            | 1.24|
| 65280| 0.00391 | 0.00781        | 4.004e−04                            | 1.25|
| 261632|0.00195 | 0.00391       | 1.682e−04                            | 1.25|
| 1047552|0.00098 | 0.00195       | 7.070e−05                            | 1.25|
| 4192256|0.00049 | 0.00098       | 2.971e−05                            | 1.25|

In the calculation of the errors of the space-time Galerkin approximation (80), we truncate the series (89) at $\eta = 1000$. For the spatial discretization, we choose a uniform initial mesh with mesh width $h_x$ and apply a uniform refinement strategy.

In the first test, we use a temporal mesh with mesh width $k_{\text{max}} = k_1 = \cdots = k_m$ and linear polynomials, i.e., $p = (1, \ldots, 1) \in \mathbb{N}^m$. The errors and the estimated orders of convergence (eoc) are reported in Table 1. We observe a reduced order of convergence, as the compatibility condition between the right-hand side $g_1 \equiv 1$ and the homogeneous initial condition is not satisfied. Note that the forcing $g_1 \equiv 1$ satisfies the temporal analytic regularity (39) for any $\varepsilon \in (0, 1/2)$ with $\delta = 1$ and a constant $C = C(\varepsilon)$ depending on $\varepsilon$.

In the second test, we use the temporal $hp$-approximation of Section 4.1. For this purpose, we apply a uniform refinement strategy for the spatial discretization, i.e., the number $N$ of degrees of freedom in the spatial discretization doubles with each uniform refinement. Then, corresponding to a given spatial discretization with parameter $N$, we choose the temporal mesh as in (56) with subdivision parameter $\sigma = 0.31$, slope parameter $\mu_{hp} = 2.0$, numbers of elements $m_1 = \lfloor 1.4 \cdot \ln N \rfloor$, $m_2 = 1$, and temporal polynomial degrees $p \in \mathbb{N}^m$ as in (58). This choice of the discretization parameters fulfills condition (71) with $\mu_{hp} = 2.0 > \frac{345}{31\sqrt{31}} \approx 1.99883$ and condition (72) with $m_2 = 1 > \frac{5}{2\sqrt{31}} \approx 0.45$. In addition, this choice balances the terms of the error bound (75), i.e., the total number of degrees of freedom $MN$ behaves like in (76). The numerical results reported in Figure 1 confirm Theorem 5.6.

6.3. Numerical examples in 2D

We present numerical examples in the two-dimensional spatial L-shaped domain

$$D = (-1,1)^2 \setminus [0,1]^2 \subset \mathbb{R}^2,$$

and final time $T = 2$, i.e., $Q = J \times D = (0,2) \times D \subset \mathbb{R}^3$.

6.3.1. Spatial meshes

For the spatial discretization, we consider uniformly refined meshes, see Figure 2, or meshes with corner-refinements towards the origin, where in both cases, the mesh width $h_x$ decreases by a factor 2 with each refinement.

As pointed out in Section 3.2.2, spatial meshes with corner-refinements towards the origin are needed to ensure second-order convergence in $L^2(D)$ for $\mathbb{P}^1$-FEM approximations in $D$. For a given maximal mesh width $h_x > 0$,
Figure 1. Numerical results with the space-time Galerkin approximation (80) for the 1D example with the right-hand side $g_1 \equiv 1$ and solution $u_1$ in (89), for a spatial uniform mesh refinement and temporal $\mathbb{P}^1$-FEM approximations with uniform mesh refinement or with temporal $h$-$p$-FEM with geometric partition of $J$ with subdivision parameter $\sigma = 0.31$, slope parameter $\mu_{hp} = 2.0$, numbers of elements $m_1 = \lfloor 1.4 \cdot \ln N \rfloor$, $m_2 = 1$, and temporal polynomial degrees $p \in \mathbb{N}^m$ as in (58).

Figure 2. Spatial meshes with uniform refinement strategy: starting mesh and mesh after one refinement step.

we construct spatial meshes $T^N_\beta$ with corner-refinements towards the origin fulfilling the grading condition

$$\forall \omega \in T^N_\beta : \quad h_{x,\omega} \sim \begin{cases} h_x^{1/\beta}, & \text{dist}(\omega, \mathbf{0}) = 0, \\ h_x \cdot \text{dist}(\omega, \mathbf{0})^{1-\beta}, & 0 < \text{dist}(\omega, \mathbf{0}) \leq R, \\ h_x, & \text{dist}(\omega, \mathbf{0}) > R, \end{cases}$$

(90)

where the mesh grading parameters $\beta \in (0, 1]$ and $R > 0$ are fixed. Here, $h_{x,\omega}$ is the spatial mesh width of the triangle $\omega \in T^N_\beta$, and dist$(\omega, \mathbf{0})$ is the distance of the triangle $\omega \in T^N_\beta$ from the origin $\mathbf{0}$. To get a sequence of these graded spatial meshes, we halve the maximal mesh width $h_x$ and use the newest vertex bisection for the refinement, see Remark 4.4. Figure 3 shows the spatial graded meshes for the first four levels of refinement with mesh grading parameters $\beta = 0.6$ and $R = 0.25$, which are used in the remainder of this section.
6.3.2. Spatially singular solution

We consider the manufactured solution

$$u_2(t, x_1, x_2) = u_{\text{reg}}(t, x_1, x_2) + te^{-t}\eta(x_1, x_2) \cdot r(x_1, x_2)^{2/3} \cdot \sin\left(\frac{2}{3}\left(\arg(x_1, x_2) - \frac{\pi}{2}\right)\right)$$  (91)

for \((t, x_1, x_2) \in \overline{Q}\) with the smooth part

$$u_{\text{reg}}(t, x_1, x_2) = \frac{1}{100} t \sin(\pi x_1) \sin(\pi x_2) e^{-t(x_1-\frac{1}{4})^2 - t(x_2+\frac{1}{4})^2}, \quad (t, x_1, x_2) \in \overline{Q},$$  (92)

where \(r(x_1, x_2) \in [0, \infty)\) is the radial coordinate, \(\arg(x_1, x_2) \in (0, 2\pi]\) is the angular coordinate, and the cutoff function \(\eta \in C^2(\mathbb{R}^2)\) is given by

$$\eta(x_1, x_2) := \begin{cases} 1, & r(x_1, x_2) \leq 1/4, \\ \frac{27}{8} - \frac{135}{4} r(x_1, x_2) + 180 r(x_1, x_2)^2 \\ -440 r(x_1, x_2)^3 + 480 r(x_1, x_2)^4 - 192 r(x_1, x_2)^5, & 1/4 < r(x_1, x_2) \leq 3/4, \\ 0, & 3/4 < r(x_1, x_2). \end{cases}$$  (93)

Note that the solution \(u_2\) is smooth in time but has a corner singularity in space, which leads to reduced convergence rates, when the spatial meshes are refined uniformly. Hence, we use the spatial graded meshes as in Figure 3 in order to recover maximal convergence rates. We point out that, in numerical tests not reported here, we have verified that, for a Poisson problem with a solution of regularity as the regularity in space of \(u_2\) in (91), one obtains for the \(L^2(D)\) error convergence rates \(N^{-2/3} \sim h_x^{4/3}\) with uniform meshes, and \(N^{-1} \sim h_x^2\) with the considered graded meshes.

For the temporal discretizations, we use \(p^1\)-FEM approximation on uniformly refined meshes, or \(p\)-FEM for a fixed number \(m = 4\) of elements. In connection with the spatial uniform or graded meshes as in Figures 2, 3,
Figure 4. Numerical results with the space-time Galerkin approximation (80) for the 2D example with the singular-in-space solution $u_2$ in (91), for all combinations of uniform mesh refinement or graded meshes in space (with grading parameter $\beta = 0.6$), and $P^1$-FEM with uniform mesh refinement or $p$-FEM in time. For the $p$-FEM in time, for a spatial discretization parameter $N$, we use a fixed mesh with $m = 4$ elements and polynomial degrees $p = (p, p, p, p)$ with $p = \lfloor \ln \frac{N}{2} \rfloor$ respectively, we investigate four possibilities: (i) uniform mesh refinement both in space and in time, (ii) uniform mesh refinement in space and $p$-FEM in time, (iii) graded meshes in space and uniform mesh refinement in time, (iv) graded meshes in space and $p$-FEM in time. For all four cases, the numerical results for the space-time Galerkin approximation (80) of the solution $u_2$ are reported in Figure 4. For a given spatial discretization parameter $N$ and $m = 4$ temporal elements, we choose the temporal polynomial degrees $p = (p, p, p, p)$ with $p = \lfloor \ln \frac{N}{2} \rfloor$. This choice of the discretization parameters balances the terms of the error bound (77). Hence, the total number of degrees of freedom $M_N$ behaves like in (78). The numerical results in Figure 4 confirm Remark 5.7.

6.3.3. Singular solution

We consider the singular solution

$$u_3(t, x_1, x_2) = u_{\text{reg}}(t, x_1, x_2) + t^{3/5} e^{-t} \eta(x_1, x_2) \cdot r(x_1, x_2)^{2/3} \cdot \sin \left( \frac{2}{3} \left( \text{arg}(x_1, x_2) - \frac{\pi}{2} \right) \right)$$

for $(t, x_1, x_2) \in \overline{Q}$ with the smooth part $u_{\text{reg}}$ in (92), the radial coordinate $r(x_1, x_2) \in [0, \infty)$, the angular coordinate $\text{arg}(x_1, x_2) \in (0, 2\pi]$, and the cutoff function $\eta \in C^2(\mathbb{R}^2)$ in (93). This solution has a temporal singularity at $t = 0$. We observe that the corresponding right-hand side $g_3$ does not fulfill the temporal analytic regularity (39). On the other hand, solutions with a singular behavior as $u_3$ are possible even for sources $g$, which satisfy the condition (39). As closed-form representations of such singular solutions do not seem to be available, we perform our numerical tests with the manufactured solution $u_3$ in (94). Furthermore, the solution $u_3$ has the same spatial singularity as $u_2$ in (91). Thus, in order to get the full convergence rates, we use the graded meshes in Figure 3 for the spatial discretization, and a temporal $hp$-FEM. We investigate four possibilities: (i) uniform mesh refinement both in space and in time, (ii) uniform mesh refinement in space and $hp$-FEM in time, (iii) graded meshes in space and uniform mesh refinement in time, (iv) graded meshes in space and $hp$-FEM in time. For all four cases, the numerical results for the space-time Galerkin approximation (80) of the solution $u_3$ are reported in Figure 5. For a given spatial discretization parameter $N$, we choose the temporal mesh as in
Figure 5. Numerical results with the space-time Galerkin approximation (80) for the 2D example with the singular solution $u_3$ in (94), for all combinations of uniform mesh refinement or graded meshes in space (with grading parameter $\beta = 0.6$), and $P^1$-FEM with uniform mesh refinement or $hp$-FEM in time. For the $hp$-FEM in time, for a spatial discretization of parameter $N$, we use a geometric temporal mesh with subdivision parameter $\sigma = 0.17$, slope parameter $\mu_{hp} = 1.0$, numbers of elements $m_1 = \lfloor 2.2 \cdot \ln N \rfloor$, $m_2 = 1$, and temporal polynomial degrees $p \in \mathbb{N}^m$ as in (58).

(56) with subdivision parameter $\sigma = 0.17$, slope parameter $\mu_{hp} = 1.0$, numbers of elements $m_1 = \lfloor 2.2 \cdot \ln N \rfloor$, $m_2 = 1$, and temporal polynomial degrees $p \in \mathbb{N}^m$ as in (58). This choice of the discretization parameters balances the terms of the error bound (75). Hence, the total number of degrees of freedom $MN$ behaves like in (76). The numerical results in Figure 5 are in accordance with Theorem 5.6, when the temporal analytic regularity condition (39), and hence, the conditions on parameters $\mu_{hp}$, $m_2$, i.e., (71), (72), are ignored.

7. Conclusion

Based on a variational space-time formulation of the IBVP (1)–(3), we analyzed tensorized discretization consisting of an exponentially convergent time-discretization of $hp$-type, combined with a first-order Lagrangian FEM in the spatial domain, with corner-mesh refinement to account for the presence of spatial singularities. Stability of the considered discretization scheme is achieved by Hilbert-transforming the temporal $hp$-trial spaces. Details on the efficient, exponentially accurate, numerical realization of this transformation were presented. Several numerical examples in space dimension $d = 2$ in nonconvex polygonal domains confirmed the asymptotic error bounds. In effect, the overall number of degrees of freedom scales essentially as those for one instance of the spatial problem.

The presented proof of time-analyticity via eigenfunction expansions is limited to self-adjoint, elliptic spatial differential operators. Nonselfadjoint spatial operators which are $t$-independent allow similar analytic regularity results via semigroup theory (see, e.g., [29]).

The adopted space-time formulation and its operator perspective and the error analysis extend verbatim to self-adjoint, elliptic spatial operators of positive order. Also, certain nonlinear evolution equations allow for corresponding formulations (see, e.g., [32]). Moreover, transmission problems with piecewise Lipschitz coefficients in the spatial operators can be covered (with the local mesh refinement also at multi-material interface points).

The present error analysis with the same convergence rates is readily extended to a coefficient in the temporal derivative that is time-dependent and analytic in $[0, T]$. 
We finally remark that the presently adopted space-time variational formulation will also allow for \(a \text{ posteriori}\) time-discretization error estimation, which is reliable and robust uniformly with respect to \(p\). Details shall be developed elsewhere.

**Appendix A. Some properties of \(H^{1/2}_0(a,b)\)**

In this appendix, we provide proof of Lemmas 2.1 and 2.2 of Section 2.1 concerning the Sobolev space \(H^{1/2}_0(a,b)\), and state Poincaré and interpolation inequalities in \(H^{1/2}_0(a,b)\).

This result of Lemma 2.1 is well-known, but we need to make explicit the dependency of the involved constants on the interval \((a,b)\), which is essential for the derivation of the temporal hp-error estimates in Section 5. For simplicity, we restrict to the case of real-valued functions \(v: (a,b) \to \mathbb{R}\). All results and proofs can be generalized straightforwardly to \(X\)-valued functions \(v: (a,b) \to X\) for a Hilbert space \(X\). We introduce the following notation. For the classical Sobolev space

\[
H^{1/2}(\mathbb{R}) = (H^1(\mathbb{R}), L^2(\mathbb{R}))_{1/2,2},
\]

where \(H^1(\mathbb{R})\) is equipped with the norm \(\|\cdot\|_{H^1(\mathbb{R})} = (\|\cdot\|_{L^2(\mathbb{R})}^2 + \|\partial_t \cdot\|_{L^2(\mathbb{R})}^2)^{1/2}\), we consider the interpolation norm \(\|\cdot\|_{H^{1/2}(\mathbb{R})}\) and the Slobodetskii norm

\[
\|v\|_{H^{1/2}(\mathbb{R})} := \left(\|v\|_{L^2(\mathbb{R})}^2 + |v|^2_{H^{1/2}(\mathbb{R})}\right)^{1/2}
\]

for \(v \in H^{1/2}(\mathbb{R})\), with

\[
|v|_{H^{1/2}(\mathbb{R})} := \left(\int_{-\infty}^\infty \int_{-\infty}^\infty \frac{|v(s) - v(t)|^2}{|s - t|^2} \, ds \, dt\right)^{1/2}.
\]

**Proof of Lemma 2.1.** The equivalence of norms is proven, e.g., [25]. We give more details about the norm equivalence constants. For this purpose, we introduce an extension operator and establish bounds on its norm. Define \(\mathcal{E}_1: H^1_0(a,b) \to H^1(\mathbb{R})\),

\[
\mathcal{E}_1 v(t) := \begin{cases} v(t), & t \in [a,b], \\ v(2b - t), & t \in (b, 2b - a], \\ 0, & \text{otherwise} \end{cases}
\]

for \(v \in H^1_0(a,b)\). The mapping \(\mathcal{E}_0: L^2(a,b) \to L^2(\mathbb{R})\) is defined for \(v \in L^2(a,b)\) as

\[
\mathcal{E}_0 v(t) := \begin{cases} v(t), & t \in (a,b), \\ v(2b - t), & t \in (b, 2b - a), \\ 0, & \text{otherwise} \end{cases}
\]

Evidently, \(\mathcal{E}_1 v = \mathcal{E}_0 v\) for \(v \in H^1_0(a,b)\). Next, for \(v \in L^2(a,b)\),

\[
\|\mathcal{E}_0 v\|_{L^2(\mathbb{R})}^2 = \int_a^b |v(t)|^2 \, dt + \int_b^{2b-a} |v(2b-t)|^2 \, dt = 2\|v\|_{L^2(a,b)}^2.
\]

and, for \(v \in H^1_0(a,b)\),

\[
\|\partial_t \mathcal{E}_1 v\|_{L^2(\mathbb{R})}^2 = \int_a^b |\partial_t v(t)|^2 \, dt + \int_b^{2b-a} |\partial_t v(2b-t)|^2 \, dt = 2\|\partial_t v\|_{L^2(a,b)}^2.
\]
Hence, for \( v \in H^1_0(a, b) \), it holds true that

\[
\|\mathcal{E}_1 v\|_{H^1(\mathbb{R})}^2 = 2\|v\|^2_{H^1_0(a, b)} \leq 2 \left( 1 + \frac{4(b-a)^2}{\pi^2} \right) \|\partial_t v\|^2_{L^2(a,b)},
\]

where the Poincaré inequality (see Lem. A.1 below) is used in the last step. Interpolation yields an operator \( \mathcal{E}_{1/2} : H^{1/2}_0(a, b) \to H^{1/2}(\mathbb{R}) \) with \( \mathcal{E}_{1/2} v = \mathcal{E}_0 v \) for \( v \in H^{1/2}_0(a, b) \) and

\[
\forall v \in H^{1/2}_0(a, b) : \quad \|\mathcal{E}_{1/2} v\|_{H^{1/2}(\mathbb{R})} \leq 2 \sqrt{1 + \frac{4(b-a)^2}{\pi^2}} \|v\|^2_{H^{1/2}_0(a, b)}. \tag{A.1}
\]

Next, we estimate \( \|\mathcal{E}_{1/2} v\|_{H^{1/2}(\mathbb{R})} \) for \( v \in H^{1/2}_0(a, b) \). For this purpose, we compute

\[
\|\mathcal{E}_{1/2} v\|_{H^{1/2}(\mathbb{R})}^2 = \int_a^\infty \int_a^\infty (\varepsilon_1/2 v(t+s))_a b (\varepsilon_1/2 v(t+s))_a b d(s, t) + 0
\]

\[
= \int_a^\infty \int_a^\infty (\varepsilon_1/2 v(t+s))_a b (\varepsilon_1/2 v(t+s))_a b d(s, t) + 0
\]

\[
= \int_a^\infty \frac{\mathcal{E}_{1/2} v(t)}{t-a} dt \tag{A.2}
\]

for \( v \in H^{1/2}_0(a, b) \), where the seminorm \( |\cdot|_{H^{1/2}(\mathbb{R})} \) is defined by (9) with \( b = \infty \). The integral in the bound (A.2) is finite due to \( v \in H^{1/2}_0(a, b) \), cf. (10). Thus, we get

\[
\|\mathcal{E}_{1/2} v\|_{H^{1/2}(\mathbb{R})}^2 \leq 2\|v\|^2_{L^2(a,b)} + \int_a^\infty \int_a^\infty (\varepsilon_1/2 v(t+a))_a b (\varepsilon_1/2 v(t+a))_a b d(s, t) + 0
\]

\[
= 2\|v\|^2_{L^2(a,b)} + \int_a^\infty \int_a^\infty (\varepsilon_1/2 v(t+a))_a b (\varepsilon_1/2 v(t+a))_a b d(s, t) + 2 \int_a^\infty \int_a^\infty \frac{\mathcal{E}_{1/2} v(t)}{t-a} dt.
\]

The third term on the right side is bounded by

\[
2 \int_a^b \int_a^b (\varepsilon_1/2 v(t+a))_a b (\varepsilon_1/2 v(t+a))_a b d(s, t) + 2 \int_a^\infty \int_a^\infty (\varepsilon_1/2 v(t+a))_a b (\varepsilon_1/2 v(t+a))_a b d(s, t)
\]

\[
= 2 \int_a^b \int_a^b \frac{\mathcal{E}_{1/2} v(t+a)}{t-a} dt + 2 \int_a^\infty \int_a^\infty \frac{\mathcal{E}_{1/2} v(t+a)}{t-a} dt,
\]

the fourth term is

\[
\int_a^\infty \int_a^\infty (\varepsilon_1/2 v(t+a))_a b (\varepsilon_1/2 v(t+a))_a b d(s, t) + 2 \int_a^\infty \int_a^\infty \frac{\mathcal{E}_{1/2} v(t+a)}{t-a} dt
\]

\[
= \int_a^\infty \int_a^\infty (\varepsilon_1/2 v(t+a))_a b (\varepsilon_1/2 v(t+a))_a b d(s, t) + 2 \int_a^\infty \int_a^\infty \frac{\mathcal{E}_{1/2} v(t+a)}{t-a} dt
\]

\[
= \|v\|^2_{H^{1/2}(a,b)} + \int_a^b \frac{\mathcal{E}_{1/2} v(t)}{t-a} dt = \|v\|^2_{H^{1/2}(a,b)} + \int_a^b \frac{\mathcal{E}_{1/2} v(t)}{t-a} dt,
\]

whereas for the fifth term, we have

\[
2 \int_a^b \frac{\mathcal{E}_{1/2} v(t)}{t-a} dt.
\]
Using the above estimates gives for all $v \in H^{1/2}_0(a,b)$

$$
\| \mathcal{E}_{1/2} v \|_{H^{1/2}_0(\mathbb{R})}^2 \leq 2 \| v \|_{L^2(a,b)}^2 + 4 |v|^2_{H^{1/2}(a,b)} + 8 \int_a^b \frac{|v(t)|^2}{t-a} \, dt \leq 8 \| v \|_{H^{1/2}_0(a,b)}^2.
$$

With these properties, we have for all $v \in H^{1/2}_0(a,b)$ the lower bound in the norm equivalence:

$$
\| v \|_{H^{1/2}_0(a,b)} \leq \mathcal{E}_{1/2} v \|_{H^{1/2}_0(\mathbb{R})} \leq \frac{1}{C_{R,1}} \mathcal{E}_{1/2} v \|_{H^{1/2}_0(\mathbb{R})} \leq \frac{2\sqrt{2}}{C_{R,1}} \| v \|_{H^{1/2}_0(a,b)}.
$$

Here, the first inequality is proven by interpolation, the second estimate follows from constants $C_{R,1}, C_{R,2} > 0$, see Theorem B.7 of [25] and Lemma 4.1 of [13].

For the upper bound, relations (A.2), (A.3) and (A.1) yield

$$
\| v \|_{H^{1/2}_0(a,b)}^2 \leq \| \mathcal{E}_{1/2} v \|_{L^2(\mathbb{R})}^2 + \| (\mathcal{E}_{1/2} v)_{(a,\infty)} \|_{H^{1/2}(a,\infty)}^2 + \int_a^\infty \frac{\| \mathcal{E}_{1/2} v(t) \|_2^2}{t-a} \, dt
$$

$$
= \| \mathcal{E}_{1/2} v \|_{L^2(\mathbb{R})}^2 + \frac{1}{2} \| (\mathcal{E}_{1/2} v)_{(a,\infty)} \|_{H^{1/2}(a,\infty)}^2 + \frac{1}{2} \| \mathcal{E}_{1/2} v \|_{H^{1/2}(\mathbb{R})}^2
$$

$$
\leq \| \mathcal{E}_{1/2} v \|_{H^{1/2}_0(\mathbb{R})}^2 \leq (C_{R,2})^2 \| \mathcal{E}_{1/2} v \|_{H^{1/2}_0(\mathbb{R})}^2 \leq (C_{R,2})^2 2\sqrt{1 + \frac{4(b-a)^2}{\pi^2}} \| v \|_{H^{1/2}_0(a,b)}^2,
$$

i.e., the assertion is proven. \qed

The following proof of Lemma 2.2 restricts the argument of [14] to our particular case.

**Proof of Lemma 2.2.** Let $v \in H^{1/2}(a,b)$ and $\tau \in (a,b)$ be given. Then, we split the integral in the definition (9) as follows:

$$
|v|^2_{H^{1/2}(a,b)} = \int_a^\tau \int_a^b \left( \cdots \right) \, ds \, dt + \int_a^b \int_a^\tau \left( \cdots \right) \, ds \, dt
$$

$$
= \int_a^\tau \int_a^\tau \left( \cdots \right) \, ds \, dt + 2 \int_a^\tau \int_a^b \left( \cdots \right) \, ds \, dt + \int_a^b \int_a^\tau \left( \cdots \right) \, ds \, dt
$$

$$
= |v|^2_{H^{1/2}(a,\tau)} + 2 \int_a^\tau \int_a^b \left( \cdots \right) \, ds \, dt + |v|^2_{H^{1/2}(\tau,b)}.
$$

For the integral on the right side, we get

$$
2 \int_a^\tau \int_a^b \frac{|v(s) - v(t)|^2}{|s-t|^2} \, ds \, dt \leq \frac{4}{\tau} \int_a^\tau \int_a^b \frac{|v(s)|^2}{|s-t|^2} \, ds \, dt + 4 \int_a^\tau \int_a^b \frac{|v(t)|^2}{|s-t|^2} \, ds \, dt
$$

$$
\leq 4 \int_a^\tau \frac{|v(s)|^2}{|s-\tau|^2} \, ds + 4 \int_a^\tau \frac{|v(t)|^2}{(s-t)^{-1} - (s-a)^{-1}} \, ds + \int_a^\tau \frac{|v(t)|^2}{(\tau-t)^{-1} - (b-t)^{-1}} \, dt
$$

Thus, the assertion follows. \qed
Lemma A.1. For \( a, b \in \mathbb{R} \), \( a < b \), the Poincaré inequalities

\[
\forall v \in H^{1/2}_0(a, b) : \quad \|v\|_{L^2(a, b)} \leq \sqrt{\frac{2(b-a)}{\pi}} \|v\|_{H^{1/2}_0(a, b)},
\]

\[
\forall v \in H^1_0(a, b) : \quad \|v\|_{H^{1/2}_0(a, b)} \leq \frac{2(b-a)}{\pi} \|\partial_t v\|_{L^2(a, b)},
\]

\[
\forall v \in H^1(a, b) : \quad \|v\|_{L^2(a, b)} \leq \frac{2(b-a)}{\pi} \|\partial_t v\|_{L^2(a, b)}
\]

hold true, where the constants are sharp.

Proof. By interpolation, we have the Fourier series representations

\[
\|v\|_{L^2(a, b)}^2 = \sum_{k=0}^{\infty} |v_k|^2, \quad \|v\|_{H^{1/2}_0(a, b)}^2 = \sum_{k=0}^{\infty} \lambda_k |v_k|^2, \quad \|\partial_t v\|_{L^2(a, b)}^2 = \sum_{k=0}^{\infty} \lambda_k |v_k|^2
\]

(A.4)

with coefficients \( v_k \) as in (5) and eigenvalues \( \lambda_k = \frac{\pi^2(2k+1)^2}{4(b-a)^2} \) of the eigenvalue problem (6). Hence, all Poincaré inequalities follow from these representations. The constants are sharp since for \( v \) with \( v_0 \neq 0 \) and \( v_k = 0 \) for \( k \in \mathbb{N} \), equality holds true.

\[\square\]

Lemma A.2. For \( a, b \in \mathbb{R} \) with \( a < b \), the interpolation estimate

\[
\forall v \in H^1_0(a, b) : \quad \|v\|_{H^{1/2}_0(a, b)} \leq \sqrt{\|v\|_{L^2(a, b)} \|\partial_t v\|_{L^2(a, b)}}
\]

holds true, where \( \|v\|_{H^{1/2}_0(a, b)} \) denotes the interpolation norm (4).

Proof. Using the Cauchy–Schwarz inequality, the assertion follows immediately from the Fourier representations (A.4).

\[\square\]

Appendix B. Proof of Lemma 3.6

Let \( b \in (0, T] \) be fixed. According to (41) for \( l = 0 \), the estimate

\[
\forall t > 0 : \quad \|E(t)\|_{L(X_\varepsilon, X_2)} \leq \frac{1}{\sqrt{2\pi}} \left( \frac{1}{2} \right)^{2-\varepsilon} \Gamma(3-\varepsilon)t^{-2+\varepsilon}.
\]

holds true. The logarithmic convexity of the gamma function gives \( \Gamma(3-\varepsilon) = \Gamma(2\varepsilon + 3(1-\varepsilon)) \leq \Gamma(2 \varepsilon) \Gamma(3) \Gamma(1-\varepsilon) = 2^{1-\varepsilon} \) and we obtain

\[
\forall t > 0 : \quad \|E(t)\|_{L(X_\varepsilon, X_2)} \leq \frac{1}{\sqrt{2\pi}} \left( \frac{1}{2} \right)^{t^{-2+\varepsilon}} = \frac{1}{\sqrt{8\pi}} t^{-1+\varepsilon/2}.
\]

(B.1)

The solution \( u \) admits the representation

\[
u(t) = \int_0^t E(\tau) g(t-\tau) \, d\tau, \quad 0 \leq t \leq b, \quad \text{(B.2)}
\]

see (38), and for \( t \in [0, b] \), it follows that

\[
\|u(t)\|_{X_2} \leq \int_0^t \|E(\tau)g(t-\tau)\|_{X_2} \, d\tau \leq \int_0^t \|E(\tau)\|_{L(X_\varepsilon, X_2)} \|g(t-\tau)\|_{X_\varepsilon} \, d\tau
\]

\[
\leq \frac{1}{\sqrt{8\pi}} C_g \int_0^t \tau^{-1+\varepsilon/2} \, d\tau = \frac{1}{\sqrt{8\pi}} C_g \frac{1}{\varepsilon} t^{\varepsilon/2} = \frac{\sqrt{2}}{\pi} \frac{1}{\varepsilon} t^{\varepsilon/2}.
\]
We estimate the three terms of $\|u\|_{H^{1/2}_{0}(\{(0,b):X_2\}$ expressed as in (10).

**First term:** from the previous bound for $\|u(t)\|_{X_2}$, we derive

$$
\int_0^b \|u(t)\|^2_{X_2} \frac{dt}{t} \leq 2 \left( \frac{1}{\epsilon^2} \frac{1}{\sqrt{\pi}} \int_0^b t^{1-\epsilon} \frac{dt}{t} \right) = 2 \left( \frac{1}{\epsilon^2} \frac{1}{\sqrt{\pi}} b^{1+\epsilon} \right).
$$

**Third term:** similarly, we obtain

$$
\int_0^b \frac{\|u(t)\|^2_{X_2}}{t} \frac{dt}{t} \leq 2 \left( \frac{1}{\epsilon^2} \frac{1}{\sqrt{\pi}} \int_0^b t^{-\epsilon-1} \frac{dt}{t} \right) = 2 \left( \frac{1}{\epsilon^2} \frac{1}{\sqrt{\pi}} b^{-\epsilon} \right).
$$

**Second term:** recalling (9), we need to estimate $\int_0^b \int_0^b \frac{\|u(s)-u(t)\|^2_{X_2}}{|s-t|^2} \frac{ds}{s} \frac{dt}{t}$.

For $b \geq s \geq t \geq 0$, we have

$$
\int_0^b \int_0^b \frac{\|u(s)-u(t)\|^2_{X_2}}{|s-t|^2} \frac{ds}{s} \frac{dt}{t} \leq \left( \frac{1}{\sqrt{8\pi}} \frac{1}{\epsilon} \int_0^b \frac{\|E(s-t)\|_{L^2_{s,t}(X_2)}}{|s-t|^2} \frac{ds}{s} \frac{dt}{t} \right) \leq \left( \frac{1}{\sqrt{8\pi}} \frac{1}{\epsilon} \int_0^b \frac{\|E(s-t)\|_{L^2_{s,t}(X_2)}}{|s-t|^2} \frac{ds}{s} \frac{dt}{t} \right)
$$

Analogously, for $b \geq t \geq s \geq 0$, the estimate

$$
\int_0^b \int_0^b \frac{\|u(s)-u(t)\|^2_{X_2}}{|s-t|^2} \frac{ds}{s} \frac{dt}{t} \leq \left( \frac{1}{\sqrt{8\pi}} \frac{1}{\epsilon} \int_0^b \frac{\|E(s-t)\|_{L^2_{s,t}(X_2)}}{|s-t|^2} \frac{ds}{s} \frac{dt}{t} \right)
$$

holds true. We conclude that

$$
\int_0^b \int_0^b \frac{\|u(s)-u(t)\|^2_{X_2}}{|s-t|^2} \frac{ds}{s} \frac{dt}{t} \leq \left( \frac{1}{\sqrt{8\pi}} \frac{1}{\epsilon} \int_0^b \frac{\|E(s-t)\|_{L^2_{s,t}(X_2)}}{|s-t|^2} \frac{ds}{s} \frac{dt}{t} \right)
$$

and

$$
\int_0^b \int_0^b \frac{\|u(s)-u(t)\|^2_{X_2}}{|s-t|^2} \frac{ds}{s} \frac{dt}{t} \leq \left( \frac{1}{\sqrt{8\pi}} \frac{1}{\epsilon} \int_0^b \frac{\|E(s-t)\|_{L^2_{s,t}(X_2)}}{|s-t|^2} \frac{ds}{s} \frac{dt}{t} \right)
$$
Conclusion of the proof: by combining the bounds of the three terms, we arrive at the a priori estimate
\[
\| u \|_{H^{1/2}_0((0,b);X_2)} \leq \sqrt{\frac{2}{\pi \varepsilon}} \ v^{1/2} \left( \frac{b}{1 + \varepsilon} + \frac{3}{\varepsilon} + \frac{4b^2}{(\varepsilon + 1)(\varepsilon + 2)} \right)^{1/2} C_g,
\]
which gives the assertion.

Appendix C. Proof of Lemma 5.2

This proof is a slight modification of the proof of Lemma 3.4 from [12]. We use Stirling’s inequalities
\[
\forall x > 0: \quad \sqrt{2\pi} x^{x-1/2} e^{-x} \leq \Gamma(x) \leq \sqrt{2\pi} x^{x-1/2} e^{-x} e^{1/x}.
\]
For \( j \geq 1 \), (C.1) yields
\[
\frac{\Gamma([\mu j] - j + 1)}{\Gamma([\mu j] + j + 1)} \leq \frac{\Gamma([\mu j] - j + 1)}{\Gamma([\mu j] + j)} \leq \frac{\mu j}{\mu j + 1} \frac{\mu j - j + 1/2}{\mu j + 1/2} e^{(\mu j - j + 1)} \left( 1 + \frac{4 \mu j}{(\mu j + 1)(\mu j + 2)} \right) \leq \frac{2 \mu j}{e} \left( \frac{e}{\mu j} \right)^{2j}
\]
and
\[
\Gamma(j + 3)^2 = \left( \frac{j + 1}{2} \right)^3 (j + 1)^2 j^2 \Gamma(j)^2 \leq j^6 \cdot 72 \pi j^{2j-1} e^{-2j} e^{\pi} \leq 144 \pi j^5 j^{2j} e^{-2j}.
\]
Thus, we have
\[
\forall j \in \mathbb{N}: \quad \alpha^{2j} \frac{\Gamma([\mu j] - j + 1)}{\Gamma([\mu j] + j + 1)} \Gamma(j + 3)^2 \leq \frac{288 \pi \mu}{e} j^6 \left( \frac{\alpha}{\mu} \right)^{2j}.
\]
Hence, we conclude that
\[
\sum_{j=0}^{m} \alpha^{2j} \frac{\Gamma([\mu j] - j + 1)}{\Gamma([\mu j] + j + 1)} \Gamma(j + 3)^2 = 4 + \sum_{j=1}^{m} \alpha^{2j} \frac{\Gamma([\mu j] - j + 1)}{\Gamma([\mu j] + j + 1)} \Gamma(j + 3)^2 \leq 4 + \frac{288 \pi \mu}{e} \sum_{j=1}^{\infty} j^6 \left( \frac{\alpha}{\mu} \right)^{2j} < \infty,
\]
since the ratio test gives
\[
\lim_{j \to \infty} \frac{(j + 1)^6 \left( \frac{\alpha}{\mu} \right)^{2(j+1)}}{j^6 \left( \frac{\alpha}{\mu} \right)^{2j}} = \left( \frac{\alpha}{\mu} \right)^2 < 1,
\]
i.e., the assertion is proven.

Acknowledgements. This work was performed in autumn 2021 when ChS visited Uni Vienna while on sabbatical leave from ETH, and completed at the 2021 Söllerhaus Workshop in October 2021 in Hirschegg, Austria. Excellent working conditions there are acknowledged. IP has been funded by the Austrian Science Fund (FWF) through project F 65 “Taming Complexity in Partial Differential Systems” and project P 33477-N.

References

[1] J.H. Adler and V. Nistor, Graded mesh approximation in weighted Sobolev spaces and elliptic equations in 2D. Math. Comp. 84 (2015) 2191–2230.
[2] B. Ammann and V. Nistor, Weighted Sobolev spaces and regularity for polyhedral domains. Comput. Methods Appl. Mech. Eng. 196 (2007) 3630–3659.
[3] R. Andreev, Stability of sparse space-time finite element discretizations of linear parabolic evolution equations. IMA J. Numer. Anal. 33 (2013) 242–260.
[37] R. Stevenson and J. Westerdiep, Stability of Galerkin discretizations of a mixed space-time variational formulation of parabolic evolution equations. *IMA J. Numer. Anal.* **41** (2021) 28–47.

[38] V. Thomée, Galerkin Finite Element Methods for Parabolic Problems. Vol. 25 of *Springer Series in Computational Mathematics*, 2nd edition. Springer-Verlag, Berlin (2006).

[39] H. Triebel, Interpolation Theory, Function Spaces, Differential Operators, 2nd edition. Johann Ambrosius Barth, Heidelberg (1995).

[40] M. Zank, An exact realization of a modified Hilbert transformation for space-time methods for parabolic evolution equations. *Comput. Methods Appl. Math.* **21** (2021) 479–496.

---

**Subscribe to Open (S2O)**

A fair and sustainable open access model

This journal is currently published in open access under a Subscribe-to-Open model (S2O). S2O is a transformative model that aims to move subscription journals to open access. Open access is the free, immediate, online availability of research articles combined with the rights to use these articles fully in the digital environment. We are thankful to our subscribers and sponsors for making it possible to publish this journal in open access, free of charge for authors.

**Please help to maintain this journal in open access!**

Check that your library subscribes to the journal, or make a personal donation to the S2O programme, by contacting subscribers@edpsciences.org

More information, including a list of sponsors and a financial transparency report, available at: [https://www.edpsciences.org/en/maths-s2o-programme](https://www.edpsciences.org/en/maths-s2o-programme)