Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Seminar for Applied Mathematics

Exponential Convergence of hp-Time-Stepping in Space-Time Discretizations of Parabolic PDEs

I. Perugia and Ch. Schwab and M. Zank

Research Report No. 2022-09
March 2022

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

Funding: Austrian Science Fund (FWF) through project F~65 "Taming Complexity in Partial Differential Systems" and project P~33477-N

AMS Subject Classification: primary 65N30

EXPONENTIAL CONVERGENCE OF hp-TIME-STEPPING IN SPACE-TIME DISCRETIZATIONS OF PARABOLIC PDES

ILARIA PERUGIA, CHRISTOPH SCHWAB, AND MARCO ZANK

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 hp-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.

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., [18]), and wavelet-based compressive schemes (e.g., [19, 30] 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,

---

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

Key words and phrases. Parabolic IBVP, Space-Time Methods, hp-FEM, Exponential Convergence.

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.

Bochner-space perspective used to establish 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 accomodation of efficient, compressive space-time numerical schemes. We mention only [30, 3, 32, 20, 9, 34, 25, 14, 15, 36] and the references there. A comprehensive account of the numerical analysis of *fixed order time-discretizations* is provided in [37] 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 $\dot{H}^s(\mathrm{D})$, which impose implicit boundary compatibilities of regular data, see [37, Chapter 19].

The presently investigated time-discretization approach is based on the space-time variational formulation in [34]. 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 [34, 35, 39, 21, 33]. We briefly recapitulate it here, and refer to [34] for full development of details. The compressive aspect is here realized by the *hp*-time discretization for this formulation.

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

$$Bu := \partial_t u + A(\partial_x)u = g \quad \text{in} \quad (0, T) \times \mathrm{D}. \tag{1.1}$$

Here, the forcing function $g : Q \to \mathbb{R}$ is assumed to belong to $\mathcal{A}([0, T]; L^2(\mathrm{D}))$, i.e., it is analytic as a map from $[0, T]$ into $L^2(\mathrm{D})$. 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 $A \in L^\infty(\mathrm{D}; \mathbb{R}^{d\times d})$ being a symmetric, positive definite matrix function of $x \in \mathrm{D}$ which does not depend on the temporal variable t. The PDE (1.1) is completed by initial condition

$$u|_{t=0} = u_0\ , \tag{1.2}$$

and by mixed boundary conditions

$$\gamma_0(u) = u_D \quad \text{on} \quad \Gamma_D\ , \quad \gamma_1(u) = u_N \quad \text{on} \quad \Gamma_N\ . \tag{1.3}$$

Here, Γ_D and Γ_N denote a partitioning of $\Gamma = \partial\mathrm{D}$ into a Dirichlet and a Neumann part, γ_0 denotes the Dirichlet trace map, and γ_1 denotes the conormal trace operator, given (in strong form) by $\gamma_1(v) = n_x \cdot (A(x)\nabla_x v)|_\Gamma$, with $\Gamma = \partial\mathrm{D}$ denoting the boundary of D, and $n_x \in L^\infty(\Gamma; \mathbb{R}^d)$ the exterior unit normal vector field on Γ.

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* (1.1)–(1.3) *is linear, superposition for a sufficiently regular function $U(x, t)$ in Q, which satisfies* (1.2) *and* (1.3)*, will imply that the function $u - U$ will solve* (1.1)–(1.3) *with $g - BU$ in place of g in* (1.1)*, and with homogeneous initial and boundary data in* (1.2) *and* (1.3)*. All regularity hypotheses which we will impose below on the source term g in* (1.1) *(in particular, time-analyticity* (3.9)*) 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 $W \subset L^2(\mathrm{D})$ on the bounded spatial domain $\mathrm{D} \subset \mathbb{R}^d$.

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 $\mathrm{D} \subset \mathbb{R}^2$, 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 $W^{1,\infty}(\mathrm{D}; \mathbb{R}^{2\times 2})$. 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: $\mathbb{N} = \{1, 2, \dots\}$ 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 $L^q(\mathrm{D})$ is adopted for Lebesgue spaces of q-integrable functions $u : \mathrm{D} \to \mathbb{R}$ over some (bounded) domain D in the Euclidean space $\mathbb{R}^d$. For nonnegative integers k, Hilbertian Sobolev spaces (where $q = 2$) on such domains D are denoted by $H^k(\mathrm{D})$. For $k = 0$, as usual, $H^0(\mathrm{D}) = L^2(\mathrm{D})$. 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 $Q := J \times \mathrm{D}$, with the finite time interval $J := (0, T)$ and the bounded spatial domain $\mathrm{D} \subset \mathbb{R}^d$.

2.1. **Function Spaces.**

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

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

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

For any integer $k \geq 1$, we denote by H_0^k the closed subspace of H^k of functions with homogeneous boundary values in the sense of closure of C_0^∞ with respect to the norm of H^k. For instance, H_0^1 denotes the closed nullspace of the Dirichlet trace operator γ_0.

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

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

Evidently, for $\Gamma_D \subset \Gamma$, $H_0^1(\mathrm{D}) = H_\Gamma^1(\mathrm{D}) \subset H^1_{\Gamma_D}(\mathrm{D}) \subset H^1(\mathrm{D})$.

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 $v \colon (a, b) \to \mathbb{R}$. All results and proofs can be generalized straightforwardly to X-valued functions $v \colon (a, b) \to X$ for a Hilbert space X, i.e., Bochner–Sobolev spaces. We write

$$H^1_{0,}(a, b) = H^1_{\{a\}}(a, b) = \{v \in H^1(a, b) |\ v(a) = 0\},$$
$$H^1_{,0}(a, b) = H^1_{\{b\}}(a, b) = \{v \in H^1(a, b) |\ v(b) = 0\} \ .$$

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 $H^1_{,0}(a, b)$, whereas the space $H^1(a, b)$ is endowed with the norm $\|\circ\|_{H^1(a,b)} = (\|\circ\|^2_{L^2(a,b)} + \|\partial_t \circ\|^2_{L^2(a,b)})^{1/2}$.

Fractional order spaces shall be defined by interpolation, via the real method of interpolation (see, e.g., [38, Chapter 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 [34, Section 2.3] for $(a, b) = (0, T)$) defined by

$$\|v\|_{H^{1/2}_{0,}(a,b)} := \left(\sum_{k=0}^\infty \frac{\pi(2k+1)}{2(b-a)} |v_k|^2 \right)^{1/2}, \quad v \in H^{1/2}_{0,}(a, b), \tag{2.1}$$

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

$$z(t) = \sum_{k=0}^\infty z_k V_k(t), \quad z_k = \int_a^b z(s) V_k(s) \mathrm{d}s, \ k \in \mathbb{N}_0, \tag{2.2}$$

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

$$-\partial_{tt} V_k(t) = \lambda_k V_k(t) \ \text{ for } t \in (a, b), \quad V_k(a) = \partial_t V_k(b) = 0, \quad \|V_k\|_{L^2(a,b)} = 1. \tag{2.3}$$

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
(2.4)
$$v(t) = \sum_{k=0}^{\infty} v_k \sqrt{\frac{2}{T}} \sin\left(\left(\frac{\pi}{2} + k\pi\right)\frac{t}{T}\right),\ v_k = \int_0^T v(s)\sqrt{\frac{2}{T}} \sin\left(\left(\frac{\pi}{2} + k\pi\right)\frac{s}{T}\right) \mathrm{d}s.$$

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 [34, Section 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 $w_k = \int_0^T w(s)\sqrt{\frac{2}{T}} \cos\left(\left(\frac{\pi}{2} + k\pi\right)\frac{s}{T}\right) \mathrm{d}s$.

To prove exponential convergence of a temporal *hp*-discretization, we need further investigations of the Sobolev space $H^{1/2}_{0,}(a,b)$ and its norm $\|\circ\|_{H^{1/2}_{0,}(a,b)}$. For this purpose, let the classical Sobolev space $H^{1/2}(a,b)$ be endowed with the Slobodetskii norm [24, p. 74]

$$\|\|v\|\|_{H^{1/2}(a,b)} := \left(\|v\|^2_{L^2(a,b)} + |v|^2_{H^{1/2}(a,b)}\right)^{1/2} \tag{2.5}$$

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

$$|v|_{H^{1/2}(a,b)} := \left(\int_a^b\int_a^b \frac{|v(s)-v(t)|^2}{|s-t|^2}\mathrm{d}s\mathrm{d}t\right)^{1/2}. \tag{2.6}$$

With the Slobodetskii norm (2.5), we endow $H^{1/2}_{0,}(a,b)$ with the norm

$$\|\|v\|\|_{H^{1/2}_{0,}(a,b)} := \left(\|v\|^2_{L^2(a,b)} + |v|^2_{H^{1/2}(a,b)} + \int_a^b \frac{|v(t)|^2}{t-a}\mathrm{d}t\right)^{1/2} \tag{2.7}$$

for $v \in H^{1/2}_{0,}(a,b)$. We have the following equivalence result for the norms defined in (2.1) and (2.7), which is proven, e.g., in [24] (see the proof in Appendix A for the characterization of the equivalence constants).

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

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

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

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 [13] for a more general localization result.

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

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

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 (1.1) is taken by the nonlocal operator $\mathcal{H}_T \in \mathcal{L}(L^2(J), L^2(J))$, which is defined by

$$(\mathcal{H}_T v)(t) := \sum_{k=0}^{\infty} v_k \sqrt{\frac{2}{T}} \cos\left(\left(\frac{\pi}{2} + k\pi\right)\frac{t}{T}\right), \quad t \in J. \tag{2.8}$$

Here, $v \in L^2(J)$ and its Fourier coefficients $v_k = \int_0^T v(s)\sqrt{\frac{2}{T}} \sin\left(\left(\frac{\pi}{2} + k\pi\right)\frac{s}{T}\right) \mathrm{d}s$ are represented as in (2.4). We collect some properties of $\mathcal{H}_T$.

Proposition 2.3 ([34, Section 2.4], [35, 39])**.** *The modified Hilbert transformation $\mathcal{H}_T$ defined in* (2.8) *is a linear isometry as mapping*

$$\mathcal{H}_T \colon H^\nu_{0,}(J) \to H^\nu_{,0}(J) \quad \textit{for } \nu \in \{0, 1/2, 1\} \tag{2.9}$$

and is $H^{1/2}_{0,}(J)$-elliptic, satisfying

$$\forall v \in H^{1/2}_{0,}(J): \quad \langle \partial_t v, \mathcal{H}_T v\rangle_{L^2(J)} = \|v\|^2_{H^{1/2}_{0,}(J)}. \tag{2.10}$$

Additionally, $\mathcal{H}_T$ fulfills the following properties:

$$\forall v, w \in H^{1/2}_{0,}(J): \quad \langle \partial_t w, \mathcal{H}_T v\rangle_{L^2(J)} = \langle \mathcal{H}_T w, \partial_t v\rangle_{L^2(J)} = \langle w, v\rangle_{H^{1/2}_{0,}(J)}\ , \tag{2.11}$$

$$\forall w \in H^1_{0,}(J), \forall v \in L^2(J): \quad \langle \partial_t \mathcal{H}_T w, v\rangle_{L^2(J)} = -\langle \mathcal{H}_T^{-1} \partial_t w, v\rangle_{L^2(J)}\ , \tag{2.12}$$

$$\forall v, w \in L^2(J): \quad \langle \mathcal{H}_T v, w\rangle_{L^2(J)} = \langle v, \mathcal{H}_T^{-1} w\rangle_{L^2(J)}\ , \tag{2.13}$$

$$\forall v \in L^2(J): \quad \langle v, \mathcal{H}_T v\rangle_{L^2(J)} \geq 0\ , \tag{2.14}$$

$$\forall \nu \in \{1/2, 1\}, \forall v \in H^\nu_{0,}(J), v \neq 0: \quad \langle v, \mathcal{H}_T v\rangle_{L^2(J)} > 0\ . \tag{2.15}$$

Remark 2.4. *We remark that* (2.9)–(2.15) *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* [11] *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

$$\partial_t u = f \text{ in } J, \quad u(0) = 0\ .$$

A weak formulation relevant for treatment of IBVP (1.1) is to find $u \in H^{1/2}_{0,}(J)$ such that

$$\forall w \in H^{1/2}_{,0}(J): \ \langle \partial_t u, w\rangle_{L^2(J)} = \langle f, w\rangle_{L^2(J)} \tag{2.16}$$

for given $f \in [H^{1/2}_{,0}(J)]'$. Here, $\langle \circ, \circ\rangle_{L^2(J)}$ 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 (2.16) is inf-sup stable:

$$\inf_{0 \neq u \in H^{1/2}_{0,}(J)} \sup_{0 \neq w \in H^{1/2}_{,0}(J)} \frac{\langle \partial_t u, w\rangle_{L^2(J)}}{\|u\|_{H^{1/2}_{0,}(J)} \|w\|_{H^{1/2}_{,0}(J)}} \geq 1\ . \tag{2.17}$$

This is shown in [34, Rem. 2.10] 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 (2.16) then admits a unique solution $u \in H^{1/2}_{0,}(J)$.

For the derivation of a space-time variational formulation of (1.1), it is useful to consider a *parametric IVP*: for a given parameter $\mu \geq 0$ (eventually in the spectrum of the spatial operator of (1.1)) 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

$$\forall w \in H^{1/2}_{,0}(J) : \langle \partial_t u, w\rangle_{L^2(J)} + \mu\langle u, w\rangle_{L^2(J)} = \langle f, w\rangle_{L^2(J)} \,. \tag{2.18}$$

A *Bubnov–Galerkin variational form* with equal trial and test function spaces is to find $u \in H^{1/2}_{0,}(J)$ such that

$$\forall v \in H^{1/2}_{0,}(J) : \langle \partial_t u, \mathcal{H}_T v\rangle_{L^2(J)} + \mu\langle u, \mathcal{H}_T v\rangle_{L^2(J)} = \langle f, \mathcal{H}_T v\rangle_{L^2(J)} \,. \tag{2.19}$$

Both formulations (2.18) and (2.19) admit unique solutions due to (2.17) and (2.14).

2.4. **Temporal hp-Discretization.** To discretize (2.19), 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* $\boldsymbol{p} := (p_j)_{j=1}^m \in \mathbb{N}^m$. We define

$$V^M_t = S^{\boldsymbol{p},1}_{0,}(J;\mathcal{G}) := \{v \in H^1_{0,}(J) : v_{|_{I_j}} \in \mathbb{P}^{p_j},\ I_j \in \mathcal{G}\} \,. \tag{2.20}$$

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

We restrict (2.19) to V^M_t to obtain the temporal hp-approximation: find $u^M_t \in V^M_t$ such that

$$\forall v \in V^M_t : \langle \partial_t u^M_t, \mathcal{H}_T v\rangle_{L^2(J)} + \mu\langle u^M_t, \mathcal{H}_T v\rangle_{L^2(J)} = \langle f, \mathcal{H}_T v\rangle_{L^2(J)} \,. \tag{2.21}$$

Due to the inf-sup stability (2.17) and $\mu \geq 0$, the discretization (2.21) is well-posed with inf-sup constant independent of $\mathcal{G}$ and of $\boldsymbol{p}$. Its numerical implementation will require, similar to [34, 11], the efficient evaluation of $\mathcal{H}_T v$ 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 (1.1). Its variational form reads: given a source term $f \in L^2(\mathrm{D})$, find

$$w \in H^1_{\Gamma_D}(\mathrm{D}) \text{ such that } \forall v \in H^1_{\Gamma_D}(\mathrm{D}) : a(w, v) = \langle f, v\rangle_{L^2(\mathrm{D})} \,. \tag{2.22}$$

Here, $a(w, v) = \int_{\mathrm{D}} A(x)\nabla_x w(x) \cdot \nabla_x v(x)\mathrm{d}x$. We assume uniform positive definiteness of A:

$$a_{\min} := \operatorname*{ess\,inf}_{x\in\mathrm{D}} \inf_{0\neq\xi\in\mathbb{R}^d} \frac{\xi^\top A(x)\xi}{\xi^\top \xi} > 0 \,. \tag{2.23}$$

With assumption (2.23), we have

$$\forall w \in H^1_{\Gamma_D}(\mathrm{D}): \quad a(w,w) \geq a_{\min}\|\nabla_x w\|^2_{L^2(\mathrm{D})} \geq a_{\min} c \|w\|^2_{H^1(\mathrm{D})}$$

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 $v,w \in H^1(\mathrm{D})$ ensure that the corresponding eigenvalue problem to find

$$0 \neq \phi \in H^1_{\Gamma_D}(\mathrm{D}),\ \mu \in \mathbb{R}: \ \ \forall v \in H^1_{\Gamma_D}(\mathrm{D}): \ a(\phi, v) = \mu \langle \phi, v\rangle_{L^2(\mathrm{D})} \tag{2.24}$$

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 ϕ_k orthonormal in $L^2(\mathrm{D})$ and orthogonal in $H^1_{\Gamma_D}(\mathrm{D})$, and with μ_k accumulating only at ∞. In view of the forthcoming analysis, *in what follows, we endow* $H^1_{\Gamma_D}(\mathrm{D})$ *with the "energy" norm* $a(\circ,\circ)^{1/2}$. We remark that, for $v \in H^1_{\Gamma_D}(\mathrm{D})$, $a(v,v) = \sum_{i=1}^{\infty} \mu_i |v_i|^2$, where $v_i = \langle v, \phi_i\rangle_{L^2(\mathrm{D})}$.

The space-time variational formulation of (1.1) is based on the intersection space

$$H^{1,1/2}_{\Gamma_D;0,}(Q) := \left(L^2(J) \otimes H^1_{\Gamma_D}(\mathrm{D})\right) \cap \left(H^{1/2}_{0,}(J) \otimes L^2(\mathrm{D})\right)\ ,$$

which we equip with the corresponding sum norm. The space $H^{1,1/2}_{\Gamma_D;,0}(Q)$ is defined analogously. Proceeding as in [34, Thm. 3.2], the *initial-boundary value problem* (1.1)–(1.3) *is set as a well-posed operator equation.*

Theorem 2.5. *Consider* (1.1)–(1.3) *with homogeneous data* $u_0 = 0$ *in* (1.2) *and* $u_D, u_N = 0$ *in* (1.3)*. Assume* $|\Gamma_D| > 0$ *if* $d = 2,3$ *or* $\Gamma_D \neq \emptyset$ *if* $d = 1$*, and that the coefficient* $A \in L^\infty(\mathrm{D}; \mathbb{R}^{d\times d}_{\mathrm{sym}})$ *satisfies* (2.23)*.*

Then, the space-time variational formulation of (1.1) *to find* $u \in H^{1,1/2}_{\Gamma_D;0,}(Q)$ *such that*

$$\forall v \in H^{1,1/2}_{\Gamma_D;,0}(Q): \ \langle \partial_t u, v\rangle_{L^2(Q)} + \langle A\nabla_x u, \nabla_x v\rangle_{L^2(Q)} = \langle g, v\rangle_{L^2(Q)} \tag{2.25}$$

induces an isomorphism

$$B := \partial_t + A(\partial_x) \in \mathcal{L}_{\mathrm{iso}}(H^{1,1/2}_{\Gamma_D;0,}(Q), [H^{1,1/2}_{\Gamma_D;,0}(Q)]')\ .$$

In particular, for every $g \in [H^{1,1/2}_{\Gamma_D;,0}(Q)]'$*, IBVP* $Bu = g$ *in* (1.1) *admits a unique solution* $u \in H^{1,1/2}_{\Gamma_D;0,}(Q)$*.*

We remark that $\langle \circ, \circ\rangle_{L^2(Q)}$ denotes the inner product in $L^2(Q)$ and as continuous extension of it, also the duality pairing with respect to $[H^{1,1/2}_{\Gamma_D;,0}(Q)]'$ and $H^{1,1/2}_{\Gamma_D;,0}(Q)$. The *space-time discretization* of (2.25) is straightforward: for any conforming, spatial finite element subspace $V_x^N \subset H^1_{\Gamma_D}(\mathrm{D})$ of finite dimension N, and for the temporal *hp*-subspace $V_t^M \subset H^{1/2}_{0,}(J)$ introduced in (2.20), we restrict (2.25) to the space-time approximation space

$$V_t^M \otimes V_x^N \subset H^{1,1/2}_{\Gamma_D;0,}(Q)\ . \tag{2.26}$$

That is, we seek an approximate solution $u^{MN} \in V_t^M \otimes V_x^N$ such that

$$\langle \partial_t u^{MN}, v\rangle_{L^2(Q)} + \langle A\nabla_x u^{MN}, \nabla_x v\rangle_{L^2(Q)} = \langle g, v\rangle_{L^2(Q)} \tag{2.27}$$

holds true for all $v \in (\mathcal{H}_T V_t^M) \otimes V_x^N \subset H^{1,1/2}_{\Gamma_D;,0}(Q)$.

For these choices of test function spaces and for *any* subspace $V_x^N \subset V$ of finite dimension N, as in [34, Sect. 3], existence and uniqueness of the discrete solution $u^{MN} \in V_t^M \otimes V_x^N \subset H^{1,1/2}_{\Gamma_D;0,}(Q)$ of (2.27) follow from the *continuous inf-sup condition*

$$\inf_{0\neq u \in H^{1,1/2}_{\Gamma_D;0,}(Q)} \sup_{0\neq w \in H^{1,1/2}_{\Gamma_D;,0}(Q)} \frac{\langle \partial_t u, w\rangle_{L^2(Q)} + \langle A\nabla_x u, \nabla_x w\rangle_{L^2(Q)}}{\|u\|_{H^{1,1/2}_{\Gamma_D;0,}(Q)} \|w\|_{H^{1,1/2}_{\Gamma_D;,0}(Q)}} \geq \frac{1}{2} \;.$$

With $H^1_{\Gamma_D}(\mathrm{D})$ endowed with the $a(\circ,\circ)^{1/2}$ norm, the proof of this condition with constant independent of A follows *verbatim* that of [34, Thm. 3.2, Cor. 3.3] for the case $A = \mathbb{I}$. Evidently, the stability of the discrete problem is a consequence of the choice of the test function space $\mathcal{H}_T V_t^M$, whose efficient numerical realization will be discussed in Section 6.

3. Regularity

To obtain convergence rate bounds, we address the regularity of the solution $u \in H^{1,1/2}_{\Gamma_D;0,}(Q)$. We consider separately the temporal and spatial regularity. The solution operator to the parabolic equation (1.1) being an analytic semigroup, for time-analytic forcing g in (1.1) 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 [27, 11] 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 $u(t) := u(t,\circ) \in L^2(\mathrm{D})$. To this end, we recall the eigenvalue problem (2.24). Setting $H := L^2(\mathrm{D})$ and thus denoting by $\langle\circ,\circ\rangle_H$ the $L^2(\mathrm{D})$ inner product, the solution $u(t)$ of (1.1) at time $t > 0$ for $g = 0$, $u_D = u_N = 0$, and for initial data $u_0 \in H$ may be written as

$$u(t) = E(t)u_0 := \sum_{i=1}^{\infty} \exp(-\mu_i t)\langle u_0, \phi_i\rangle_H \phi_i \tag{3.1}$$

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

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

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

$$X_r := \{v \in H : \|v\|^2_{X_r} := \sum_{i=1}^{\infty} \mu_i^r |v_i|^2 < \infty\} \;. \tag{3.2}$$

Here, $v_i = \langle v, \phi_i\rangle_H$ denotes the i-th coefficient in the eigenfunction expansion of v (recall from (2.24) that the sequence $\{\phi_i\}_{i\geq 1}$ was assumed to be an orthonormal basis of $H = X_0$). We remark that the norm $\|\circ\|_{X_1}$ is the energy-norm on the space $V = H^1_{\Gamma_D}(\mathrm{D})$, due to

$$\forall v \in V: \quad \|v\|^2_{X_1} = a(v,v) = \sum_{i=1}^{\infty} \mu_i |v_i|^2 \;.$$

For $|\Gamma_D| > 0$ if $d = 2,3$ or $\Gamma_D \neq \emptyset$ if $d = 1$, $\|\circ\|_{X_1}$ is equivalent to the $H^1(\mathrm{D})$ norm on V and the norm bounds $\|v\|_{X_r} \leq c\|v\|_{X_{r'}}$ for $r' \geq r$ follow from (3.2) and the assumed enumeration of the real eigenvalues $\mu_i > 0$ with $\mu_i \uparrow \infty$ as $i \uparrow \infty$:

$$\|v\|^2_{X_r} = \sum_{i=1}^{\infty} \mu_i^r |v_i|^2 \leq \left(\sup_{m\in\mathbb{N}} \mu_m^{r-r'}\right) \sum_{i=1}^{\infty} \mu_i^{r'} |v_i|^2 \leq \mu_1^{r-r'} \|v\|^2_{X_{r'}} . \tag{3.3}$$

For $\theta, r \geq 0$ and for any $t > 0$, $E(t)$ in (3.1) belongs to $\mathcal{L}(X_\theta, X_r)$. In fact, for any $t > 0$ and $v \in X_\theta$, we have

$$\|E(t)v\|_{X_r}^2 = \sum_{i=1}^{\infty} \mu_i^r \exp(-2\mu_i t)|v_i|^2 = \sum_{i=1}^{\infty} \mu_i^{r-\theta} \exp(-2\mu_i t)\mu_i^\theta |v_i|^2. \tag{3.4}$$

For $\theta \geq r \geq 0$, identity (3.4) 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

$$\forall \theta \geq r \geq 0,\ \forall t > 0: \quad \|E(t)\|_{\mathcal{L}(X_\theta, X_r)}^2 \leq \mu_1^{-(\theta-r)} \exp(-2\mu_1 t)\ . \tag{3.5}$$

For $r \geq \theta \geq 0$, for any $t > 0$ and $v \in X_\theta$, identity (3.4) implies

$$\|E(t)v\|_{X_r}^2 \leq \sup_{i \in \mathbb{N}} \{\mu_i^{r-\theta} \exp(-2\mu_i t)\} \sum_{i=1}^{\infty} \mu_i^\theta |v_i|^2 =: G_{r-\theta}(t)\|v\|_{X_\theta}^2\ . \tag{3.6}$$

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

$$\forall t > 0: \quad G_{2\sigma}(t) \leq G_{\max}(\sigma, t) := [\mu_*^\sigma \exp(-\mu_* t)]^2 = \left(\frac{\sigma}{t\mathrm{e}}\right)^{2\sigma}\ . \tag{3.7}$$

Inserting (3.7) with $\sigma = (r-\theta)/2 > 0$ into (3.6), we arrive at

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

The exponential decay of the Fourier coefficients for $t > 0$ implied by the exponential weighting $\exp(-\mu_i t)$ 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 (1.1) for $u_0 = 0$ and $u_D = u_N = 0$ with the Duhamel representation (see, e.g., [26])

$$u(t) = \int_0^t E(t-s)g(s)\mathrm{d}s, \quad 0 < t \leq T\ . \tag{3.8}$$

We work under the following *time-analyticity assumption* on the forcing g in (1.1): There exist constants $C > 0$ and $\delta \geq 1$ such that, for some $\varepsilon \in (0,1)$, we have

$$\forall\, l \in \mathbb{N}_0: \quad \sup_{0 \leq t \leq T} \|g^{(l)}(t)\|_{X_\varepsilon} \leq C\delta^l \Gamma(l+1), \tag{3.9}$$

where $\Gamma(\circ)$ denotes the gamma function fulfilling $\Gamma(l) = (l-1)!$ for all $l \in \mathbb{N}$. *Formally differentiating* (3.8) l-times with respect to t, upon writing it equivalently as $u(t) = \int_0^t E(s)g(t-s)\mathrm{d}s$, gives

$$\frac{\mathrm{d}^l}{\mathrm{d}t^l} u(t) = \sum_{i=0}^{l-1} E^{(i)}(t) g^{(l-1-i)}(0) + \int_0^t E(s) g^{(l)}(t-s)\mathrm{d}s\ , \quad l \in \mathbb{N}, t > 0\ . \tag{3.10}$$

The right limits at $t = 0$ of the time-derivatives of the forcing g in (1.1) contribute to the time-regularity. We estimate the norm of the operators $E^{(l)}(t)$ in $\mathcal{L}(X_\theta, X_r)$.

Lemma 3.1. *For $r \geq \theta \geq 0$, we have*
(3.11)
$$\forall l \in \mathbb{N}_0,\ \forall t > 0: \quad \|E^{(l)}(t)\|^2_{\mathcal{L}(X_\theta, X_r)} \leq \frac{1}{\sqrt{2\pi}} \left(\frac{1}{2}\right)^{2l+r-\theta} \Gamma(2l+1+r-\theta)\, t^{-2l-(r-\theta)}\ .$$

Proof. For $v \in H = X_0$, 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 (3.1) yields (with formal, term-by-term differentiation)
$$\frac{\mathrm{d}^l}{\mathrm{d}t^l} v(t) = \sum_{i=1}^{\infty} (-\mu_i)^l \exp(-\mu_i t) v_i \phi_i$$
with convergence in H for arbitrary, fixed $t > 0$. Therefore
$$\forall t > 0: \quad \|v^{(l)}(t)\|^2_{X_r} = \sum_{i=1}^{\infty} \mu_i^{2l+r-\theta} \exp(-2\mu_i t) \mu_i^\theta |v_i|^2\ .$$
It follows from (3.7) that for every $t > 0$
$$\|v^{(l)}(t)\|^2_{X_r} \leq G_{2(l+[r-\theta]/2)}(t) \|v\|^2_{X_\theta} \leq G_{\max}(l + [r-\theta]/2, t) \|v\|^2_{X_\theta}\ .$$
Therefore, for every $v \in X_\theta$ and every $r \geq \theta \geq 0$, we have
$$\begin{aligned} \forall l \in \mathbb{N}, t > 0: \quad \|v^{(l)}(t)\|^2_{X_r} &\leq \left(\frac{2l + r - \theta}{2t\mathrm{e}}\right)^{2l+r-\theta} \|v\|^2_{X_\theta} \\ &= \left(\frac{1}{2}\right)^{2l+r-\theta} \left(\frac{2l+r-\theta}{\mathrm{e}}\right)^{2l+r-\theta} t^{-2l-(r-\theta)} \|v\|^2_{X_\theta}\ . \end{aligned}$$
For $x \in \mathbb{R}_+$, the Stirling's formula (C.1) states $\sqrt{2\pi}\, x^{x-1/2} \mathrm{e}^{-x} \leq \Gamma(x)$, which implies $(x/\mathrm{e})^x \leq \frac{1}{\sqrt{2\pi}} x^{-1/2} \Gamma(x+1)$. With $x = 2l + r - \theta$, this gives the claimed bound, as $(2l + r - \theta)^{-1/2} \leq 1$. ☐

Lemma 3.2. *Assume* (3.9) *with some $\varepsilon \in (0,1)$ and some $\delta \geq 1$. For $r \in [0,2]$, there exists a constant $C > 0$ (independent of δ, l, t) such that, for every $l \in \mathbb{N}_0$ and $t > 0$, we have*
$$\|u^{(l)}(t)\|_{X_r} \leq C\delta^l \Gamma(l+1) \left(t^{(2-r+\min\{r,\varepsilon\})/2} + \sum_{i=0}^{l-1} t^{-i-r/2+\varepsilon/2} \right)\ . \tag{3.12}$$
For $l = 0$, this bound is valid without the sum.

Proof. From (3.10), we estimate for every $0 < t \leq T$
$$\begin{aligned} \|u^{(l)}(t)\|_{X_r} \leq \sum_{i=0}^{l-1} \|E^{(i)}(t)\|_{\mathcal{L}(X_\varepsilon, X_r)} \|g^{(l-i-1)}(0)\|_{X_\varepsilon} \\ + \int_0^t \|E(s)\|_{\mathcal{L}(X_\varepsilon, X_r)} \|g^{(l)}(t-s)\|_{X_\varepsilon} \mathrm{d}s. \end{aligned}$$

To estimate the sum, we use (3.11) with $\theta = \varepsilon$ and assumption (3.9) and obtain

$$\begin{aligned}
&\sum_{i=0}^{l-1} \|E^{(i)}(t)\|_{\mathcal{L}(X_\varepsilon, X_r)} \|g^{(l-i-1)}(0)\|_{X_\varepsilon} \\
&\quad \leq \sum_{i=0}^{l-1} C \left(\frac{1}{2}\right)^{i+r/2-\varepsilon/2} \Gamma(2i+1+r-\varepsilon)^{1/2} t^{-i-r/2+\varepsilon/2} \delta^{l-1-i} \Gamma(l-i) \\
&\quad \leq C\delta^{l-1} \sum_{i=0}^{l-1} \left(\frac{1}{2}\right)^{i+r/2-\varepsilon/2} \Gamma(2i+1+r-\varepsilon)^{1/2} \Gamma(l-i)\, t^{-i-r/2+\varepsilon/2} \\
&\quad \leq C\delta^{l-1} \sum_{i=0}^{l-1} \Gamma(i+1+r/2-\varepsilon/2)\Gamma(l-i)\, t^{-i-r/2+\varepsilon/2} \\
&\quad \leq C\delta^{l-1} \Gamma(l+1) \sum_{i=0}^{l-1} t^{-i-r/2+\varepsilon/2} \ ,
\end{aligned}$$

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 (3.9) with $\varepsilon \in (0,1)$ and (3.11) with $l = 0$, $\theta = \varepsilon > 0$ and $\varepsilon \leq r \leq 2$, and obtain, for every $l \in \mathbb{N}_0$,

$$\begin{aligned}
\int_0^t \|E(s)\|_{\mathcal{L}(X_\varepsilon, X_r)} \|g^{(l)}(t-s)\|_{X_\varepsilon} \mathrm{d}s &\leq C\delta^l \Gamma(l+1) \int_0^t s^{-(r-\varepsilon)/2} \mathrm{d}s \\
&= CC_\varepsilon \delta^l \Gamma(l+1) t^{(2-r+\varepsilon)/2}
\end{aligned}$$

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

$$\begin{aligned}
\int_0^t \|E(s)\|_{\mathcal{L}(X_\varepsilon, X_r)} \|g^{(l)}(t-s)\|_{X_\varepsilon} \mathrm{d}s &\leq C\delta^l \Gamma(l+1) \int_0^t \mu_1^{-(\varepsilon-r)/2} \exp(-\mu_1 s) \mathrm{d}s \\
&\leq C\widetilde{C}_\varepsilon \delta^l \Gamma(l+1) t \ ,
\end{aligned}$$

where the bound (3.5) is used ($\widetilde{C}_\varepsilon = \mu_1^{-(\varepsilon-r+2)/2}$). This completes the proof of the assertion. □

Remark 3.3. *For $0 \leq r < 2$, the preceding result is valid under hypothesis* (3.9) *with $\varepsilon = 0$, as used, e.g., in* [27]*, but with $C(r) \uparrow \infty$ as $r \uparrow 2$.*

Lemma 3.4. *Assume* (3.9) *with some $\varepsilon \in (0,1)$ and some $\delta \geq 1$. Let u be the solution of* (2.25)*. For $T \geq b > a \geq 1$, the estimate*

$$\forall l \in \mathbb{N}_0: \quad \left(\int_a^b \|u^{(l)}(t)\|_{X_2}^2 \mathrm{d}t \right)^{1/2} \leq \delta^l \Gamma(l+2) C(\varepsilon, a, b) \tag{3.13}$$

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

$$\left(\int_0^T \|u(t)\|_H^2 \mathrm{d}t \right)^{1/2} \leq CT^{3/2}, \quad \left(\int_0^T \|u'(t)\|_H^2 \mathrm{d}t \right)^{1/2} \leq C\delta\, (T^{1+\varepsilon} + T^3)^{1/2} \tag{3.14}$$

and $u \in L^2(J;X_2)$ *with*

$$\|u\|_{L^2(J;X_2)} \le CT^{\varepsilon/2+1/2}, \tag{3.15}$$

where the constant $C>0$ *is independent of* ε, δ *and* T.

Proof. The bound (3.13) follows from (3.12). For the estimates (3.14) and (3.15), we use (3.12) for $r=0$ with $l=0$ or $l=1$, and $r=2$ with $l=0$, respectively. □

Proposition 3.5. *Assume* (3.9) *with some* $\varepsilon \in (0,1)$ *and some* $\delta \ge 1$. *For* $r \in [0,2]$, *there exists a constant* $C>0$ *(independent of* l, δ, a, b, t, q*) such that the solution* u *of* (2.25) *satisfies*

$$\forall l \in \mathbb{N}_0,\ \forall t \in (0,\min\{1,T\}]:\ \|u^{(l)}(t)\|_{X_r} \le C\delta^l\Gamma(l+2)t^{-l+1-r/2+\varepsilon/2}\ , \tag{3.16}$$

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

$$\forall l \in \mathbb{N}, l \ge 2:\quad \left(\int_a^b \|u^{(l)}(t)\|_{X_r}^2 \mathrm{d}t\right)^{1/2} \le C\delta^l\Gamma(l+2)a^{-l+3/2-r/2+\varepsilon/2}\ , \tag{3.17}$$

and, for arbitrary $q \ge 2$,

$$\|u\|_{H^q((a,b);X_r)} \le C\delta^q\Gamma(q+3)a^{-q+3/2-r/2+\varepsilon/2}\ . \tag{3.18}$$

Proof. The bound (3.16) follows from (3.12). Estimate (3.17) is obtained by integrating the pointwise bound (3.16). The Sobolev bound (3.18) follows by interpolation. □

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* (3.9) *with some* $\varepsilon \in (0,1)$ *and some* $\delta \ge 1$ *for* $l=0,1$. *Then, for* $b \in (0,T]$, *the solution* u *of* (2.25) *belongs to* $H^{1/2}_{0,}((0,b);X_2)$ *and the estimate*

$$|||u|||_{H^{1/2}_{0,}((0,b);X_2)} \le \sqrt[4]{\frac{2}{\pi}}\,\frac{1}{\varepsilon}\,b^{\varepsilon/2}\left(\frac{b}{1+\varepsilon}+\frac{3}{\varepsilon}+\frac{4b^2}{(\varepsilon+1)(\varepsilon+2)}\right)^{1/2} C_g$$

holds true, with

$$C_g := \|g\|_{W^{1,\infty}((0,b);X_\varepsilon)} = \max\left\{\sup_{0\le t\le b}\|g(t)\|_{X_\varepsilon},\ \sup_{0\le t\le b}\|g'(t)\|_{X_\varepsilon}\right\}. \tag{3.19}$$

Remark 3.7. *The assertion of Lemma 3.6 can be generalized. For this purpose, define the interpolation space* $H^\theta_{0,}(a,b) := (H^1_{0,}(a,b), L^2(a,b))_{\theta,2}$ *for* $\theta \in [1/2,1]$ *with the usual Slobodetskii norm* $|||\circ|||_{H^\theta(a,b)}$ *as in* [24, p. 74], *where* $a<b$, $a,b \in \mathbb{R}$. *Then, under the assumption of Lemma 3.6, we have*

$$|||u|||_{H^\theta_{0,}((0,b);X_2)} \le C(b,\varepsilon,\theta,g)$$

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 $x \in \mathrm{D}$. For (1.1), this regularity is, of course, dependent on the temporal variable t, and the spaces X_r defined in (3.2) via eigensystems, which are intrinsic to the spatial operator (2.22) with (2.23), 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 $\mathbb{P}^1$-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 (3.2). 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 $\mathrm{D} \subset \mathbb{R}^d$ is a bounded Lipschitz domain. Assume further that $A \in L^\infty(\mathrm{D}; \mathbb{R}^{d\times d}_{\mathrm{sym}})$ is uniformly positive definite in the sense that (2.23) is satisfied. Then, $X_0 = L^2(\mathrm{D})$ and $X_1 \simeq H^1_{\Gamma_D}(\mathrm{D})$ and for $0 < r < 1$, $X_r \simeq (L^2(\mathrm{D}), H^1_{\Gamma_D}(\mathrm{D}))_{r,2}$.*

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 (2.22), with assumption (2.23) in place. In addition, we assume

$$f \in L^2(\mathrm{D}), \quad A \in W^{1,\infty}(\mathrm{D}; \mathbb{R}^{d\times d}_{\mathrm{sym}}). \tag{3.20}$$

Then, eigenfunction expansions of $f \in L^2(\mathrm{D})$ imply that the unique solution $u \in X_1$ of (2.22) belongs to X_2. Furthermore, the solution operator is bijective, since from (3.2) and (3.20) it follows that

$$\|u\|^2_{X_2} = \sum_{k=1}^{\infty} \mu_k^2 |u_k|^2 = \|A(\partial_x)u\|^2_H = \sum_{k=1}^{\infty} |f_k|^2 = \|f\|^2_{L^2(\mathrm{D})}. \tag{3.21}$$

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 (3.21), $X_2 = (A(\partial_x))^{-1}L^2(\mathrm{D})$. To characterize elements in X_2, we use the elliptic regularity of the BVP (2.22) with time-independent data $f \in L^2(\mathrm{D})$ in standard (if $d = 1$) or corner-weighted (if $d \geq 2$) Sobolev spaces in $\mathrm{D} \subset \mathbb{R}^d$.

3.2.1. **Case** $d = 1$. The spatial domain D is an open, bounded and connected interval, and, by (3.20), the diffusion coefficient is a scalar $a \in W^{1,\infty}(\mathrm{D})$ such that (2.23) is satisfied.

Standard elliptic regularity results imply that there exists a constant $c > 0$ such that, for every $f \in L^2(\mathrm{D})$, the solution $u = A(\partial_x)^{-1}f$ belongs to $H^2(\mathrm{D})$ and satisfies $\|v\|_{H^2(\mathrm{D})} \leq c\|f\|_{L^2(\mathrm{D})}$. This, combined with (3.21), gives that $X_2 \subset H^2(\mathrm{D})$ and

$$\forall v \in X_2 : \quad \|v\|_{H^2(\mathrm{D})} \leq c\|v\|_{X_2} . \tag{3.22}$$

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_{sub} disjoint, open and connected subintervals $\mathcal{D} = \{\mathrm{D}_i\}_{i=1}^{n_{\mathrm{sub}}}$ and denote the corresponding broken Sobolev spaces $W^{1,\infty}(\mathcal{D}) = \{a \in L^\infty(\mathrm{D}) : a_{|\mathrm{D}_i} \in W^{1,\infty}(\mathrm{D}_i),\ i = 1, \ldots, n_{\mathrm{sub}}\}$ and $H^2(\mathcal{D}) := \{v \in H^1(\mathrm{D}) : v_{|\mathrm{D}_i} \in H^2(\mathrm{D}_i),\ i = 1, \ldots, n_{\mathrm{sub}}\}$. We set $\|v\|^2_{H^2(\mathcal{D})} := \|v\|^2_{H^1(\mathrm{D})} + \sum_{i=1}^{n_{\mathrm{sub}}} |v|^2_{H^2(\mathrm{D}_i)}$. We assume that the diffusion coefficient a belongs to $W^{1,\infty}(\mathcal{D})$ and satisfies (2.23). In this case, standard elliptic regularity results imply that there exists a constant $c > 0$ such that, for every $f \in L^2(\mathrm{D})$, $u = A(\partial_x)^{-1}f \in H^2(\mathcal{D})$ and $\|v\|_{H^2(\mathcal{D})} \leq c\|f\|_{L^2(\mathrm{D})}$. This,* *combined with* (3.21)*, gives* $X_2 \subset H^2(\mathcal{D})$ *and* (3.22) *is valid with* $\|v\|_{H^2(\mathcal{D})}$ *on the left side.*

3.2.2. **Case** $d = 2$. Under (3.20), for polygonal domains $\mathrm{D} \subset \mathbb{R}^2$, weak solutions of the source problem (2.22) 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* $\mathrm{D} \subset \mathbb{R}^2$ *is a bounded polygonal domain with* ≥ 3 *corners and straight sides, whose boundary* $\partial\mathrm{D}$ *is Lipschitz.*

Denote by $r_{\mathrm{D}} : \mathrm{D} \to \mathbb{R}_{\geq 0}$ *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* $a > 0$*, the Kondrat'ev corner-weighted Sobolev space* $\mathcal{K}^m_a(\mathrm{D})$ *is defined as*

$$\mathcal{K}^m_a(\mathrm{D}) := \left\{ v\colon\ \mathrm{D} \to \mathbb{R} :\ \forall|\alpha| \leq m :\ r_{\mathrm{D}}^{|\alpha|-a} \partial^\alpha v \in L^2(\mathrm{D}) \right\}\ , \tag{3.23}$$

with $\|u\|^2_{\mathcal{K}^m_a(\mathrm{D})} := \sum_{|\alpha|\leq m} \|r_{\mathrm{D}}^{|\alpha|-a} \partial^\alpha v\|^2_{L^2(\mathrm{D})}$.

The regularity result in question is a special case of [8, Thm. 4.4], which we state here for definiteness in the form required by us.

Proposition 3.11. *Assume that* $\mathrm{D} \subset \mathbb{R}^2$ *is a bounded polygon with boundary* $\partial\mathrm{D}$ *consisting of a finite number of straight sides. Consider the elliptic source problem* (2.22) *with assumptions* (2.23) *and* (3.20) *in place.*

Then, there exist $c > 0$ *and a constant* $a > 0$ *such that, for every* $f \in L^2(\mathrm{D})$*, the weak solution* $u \in X_1 = H^1_{\Gamma_D}(\mathrm{D})$ *of* (2.22) *belongs to* $\mathcal{K}^2_{a+1}(\mathrm{D})$ *and satisfies the a priori estimate*

$$\|u\|_{\mathcal{K}^2_{a+1}(\mathrm{D})} \leq c\|f\|_{L^2(\mathrm{D})}. \tag{3.24}$$

In particular, therefore, $X_2 \subset \mathcal{K}^2_{a+1}(\mathrm{D})$ *and there exists* $c > 0$ *such that*

$$\forall v \in X_2 :\quad \|v\|_{\mathcal{K}^2_{a+1}(\mathrm{D})} \leq c\|v\|_{X_2}\ . \tag{3.25}$$

Proof. Assumption (3.20) implies that $A \in \mathcal{W}^{1,\infty}(\mathrm{D})$ as defined in [8, Eqn. (5)], and that $\|A\|_{\mathcal{W}^{1,\infty}(\mathrm{D})} \leq C(\mathrm{D})\|A\|_{W^{1,\infty}(\mathrm{D})}$. We may then use [8, Thm. 4.4] with $b_i = c = 0$, $m = 1$, to conclude the *a priori* estimate

$$\|u\|_{\mathcal{K}^2_{a+1}(\mathrm{D})} \leq c\|f\|_{\mathcal{K}^0_{a-1}(\mathrm{D})}$$

for all $|a| < \eta$ for some (sufficiently small) $\eta > 0$. We assume, without loss of generality, that $0 < \eta < 1$. Then, definition (3.23) states that $f \in \mathcal{K}^0_{a-1}(\mathrm{D})$ means $r_{\mathrm{D}}^{-(a-1)} f \in L^2(\mathrm{D})$. As $-(a-1) > 0$, $r_{\mathrm{D}}^{-(a-1)} \in L^\infty(\mathrm{D})$, so that $\|f\|_{\mathcal{K}^0_{a-1}(\mathrm{D})} \leq c(a,\mathrm{D})\|f\|_{L^2(\mathrm{D})}$. The *a priori* estimate implies then (3.24). Since $\|f\|_{L^2(\mathrm{D})} = \|u\|_{X_2}$ (see (3.21)), the *a priori* estimate also implies (3.25). □

Remark 3.12. *For transmission problems in a polygonal domain* D*, with* piecewise constant, isotropic coefficients *in materials occupying a finite number* n_{sub} *of polygonal subdomains* $\mathrm{D}_i \subset \mathrm{D}$*, regularity in the weighted spaces* $\mathcal{K}^2_{a+1}(\mathrm{D})$ *with radial weights also at multi-material intersection points in* D *are stated in* [22, Theorem 3.7]*. The assumptions in* [22] *on* A *are more restrictive than just* (2.23) *and* $A \in W^{1,\infty}(\mathcal{D}; \mathbb{R}^{d\times d}_{\mathrm{sym}})$*. The regularity result in* [22, Theorem 3.7] *with* $m = 1$ *will imply for* $u \in X_2$ *a splitting* $u = u_{\mathrm{reg}} + w_s$*, with the bound* (3.24) *for* $u_{\mathrm{reg}}|_{\mathrm{D}_i}$ *on each subdomain* D_i*, and with* w_s *in a finite-dimensional space* W_s*, see* [22, Sect. 3.2].

3.2.3. **Case** $d = 3$. Proposition 3.11 remains valid in space dimension $d = 3$. To detail a precise statement, we still assume (3.20). Then, [2, Theorem 1.1] implies (3.24) and (3.25) in bounded, polyhedral domains $\mathrm{D} \subset \mathbb{R}^3$ with Lipschitz boundary $\partial\mathrm{D}$ consisting of a finite number of plane faces. Similar results are shown in [23] and, for the Poisson equation with $\Gamma = \Gamma_D$, in [6, Theorem 1.2] (with $\mu = 1$ 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 (2.25). In order to use the tensor product construction of subspaces in (2.26), we specify the choice of temporal subspaces $V_t^M \subset H^{1/2}_{0,}(J)$ for the temporal domain $J = (0,T)$. In the spatial domain D, $V_x^N \subset H^1_{\Gamma_D}(\mathrm{D})$ will be specified in Section 4.2 below.

4.1. ***hp*-Approximation in $\overline{J} = [0,T]$.** To specify the hp-subspace $V_t^M \subset H^{1/2}_{0,}(J)$ in (2.26), 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

$$t_j := \begin{cases} 0, & j = 0, \\ T_1\sigma^{m_1-j}, & j \in \{1,\dots,m_1\}, \\ \frac{T-T_1}{m_2}\cdot(j-m_1)+T_1, & j \in \{m_1+1,\dots,m_1+m_2\}, \qquad \text{if } m_2 > 0, \end{cases} \tag{4.1}$$

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

$$k_j = \begin{cases} T_1\sigma^{m_1-1}, & j = 1, \\ T_1\sigma^{m_1-j}(1-\sigma), & j \in \{2,\dots,m_1\}, \\ k_T := \frac{T-T_1}{m_2}, & j \in \{m_1+1,\dots,m_1+m_2\}, \qquad \text{if } m_2 > 0. \end{cases} \tag{4.2}$$

Note that the splitting of $\overline{J} = [0,T]$ 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 $b \le T_1 = \min\{1,T\}$ 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 $\mathcal{G}^m_\sigma = \{I_j\}_{j=1}^m$ of $J = (0,T)$. On $\mathcal{G}^m_\sigma$, we introduce the distribution $\boldsymbol{p} = (p_1,\dots,p_m) \in \mathbb{N}^m$ of polynomial degrees as follows: For a given slope parameter $\mu_{\mathrm{hp}} \in \mathbb{R}$, $\mu_{\mathrm{hp}} \ge 1$, we set

$$p_j := \begin{cases} 1, & j = 1, \\ \lfloor \mu_{\mathrm{hp}} j \rfloor, & j \in \{2,\dots,m_1\}, \\ p_T := \lfloor \mu_{\mathrm{hp}} m_1 \rfloor, & j \in \{m_1+1,\dots,m_1+m_2\}, \qquad \text{if } m_2 > 0, \end{cases} \tag{4.3}$$

where $\lfloor \circ \rfloor$ denotes the floor function. Again, in the case $m_2 = 0$, the last line is omitted. Thus, we set $S^{\boldsymbol{p},1}(J;\mathcal{G}^m_\sigma) := \{v \in C^0(\overline{J}) : v_{|I_j} \in \mathbb{P}^{p_j}\}$, and the temporal subspace V_t^M in (2.26) is defined as

$$S^{\boldsymbol{p},1}_{0,}(J;\mathcal{G}^m_\sigma) := \{v \in S^{\boldsymbol{p},1}(J;\mathcal{G}^m_\sigma) : v(0) = 0\} \subset H^{1/2}_{0,}(J)\,. \tag{4.4}$$

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

$$M = \dim(S_{0,}^{\boldsymbol{p},1}(J;\mathcal{G}_\sigma^m)) = \sum_{j=1}^{m} p_j.$$

We introduce the temporal quasi-interpolant $\Pi_{\mathcal{G}_\sigma^m}^{\boldsymbol{p},1} v$ for a sufficiently smooth function $v\colon [0,T] \to \mathbb{R}$ by

$$\left(\Pi_{\mathcal{G}_\sigma^m}^{\boldsymbol{p},1} v\right)(t) := \begin{cases} v(t_1)t/t_1, & t \in \overline{I_1} \\ v(t_{j-1}) + \int_{t_{j-1}}^{t} (\Pi_{L^2(I_j)}^{p_j-1} v')(\xi)\mathrm{d}\xi, & t \in \overline{I_j},\ j \in \{2,\dots,m\}, \end{cases} \tag{4.5}$$

where $\Pi_{L^2(I_j)}^{p_j-1}$ denotes the $L^2(I_j)$ projection onto $\mathbb{P}^{p_j-1}$. As (4.5) uses point values of the interpolated function, $\Pi_{\mathcal{G}_\sigma^m}^{\boldsymbol{p},1}$ is only defined on a subspace of the continuous functions $C^0(\overline{J})$. Note that the nodal property

$$\forall j \in \{0,\dots,m\} : \quad \left(\Pi_{\mathcal{G}_\sigma^m}^{\boldsymbol{p},1} v\right)(t_j) = v(t_j) \tag{4.6}$$

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 [29, Chapter 3]. 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 $\hat{\Pi}_1^p\colon H^1(\hat{I}) \to \mathbb{P}^p(\hat{I})$ exists such that, for all $v \in H^{r+1}(\hat{I})$ with some $r \in \mathbb{N}$,*

$$\|v' - (\hat{\Pi}_1^p v)'\|_{L^2(\hat{I})}^2 \leq \frac{(p-s)!}{(p+s)!}\|v^{(s+1)}\|_{L^2(\hat{I})}^2 \tag{4.7}$$

and

$$\|v - \hat{\Pi}_1^p v\|_{L^2(\hat{I})}^2 \leq \frac{1}{p(p+1)}\frac{(p-s)!}{(p+s)!}\|v^{(s+1)}\|_{L^2(\hat{I})}^2 \tag{4.8}$$

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

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

We remark that the projectors $\hat{\Pi}_1^p$ for $p \geq 1$ are given by

$$\left(\hat{\Pi}_1^p v\right)(t) := v(-1) + \int_{-1}^{t} \hat{\Pi}_0^{p-1}(v')(\xi)\mathrm{d}\xi\ , \quad t \in \hat{I}\ ,$$

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

For $I_j \in \mathcal{G}_\sigma^m$ with $j \geq 2$, the global quasi-interpolation projectors $\Pi_{\mathcal{G}_\sigma^m}^{\boldsymbol{p},1}$ are obtained by transporting $\hat{\Pi}_1^{p_j}$ from $\hat{I}$ to $I_j \in \mathcal{G}_\sigma^m$ via affine transformations $T_j : \hat{I} \to I_j$, resulting in local projections $\Pi_{1,j}^{p_j}$.

We scale the projection error bounds (4.7) and (4.8) to I_j, and apply them to strongly measurable maps $v\colon I_j \to X$ 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 $I_j \in \mathcal{G}^m_\sigma$ with $j \geq 2$ with time-step size $k_j = |I_j|$, and for every $p \in \mathbb{N}$, there exists a projector $\Pi^p_{1,j} : H^1(I_j; X) \to \mathbb{P}^p(I_j; X)$ such that, for every $v \in H^{r+1}(I_j; X)$ with some $r \in \mathbb{N}$, the error bounds*

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

and

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

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

$$\left(\Pi^p_{1,j} v\right)(t) = v(t) \ \text{ in } X \ \text{ for } t \in \partial I_j = \{t_{j-1}, t_j\}\ .$$

4.2. **$\mathbb{P}^1$-FEM Approximation in** D. We consider the choice of subspaces $V_x^N \subset H^1_{\Gamma_D}(\mathrm{D})$ in (2.26) as standard, conforming $\mathbb{P}^1$-Lagrangian finite elements on simplicial meshes $\mathcal{T}$ of D. We denote by $S^1(\mathrm{D}; \mathcal{T})$ the space of continuous, piecewise linear functions on $\mathcal{T}$, and further, we define the closed subspace

$$S^1_{\Gamma_D}(\mathrm{D}; \mathcal{T}) := S^1(\mathrm{D}; \mathcal{T}) \cap H^1_{\Gamma_D}(\mathrm{D}) \subset H^1_{\Gamma_D}(\mathrm{D}). \tag{4.9}$$

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

$$\forall v \in X_2 : \quad \|v - I^N v\|_{L^2(\mathrm{D})} + N^{-1} \|v - I^N v\|_{H^1(\mathrm{D})} \leq c N^{-2} \|v\|_{H^2(\mathrm{D})}\ . \tag{4.10}$$

With (3.22), for any $f \in L^2(\mathrm{D})$, we also have that the solution $u = A(\partial_x)^{-1} f$ satisfies

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

Remark 4.3. *For transmission problems with diffusion coefficient $a \in W^{1,\infty}(\mathcal{D})$ as in Remark 3.9, assuming that $\mathcal{T}$ is compatible with the partition $\mathcal{D}$ (i.e., the set of nodes of $\mathcal{T}$ includes all interfaces in $\mathcal{D}$), the nodal interpolant $I^N : C^0(\overline{\mathrm{D}}) \to S^1(\mathrm{D}; \mathcal{T})$ satisfies* (4.10) *with $\|v\|_{H^2(\mathcal{D})}$ instead of $\|v\|_{H^2(\mathrm{D})}$ on the right side. The subsequent estimate for $u = A(\partial_x)^{-1} f$, $f \in L^2(\mathrm{D})$, follows from* (3.22) *with $\|v\|_{H^2(\mathcal{D})}$ on the left side (see Remark 3.9).*

4.2.2. **Case** $d = 2$. $\mathrm{D} \subset \mathbb{R}^2$ is a polygon with a finite number of corners and straight sides. We assume furthermore that each entire side Γ_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., [5, 4, 1] and the references there), functions $u \in \mathcal{K}^2_{a+1}(\mathrm{D})$ allow for rate-optimal approximation in $H^1(\mathrm{D})$ and $L^2(\mathrm{D})$ norms in terms of continuous, piecewise linear nodal Lagrangian FEM in D, on regular, simplicial partitions $\mathcal{T}^N_\beta$ (see, e.g., [5, 4, 1] 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 $\mathcal{K}^2_{a+1}(\mathrm{D}) \subset C(\overline{\mathrm{D}})$ (see, e.g., [5]), the nodal interpolation operator I^N_β is well-defined for $u \in \mathcal{K}^2_{a+1}(\mathrm{D})$. Also, for $u \in \mathcal{K}^2_{a+1}(\mathrm{D}) \cap H^1_{\Gamma_D}(\mathrm{D})$, the interpolants $I^N_\beta u$ satisfy exactly the homogeneous Dirichlet boundary conditions on Γ_D. Furthermore, for suitably strong mesh grading as expressed by the parameter β (depending on D, and the corner angles at the vertices of D), the interpolants $I_\beta^N u$ of $u \in \mathcal{K}^2_{a+1}(\mathrm{D})$ converge at optimal rates under mesh refinement: there exists a constant $c > 0$ such that, for all $N = \dim(S^1_{\Gamma_D}(\mathrm{D};\mathcal{T}^N_\beta)) \in \mathbb{N}$,

(4.11)

$$\|u - I_\beta^N u\|_{L^2(\mathrm{D})} + N^{-\frac{1}{2}}\|u - I_\beta^N u\|_{H^1(\mathrm{D})} \le cN^{-1}\|u\|_{\mathcal{K}^2_{a+1}(\mathrm{D})} \le cN^{-1}\|f\|_{L^2(\mathrm{D})}\,.$$

Here, we used (3.24) in the last step.

Remark 4.4. *The interpolation error bound* (4.11) *is based on the graded mesh family* $\{\mathcal{T}^N_\beta\}_{N\ge1}$*. The bound* (4.11) *also holds on families of bisection tree meshes, as shown in* [16, Theorems 5.1, 2.1]*. 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* (2.23)*, piecewise smooth on a finite partition* $\{\mathrm{D}_i\}_{i=1}^{n_{\mathrm{sub}}}$ *of* D *in straight-sided polygons* D_i*, the results in* [22, Theorem 3.7] *imply that with graded meshes in each* D_i *with grading towards multimaterial intersection points, the interpolation error bound* (4.11) *is based on the graded mesh family* $\{\mathcal{T}^N_\beta\}_{N\ge1}$ *still remains true by approximating* u_{reg} *and* w_s *in the decomposition of* [22, Theorem 3.7] *separately.*

4.2.3. **Case** $d = 3$. Only partial extensions of (4.11) to space dimension $d = 3$ are available. We indicate the argument in one particular case. Specifically, we assume (2.23), (3.20) and, in addition, that $A(x) = a(x)\mathbb{I}$, with $a \in W^{1,\infty}(\mathrm{D})$. Furthermore, we assume that $\Gamma_D = \Gamma$, i.e., we consider homogeneous Dirichlet boundary conditions on the entire Γ. 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 (4.11) (with rate $N^{-1/3}$ instead of $N^{-1/2}$) is stated in [6, Theorem 2.1] 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 (2.27) with $V_t^M = S_{0,}^{\boldsymbol{p},1}(J;\mathcal{G}^m_\sigma)$ as defined in (4.4) and with $V_x^N = S^1_{\Gamma_D}(\mathrm{D};\mathcal{T}^N_\beta)$ as given in (4.9), where $\beta = 1$ in the case $d = 1$.

We will require the temporal $H^{1/2}_{0,}(J)$ projector $Q_t^{1/2}$ onto V_t^M and the spatial $H^1_{\Gamma_D}(\mathrm{D})$ "Ritz" projector Q^1_x into V_x^N. Being orthogonal projections, they are stable, i.e., $\|Q_t^{1/2}v\|_{H^{1/2}_{0,}(J)} \le \|v\|_{H^{1/2}_{0,}(J)}$, $\|Q^1_x v\|_{X_1} \le \|v\|_{X_1}$, and optimal in the respective spaces, i.e.,

$$\|v - Q_t^{1/2}v\|_{H^{1/2}_{0,}(J)} = \min_{w\in V_t^M}\|v-w\|_{H^{1/2}_{0,}(J)} \text{ and } \|v - Q^1_x v\|_{X_1} = \min_{w\in V_x^N}\|v-w\|_{X_1}\,.$$

Here, we recall that $X_1 = H^1_{\Gamma_D}(\mathrm{D})$ denotes the "energy" space with norm given by $\|v\|_{X_1} := a(v,v)^{1/2}$. Hence, we may write (for sufficiently regular arguments v)

$$\|v - Q^1_x v\|_{X_1} \le c\|v - I^N_\beta v\|_{H^1(\mathrm{D})} \tag{5.1}$$

with a constant $c > 0$ depending on D and on the coefficient A. Assuming a sufficiently strong corner-mesh refinement in D in the case $d = 2$, an Aubin–Nitsche duality argument, together with (4.10) and (3.22) if $d = 1$, or (4.11) and (3.25) if $d = 2$, implies that there exists a constant $c > 0$ such that, for all $N = \dim(S^1_{\Gamma_D}(\mathrm{D}; \mathcal{T}^N_\beta))$ and all $w \in X_2$ (see, e.g., [5, Thm. 5.2]),

$$\|w - Q^1_x w\|_{L^2(\mathrm{D})} \le c N^{-2/d} \|w\|_{X_2} \,. \tag{5.2}$$

The optimality of the temporal projection $Q^{1/2}_t$ in $H^{1/2}_{0,}(J)$ also implies

$$\|v - Q^{1/2}_t v\|_{H^{1/2}_{0,}(J)} \le \|v - \Pi^{\boldsymbol{p},1}_{\mathcal{G}^m_\sigma} v\|_{H^{1/2}_{0,}(J)} \tag{5.3}$$

for a sufficiently regular $v \colon J \to \mathbb{R}$. Here, $\Pi^{\boldsymbol{p},1}_{\mathcal{G}^m_\sigma}$ is the temporal quasi-interpolant of Subsection 4.1. Proceeding as in the proof of [34, Theorem 3.4], we obtain the following estimate (see [34, p. 175 bottom]).

Lemma 5.1. *Let u and u^{MN} be the solutions to* (2.25) *and* (2.27)*, respectively. We have*

$$\begin{aligned} \|u - u^{MN}\|_{H^{1/2}_{0,}(J;L^2(\mathrm{D}))} &\le \|u - Q^{1/2}_t u\|_{H^{1/2}_{0,}(J;L^2(\mathrm{D}))} \\ &\quad + \|u - Q^1_x u\|_{H^{1/2}_{0,}(J;L^2(\mathrm{D}))} + \left\|(I - Q^{1/2}_t)(I - Q^1_x) u\right\|_{H^{1/2}_{0,}(J;L^2(\mathrm{D}))} \\ &\quad + \|u - Q^1_x u\|_{H^{1/2}_{0,}(J;L^2(\mathrm{D}))} + \|A(\partial_x)(u - Q^{1/2}_t u)\|_{[H^{1/2}_{,0}(J;L^2(\mathrm{D}))]'} \,. \end{aligned} \tag{5.4}$$

We combine (5.1)–(5.4) with the preceding regularity, proven in Section 3, and the approximation properties of the projections $Q^{1/2}_t$, Q^1_x to obtain our main convergence rate bound. For this purpose, we address **Term**1 through **Term**5 in the upper bound (5.4). To this end, we use that the solution u to (2.25) belongs to $H^{1/2}_{0,}(J; X_2)$, which was proven in Lemma 3.6.

We start by deriving upper bounds for **Term**1 and **Term**5. We have $L^2(Q) \simeq [L^2(Q)]' \hookrightarrow [H^{1/2}_{,0}(J; L^2(\mathrm{D}))]'$ and $H^{1/2}_{0,}(J; X_2) \hookrightarrow L^2(J; X_2)$ with continuous and dense injections. This, together with (3.21), gives the following bound for **Term**5:

$$\begin{aligned} \|A(\partial_x)(u - Q^{1/2}_t u)\|_{[H^{1/2}_{,0}(J;L^2(\mathrm{D}))]'} &\le \tilde{c}(T) \|A(\partial_x)(u - Q^{1/2}_t u)\|_{L^2(Q)} \\ &= \tilde{c}(T) \|u - Q^{1/2}_t u\|_{L^2(J;X_2)} \\ &\le c(T) \|u - Q^{1/2}_t u\|_{H^{1/2}_{0,}(J;X_2)} \,. \end{aligned}$$

Using estimate (3.3) yields that **Term**1 can be bounded by

$$\|u - Q^{1/2}_t u\|_{H^{1/2}_{0,}(J;L^2(\mathrm{D}))} \le c \|u - Q^{1/2}_t u\|_{H^{1/2}_{0,}(J;X_2)}$$

with a constant $c > 0$, i.e., for both **Term**1 and **Term**5, we need an estimate of the term $\|u - Q^{1/2}_t u\|_{H^{1/2}_{0,}(J;X_2)}$. For this purpose, we use the temporal quasi-interpolant $\Pi^{\boldsymbol{p},1}_{\mathcal{G}^m_\sigma}$ of Subsection 4.1 and the inequality (5.3). First, note that $\Pi^{\boldsymbol{p},1}_{\mathcal{G}^m_\sigma} u$ is well-defined since $u \colon [0, T] \to X_2$ is continuous, see estimate (3.16) for $l = 0$, $r = 2$, and since $u \colon [0, T] \to X_2$ is smooth for $t > 0$ due to Lemma 3.2. Second, we have $u \in H^{1/2}_{0,}(J; X_2)$ because of Lemma 3.6, hence $u - \Pi^{\boldsymbol{p},1}_{\mathcal{G}^m_\sigma} u \in H^{1/2}_{0,}(J; X_2)$. Thus, it remains to estimate $\|u - \Pi^{\boldsymbol{p},1}_{\mathcal{G}^m_\sigma} u\|_{H^{1/2}_{0,}(J;X_2)}$, 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 α, μ, but independent of m such that*

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

Proof. The proof is based on [11, Lemma 3.4], see Appendix C. □

Lemma 5.3. *Assume (3.9) with some $\varepsilon \in (0,1)$ and some $\delta \geq 1$. Let the grading parameter $\sigma \in (0,1)$ be given. Choose the slope parameter $\mu_{\mathrm{hp}} \geq 1$ such that*

$$\mu_{\mathrm{hp}} > \frac{(1-\sigma)\delta}{2\sigma^{(3+\varepsilon)/2}}, \tag{5.5}$$

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

$$m_2 \begin{cases} = 0, & T \leq 1, \\ > \frac{T-T_1}{4} \cdot \delta\sigma^{-\frac{1+\varepsilon}{2\lfloor \mu_{\mathrm{hp}} \rfloor}}, & T > 1, \end{cases} \tag{5.6}$$

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 (4.1) with time-step sizes k_j in (4.2), and the temporal order distribution $\boldsymbol{p} \in \mathbb{N}^m$ defined by (4.3), lead to the error bound

$$\|u - \Pi_{\mathcal{G}_\sigma^m}^{\boldsymbol{p},1} u\|^2_{H_{0,}^{1/2}((t_2,T);X_2)} \leq C\sigma^{\varepsilon m_1},$$

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

Proof. Set $w = u - \Pi_{\mathcal{G}_\sigma^m}^{\boldsymbol{p},1} u$. Since $w \in H_{0,}^1((t_2,T);X_2)$, see the nodal property (4.6), the interpolation estimate (Lemma A.2) yields

$$\|w\|^2_{H_{0,}^{1/2}((t_2,T);X_2)} \leq \|\partial_t w\|_{L^2((t_2,T);X_2)} \|w\|_{L^2((t_2,T);X_2)}.$$

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 $\lambda = \frac{1-\sigma}{\sigma}$, the time-step size fulfills $k_j = t_j - t_{j-1} = t_{j-1}\lambda$ for $j = 2, \ldots, m_1$. Lemma 4.2 with $p_j = \lfloor \mu_{\mathrm{hp}} j \rfloor$, $s_j = j$ and estimate (3.17) in Proposition 3.5 yield

$$\begin{aligned}
&\|\partial_t w\|^2_{L^2((t_2,T_1);X_2)} = \sum_{j=3}^{m_1} \|\partial_t w\|^2_{L^2(I_j;X_2)} \\
&\leq C \sum_{j=3}^{m_1} \frac{(\lfloor \mu_{\mathrm{hp}} j \rfloor - j)!}{(\lfloor \mu_{\mathrm{hp}} j \rfloor + j)!} \left(\frac{\lambda}{2}\right)^{2j} t_{j-1}^{2j} \delta^{2(j+1)} \Gamma(j+3)^2 t_{j-1}^{-2(j+1)+1+\varepsilon} \\
&\overset{(4.1)}{=} C\delta^2 T_1^{-1+\varepsilon} \sigma^{(m_1+1)(-1+\varepsilon)} \sum_{j=3}^{m_1} \frac{\Gamma(\lfloor \mu_{\mathrm{hp}} j \rfloor - j + 1)}{\Gamma(\lfloor \mu_{\mathrm{hp}} j \rfloor + j + 1)} \left(\frac{\lambda\delta}{2\sigma^{(-1+\varepsilon)/2}}\right)^{2j} \Gamma(j+3)^2 \\
&\leq C_1 \sigma^{m_1(-1+\varepsilon)},
\end{aligned}$$

where, in the last step, Lemma 5.2 is applied for $\mu_{\text{hp}} > \alpha = \frac{(1-\sigma)\delta}{2\sigma^{(3+\varepsilon)/2}} \geq \frac{(1-\sigma)\delta}{2\sigma^{(1+\varepsilon)/2}} = \frac{\lambda\delta}{2\sigma^{(-1+\varepsilon)/2}}$ with (5.5) and the constant $C_1 > 0$ is independent of m_1. In the same way, we get from Lemma 5.2 for $\mu_{\text{hp}} > \alpha = \frac{(1-\sigma)\delta}{2\sigma^{(3+\varepsilon)/2}} = \frac{\lambda\delta}{2\sigma^{(1+\varepsilon)/2}}$ with (5.5) that

$$\begin{aligned}\|w\|^2_{L^2((t_2,T_1);X_2)} &= \sum_{j=3}^{m_1} \|w\|^2_{L^2(I_j;X_2)} \\ &\leq C\sigma^{m_1(1+\varepsilon)} \sum_{j=3}^{m_1} \frac{\Gamma(\lfloor\mu_{\text{hp}}j\rfloor - j + 1)}{\Gamma(\lfloor\mu_{\text{hp}}j\rfloor + j + 1)} \left(\frac{\lambda\delta}{2\sigma^{(1+\varepsilon)/2}}\right)^{2j} \Gamma(j+3)^2 \leq C_2\sigma^{m_1(1+\varepsilon)}\end{aligned}$$

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 $p_j = s_j = p_T := \lfloor\mu_{\text{hp}} m_1\rfloor$ and $k_j = k_T$, estimate (3.13) in Lemma 3.4, and the Stirling's formula $\sqrt{2\pi n}\left(\frac{n}{\mathrm{e}}\right)^n < n! < \sqrt{2\pi n}\left(\frac{n}{\mathrm{e}}\right)^n \mathrm{e}^{\frac{1}{12n}}$ with $\mathrm{e}^{\frac{1}{6n}} < 2$, we get

$$\begin{aligned}\|\partial_t w\|^2_{L^2((1,T);X_2)} &= \sum_{j=m_1+1}^{m} \|\partial_t w\|^2_{L^2(I_j;X_2)} \\ &\leq C\frac{1}{(2p_T)!}\left(\frac{k_T}{2}\right)^{2p_T} \underbrace{\sum_{j=m_1+1}^{m} \|\partial_t^{(p_T+1)} u\|^2_{L^2(I_j;X_2)}}_{=\|\partial_t^{(p_T+1)} u\|^2_{L^2((1,T);X_2)}} \\ &\leq C\delta^2 \frac{1}{(2p_T)!}\left(\frac{k_T\delta}{2}\right)^{2p_T} \Gamma(p_T+3)^2 C(\varepsilon,1,T)^2 \\ &\leq C\delta^2 C(\varepsilon,1,T)^2 (p_T+2)^4 \frac{4\pi p_T\left(\frac{p_T}{\mathrm{e}}\right)^{2p_T}}{\sqrt{2\pi}\sqrt{2p_T}\left(\frac{2p_T}{\mathrm{e}}\right)^{2p_T}} \left(\frac{k_T\delta}{2}\right)^{2p_T} \\ &= C2\sqrt{\pi}\delta^2 C(\varepsilon,1,T)^2 (p_T+2)^4 \sqrt{p_T}\left(\frac{k_T\delta}{4}\right)^{2p_T} \leq C_3\sigma^{m_1(1+\varepsilon)}\end{aligned}$$

with a constant $C_3 > 0$ independent of m_1. In the last step, due to (5.6), we use that a constant $q \in (0,1)$ exists such that

$$k_T = \frac{T - T_1}{m_2} = q\frac{4}{\delta}\sigma^{\frac{m_1(1+\varepsilon)}{2\lfloor\mu_{\text{hp}}\rfloor m_1}} \leq q\frac{4}{\delta}\sigma^{\frac{m_1(1+\varepsilon)}{2\lfloor\mu_{\text{hp}} m_1\rfloor}} = q\frac{4}{\delta}\sigma^{\frac{m_1(1+\varepsilon)}{2p_T}}$$

and therefore, $(p_T+2)^4\sqrt{p_T}q^{2p_T} \to 0$ as $p_T \to \infty$. Analogously, we obtain

$$\|w\|^2_{L^2((1,T);X_2)} \leq C_4\sigma^{m_1(1+\varepsilon)}$$

with a constant $C_4 > 0$ independent of m_1.

With all estimates above, we conclude that

$$\begin{aligned}\|w\|^2_{H^{1/2}_{0,}((t_2,T);X_2)} &\leq \|\partial_t w\|_{L^2((t_2,T);X_2)}\|w\|_{L^2((t_2,T);X_2)} \\ &\leq \sqrt{C_1\sigma^{m_1(-1+\varepsilon)} + C_3\sigma^{m_1(1+\varepsilon)}}\sqrt{C_2\sigma^{m_1(1+\varepsilon)} + C_4\sigma^{m_1(1+\varepsilon)}} \leq C_{\text{est}}\sigma^{\varepsilon m_1},\end{aligned}$$

where $C_{\text{est}} > 0$ is independent of m_1. □

Lemma 5.4. *Under the assumptions of Lemma 5.3, the estimate*

$$\|u - \Pi^{\boldsymbol{p},1}_{\mathcal{G}^m_\sigma} u\|_{H^{1/2}_{0,}(J;X_2)} \leq C\exp(-b\sqrt{M})$$

holds true with a constant C independent of b and M, where $b = -\varepsilon \ln \sigma / \sqrt{8\mu_{\mathrm{hp}}} > 0$ and $M = \dim(S_{0,}^{\boldsymbol{p},1}(J;\mathcal{G}_\sigma^m)) \le 2\mu_{\mathrm{hp}} m_1^2 \le 2\mu_{\mathrm{hp}} m^2$.

Proof. Set $w = u - \Pi_{\mathcal{G}_\sigma^m}^{\boldsymbol{p},1} u$. Then, for X_2-valued functions, the norm equivalence in Lemma 2.1 and the localization in Lemma 2.2 for $a=0$, $b=T$, $\tau=t_2$ yield

$$
\begin{aligned}
(C_{\mathrm{Int},1})^2 \|w\|^2_{H_{0,}^{1/2}(J;X_2)} &\le |||w|||^2_{H_{0,}^{1/2}(J;X_2)} \\
&\le \|w\|^2_{L^2(J;X_2)} + |w|^2_{H^{1/2}((0,t_2);X_2)} + 4\int_0^{t_2} \frac{\|w(t)\|^2_{X_2}}{t_2 - t}\mathrm{d}t \\
&\quad + 4\int_{t_2}^T \frac{\|w(s)\|^2_{X_2}}{s - t_2}\mathrm{d}s + |w|^2_{H^{1/2}((t_2,T);X_2)} + \int_0^T \frac{\|w(t)\|^2_{X_2}}{t}\mathrm{d}t \\
&\le |||w|||^2_{H_{0,}^{1/2}((0,t_2);X_2)} + 4\int_0^{t_2} \frac{\|w(t)\|^2_{X_2}}{t_2 - t}\mathrm{d}t + 5|||w|||^2_{H_{0,}^{1/2}((t_2,T);X_2)},
\end{aligned}
$$

where we used the definition (2.7) of the triple norm and the bound $\int_{t_2}^T \frac{\|w(t)\|^2_{X_2}}{t}\mathrm{d}t \le \int_{t_2}^T \frac{\|w(s)\|^2_{X_2}}{s-t_2}\mathrm{d}s$. Next, we estimate the three terms on the right side.

First term: The triangle inequality, Lemma 2.1, Lemma 3.6, the Poincaré inequality (Lemma A.1), definition (4.5), and estimates (3.16), (3.17) yield

$$
\begin{aligned}
&|||w|||^2_{H_{0,}^{1/2}((0,t_2);X_2)} \le 2|||u|||^2_{H_{0,}^{1/2}((0,t_2);X_2)} + 2\left|\left|\left|\Pi_{\mathcal{G}_\sigma^m}^{\boldsymbol{p},1} u\right|\right|\right|^2_{H_{0,}^{1/2}((0,t_2);X_2)} \\
&\le C t_2^\varepsilon + 2(C_{\mathrm{Int},2})^2 \sqrt{1 + \frac{4t_2^2}{\pi^2}} \underbrace{\|\Pi_{\mathcal{G}_\sigma^m}^{\boldsymbol{p},1} u\|^2_{H_{0,}^{1/2}((0,t_2);X_2)}}_{\le \frac{2}{\pi}(t_2 - 0)\|\partial_t \Pi_{\mathcal{G}_\sigma^m}^{\boldsymbol{p},1} u\|^2_{L^2((0,t_2);X_2)}} \\
&\le C T_1^\varepsilon \sigma^{-2\varepsilon} \sigma^{\varepsilon m_1} + C t_2 \Big[\underbrace{\int_0^{t_1} \frac{\|u(t_1)\|^2_{X_2}}{t_1^2}\mathrm{d}t}_{\le C t_1^{-1+\varepsilon}} + \underbrace{\int_{t_1}^{t_2} \|\Pi_{L^2(I_2)}^{p_2-1} \partial_t u(t)\|^2_{X_2}\mathrm{d}t}_{\le \|\partial_t u\|^2_{L^2(I_2;X_2)} \le C\delta^2 t_1^{-1+\varepsilon}} \Big] \le C_1 \sigma^{\varepsilon m_1},
\end{aligned}
$$

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

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

Second term: With the bound (3.16), the nodal property (4.6) and $k_1 = t_1 = T_1 \sigma^{m_1 - 1}$, $k_2 = T_1 \sigma^{m_1 - 2}(1 - \sigma)$, we find

$$
\begin{aligned}
4\int_0^{t_2} \frac{\|w(t)\|^2_{X_2}}{t_2 - t}\mathrm{d}t &= 4\int_0^{t_1} \frac{\|w(t)\|^2_{X_2}}{t_2 - t}\mathrm{d}t + 4\int_{t_1}^{t_2} \frac{\|w(t)\|^2_{X_2}}{t_2 - t}\mathrm{d}t \\
&= 4\int_0^{t_1} \frac{\|u(t) - u(t_1)t/t_1\|^2_{X_2}}{t_2 - t}\mathrm{d}t + 4\int_{t_1}^{t_2} \frac{\|\int_t^{t_2} \partial_t w(\xi)\mathrm{d}\xi\|^2_{X_2}}{t_2 - t}\mathrm{d}t \\
&\le \frac{8}{k_2}\int_0^{t_1} \|u(t)\|^2_{X_2}\mathrm{d}t + \frac{8}{k_2}\int_0^{t_1} \|u(t_1)\|^2_{X_2} \frac{t^2}{k_1^2}\mathrm{d}t + 4\int_{t_1}^{t_2} \frac{\left[\int_t^{t_2} \|\partial_t w(\xi)\|_{X_2}\mathrm{d}\xi\right]^2}{t_2 - t}\mathrm{d}t \\
&\le \frac{C}{k_2}\int_0^{t_1} t^\varepsilon \mathrm{d}t + \frac{C t_1^\varepsilon}{k_1^2 k_2}\int_0^{t_1} t^2 \mathrm{d}t + 4\int_{t_1}^{t_2} \|\partial_t w\|^2_{L^2((t,t_2);X_2)}\mathrm{d}t \\
&\le 2C \frac{T_1^\varepsilon \sigma^{1-\varepsilon}}{1 - \sigma} \sigma^{\varepsilon m_1} + 4k_2 \|\partial_t w\|^2_{L^2(I_2;X_2)} \le C_2 \sigma^{\varepsilon m_1},
\end{aligned}
$$

with a constant $C_2 > 0$ independent of m_1, where in the last step we have used the estimate (3.17). This yields

$$4k_2\|\partial_t u\|^2_{L^2(I_2;X_2)} \leq CT_1\sigma^{m_1-2}(1-\sigma)t_1^{-1+\varepsilon} = CT_1^{\varepsilon}\frac{1-\sigma}{\sigma^{1+\varepsilon}}\sigma^{\varepsilon m_1}.$$

Third term: Lemma 2.1 and Lemma 5.3 give

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

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

Conclusion of the proof: As the temporal number of degrees of freedom M fulfills

$$M \leq \sum_{j=1}^{m_1}\lfloor\mu_{\mathrm{hp}}j\rfloor + \lfloor\mu_{\mathrm{hp}}m_1\rfloor m_2 \leq \mu_{\mathrm{hp}}\frac{m_1(m_1+1)}{2} + \mu_{\mathrm{hp}}m_1^2 \leq 2\mu_{\mathrm{hp}}m_1^2 \tag{5.7}$$

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

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

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

As Lemma 5.4 implies exponential convergence bounds on **Term**1 and **Term**5, it remains to treat **Terms**2–4 in (5.4). **Term**2 and **Term**4 are identical. We focus on **Term**3. Using that $Q_t^{1/2}$ is a projector in the Hilbert space $H^{1/2}_{0,}(J)$, the triangle inequality gives

$$\left\|(I-Q_t^{1/2})(I-Q_x^1)u\right\|_{H^{1/2}_{0,}(J;L^2(\mathrm{D}))} \leq 2\left\|u-Q_x^1u\right\|_{H^{1/2}_{0,}(J;L^2(\mathrm{D}))}.$$

Thus, **Term**3 can be estimated in the same way as **Term**2 and **Term**4. Using

$$H^{1/2}_{0,}(0,T;L^2(\mathrm{D})) \simeq H^{1/2}_{0,}(J)\otimes L^2(\mathrm{D}) \simeq L^2(\mathrm{D})\otimes H^{1/2}_{0,}(J) \simeq L^2(\mathrm{D};H^{1/2}_{0,}(J)),$$

we may use the $L^2(\mathrm{D})$ error bound (5.2) 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

$$\left\|u-Q_x^1u\right\|_{H^{1/2}_{0,}(J;L^2(\mathrm{D}))} \leq cN^{-2/d}\left\|u\right\|_{H^{1/2}_{0,}(J;X_2)} \leq CN^{-2/d}, \tag{5.8}$$

with a constants $c > 0$, $C > 0$ independent of N.

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

Theorem 5.5. *Let the space dimension d be either $d = 1$ or $d = 2$. Assume that the diffusion coefficient $A \in W^{1,\infty}(\mathrm{D};\mathbb{R}^{d\times d}_{\mathrm{sym}})$ is uniformly positive definite, i.e., that* (2.23) *is satisfied, and that the forcing g in* (1.1) *satisfies the temporal analytic regularity* (3.9)*. Furthermore, assume that the assumptions of Lemma 5.3 on the temporal mesh $\mathcal{G}^m_\sigma$ in* (4.1) *with $\mu_{\mathrm{hp}} \geq 1$ and $m_2 \in \mathbb{N}_0$ fulfilling* (5.5) *and* (5.6)*, respectively, and the temporal order distribution $\boldsymbol{p} \in \mathbb{N}^m$ in* (4.3) *are satisfied.*

Then the space-time Galerkin approximation (2.27) *admits a unique solution $u^{MN} \in S^{\boldsymbol{p},1}_{0,}(J;\mathcal{G}^m_\sigma)\otimes S^1_{\Gamma_D}(\mathrm{D};\mathcal{T}^N_\beta)$ with the temporal hp-FE space $S^{\boldsymbol{p},1}_{0,}(J;\mathcal{G}^m_\sigma)$ of dimension $M = \dim(S^{\boldsymbol{p},1}_{0,}(J;\mathcal{G}^m_\sigma))$ as defined in* (4.4)*, and with the spatial FE space $S^1_{\Gamma_D}(\mathrm{D};\mathcal{T}^N_\beta)$ of continuous, piecewise linear FEM on a sequence of suitably graded, regular triangulations $\{\mathcal{T}^N_\beta\}_N$ in* D *($\beta = 1$, i.e., quasi-uniform partitions, if $d = 1$) of dimension $N = \dim(S^1_{\Gamma_D}(\mathrm{D};\mathcal{T}^N_\beta))$.*

Moreover, a constant $C > 0$ (independent of M and N) exists such that the space-time discretization (2.27) *based on these spaces satisfies the error bound*

$$\|u - u^{MN}\|_{H^{1/2}_{0,}(J;L^2(\mathrm{D}))} \leq C\left(\exp(-b\sqrt{M}) + N^{-2/d}\right) \tag{5.9}$$

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

Proof. Existence and uniqueness of the solution u^{MN} were established at the end of Section 2.5. Estimate (5.9) follows from Lemma 5.1, taking into account Lemma 5.4, and estimate (5.8). □

Balancing the terms in the upper bound (5.9) results in

$$M \simeq O\left((\log N)^2\right) \qquad \text{or} \qquad m_1 \simeq O(\log N),$$

where $M \leq 2\mu_{\mathrm{hp}} m_1^2 \leq 2\mu_{\mathrm{hp}} m^2$, see (5.7). Then, the *number of degrees of freedom for the space-time discretization* behaves, as $N \to \infty$, as

$$MN \simeq O\left(N(\log N)^2\right), \tag{5.10}$$

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 5.6. *Theorem 5.5 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.,* [10, Chapter 12]*) imply that for any constant number of temporal elements $m \in \mathbb{N}$ (e.g., $m = 1$) with temporal polynomial degrees $\boldsymbol{p} = (p, \ldots, p) \in \mathbb{N}^m$ with $p \in \mathbb{N}$, temporal exponential convergence follows when $p \to \infty$ (p-method). Under the otherwise exact same assumptions as in Theorem 5.5, one obtains in place of* (5.9) *the error bound*

$$\|u - u^{MN}\|_{H^{1/2}_{0,}(J;L^2(\mathrm{D}))} \leq C\left(\exp(-bp) + N^{-2/d}\right) \tag{5.11}$$

with $M = mp$ and constants $b > 0$, $C > 0$ independent of p and N. This allows to improve (5.10) *to*

$$MN \simeq O(N \log N). \tag{5.12}$$

6. Numerical Experiments

In this section, we present numerical examples for the space-time Galerkin approximation (2.27) of the heat equation with homogeneous Dirichlet conditions

$$\partial_t u - \Delta_x u = g \quad \text{in } Q, \quad u|_{t=0} = 0, \quad \gamma_0(u) = 0 \quad \text{on } \partial\mathrm{D}, \tag{6.1}$$

i.e., $A(\partial_x) = -\Delta_x$ in (1.1), $u_0 = 0$ in (1.2) and $u_D = 0$ with $\Gamma_D = \Gamma = \partial\mathrm{D}$ in (1.3). We use globally continuous functions, which are piecewise linear in space and piecewise polynomials of higher-order in time, see Theorem 5.5. We start by deriving the algebraic linear system associated with (2.27), and by describing the realization of the operator $\mathcal{H}_T$ for a temporal hp-FEM.

For (6.1), we solve the discrete space-time variational formulation to find $u^{MN} \in S^{\boldsymbol{p},1}_{0,}(J; \mathcal{G}^m_\sigma) \otimes S^1_{\Gamma_D}(\mathrm{D}; \mathcal{T}^N_\beta)$ such that

$$\langle \partial_t u^{MN}, v \rangle_{L^2(Q)} + \langle \nabla_x u^{MN}, \nabla_x v \rangle_{L^2(Q)} = \langle \Pi^{MN} g, v \rangle_{L^2(Q)} \tag{6.2}$$

is satisfied for all $v \in (\mathcal{H}_T S_{0,}^{\boldsymbol{p},1}(J;\mathcal{G}_\sigma^m)) \otimes S_{\Gamma_D}^1(\mathrm{D};\mathcal{T}_\beta^N)$. Here, we use the notation of Section 4 with

$$S_{0,}^{\boldsymbol{p},1}(J;\mathcal{G}_\sigma^m) =: V_t^M := \operatorname{span}\{\varphi_l\}_{l=1}^M, \tag{6.3}$$

and

$$S_{\Gamma_D}^1(\mathrm{D};\mathcal{T}_\beta^N) =: V_x^N := \operatorname{span}\{\psi_i\}_{i=1}^N,$$

where the functions φ_l are basis functions in time, and the functions ψ_i are the usual nodal basis functions in space. The total number of degrees of freedom is

$$MN = \dim\left(S_{0,}^{\boldsymbol{p},1}(J;\mathcal{G}_\sigma^m) \otimes S_{\Gamma_D}^1(\mathrm{D};\mathcal{T}_\beta^N)\right).$$

In addition, for an easier implementation, we approximate the right-hand side $g \in L^2(Q)$ by $g \approx \Pi^{MN} g$, where $\Pi^{MN}\colon L^2(Q) \to S^{\boldsymbol{p},1}(J;\mathcal{G}_\sigma^m) \otimes S^1(\mathrm{D};\mathcal{T}_\beta^N)$ is the space-time $L^2(Q)$ projection, namely $\Pi^{MN} g \in S^{\boldsymbol{p},1}(J;\mathcal{G}_\sigma^m) \otimes S^1(\mathrm{D};\mathcal{T}_\beta^N)$ is such that

$$\forall w \in S^{\boldsymbol{p},1}(J;\mathcal{G}_\sigma^m) \otimes S^1(\mathrm{D};\mathcal{T}_\beta^N):\ \langle \Pi^{MN} g, w\rangle_{L^2(Q)} = \langle g, w\rangle_{L^2(Q)}. \tag{6.4}$$

Note that the spaces $S^1(\mathrm{D};\mathcal{T}_\beta^N)$ and $S^{\boldsymbol{p},1}(J;\mathcal{G}_\sigma^m)$ 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 $k_{\max} = \max_{j=1,\dots,m} k_j$, the spatial mesh width by h_x, and the space-time mesh width by $h_{xt} = \max\{k_{\max}, h_x\}$. The space-time error $\|u - u^{MN}\|_{H_{0,}^{1/2}(J;L^2(\mathrm{D}))}$ mandates the numerical evaluation of the fractional order norm $\|\circ\|_{H_{0,}^{1/2}(J;L^2(\mathrm{D}))}$. In order to overcome this problem, we introduce the quantity

$$[v]_{H_{0,}^{1/2}(J;L^2(\mathrm{D}))} := \sqrt{\|v\|_{L^2(Q)} \cdot \|\partial_t v\|_{L^2(Q)}},$$

which is defined for $v \in H_{0,}^1(J;L^2(\mathrm{D}))$, and observe that, provided that $u \in H_{0,}^1(J;L^2(\mathrm{D}))$,

$$\|u - u^{MN}\|_{H_{0,}^{1/2}(J;L^2(\mathrm{D}))} \le [u - u^{MN}]_{H_{0,}^{1/2}(J;L^2(\mathrm{D}))},$$

due to the interpolation estimate (Lemma A.2). Therefore, in the experiments below, instead of the space-time error $\|u - u^{MN}\|_{H_{0,}^{1/2}(J;L^2(\mathrm{D}))}$, we consider its upper bound $[u - u^{MN}]_{H_{0,}^{1/2}(J;L^2(\mathrm{D}))}$, which can be numerically evaluated via *local* integration.

The fully discrete, space-time variational formulation (6.2) is equivalent to the global linear system

$$B^{MN}\boldsymbol{u} = \boldsymbol{G}, \tag{6.5}$$

with the system matrix

$$B^{MN} = A_t^{\mathcal{H}_T} \otimes M_x + M_t^{\mathcal{H}_T} \otimes A_x \in \mathbb{R}^{M\cdot N \times M\cdot N},$$

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

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

for $i,j = 1,\dots,N$, and $M_t^{\mathcal{H}_T} \in \mathbb{R}^{M\times M}$ and $A_t^{\mathcal{H}_T} \in \mathbb{R}^{M\times M}$ are defined by

$$M_t^{\mathcal{H}_T}[k,l] := \langle \varphi_l, \mathcal{H}_T \varphi_k\rangle_{L^2(J)}, \quad A_t^{\mathcal{H}_T}[k,l] := \langle \partial_t \varphi_l, \mathcal{H}_T \varphi_k\rangle_{L^2(J)} \tag{6.6}$$

for $k, l = 1, \ldots, M$. Note that, due to the nonlocality of $\mathcal{H}_T$, the matrices $M_t^{\mathcal{H}_T}$ and $A_t^{\mathcal{H}_T}$ are densely populated. Furthermore, the temporal stiffness matrix $A_t^{\mathcal{H}_T}$ is symmetric (due to (2.11)) and positive definite (due to (2.10)), whereas $M_t^{\mathcal{H}_T}$ is nonsymmetric and positive definite (due to (2.15)). The assembling of the matrices $M_t^{\mathcal{H}_T}$ and $A_t^{\mathcal{H}_T}$ is described in Subsection 6.1 below. For the right-hand side $\boldsymbol{G}$, the integrals for computing the projection $\Pi^{MN} g$ in (6.4) are calculated by using high-order quadrature rules. The global linear system (6.5) is solved in MATLAB by using the Bartels-Stewart method with real-Schur decomposition, see [21, Algorithm 4.1]. 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 $M_t^{\mathcal{H}_T}$ and $A_t^{\mathcal{H}_T}$ in (6.6). The crucial point is the realization of the modified Hilbert transformation $\mathcal{H}_T$, for which different possibilities exist, see [35, 39]. In particular, for a uniform degree vector $\boldsymbol{p} := (p, p, \ldots, p)$ with a fixed, low polynomial degree $p \in \mathbb{N}$, e.g., $p = 1$ or $p = 2$, the matrices $M_t^{\mathcal{H}_T}$ and $A_t^{\mathcal{H}_T}$ in (6.6) can be calculated using a series expansion based on the *Legendre chi function*, which converges very fast, independently of the temporal mesh widths; see [39, Subsection 2.2]. As for the temporal *hp*-FEM the degree vector $\boldsymbol{p}$ is not uniform, it is convenient to apply numerical quadrature rules to numerically approximate the matrix entries.

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

$$(\mathcal{H}_T v)(t) = -\frac{2}{\pi} v(0) \ln \tan \frac{\pi t}{4T} - \frac{1}{\pi} \int_0^T \ln \left[\tan \frac{\pi (s+t)}{4T} \tan \frac{\pi \, |t-s|}{4T} \right] \partial_t v(s) \mathrm{d}s,$$

$t \in J$, $v \in H^1(J)$, as a weakly singular integral, see [35, Lemma 2.1], we have

$$\begin{aligned} A_t^{\mathcal{H}_T}[k,l] &= \langle \partial_t \varphi_l, \mathcal{H}_T \varphi_k \rangle_{L^2(J)} \\ &= -\frac{1}{\pi} \int_0^T \partial_t \varphi_l(t) \int_0^T \ln \left[\tan \frac{\pi (s+t)}{4T} \tan \frac{\pi \, |t-s|}{4T} \right] \partial_t \varphi_k(s) \, \mathrm{d}s \, \mathrm{d}t \end{aligned} \tag{6.7}$$

and

$$\begin{aligned} M_t^{\mathcal{H}_T}[k,l] &= \langle \varphi_l, \mathcal{H}_T \varphi_k \rangle_{L^2(J)} \\ &= -\frac{1}{\pi} \int_0^T \varphi_l(t) \int_0^T \ln \left[\tan \frac{\pi (s+t)}{4T} \tan \frac{\pi \, |t-s|}{4T} \right] \partial_t \varphi_k(s) \, \mathrm{d}s \, \mathrm{d}t \end{aligned} \tag{6.8}$$

for $k, l = 1, \ldots, M$, with the temporal basis functions φ_l in (6.3). In the following, we only describe how to compute the matrix entries $M_t^{\mathcal{H}_T}[k,l]$ in (6.8), since the matrix entries $A_t^{\mathcal{H}_T}[k,l]$ in (6.7) can be computed in the same way.

The matrix entries $M_t^{\mathcal{H}_T}[k,l]$ in (6.8) are computed element-wise for the partition $\mathcal{G}_\sigma^m = \{I_j\}_{j=1}^m$ 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 $i, j \in \{1, \ldots, m\}$ and related local polynomial degrees $p_i, p_j \in \mathbb{N}$. We define the local matrix $M_t^{\mathcal{H}_T, i, j} \in$

$\mathbb{R}^{(p_i+1)\times(p_j+1)}$ by

$$M_t^{\mathcal{H}_T,i,j}[\kappa,\ell] = -\frac{1}{\pi}\int_{t_{j-1}}^{t_j} \varphi_{\alpha(\ell,j)}(t)\int_{t_{i-1}}^{t_i} \ln\left[\tan\frac{\pi(s+t)}{4T}\tan\frac{\pi\,|t-s|}{4T}\right]\partial_t\varphi_{\alpha(\kappa,i)}(s)\,\mathrm{d}s\,\mathrm{d}t \tag{6.9}$$

for $\kappa = 1,\dots,p_i+1$ and $\ell = 1,\dots,p_j+1$. Here, $\alpha(\kappa,i) \in \{0,1,\dots,M\}$ is the global index related to the local index κ for the time interval I_i; similarly for $\alpha(\ell,j)$. Notice that the function φ_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

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

$\xi\in[-1,1]$, where L_ℓ denotes the ℓ-th Legendre polynomial on $[-1,1]$, see [29, Chapter 3]. With these shape functions and the affine transformation $T_\iota\colon [-1,1]\to[t_{\iota-1},t_\iota]$ for $\iota\in\{1,\dots,m\}$, the entries (6.9) of the local matrix $M_t^{\mathcal{H}_T,i,j}$ are

$$M_t^{\mathcal{H}_T,i,j}[\kappa,\ell] = -\frac{k_j}{2\pi}\int_{-1}^{1} N_\ell(\eta)\int_{-1}^{1}\ln\left[\tan\frac{\pi(T_i(\xi)+T_j(\eta))}{4T}\tan\frac{\pi\,|T_j(\eta)-T_i(\xi)|}{4T}\right]N_\kappa'(\xi)\,\mathrm{d}\xi\,\mathrm{d}\eta \tag{6.10}$$

for $\kappa = 1,\dots,p_i+1$ and $\ell=1,\dots,p_j+1$, where $k_j = |t_j - t_{j-1}|$ is the length of the time interval I_j. To compute the integrals in (6.10), we split these integrals into regular and singular parts, see [35, Subsection 3.1]. For the regular parts, a tensor Gauss quadrature is applied. In [35, Subsection 3.1], 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 [17, Eq. (1.6), (1.7)], 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 (6.6) 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 $\mathrm{D}=(0,1)\subset\mathbb{R}$ with final time $T=2$, i.e., $Q = J\times\mathrm{D} = (0,2)\times(0,1)\subset\mathbb{R}^2$. We choose the constant right-hand side $g_1\equiv 1$, for which the solution to problem (6.1) is given by the Fourier series

$$u_1(t,x) = \sum_{\eta=1}^{\infty}\frac{4-4e^{-\pi^2(2\eta-1)^2 t}}{\pi^3(2\eta-1)^3}\sin(\pi(2\eta-1)x),\quad (t,x)\in\overline{Q}. \tag{6.11}$$

In the calculation of the errors of the space-time Galerkin approximation (6.2), we truncate the series (6.11) 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 $k_{\max} = k_1 = \dots = k_m$ and linear polynomials, i.e., $\boldsymbol{p} = (1,\dots,1)\in\mathbb{N}^m$. 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 (3.9) for any $\varepsilon \in (0, 1/2)$ with $\delta = 1$ and a constant $C = C(\varepsilon)$ depending on ε. In the second test, we use the temporal hp-

MN	h_x	$k_{\max}$	$[u_1 - u_1^{MN}]_{H_{0,}^{1/2}(J;L^2(\mathrm{D}))}$	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

TABLE 1. Numerical results with the space-time Galerkin approximation (6.2) for the 1D example with the right-hand side $g_1 \equiv 1$ and solution u_1 in (6.11), for a uniform mesh refinement strategy and piecewise linear polynomials both in space and time.

approximation of Subsection 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 (4.1) with grading parameter $\sigma = 0.31$, slope parameter $\mu_{\mathrm{hp}} = 2.0$, numbers of elements $m_1 = \lfloor 1.4 \cdot \ln N \rfloor$, $m_2 = 1$, and temporal polynomial degrees $\boldsymbol{p} \in \mathbb{N}^m$ as in (4.3). This choice of the discretization parameters fulfills condition (5.5) with $\mu_{\mathrm{hp}} = 2.0 > \frac{345}{31\sqrt{31}} \approx 1.99883$ and condition (5.6) with $m_2 = 1 > \frac{5}{2\sqrt{31}} \approx 0.45$. In addition, this choice balances the terms of the error bound (5.9), i.e., the total number of degrees of freedom MN behaves like in (5.10). The numerical results reported in Figure 1 confirm Theorem 5.5.

6.3. **Numerical Examples in 2D.** We present numerical examples in the two-dimensional spatial L-shaped domain

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

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

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 $L^2(\mathrm{D})$ for $\mathbb{P}^1$-FEM approximations in D. For a given maximal mesh width $h_x > 0$, we construct spatial meshes $\mathcal{T}_\beta^N$ with corner-refinements towards the origin fulfilling the grading

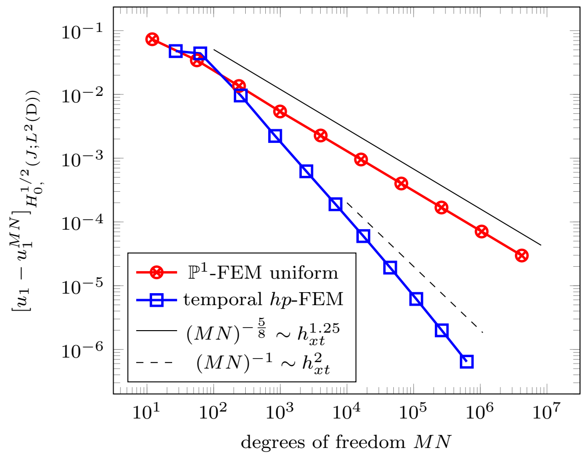

FIGURE 1. Numerical results with the space-time Galerkin approximation (6.2) for the 1D example with the right-hand side $g_1 \equiv 1$ and solution u_1 in (6.11), for a spatial uniform mesh refinement and temporal $\mathbb{P}^1$-FEM approximations with uniform mesh refinement or with temporal hp-FEM with geometric partition of J with grading parameter $\sigma = 0.31$, slope parameter $\mu_{\mathrm{hp}} = 2.0$, numbers of elements $m_1 = \lfloor 1.4 \cdot \ln N \rfloor$, $m_2 = 1$, and temporal polynomial degrees $\boldsymbol{p} \in \mathbb{N}^m$ as in (4.3).

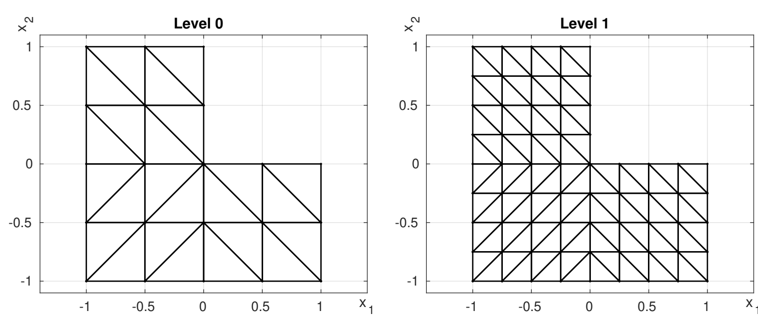

FIGURE 2. Spatial meshes with uniform refinement strategy: starting mesh and mesh after one refinement step.

condition

$$\forall \omega \in \mathcal{T}_\beta^N: \quad h_{x,\omega} \sim \begin{cases} h_x^{1/\beta}, & \mathrm{dist}(\omega, \mathbf{0}) = 0, \\ h_x \cdot \mathrm{dist}(\omega, \mathbf{0})^{1-\beta}, & 0 < \mathrm{dist}(\omega, \mathbf{0}) \le R, \\ h_x, & \mathrm{dist}(\omega, \mathbf{0}) > R, \end{cases} \tag{6.12}$$

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 $\omega \in \mathcal{T}_\beta^N$, and $\mathrm{dist}(\omega, \mathbf{0})$ is the distance of the triangle $\omega \in \mathcal{T}_\beta^N$ 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

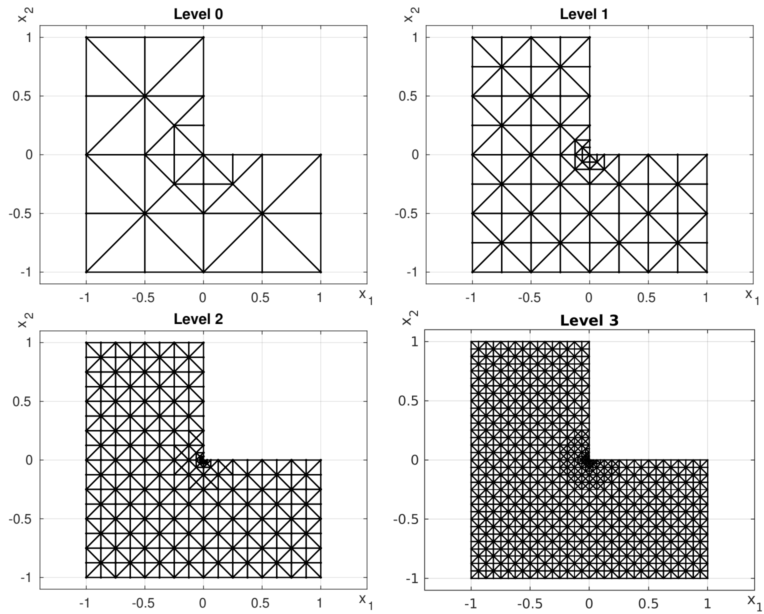

FIGURE 3. Spatial meshes with corner-refinements towards the origin fulfilling the grading condition (6.12) with parameters $\beta = 0.6$ and $R = 0.25$.

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

$$u_2(t, x_1, x_2) = u_{\mathrm{reg}}(t, x_1, x_2) + t\mathrm{e}^{-t}\eta(x_1, x_2)\cdot r(x_1, x_2)^{2/3}\cdot \sin\left(\frac{2}{3}\left(\arg(x_1, x_2) - \frac{\pi}{2}\right)\right) \tag{6.13}$$

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

$$u_{\mathrm{reg}}(t, x_1, x_2) = \frac{1}{100} t \sin(\pi x_1)\sin(\pi x_2)\mathrm{e}^{-t\left(x_1 - \frac{1}{4}\right)^2 - t\left(x_2 + \frac{1}{4}\right)^2}, \quad (t, x_1, x_2) \in \overline{Q}, \tag{6.14}$$

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

$$\eta(x_1, x_2) := \begin{cases} 1, & r(x_1, x_2) \leq 1/4, \\ \frac{27}{8} - \frac{135}{4} r(x_1, x_2) + 180 r(x_1, x_2)^2 & \\ \quad -440 r(x_1, x_2)^3 + 480 r(x_1, x_2)^4 & \\ \quad -192 r(x_1, x_2)^5, & 1/4 < r(x_1, x_2) \leq 3/4, \\ 0, & 3/4 < r(x_1, x_2). \end{cases} \tag{6.15}$$

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

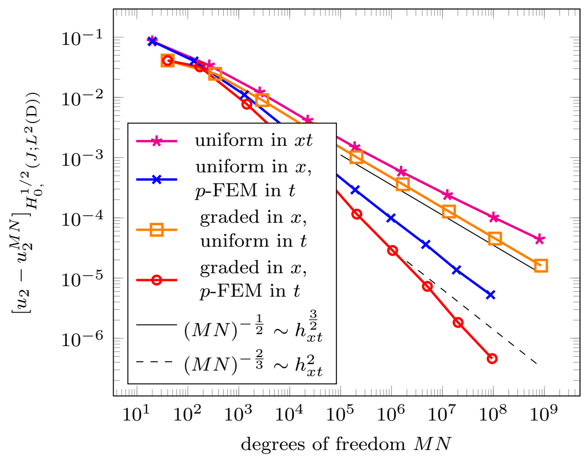

FIGURE 4. Numerical results with the space-time Galerkin approximation (6.2) for the 2D example with the singular-in-space solution u_2 in (6.13), for all combinations of uniform mesh refinement or graded meshes in space (with grading parameter $\beta = 0.6$), and $\mathbb{P}^1$-FEM with uniform mesh refinement or p-FEM in time. For the p-FEM in time, for a spatial discretization of parameter N, we use a fixed mesh with $m = 4$ elements and polynomial degrees $\boldsymbol{p} = (p,p,p,p)$ with $p = \lfloor \frac{\ln N}{2} \rfloor$.

reported here, we have verified that, for a Poisson problem with a solution of regularity as the regularity in space of u_2 in (6.13), one obtains for the $L^2(\mathrm{D})$ 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 $\mathbb{P}^1$-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 Figure 2, Figure 3, 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 (6.2) 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 $\boldsymbol{p} = (p,p,p,p)$ with $p = \lfloor \frac{\ln N}{2} \rfloor$. This choice of the discretization parameters balances the terms of the error bound (5.11). Hence, the total number of degrees of freedom MN behaves like in (5.12). The numerical results in Figure 4 confirm Remark 5.6.

6.3.3. *Singular Solution.* We consider the singular solution

$$u_3(t,x_1,x_2) = u_{\mathrm{reg}}(t,x_1,x_2) + t^{3/5}\mathrm{e}^{-t}\eta(x_1,x_2) \cdot r(x_1,x_2)^{2/3} \cdot \sin\left(\frac{2}{3}\left(\arg(x_1,x_2) - \frac{\pi}{2}\right)\right) \tag{6.16}$$

for $(t, x_1, x_2) \in \overline{Q}$ with the smooth part u_{reg} in (6.14), the radial coordinate $r(x_1, x_2) \in [0, \infty)$, the angular coordinate $\arg(x_1, x_2) \in (0, 2\pi]$, and the cutoff function $\eta \in C^2(\mathbb{R}^2)$ in (6.15). 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 (3.9). On the other hand, solutions with a singular behavior as u_3 are possible even for sources g, which satisfy the condition (3.9). 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 (6.16). Furthermore, the solution u_3 has the same spatial singularity as u_2 in (6.13). 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 (6.2) of the solution u_3 are reported in Figure 5. For a given spatial discretization parameter N, we choose the temporal mesh as in (4.1) with grading parameter $\sigma = 0.17$, slope parameter $\mu_{\mathrm{hp}} = 1.0$, numbers of elements $m_1 = \lfloor 2.2 \cdot \ln N \rfloor$, $m_2 = 1$, and temporal polynomial degrees $\boldsymbol{p} \in \mathbb{N}^m$ as in (4.3). This choice of the discretization parameters balances the terms of the error bound (5.9). Hence, the total number of degrees of freedom MN behaves like in (5.10). The numerical results in Figure 5 are in accordance with Theorem 5.5, when the temporal analytic regularity condition (3.9), and hence, the conditions on parameters μ_{hp}, m_2, i.e., (5.5), (5.6), are ignored.

7. Conclusion

Based on a variational space-time formulation of the IBVP (1.1)–(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., [28]).

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., [31]). 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 $[0, T]$.

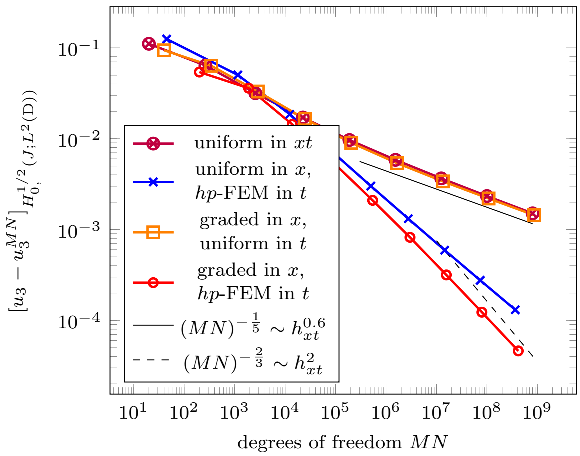

FIGURE 5. Numerical results with the space-time Galerkin approximation (6.2) for the 2D example with the singular solution u_3 in (6.16), for all combinations of uniform mesh refinement or graded meshes in space (with grading parameter $\beta = 0.6$), and $\mathbb{P}^1$-FEM with uniform mesh refinement 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 $\mu_{\mathrm{hp}} = 1.0$, numbers of elements $m_1 = \lfloor 2.2 \cdot \ln N \rfloor$, $m_2 = 1$, and temporal polynomial degrees $\boldsymbol{p} \in \mathbb{N}^m$ as in (4.3).

We finally remark that the presently adopted space-time variational formulation will also allow for *a posteriori* time-discretization error estimation, which is reliable and robust uniformly with respect to p. Details shall be developed elsewhere.

References

1. James H. Adler and Victor Nistor, *Graded mesh approximation in weighted Sobolev spaces and elliptic equations in 2D*, Math. Comp. **84** (2015), no. 295, 2191–2220. MR 3356024
2. Bernd Ammann and Victor Nistor, *Weighted Sobolev spaces and regularity for polyhedral domains*, Comput. Methods Appl. Mech. Engrg. **196** (2007), no. 37-40, 3650–3659. MR 2339991
3. Roman Andreev, *Stability of sparse space-time finite element discretizations of linear parabolic evolution equations*, IMA J. Numer. Anal. **33** (2013), no. 1, 242–260. MR 3020957
4. Thomas Apel, *Anisotropic finite elements: local estimates and applications*, Advances in Numerical Mathematics, B. G. Teubner, Stuttgart, 1999. MR 1716824
5. I. Babuška, R. B. Kellogg, and J. Pitkäranta, *Direct and inverse error estimates for finite elements with mesh refinements*, Numer. Math. **33** (1979), no. 4, 447–471. MR 553353
6. Constantin Bacuta, Victor Nistor, and Ludmil T. Zikatanov, *Improving the rate of convergence of high-order finite elements on polyhedra. I. A priori estimates*, Numer. Funct. Anal. Optim. **26** (2005), no. 6, 613–639. MR 2187917
7. ______, *Improving the rate of convergence of high-order finite elements on polyhedra. II. Mesh refinements and interpolation*, Numer. Funct. Anal. Optim. **28** (2007), no. 7-8, 775–824. MR 2347683
8. Constantin Băcuţă, Hengguang Li, and Victor Nistor, *Differential operators on domains with conical points: precise uniform regularity estimates*, Rev. Roumaine Math. Pures Appl. **62** (2017), no. 3, 383–411. MR 3711082

9. Andrea Cangiani, Zhaonan Dong, and Emmanuil H. Georgoulis, *hp-version space-time discontinuous Galerkin methods for parabolic problems on prismatic meshes*, SIAM J. Sci. Comput. **39** (2017), no. 4, A1251–A1279. MR 3672375
10. Philip J. Davis, *Interpolation and approximation*, Dover Publications, Inc., New York, 1975. MR 0380189
11. Denis Devaud, *Petrov-Galerkin space-time hp-approximation of parabolic equations in $H^{1/2}$*, IMA J. Numer. Anal. **40** (2020), no. 4, 2717–2745. MR 4167060
12. G. I. Eskin, *Boundary value problems for elliptic pseudodifferential equations*, Translations of Mathematical Monographs, vol. 52, American Mathematical Society, Providence, R.I., 1981. MR 623608
13. Birgit Faermann, *Localization of the Aronszajn-Slobodeckij norm and application to adaptive boundary element methods. I. The two-dimensional case*, IMA J. Numer. Anal. **20** (2000), no. 2, 203–234. MR 1752263
14. Thomas Führer and Michael Karkulik, *Space-time least-squares finite elements for parabolic equations*, Comput. Math. Appl. **92** (2021), 27–36. MR 4242919
15. Gregor Gantner and Rob Stevenson, *Further results on a space-time FOSLS formulation of parabolic PDEs*, ESAIM Math. Model. Numer. Anal. **55** (2021), no. 1, 283–299. MR 4216839
16. Fernando D. Gaspoz and Pedro Morin, *Approximation classes for adaptive higher order finite element approximation*, Math. Comp. **83** (2014), no. 289, 2127–2160. MR 3223327
17. Walter Gautschi, *Numerical integration over the square in the presence of algebraic/logarithmic singularities with an application to aerodynamics*, Numer. Algorithms **61** (2012), no. 2, 275–290.
18. Michael Griebel, Daniel Oeltz, and Panayot Vassilevski, *Space-time approximation with sparse grids*, SIAM J. Sci. Comput. **28** (2006), no. 2, 701–727. MR 2231727
19. Angela Kunoth and Christoph Schwab, *Analytic regularity and GPC approximation for control problems constrained by linear parametric elliptic and parabolic PDEs*, SIAM J. Control Optim. **51** (2013), no. 3, 2442–2471. MR 3064588
20. Ulrich Langer, Stephen E. Moore, and Martin Neumüller, *Space-time isogeometric analysis of parabolic evolution problems*, Comput. Methods Appl. Mech. Engrg. **306** (2016), 342–363. MR 3502571
21. Ulrich Langer and Marco Zank, *Efficient direct space-time finite element solvers for parabolic initial-boundary value problems in anisotropic Sobolev spaces*, SIAM J. Sci. Comput. **43** (2021), no. 4, A2714–A2736. MR 4295048
22. Hengguang Li, Anna Mazzucato, and Victor Nistor, *Analysis of the finite element method for transmission/mixed boundary value problems on general polygonal domains*, Electron. Trans. Numer. Anal. **37** (2010), 41–69. MR 2777235
23. Vladimir Maz'ya and Jürgen Rossmann, *Elliptic equations in polyhedral domains*, Mathematical Surveys and Monographs, vol. 162, American Mathematical Society, Providence, RI, 2010. MR 2641539
24. William McLean, *Strongly elliptic systems and boundary integral equations*, Cambridge University Press, Cambridge, 2000. MR 1742312
25. Monica Montardini, Matteo Negri, Giancarlo Sangalli, and Mattia Tani, *Space-time least-squares isogeometric method and efficient solver for parabolic problems*, Math. Comp. **89** (2020), no. 323, 1193–1227. MR 4063316
26. Amnon Pazy, *Semigroups of linear operators and applications to partial differential equations*, Applied Mathematical Sciences, vol. 44, Springer-Verlag, New York, 1983. MR 710486
27. Dominik Schötzau and Christoph Schwab, *Time Discretization of Parabolic Problems by the hp-Version of the Discontinuous Galerkin Finite Element Method*, SIAM J. Numer. Anal. **38** (2000), no. 3, 837–875.
28. Dominik Schötzau and Christoph Schwab, *hp-discontinuous Galerkin time-stepping for parabolic problems*, C. R. Acad. Sci. Paris Sér. I Math. **333** (2001), no. 12, 1121–1126. MR 1881245
29. Christoph Schwab, *p- and hp-finite element methods*, Numerical Mathematics and Scientific Computation, The Clarendon Press, Oxford University Press, New York, 1998, Theory and applications in solid and fluid mechanics. MR 1695813
30. Christoph Schwab and Rob Stevenson, *Space-time adaptive wavelet methods for parabolic evolution problems*, Math. Comp. **78** (2009), no. 267, 1293–1318. MR 2501051
31. ———, *Fractional space-time variational formulations of (Navier-) Stokes equations*, SIAM J. Math. Anal. **49** (2017), no. 4, 2442–2467. MR 3668596

32. Olaf Steinbach, *Space-time finite element methods for parabolic problems*, Comput. Methods Appl. Math. **15** (2015), no. 4, 551–566. MR 3403450
33. Olaf Steinbach and Agnese Missoni, *A note on a modified Hilbert transform*, Applicable Analysis (2022), 1–8.
34. Olaf Steinbach and Marco Zank, *Coercive space-time finite element methods for initial boundary value problems*, Electron. Trans. Numer. Anal. **52** (2020), 154–194. MR 4102914
35. ———, *A note on the efficient evaluation of a modified Hilbert transformation*, J. Numer. Math. **29** (2021), no. 1, 47–61. MR 4230416
36. Rob Stevenson and Jan Westerdiep, *Stability of Galerkin discretizations of a mixed space-time variational formulation of parabolic evolution equations*, IMA J. Numer. Anal. **41** (2021), no. 1, 28–47. MR 4205051
37. Vidar Thomée, *Galerkin finite element methods for parabolic problems*, second ed., Springer Series in Computational Mathematics, vol. 25, Springer-Verlag, Berlin, 2006. MR 2249024
38. Hans Triebel, *Interpolation theory, function spaces, differential operators*, second ed., Johann Ambrosius Barth, Heidelberg, 1995. MR 1328645
39. Marco Zank, *An exact realization of a modified Hilbert transformation for space-time methods for parabolic evolution equations*, Comput. Methods Appl. Math. **21** (2021), no. 2, 479–496. MR 4235810

Appendix A. Some Properties of $H^{1/2}_{0,}(a,b)$

In this appendix, we provide proofs of Lemma 2.1 and Lemma 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 $v\colon (a,b)\to\mathbb{R}$. All results and proofs can be generalized straightforwardly to X-valued functions $v\colon (a,b)\to X$ 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 $\|\circ\|_{H^1(\mathbb{R})} = (\|\circ\|^2_{L^2(\mathbb{R})} + \|\partial_t \circ\|^2_{L^2(\mathbb{R})})^{1/2}$, we consider the interpolation norm $\|\circ\|_{H^{1/2}(\mathbb{R})}$ and the Slobodetskii norm

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

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

$$|v|_{H^{1/2}(\mathbb{R})} := \left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{|v(s)-v(t)|^2}{|s-t|^2}\mathrm{d}s\mathrm{d}t\right)^{1/2}.$$

Proof of Lemma 2.1. The equivalence of norms is proven in, e.g., [24]. We give more details about the norm equivalence constants. For this purpose, we introduce an extension operator and establish bounds on its norm. Define $\mathcal{E}_1\colon H^1_{0,}(a,b)\to H^1(\mathbb{R})$,

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

for $v \in H^1_{0,}(a,b)$. The mapping $\mathcal{E}_0\colon\ L^2(a,b) \to L^2(\mathbb{R})$ 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 $v \in L^2(a,b)$,

$$\|\mathcal{E}_0 v\|^2_{L^2(\mathbb{R})} = \int_a^b |v(t)|^2 \,\mathrm{d}t + \int_b^{2b-a} |v(2b-t)|^2 \,\mathrm{d}t = 2\|v\|^2_{L^2(a,b)}$$

and, for $v \in H^1_{0,}(a,b)$,

$$\|\partial_t \mathcal{E}_1 v\|^2_{L^2(\mathbb{R})} = \int_a^b |\partial_t v(t)|^2 \,\mathrm{d}t + \int_b^{2b-a} |\partial_t v(2b-t)|^2 \,\mathrm{d}t = 2\|\partial_t v\|^2_{L^2(a,b)} .$$

Hence, for $v \in H^1_{0,}(a,b)$, it holds true that

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

where the Poincaré inequality (see Lemma A.1 below) is used in the last step. Interpolation yields an operator $\mathcal{E}_{1/2}\colon\ H^{1/2}_{0,}(a,b) \to H^{1/2}(\mathbb{R})$ with $\mathcal{E}_{1/2} v = \mathcal{E}_0 v$ for $v \in H^{1/2}_{0,}(a,b)$ and

$$\forall v \in H^{1/2}_{0,}(a,b): \quad \left\|\mathcal{E}_{1/2} v\right\|^2_{H^{1/2}(\mathbb{R})} \leq 2\sqrt{1 + \frac{4(b-a)^2}{\pi^2}} \|v\|^2_{H^{1/2}_{0,}(a,b)}. \tag{A.1}$$

Next, we estimate $\left|\!\left|\!\left|\mathcal{E}_{1/2} v\right|\!\right|\!\right|_{H^{1/2}(\mathbb{R})}$ for $v \in H^{1/2}_{0,}(a,b)$. For this purpose, we compute

$$\begin{aligned} \left|\mathcal{E}_{1/2} v\right|^2_{H^{1/2}(\mathbb{R})} &= \int_a^\infty \int_a^\infty () + 2\int_a^\infty \int_{-\infty}^a () + \int_{-\infty}^a \int_{-\infty}^a () \\ &= \left|(\mathcal{E}_{1/2} v)_{|(a,\infty)}\right|^2_{H^{1/2}(a,\infty)} + 2\int_a^\infty \int_{-\infty}^a \frac{|\mathcal{E}_{1/2} v(t)|^2}{|s-t|^2} \mathrm{d}s\mathrm{d}t + 0 \\ &= \left|(\mathcal{E}_{1/2} v)_{|(a,\infty)}\right|^2_{H^{1/2}(a,\infty)} + 2\int_a^\infty \frac{\left|\mathcal{E}_{1/2} v(t)\right|^2}{t-a} \mathrm{d}t \end{aligned} \tag{A.2}$$

for $v \in H^{1/2}_{0,}(a,b)$, where the seminorm $|\circ|_{H^{1/2}(a,\infty)}$ is defined by (2.6) with $b = \infty$. The integral in the bound (A.2) is finite due to $v \in H^{1/2}_{0,}(a,b)$, cf. (2.7). Thus, we get

$$\begin{aligned} \left|\!\left|\!\left|\mathcal{E}_{1/2} v\right|\!\right|\!\right|^2_{H^{1/2}(\mathbb{R})} &= 2\|v\|^2_{L^2(a,b)} + \left|(\mathcal{E}_{1/2} v)_{|(a,\infty)}\right|^2_{H^{1/2}(a,\infty)} + 2\int_a^\infty \frac{\left|\mathcal{E}_{1/2} v(t)\right|^2}{t-a} \mathrm{d}t \\ &= 2\|v\|^2_{L^2(a,b)} + |v|^2_{H^{1/2}(a,b)} + 2\int_b^\infty \int_a^b () + \int_b^\infty \int_b^\infty () + 2\int_a^\infty \frac{\left|\mathcal{E}_{1/2} v(t)\right|^2}{t-a} \mathrm{d}t. \end{aligned}$$

The third term on the right side is bounded by

$$\begin{aligned}
2\int_b^\infty \int_a^b () &= 2\int_b^{2b-a} \int_a^b \frac{|v(s)-v(2b-t)|^2}{|s-t|^2}\mathrm{d}s\mathrm{d}t + 2\int_{2b-a}^\infty \int_a^b \frac{|v(s)|^2}{|s-t|^2}\mathrm{d}s\mathrm{d}t \\
&= 2\int_a^b \int_a^b \frac{|v(s)-v(t)|^2}{\underbrace{|2b-s-t|^2}_{\geq |s-t|^2}}\mathrm{d}s\mathrm{d}t + 2\int_a^b \frac{|v(s)|^2}{\underbrace{2b-a-s}_{\geq s-a}}\mathrm{d}s \\
&\leq 2\,|v|^2_{H^{1/2}(a,b)} + 2\int_a^b \frac{|v(s)|^2}{s-a}\mathrm{d}s,
\end{aligned}$$

the fourth term is

$$\begin{aligned}
&\int_b^\infty \int_b^\infty () = \int_b^{2b-a} \int_b^{2b-a} () + 2\int_{2b-a}^\infty \int_b^{2b-a} () + \int_{2b-a}^\infty \int_{2b-a}^\infty () \\
&= \int_b^{2b-a} \int_b^{2b-a} \frac{|v(2b-s)-v(2b-t)|^2}{|s-t|^2}\mathrm{d}s\mathrm{d}t + 2\int_{2b-a}^\infty \int_b^{2b-a} \frac{|v(2b-s)|^2}{|s-t|^2}\mathrm{d}s\mathrm{d}t + 0 \\
&= |v|^2_{H^{1/2}(a,b)} + 2\int_b^{2b-a} \frac{|v(2b-s)|^2}{2b-a-s}\mathrm{d}s = |v|^2_{H^{1/2}(a,b)} + 2\int_a^b \frac{|v(s)|^2}{s-a}\mathrm{d}s,
\end{aligned}$$

whereas for the fifth term, we have

$$\begin{aligned}
2\int_a^\infty \frac{\left|\mathcal{E}_{1/2}v(t)\right|^2}{t-a}\mathrm{d}t &= 2\int_a^b \frac{|v(t)|^2}{t-a}\mathrm{d}t + 2\int_b^{2b-a} \frac{|v(2b-t)|^2}{t-a}\mathrm{d}t \\
&= 2\int_a^b \frac{|v(t)|^2}{t-a}\mathrm{d}t + 2\int_a^b \frac{|v(t)|^2}{\underbrace{2b-a-t}_{\geq t-a}}\mathrm{d}t \leq 4\int_a^b \frac{|v(t)|^2}{t-a}\mathrm{d}t.
\end{aligned}$$

Using the above estimates gives for all $v \in H^{1/2}_{0,}(a,b)$

$$\left|\!\left|\!\left|\mathcal{E}_{1/2}v\right|\!\right|\!\right|^2_{H^{1/2}(\mathbb{R})} \leq 2\|v\|^2_{L^2(a,b)} + 4\,|v|^2_{H^{1/2}(a,b)} + 8\int_a^b \frac{|v(t)|^2}{t-a}\mathrm{d}t \leq 8|\!|\!|v|\!|\!|^2_{H^{1/2}_{0,}(a,b)}\,.$$

With these properties, we have for all $v \in H^{1/2}_{0,}(a,b)$ the lower bound in the norm equivalence:

$$\|v\|_{H^{1/2}_{0,}(a,b)} \leq \left\|\mathcal{E}_{1/2}v\right\|_{H^{1/2}(\mathbb{R})} \leq \frac{1}{C_{\mathbb{R},1}}\left|\!\left|\!\left|\mathcal{E}_{1/2}v\right|\!\right|\!\right|_{H^{1/2}(\mathbb{R})} \leq \frac{2\sqrt{2}}{C_{\mathbb{R},1}}|\!|\!|v|\!|\!|_{H^{1/2}_{0,}(a,b)}\,.$$

Here, the first inequality is proven by interpolation, the second estimate follows from

$$\text{(A.3)} \qquad \forall z \in H^{1/2}(\mathbb{R}): \quad C_{\mathbb{R},1}\|z\|_{H^{1/2}(\mathbb{R})} \leq |\!|\!|z|\!|\!|_{H^{1/2}(\mathbb{R})} \leq C_{\mathbb{R},2}\|z\|_{H^{1/2}(\mathbb{R})}$$

with constants $C_{\mathbb{R},1}$, $C_{\mathbb{R},2} > 0$, see [24, Theorem B.7] and [12, Lemma 4.1].

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

$$\begin{aligned}
|||v|||^2_{H^{1/2}_{0,}(a,b)} &\le \left\|\mathcal{E}_{1/2}v\right\|^2_{L^2(\mathbb{R})} + \left|(\mathcal{E}_{1/2}v)_{|(a,\infty)}\right|^2_{H^{1/2}(a,\infty)} + \int_a^\infty \frac{\left|\mathcal{E}_{1/2}v(t)\right|^2}{t-a}\mathrm{d}t \\
&= \left\|\mathcal{E}_{1/2}v\right\|^2_{L^2(\mathbb{R})} + \frac{1}{2}\left|(\mathcal{E}_{1/2}v)_{|(a,\infty)}\right|^2_{H^{1/2}(a,\infty)} + \frac{1}{2}\left|\mathcal{E}_{1/2}v\right|^2_{H^{1/2}(\mathbb{R})} \\
&\le \left|\left|\left|\mathcal{E}_{1/2}v\right|\right|\right|^2_{H^{1/2}(\mathbb{R})} \\
&\le (C_{\mathbb{R},2})^2\left\|\mathcal{E}_{1/2}v\right\|^2_{H^{1/2}(\mathbb{R})} \le (C_{\mathbb{R},2})^2 2\sqrt{1+\frac{4(b-a)^2}{\pi^2}}\|v\|^2_{H^{1/2}_{0,}(a,b)},
\end{aligned}$$

i.e., the assertion is proven. ☐

The following proof of Lemma 2.2 restricts the argument of [13] 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 (2.6) as follows:

$$\begin{aligned}
|v|^2_{H^{1/2}(a,b)} &= \int_a^\tau \int_a^b \Big(\cdots\Big)\mathrm{d}s\mathrm{d}t + \int_\tau^b \int_a^b \Big(\cdots\Big)\mathrm{d}s\mathrm{d}t \\
&= \int_a^\tau \int_a^\tau \Big(\cdots\Big)\mathrm{d}s\mathrm{d}t + 2\int_a^\tau \int_\tau^b \Big(\cdots\Big)\mathrm{d}s\mathrm{d}t + \int_\tau^b \int_\tau^b \Big(\cdots\Big)\mathrm{d}s\mathrm{d}t \\
&= |v|^2_{H^{1/2}(a,\tau)} + 2\int_a^\tau \int_\tau^b \Big(\cdots\Big)\mathrm{d}s\mathrm{d}t + |v|^2_{H^{1/2}(\tau,b)}.
\end{aligned}$$

For the integral on the right side, we get

$$\begin{aligned}
2\int_a^\tau \int_\tau^b \frac{|v(s)-v(t)|^2}{|s-t|^2}\mathrm{d}s\mathrm{d}t &\le 4\int_a^\tau \int_\tau^b \frac{|v(s)|^2}{|s-t|^2}\mathrm{d}s\mathrm{d}t + 4\int_a^\tau \int_\tau^b \frac{|v(t)|^2}{|s-t|^2}\mathrm{d}s\mathrm{d}t \\
&= 4\int_\tau^b |v(s)|^2[(s-\tau)^{-1}-(s-a)^{-1}]\mathrm{d}s \\
&\quad + 4\int_a^\tau |v(t)|^2[(\tau-t)^{-1}-(b-t)^{-1}]\mathrm{d}t \\
&\le 4\int_\tau^b \frac{|v(s)|^2}{s-\tau}\mathrm{d}s + 4\int_a^\tau \frac{|v(t)|^2}{\tau-t}\mathrm{d}t.
\end{aligned}$$

Thus, the assertion follows. ☐

Lemma A.1. *For $a, b \in \mathbb{R}$, $a < b$, the Poincaré inequalities*

$$\begin{aligned}
\forall v \in H^{1/2}_{0,}(a,b): &\quad \|v\|_{L^2(a,b)} \le \sqrt{\frac{2(b-a)}{\pi}}\|v\|_{H^{1/2}_{0,}(a,b)}, \\
\forall v \in H^{1}_{0,}(a,b): &\quad \|v\|_{H^{1/2}_{0,}(a,b)} \le \sqrt{\frac{2(b-a)}{\pi}}\|\partial_t v\|_{L^2(a,b)}, \\
\forall v \in H^{1}_{0,}(a,b): &\quad \|v\|_{L^2(a,b)} \le \frac{2(b-a)}{\pi}\|\partial_t v\|_{L^2(a,b)}
\end{aligned}$$

hold true, where the constants are sharp.

Proof. By interpolation, we have the Fourier series representations

$$\|v\|^2_{L^2(a,b)} = \sum_{k=0}^{\infty} |v_k|^2, \quad \|v\|^2_{H^{1/2}_{0,}(a,b)} = \sum_{k=0}^{\infty} \sqrt{\lambda_k}|v_k|^2, \quad \|\partial_t v\|^2_{L^2(a,b)} = \sum_{k=0}^{\infty} \lambda_k |v_k|^2 \tag{A.4}$$

with coefficients v_k as in (2.2) and eigenvalues $\lambda_k = \frac{\pi^2(2k+1)^2}{4(b-a)^2}$ of the eigenvalue problem (2.3). 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. □

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

$$\forall v \in H^1_{0,}(a,b) : \|v\|_{H^{1/2}_{0,}(a,b)} \leq \sqrt{\|v\|_{L^2(a,b)} \|\partial_t v\|_{L^2(a,b)}}$$

holds true, where $\| \circ \|_{H^{1/2}_{0,}(a,b)}$ denotes the interpolation norm (2.1).

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

Appendix B. Proof of Lemma 3.6

Let $b \in (0, T]$ be fixed. According to (3.11) for $l = 0$, the estimate

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

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

$$\forall t > 0\colon \|E(t)\|_{\mathcal{L}(X_\varepsilon, X_2)} \leq \sqrt{\frac{1}{\sqrt{2\pi}} \left(\frac{1}{2}\right) t^{-2+\varepsilon}} = \frac{1}{\sqrt[4]{8\pi}} t^{-1+\varepsilon/2}. \tag{B.1}$$

The solution u admits the representation

$$u(t) = \int_0^t E(\tau) g(t-\tau) \mathrm{d}\tau, \quad 0 \leq t \leq b, \tag{B.2}$$

see (3.8), and for $t \in [0, b]$, it follows that

$$\begin{aligned} \|u(t)\|_{X_2} &\leq \int_0^t \|E(\tau) g(t-\tau)\|_{X_2} \mathrm{d}\tau \leq \int_0^t \|E(\tau)\|_{\mathcal{L}(X_\varepsilon, X_2)} \|g(t-\tau)\|_{X_\varepsilon} \mathrm{d}\tau \\ &\leq \frac{1}{\sqrt[4]{8\pi}} C_g \int_0^t \tau^{-1+\varepsilon/2} \mathrm{d}\tau = \frac{1}{\sqrt[4]{8\pi}} C_g \frac{2}{\varepsilon} t^{\varepsilon/2} = \sqrt[4]{\frac{2}{\pi}} C_g \frac{1}{\varepsilon} t^{\varepsilon/2}. \end{aligned}$$

We estimate the three terms of $|||u|||_{H^{1/2}_{0,}((0,b);X_2)}$ expressed as in (2.7).
First term: From the previous bound for $\|u(t)\|_{X_2}$, we derive

$$\|u\|^2_{L^2((0,b);X_2)} = \int_0^b \|u(t)\|^2_{X_2} \mathrm{d}t \leq \sqrt{\frac{2}{\pi}} C_g^2 \frac{1}{\varepsilon^2} \int_0^b t^\varepsilon \mathrm{d}t = \sqrt{\frac{2}{\pi}} C_g^2 \frac{1}{\varepsilon^2(1+\varepsilon)} b^{1+\varepsilon}.$$

Third term: Similarly, we obtain

$$\int_0^b \frac{\|u(t)\|^2_{X_2}}{t} \mathrm{d}t \leq \sqrt{\frac{2}{\pi}} C_g^2 \frac{1}{\varepsilon^2} \int_0^b t^{\varepsilon-1} \mathrm{d}t = \sqrt{\frac{2}{\pi}} C_g^2 \frac{1}{\varepsilon^3} b^\varepsilon.$$

Second term: Recalling (2.6), we need to estimate $\int_0^b\int_0^b \frac{\|u(s)-u(t)\|_{X_2}^2}{|s-t|^2}\mathrm{d}s\mathrm{d}t$. For $b\geq s\geq t\geq 0$, we have

$$\begin{aligned}
\|u(s)-u(t)\|_{X_2} &\overset{(B.2)}{=} \left\|\int_0^s E(\tau)g(s-\tau)\mathrm{d}\tau - \int_0^t E(\tau)g(t-\tau)\mathrm{d}\tau\right\|_{X_2}\\
&= \left\|\int_t^s E(\tau)g(s-\tau)\mathrm{d}\tau + \int_0^t E(\tau)g(s-\tau)\mathrm{d}\tau - \int_0^t E(\tau)g(t-\tau)\mathrm{d}\tau\right\|_{X_2}\\
&\leq \left\|\int_t^s E(\tau)g(s-\tau)\mathrm{d}\tau\right\|_{X_2} + \left\|\int_0^t E(\tau)[g(s-\tau)-g(t-\tau)]\mathrm{d}\tau\right\|_{X_2}\\
&\leq \int_t^s \|E(\tau)\|_{\mathcal{L}(X_\varepsilon,X_2)}\|g(s-\tau)\|_{X_\varepsilon}\mathrm{d}\tau\\
&\quad + \int_0^t \|E(\tau)\|_{\mathcal{L}(X_\varepsilon,X_2)}\|g(s-\tau)-g(t-\tau)\|_{X_\varepsilon}\mathrm{d}\tau\\
&\overset{(B.1)}{\leq} \frac{1}{\sqrt[4]{8\pi}}\left(\int_t^s \tau^{-1+\varepsilon/2}\|g(s-\tau)\|_{X_\varepsilon}\mathrm{d}\tau + \int_0^t \tau^{-1+\varepsilon/2}\left\|\int_{t-\tau}^{s-\tau} g'(r)dr\right\|_{X_\varepsilon}\mathrm{d}\tau\right)\\
&\overset{(3.19)}{\leq} \frac{1}{\sqrt[4]{8\pi}}\,C_g\,\frac{2}{\varepsilon}\,\Big((s^{\varepsilon/2}-t^{\varepsilon/2}) + t^{\varepsilon/2}\,(s-t)\Big).
\end{aligned}$$

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

$$\|u(s)-u(t)\|_{X_2} \leq \frac{1}{\sqrt[4]{8\pi}}\,C_g\,\frac{2}{\varepsilon}\,\Big((t^{\varepsilon/2}-s^{\varepsilon/2}) + s^{\varepsilon/2}\,(t-s)\Big)$$

holds true. We conclude that

$$\begin{aligned}
&\int_0^b\int_0^b \frac{\|u(s)-u(t)\|_{X_2}^2}{|s-t|^2}\mathrm{d}s\mathrm{d}t\\
&\qquad= \int_0^b\int_0^t \frac{\|u(s)-u(t)\|_{X_2}^2}{|s-t|^2}\mathrm{d}s\mathrm{d}t + \int_0^b\int_0^s \frac{\|u(s)-u(t)\|_{X_2}^2}{|s-t|^2}\mathrm{d}t\mathrm{d}s\\
&\qquad\leq \frac{1}{\sqrt{8\pi}}\,C_g^2\,\frac{16}{\varepsilon^2}\int_0^b\int_0^s\left(\frac{(s^{\varepsilon/2}-t^{\varepsilon/2})^2}{(s-t)^2}+t^{\varepsilon}\right)\mathrm{d}t\mathrm{d}s\\
&\qquad= \sqrt{\frac{32}{\pi}}\,\frac{1}{\varepsilon^2}\,C_g^2\left(\int_0^b s^{-1+\varepsilon}\int_0^1 \frac{(1-r^{\varepsilon/2})^2}{(1-r)^2}\mathrm{d}r\mathrm{d}s + \frac{b^{\varepsilon+2}}{(\varepsilon+1)(\varepsilon+2)}\right)\\
&\qquad\overset{\varepsilon/2\leq 1}{\leq} \sqrt{\frac{32}{\pi}}\,\frac{1}{\varepsilon^2}\,C_g^2\left(\int_0^b s^{-1+\varepsilon}\int_0^1 \frac{(1-r^{\varepsilon/2})^2}{(1-r^{\varepsilon/2})^2}\mathrm{d}r\mathrm{d}s + \frac{b^{\varepsilon+2}}{(\varepsilon+1)(\varepsilon+2)}\right)\\
&\qquad= \sqrt{\frac{32}{\pi}}\,\frac{1}{\varepsilon^2}\,C_g^2\left(\frac{b^{\varepsilon}}{\varepsilon}+\frac{b^{\varepsilon+2}}{(\varepsilon+1)(\varepsilon+2)}\right).
\end{aligned}$$

Conclusion of the proof: By combining the bounds of the three terms, we arrive at the *a priori* estimate

$$|||u|||_{H_{0,}^{1/2}((0,b);X_2)} \leq \sqrt[4]{\frac{2}{\pi}}\,\frac{1}{\varepsilon}\,b^{\varepsilon/2}\left(\frac{b}{1+\varepsilon}+\frac{3}{\varepsilon}+\frac{4b^2}{(\varepsilon+1)(\varepsilon+2)}\right)^{1/2} C_g\,,$$

which gives the assertion.

Appendix C. Proof of Lemma 5.2

This proof is a slight modification of the proof of [11, Lemma 3.4]. We use Stirling's inequalities

$$\forall x > 0: \quad \sqrt{2\pi} x^{x-1/2} \mathrm{e}^{-x} \leq \Gamma(x) \leq \sqrt{2\pi} x^{x-1/2} \mathrm{e}^{-x} \mathrm{e}^{\frac{1}{12x}}. \tag{C.1}$$

For $j \geq 1$, (C.1) yields

$$\begin{aligned}
\frac{\Gamma(\lfloor \mu j \rfloor - j + 1)}{\Gamma(\lfloor \mu j \rfloor + j + 1)} &\leq \frac{\Gamma(\mu j - j + 1)}{\Gamma(\mu j + j)} \\
&\leq \frac{\sqrt{2\pi}(\overbrace{\mu j - j + 1}^{\leq \mu j \text{ as } j \geq 1})^{\overbrace{\mu j - j + 1/2}^{\geq 0 \text{ as } \mu \geq 1}} \mathrm{e}^{-(\mu j - j + 1)} \overbrace{\mathrm{e}^{1/(12(\mu j - j + 1))}}^{\leq 2}}{\sqrt{2\pi}(\underbrace{\mu j + j}_{\geq \mu j})^{\mu j + j - 1/2} \mathrm{e}^{-(\mu j + j)}} \leq \frac{2\mu j}{\mathrm{e}} \left(\frac{\mathrm{e}}{\mu j} \right)^{2j}
\end{aligned}$$

and

$$\Gamma(j+3)^2 = (\underbrace{j+2}_{\leq 3j})^2 (\underbrace{j+1}_{\leq 2j})^2 j^2 \Gamma(j)^2 \leq j^6 \cdot 72\pi j^{2j-1} \mathrm{e}^{-2j} \underbrace{\mathrm{e}^{\frac{1}{6j}}}_{<2} \leq 144\pi j^5 j^{2j} \mathrm{e}^{-2j}.$$

Thus, we have

$$\forall j \in \mathbb{N}: \quad \alpha^{2j} \frac{\Gamma(\lfloor \mu j \rfloor - j + 1)}{\Gamma(\lfloor \mu j \rfloor + j + 1)} \Gamma(j+3)^2 \leq \frac{288\pi\mu}{\mathrm{e}} j^6 \left(\frac{\alpha}{\mu} \right)^{2j}.$$

Hence, we conclude that

$$\begin{aligned}
\sum_{j=0}^{m} \alpha^{2j} \frac{\Gamma(\lfloor \mu j \rfloor - j + 1)}{\Gamma(\lfloor \mu j \rfloor + j + 1)} \Gamma(j+3)^2 &= 4 + \sum_{j=1}^{m} \alpha^{2j} \frac{\Gamma(\lfloor \mu j \rfloor - j + 1)}{\Gamma(\lfloor \mu j \rfloor + j + 1)} \Gamma(j+3)^2 \\
&\leq 4 + \frac{288\pi\mu}{\mathrm{e}} \sum_{j=1}^{\infty} j^6 \left(\frac{\alpha}{\mu} \right)^{2j} < \infty,
\end{aligned}$$

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.

Faculty of Mathematics, University of Vienna, Vienna, Austria
Email address: `ilaria.perugia,marco.zank@univie.ac.at`

Seminar for Applied Mathematics, ETH Zürich, 8092 Zürich, Switzerland
Email address: `schwab@math.ethz.ch`