ESAIM: M2AN 55 (2021) 283–299
https://doi.org/10.1051/m2an/2020084

ESAIM: Mathematical Modelling and Numerical Analysis
www.esaim-m2an.org

FURTHER RESULTS ON A SPACE-TIME FOSLS FORMULATION OF PARABOLIC PDES

Gregor Gantner* and Rob Stevenson

Abstract. In [2019, Space-time least-squares finite elements for parabolic equations, arXiv:1911.01942] by Führer and Karkulik, well-posedness of a space-time First-Order System Least-Squares formulation of the heat equation was proven. In the present work, this result is generalized to general second order parabolic PDEs with possibly inhomogenoeus boundary conditions, and plain convergence of a standard adaptive finite element method driven by the least-squares estimator is demonstrated. The proof of the latter easily extends to a large class of least-squares formulations.

Mathematics Subject Classification. 35K20, 65M12, 65M15, 65M60.

Received May 20, 2020. Accepted December 8, 2020.

1. Introduction

Currently, there is a growing interest in simultaneous space-time methods for solving parabolic evolution equations originally introduced in [2, 3], see *e.g.*, [1, 16, 18, 20, 22, 23, 26, 28, 30, 33, 34, 40, 41]. Main reasons are that, compared to classical time marching methods, space-time methods are much better suited for a massively parallel implementation, are guaranteed to give quasi-optimal approximations from the trial space that is employed, have the potential to drive optimally converging simultanously space-time adaptive refinement routines, and they provide enhanced possibilities for reduced order modelling of parameter-dependent problems. On the other hand, space-time methods require more storage. This disadvantage however vanishes for problems of optimal control or data assimilation, for which the solution is needed simultaneously over the whole time interval anyway.

The common space-time variational formulation of a parabolic equation results in a bilinear form that is non-coercive. For the heat equation $\partial_t u - \Delta_{\mathbf{x}} u = f$, $u(0,\cdot) = u_0$ on a time-space cylinder $I \times \Omega$, where $I := (0,T)$ and $\Omega \subset \mathbb{R}^d$, with homogeneous Dirichlet boundary conditions, the corresponding operator is a boundedly invertible linear mapping between X and $Y' \times L_2(\Omega)$, where $X := L_2(I; H^1_0(\Omega)) \cap H^1(I; H^{-1}(\Omega))$ and $Y := L_2(I; H^1_0(\Omega))$. As a consequence of the non-coercivity, it requires a careful selection of the test space to arrive at a stable Petrov–Galerkin system whose solution is a quasi-best approximation from the trial space. To relax the conditions on the test space, a minimal residual Petrov–Galerkin discretization was introduced in [1]. It has an equivalent interpretation as a Galerkin discretization of an extended self-adjoint, indefinite mixed

Keywords and phrases. Parabolic PDEs, boundary conditions, space-time FOSLS, convergence of adaptive algorithm.

Korteweg-de Vries (KdV) Institute for Mathematics, University of Amsterdam, P.O. Box 94248, 1090 GE Amsterdam, The Netherlands.
*Corresponding author: g.gantner@uva.nl

system, with the Riesz lift of the residual of the primal variable from the "trial space" being an additional variable from the "test space". In [38], uniform inf-sup stability was demonstrated for both trial and test space being finite element spaces of comparable dimensions, w.r.t. general partitions of the space-time cylinder into prismatic elements, which however must be decomposable into "time-slabs". The latter means that a possibly non-uniform partition of the time interval must be global in space, which does not align with the aim to permit fully-flexible local refinements in space and time.

In the recent work [18] by Führer and Karkulik, for the aforementioned heat equation with forcing term $f \in L_2(I\times\Omega)$ and initial condition $u_0 \in L_2(\Omega)$, it was proven that with $\widetilde{U}_0 := \{\mathbf{u} \in X \times L_2(I\times\Omega)^d\colon \operatorname{div}\mathbf{u} \in L_2(I\times\Omega)\}$ equipped with the graph norm,

$$\operatorname*{argmin}_{\mathbf{u}=(u_1,\mathbf{u}_2)\in\widetilde{U}_0} \|\mathbf{u}_2 + \nabla_{\mathbf{x}} u_1\|^2_{L_2(I\times\Omega)^d} + \|\operatorname{div}\mathbf{u}_2 - f\|^2_{L_2(I\times\Omega)} + \|u(0,\cdot)-u_0\|^2_{L_2(\Omega)}$$

is a well-posed First-Order System Least-Squares (FOSLS) formulation for the pair of the solution $u = u_1$ and (minus) its spatial gradient $-\nabla_{\mathbf{x}} u = \mathbf{u}_2$. This formulation can already be found in [4] without a proof of its well-posedness though.

The FOSLS formulation from [18] has major advantages. The Euler–Lagrange equations resulting from the minimization problem correspond to a symmetric, coercive bilinear form on $\widetilde{U}_0 \times \widetilde{U}_0$, so that the Galerkin approximation from *any* conforming trial space is a quasi-best approximation from that space. In other words, there are no issues with stability or restrictions on the partitions of the space-time cylinder underlying the finite element spaces. The minimization is w.r.t. L_2-norms, so that the arising stiffness matrix is computable and sparse and can be easily computed. The least-squares functional provides an *a posteriori* estimator that is equivalent to the norm on $\widetilde{U}_0$ of the error. The squared estimator is a sum of squared local error indicators associated to the individual elements, which immediately suggests an adaptive solution method.

Considering general least-squares methods, we mention that although a least-squares estimator is efficient and reliable, and the resulting adaptive routine is generally observed to converge, even with an optimal rate, a proof of (Q-linear) convergence of such an adaptive routine has only been given for a FOSLS formulation of Poisson's equation with Dörfler marking for a bulk parameter that is sufficiently close to 1, see [13].

A disadvantage of the FOSLS method from [18] is that the graph norm on $\widetilde{U}_0$ for the error in the pair $(u, -\nabla_{\mathbf{x}} u)$ is considerably stronger than the X-norm for the error in u. This appears from the low convergence rates reported in [18] for the adaptive routine with standard Lagrange finite element spaces applied to non-smooth solutions, *e.g.*, as those that result from a discontinuity in the transition of initial and boundary data. Furthermore, as far as we know, an open problem is the development of optimal preconditioners for the space $\widetilde{U}_0$, which is an important issue in view of the fact that with space-time methods, a PDE posed on a $(d+1)$-dimensional domain has to be solved.

In the current work, we contribute to a further development of the FOSLS method from [18]. In particular,

- we show that $\widetilde{U}_0$ is isomorphic to $U_0 := \{\mathbf{u} \in L_2(I; H^1_0(\Omega)) \times L_2(I\times\Omega)^d\colon \operatorname{div}\mathbf{u} \in L_2(I\times\Omega)\}$ equipped with the graph norm (Prop. 2.1), which circumvents the dual norm incorporated in the definition of X. It is a key ingredient in the derivation of most of the other results from this work;
- we show that the FOSLS method applies to general parabolic equations of second order with homogeneous Dirichlet, homogeneous Neumann, or mixed homogeneous Dirichlet and Neumann boundary conditions (Thm. 2.3 and Prop. 2.5);
- we extend the FOSLS method to forcing functions $f \notin L_2(I\times\Omega)$ (Prop. 2.5);
- by appending an additional term to the least-squares functional measuring the squared error in the boundary data, we extend the FOSLS method to inhomogeneous Dirichlet (Thm. 2.8) or Neumann data (Thm. 2.9), where, however, the norms in which these errors are measured are not of L_2-type;
- finally, using the framework developed by Siebert [32], which particularly allows for relatively general marking strategies (Rem. 3.2), we prove plain convergence (Thm. 3.3) of the adaptive FOSLS method (Alg. 3.1) for homogeneous Dirichlet boundary conditions driven by the least-squares estimator. This convergence proof

generalizes to a large class of least-squares formulations (Rem. 3.7), including, *e.g.*, the aforementioned FOSLS formulation of the Poisson model problem. Independently, Führer and Praetorius [19] have recently used a similar proof idea to derive convergence of various least-squares formulations, excluding however the considered space-time FOSLS.

The remainder of the current section fixes some notation (Sect. 1.1), recalls abstract parabolic evolution equations (Sect. 1.2), and introduces the particular instance of parabolic PDEs of second order (Sect. 1.3) that will be considered throughout the manuscript.

1.1. Notation

In this work, by $C \lesssim D$ we will mean that C can be bounded by a multiple of D, independently of parameters on which C and D may depend. Obviously, $C \gtrsim D$ is defined as $D \lesssim C$, and $C \eqsim D$ as $C \lesssim D$ and $C \gtrsim D$.

For normed linear spaces E and F, we will denote by $\mathcal{L}(E,F)$ the normed linear space of bounded linear mappings $E \to F$, and by $\mathcal{L}\mathrm{is}(E,F)$ its subset of boundedly invertible linear mappings $E \to F$. We write $E \hookrightarrow F$ to denote that E is continuously embedded into F. For simplicity only, we exclusively consider linear spaces over the scalar field $\mathbb{R}$.

For a Hilbert space W that is densely and continuously embedded in a space of type $L_2(\Sigma)$, we mostly use the scalar product on $L_2(\Sigma)$ to denote its unique extension to the duality pairing on $W' \times W$.

1.2. Abstract parabolic evolution equation

Let V and H be separable Hilbert spaces such that $V \hookrightarrow H$ with dense and compact embedding. Identifying H with its dual, we obtain the Gelfand triple $V \hookrightarrow H \eqsim H' \hookrightarrow V'$. For almost all $t \in I := (0,T)$, let $a(t;\cdot,\cdot)$ be a bilinear form on $V \times V$ such that for any $\mu, \lambda \in V$, $t \mapsto a(t;\mu,\lambda)$ is measurable on I, and such that for some constant $\varrho \geq 0$, for *a.e.* $t \in I$, and all μ, λ,

$$\begin{aligned} |a(t;\mu,\lambda)| &\lesssim \|\mu\|_V \|\lambda\|_V && \text{(boundedness)},\\ a(t;\mu,\mu) + \varrho\|\mu\|^2 &\gtrsim \|\mu\|_V^2 && \text{(Gårding inequality)}. \end{aligned}$$

With $(A(t)\cdot)(\cdot) := a(t;\cdot,\cdot)$, we consider the *parabolic initial value problem* of finding $u\colon I \to V$ such that

$$\begin{cases} \frac{\mathrm{d}u}{\mathrm{d}t}(t) + A(t)u(t) = g(t) & \text{for } a.e.\ t \in I,\\ u(0) = u_0. \end{cases}$$

A proof of the following result is found in [29], see also Chapter IV, Section 26 of [42] and Chapter XVIII, Section 3 of [15].

Theorem 1.1. *With* $X := L_2(I;V) \cap H^1(I;V')$, $Y := L_2(I;V)$,

$$(Bu)(v) := \int_I (\partial_t u(t,\cdot))(v(t,\cdot)) + a(t;u(t),v(t))\mathrm{d}t,$$

and $\gamma_0 := u \mapsto u|_{t=0}$, *it holds that*

$$\begin{bmatrix} B \\ \gamma_0 \end{bmatrix} \in \mathcal{L}\mathrm{is}\big(X, (Y \times H)'\big),$$

with upper bounds for the norm of the operator and that of its inverse only dependent on upper bounds for the boundedness constant, the reciprocal of the constant in the Gårding inequality, and ϱ.

So for $(g,u_0) \in Y' \times H$, a well-posed variational formulation of the parabolic problem reads as finding $u \in X$ such that $(Bu, \gamma_0 u) = (g,u_0)$.

1.3. Parabolic equations of second order

For a bounded Lipschitz domain $\Omega \subset \mathbb{R}^d$ with outer normal $\mathbf{n}_\mathbf{x} \in \mathbb{R}^d$, relatively open subsets Γ_D and Γ_N of $\partial\Omega$ with $\Gamma_D \cap \Gamma_N = \emptyset$ and $\overline{\Gamma_D \cup \Gamma_N} = \partial\Omega$, $\mathbf{b} \in L_\infty(I \times \Omega)^d$, $c \in L_\infty(I \times \Omega)$, and $\mathbf{A} = \mathbf{A}^\top \in L_\infty(I \times \Omega)^{d\times d}$ uniformly positive definite, we consider the problem of finding $u\colon I \times \Omega \to \mathbb{R}$ that for given data f, ϕ, u_D, and u_0 satisfies

$$\begin{cases} \partial_t u - \operatorname{div}_\mathbf{x} \mathbf{A}\nabla_\mathbf{x} u + \mathbf{b}\cdot\nabla_\mathbf{x} u + cu = f & \text{on } I\times\Omega,\\ (\mathbf{A}\nabla_\mathbf{x} u)\cdot \mathbf{n}_\mathbf{x} = \phi & \text{on } I\times\Gamma_N,\\ u = u_D & \text{on } I\times\Gamma_D,\\ u(0,\cdot) = u_0 & \text{on } \Omega. \end{cases} \tag{1.1}$$

Taking until Section 2.2 a homogeneous Dirichlet datum $u_D = 0$, a variational formulation of (1.1) leads to a problem as in Theorem 1.1, where $V := H^1_D(\Omega) = \{u \in H^1(\Omega)\colon u|_{\Gamma_D} = 0\}$ and $H := L_2(\Omega)$, so that

$$X = L_2\left(I; H^1_D(\Omega)\right) \cap H^1\left(I; H^1_D(\Omega)'\right), \qquad Y = L_2\left(I; H^1_D(\Omega)\right),$$

the bilinear form reads as

$$a(t;\mu,\lambda) := \int_\Omega \mathbf{A}(t,\mathbf{x})\nabla\mu(\mathbf{x})\cdot\nabla\lambda(\mathbf{x}) + (\mathbf{b}(t,\mathbf{x})\cdot\nabla\mu(\mathbf{x}) + c(t,\mathbf{x})\mu(\mathbf{x}))\lambda(\mathbf{x})\,\mathrm{d}\mathbf{x},$$

and the forcing term reads as

$$g(v) := \int_{I\times\Omega} fv\,\mathrm{d}\mathbf{x}\,\mathrm{d}t + \int_{I\times\Gamma_N} \phi v\,\mathrm{d}\mathbf{s}. \tag{1.2}$$

As follows from Theorem 1.1, this variational problem is actually well-posed for *any* $g \in Y'$. For a discussion in which sense the solution of the variational problem can be interpreted as a solution of (1.1), we refer to [15], pages 524–528.

Concerning the bilinear form a, both its boundedness constant, the reciprocal of the constant in the Gårding inequality, and ϱ can be bounded in terms of upper bounds for $\|\mathbf{b}\|_{L_\infty(I\times\Omega)^d}$, $\|c\|_{L_\infty(I\times\Omega)}$, $\|\mathbf{A}\|_{L_\infty(I\times\Omega)^{d\times d}}$, and $\|\mathbf{A}^{-1}\|_{L_\infty(I\times\Omega)^{d\times d}}$.

2. Formulation as a first-order system

2.1. Homogeneous boundary conditions

For the case that $g \in L_2(I\times\Omega)$, we will derive a system for $\mathbf{u} = (u_1, \mathbf{u}_2) = (u, -\mathbf{A}\nabla_\mathbf{x} u)$ with u being the solution of the variational problem $(Bu, \gamma_0 u) = (g, u_0)$ from Section 1.3. Recall that such a problem arises from (1.1) when besides $u_D = 0$, it holds that $f \in L_2(I\times\Omega)$ and $\phi = 0$. Generally at the expense of having to solve an additional (elliptic) PDE, general $g \in Y'$ (*i.e.*, $f \notin L_2(I\times\Omega)$ and/or Neumann datum $\phi \neq 0$) will be handled as well.

Let

$$U := \{\mathbf{u} = (u_1, \mathbf{u}_2) \in L_2(I; H^1(\Omega)) \times L_2(I\times\Omega)^d\colon \operatorname{div}\mathbf{u} \in L_2(I\times\Omega)\}$$

equipped with graph norm

$$\|\mathbf{u}\|_U^2 := \|u_1\|^2_{L_2(I;H^1(\Omega))} + \|\mathbf{u}_2\|^2_{L_2(I;L_2(\Omega)^d)} + \|\operatorname{div}\mathbf{u}\|^2_{L_2(I\times\Omega)}. \tag{2.1}$$

Knowing that $\operatorname{div}\colon L_2(I\times\Omega)^{d+1} \supset \operatorname{dom}(\operatorname{div}) \to L_2(I\times\Omega)$ is a closed linear operator (a necessary condition for $H(\operatorname{div}; I\times\Omega)$ being a Hilbert space), from $L_2(I; H^1(\Omega)) \times L_2(I\times\Omega)^d \hookrightarrow L_2(I\times\Omega)^{d+1}$, it follows that $\operatorname{div}\colon L_2(I; H^1(\Omega)) \times L_2(I\times\Omega)^d \supset \operatorname{dom}(\operatorname{div}) \to L_2(I\times\Omega)$ is a closed linear operator. Together with the facts that $L_2(I; H^1(\Omega)) \times L_2(I\times\Omega)^d$ and $L_2(I\times\Omega)$ are Hilbert spaces, this shows that U is a Hilbert space.

With $\mathbf{n} = (n_t, \mathbf{n}_\mathbf{x})$ denoting the outer normal vector on the boundary of $I \times \Omega$, using that $\mathbf{u} \mapsto \mathbf{u}|_{I\times\Gamma_N} \cdot \mathbf{n} \in \mathcal{L}\big(H(\mathrm{div}; I \times \Omega), H^{\frac{1}{2}}_{00}(I \times \Gamma_N)'\big)$ we define the closed subspace U_0 of U by

$$U_0 := \{\mathbf{u} \in L_2(I; H^1_D(\Omega)) \times L_2(I \times \Omega)^d \colon \mathrm{div}\, \mathbf{u} \in L_2(I \times \Omega),\ \mathbf{u}|_{I\times\Gamma_N} \cdot \mathbf{n} = 0\}.$$

We start with showing that U_0 is isomorphic to a seemingly smaller space that was employed in [18].

Proposition 2.1. *It holds that*

$$U_0 \eqsim \widetilde{U}_0 := \{\mathbf{u} \in X \times L_2(I \times \Omega)^d \colon \mathrm{div}\, \mathbf{u} \in L_2(I \times \Omega),\ \mathbf{u}|_{I\times\Gamma_N} \cdot \mathbf{n} = 0\},$$

equipped with the graph norm

$$\|\mathbf{u}\|^2_{\widetilde{U}_0} := \|u_1\|^2_{L_2(I;H^1(\Omega))} + \|\partial_t u_1\|^2_{L_2(I;H^1_D(\Omega)')} + \|\mathbf{u}_2\|^2_{L_2(I;L_2(\Omega)^d)} + \|\mathrm{div}\, \mathbf{u}\|^2_{L_2(I\times\Omega)}.$$

This proposition is a direct consequence of the following lemma.

Lemma 2.2. *For* $\mathbf{u} \in H_{0,I\times\Gamma_N}(\mathrm{div}; I \times \Omega) := \{\mathbf{u} \in H(\mathrm{div}; I \times \Omega) \colon \mathbf{u}|_{I\times\Gamma_N} \cdot \mathbf{n} = 0\}$, *it holds that* $\partial_t u_1 \in L_2(I; H^1_D(\Omega)')$ *with*

$$\|\partial_t u_1\|_{L_2(I;H^1_D(\Omega)')} \le \sqrt{2}\|\mathbf{u}\|_{H(\mathrm{div};I\times\Omega)}.$$

Proof. For smooth $\mathbf{u} \in H_{0,I\times\Gamma_N}(\mathrm{div}; I \times \Omega)$ (for which $\mathbf{u} \cdot \mathbf{n}$ is defined in the classical pointwise sense), we have $\mathrm{div}\, \mathbf{u} = \partial_t u_1 + \mathrm{div}_\mathbf{x}\, \mathbf{u}_2$. For smooth $v \in L_2(I; H^1_D(\Omega))$, we have

$$\begin{aligned}\int_{I\times\Omega} \mathbf{u}_2 \cdot \nabla_\mathbf{x} v \,\mathrm{d}\mathbf{x}\,\mathrm{d}t &= -\int_{I\times\Omega} v\, \mathrm{div}_\mathbf{x}\, \mathbf{u}_2 \,\mathrm{d}\mathbf{x}\,\mathrm{d}t + \int_{I\times\Gamma_N} \mathbf{u}_2 \cdot \mathbf{n}_\mathbf{x}\, v \,\mathrm{d}s \\ &= -\int_{I\times\Omega} v\, \mathrm{div}_\mathbf{x}\, \mathbf{u}_2 \,\mathrm{d}\mathbf{x}\,\mathrm{d}t + \int_{I\times\Gamma_N} \mathbf{u} \cdot \mathbf{n}\, v \,\mathrm{d}s \\ &= -\int_{I\times\Omega} v\, \mathrm{div}_\mathbf{x}\, \mathbf{u}_2 \,\mathrm{d}\mathbf{x}\,\mathrm{d}t.\end{aligned}$$

Since the set of such v is dense in $L_2(I; H^1_D(\Omega))$, we conclude

$$\begin{aligned}\|\partial_t u_1\|_{L_2(I;H^1_D(\Omega)')} &\le \|\mathrm{div}\, \mathbf{u}\|_{L_2(I;H^1_D(\Omega)')} + \|\mathrm{div}_\mathbf{x}\, \mathbf{u_2}\|_{L_2(I;H^1_D(\Omega)')} \\ &\le \|\mathrm{div}\, \mathbf{u}\|_{L_2(I\times\Omega)} + \|\mathbf{u}_2\|_{L_2(I\times\Omega)^d} \le \sqrt{(2)}\|\mathbf{u}\|_{H(\mathrm{div};I\times\Omega)}.\end{aligned}$$

Since the set of such $\mathbf{u}$ is dense in $H_{0,I\times\Gamma_N}(\mathrm{div}; I \times \Omega)$, the proof is completed. For $\partial(I \times \Omega)$ instead of $I \times \Gamma_N$, the corresponding density result is well-known. The proof of Theorem 2.6 from [21] easily generalizes to $I \times \Gamma_N$ using that the term $l_{d+2} \in H^1(I \times \Omega)$ from there additionally satisfies that $l_{d+2}|_{\partial(I\times\Omega)\setminus\overline{I\times\Gamma_N}} = 0$ as $l_{d+2}|_{\partial(I\times\Omega)}$ is orthogonal to $\mathbf{u} \cdot \mathbf{n}$ for all smooth $\mathbf{u} \in H_{0,I\times\Gamma_N}(\mathrm{div}; I \times \Omega)$. □

The following theorem generalizes [18], see Remark 2.6 for a discussion.

Theorem 2.3 (Homogeneous Dirichlet)**.** *It holds that*

$$\begin{aligned}G\colon (u_1, \mathbf{u}_2) \mapsto (\mathbf{u}_2 + \mathbf{A}\nabla_\mathbf{x} u_1, \mathrm{div}\, \mathbf{u} - \mathbf{b} \cdot \mathbf{A}^{-1}\mathbf{u}_2 + c u_1, u_1(0, \cdot)) \\ \in \mathcal{L}\mathrm{is}(U_0, L_2(I \times \Omega)^d \times L_2(I \times \Omega) \times L_2(\Omega)).\end{aligned}$$

Remark 2.4. Analogously, one can prove the same result for $(u_1, \mathbf{u}_2) \mapsto (\mathbf{u}_2 + \mathbf{A}\nabla_\mathbf{x} u_1, \mathrm{div}\, \mathbf{u} + \mathbf{b} \cdot \nabla_\mathbf{x} u_1 + c u_1, u_1(0, \cdot))$.

Proof. Boundedness of G follows from the definition of U_0, and the fact that $X \hookrightarrow C(\bar{I}; L_2(\Omega))$ ([24], Chap. 1, Thm. 3.1) in combination with Proposition 2.1.

As we have seen in the proof of Lemma 2.2, for $\mathbf{u} \in U_0$ and $v \in L_2(I; H^1_D(\Omega))$, it holds that $(-\nabla'_{\mathbf{x}}\mathbf{u}_2)(v) = \int_{I\times\Omega} v \operatorname{div}_{\mathbf{x}} \mathbf{u}_2 \,\mathrm{d}\mathbf{x}\,\mathrm{d}t$. From Theorem 1.1 we infer that

$$\|u_1\|_{L_2(I;H^1(\Omega))} \le \|u_1\|_X \lesssim \|Bu_1\|_{L_2(I;H^1_D(\Omega)')} + \|u_1(0,\cdot)\|_{L_2(\Omega)},$$

where

$$\begin{aligned}
\|Bu_1\|_{L_2(I;H^1_D(\Omega)')} &= \|\partial_t u_1 + \nabla'_{\mathbf{x}}\mathbf{A}\nabla_{\mathbf{x}} u_1 + \mathbf{b}\cdot\nabla_{\mathbf{x}} u_1 + c u_1\|_{L_2(I;H^1_D(\Omega)')} \\
&\le \|\partial_t u_1 - \nabla'_{\mathbf{x}}\mathbf{u}_2 + \mathbf{b}\cdot\nabla_{\mathbf{x}} u_1 + c u_1\|_{L_2(I;H^1_D(\Omega)')} + \|\nabla'_{\mathbf{x}}(\mathbf{u}_2 + \mathbf{A}\nabla_{\mathbf{x}} u_1)\|_{L_2(I;H^1_D(\Omega)')} \\
&\lesssim \|\operatorname{div}\mathbf{u} + \mathbf{b}\cdot\nabla_{\mathbf{x}} u_1 + c u_1\|_{L_2(I\times\Omega)} + \|\mathbf{u}_2 + \mathbf{A}\nabla_{\mathbf{x}} u_1\|_{L_2(I\times\Omega)^d} \\
&\lesssim \|\operatorname{div}\mathbf{u} - \mathbf{b}\cdot\mathbf{A}^{-1}\mathbf{u}_2 + c u_1\|_{L_2(I\times\Omega)} + \|\mathbf{u}_2 + \mathbf{A}\nabla_{\mathbf{x}} u_1\|_{L_2(I\times\Omega)^d}.
\end{aligned}$$

From

$$\begin{aligned}
\|\mathbf{u}_2\|_{L_2(I\times\Omega)^d} &\le \|\mathbf{u}_2 + \mathbf{A}\nabla_{\mathbf{x}} u_1\|_{L_2(I\times\Omega)^d} + \|\mathbf{A}\nabla_{\mathbf{x}} u_1\|_{L_2(I\times\Omega)^d} \\
&\lesssim \|\mathbf{u}_2 + \mathbf{A}\nabla_{\mathbf{x}} u_1\|_{L_2(I\times\Omega)^d} + \|u_1\|_{L_2(I;H^1(\Omega))},
\end{aligned}$$

and

$$\|\operatorname{div}\mathbf{u}\|_{L_2(I\times\Omega)} \lesssim \|\operatorname{div}\mathbf{u} - \mathbf{b}\cdot\mathbf{A}^{-1}\mathbf{u}_2 + c u_1\|_{L_2(I\times\Omega)} + \|\mathbf{u}\|_{L_2(I\times\Omega)^{d+1}},$$

we conclude that $\|\mathbf{u}\|_U \lesssim \|G\mathbf{u}\|_{L_2(I\times\Omega)^d\times L_2(I\times\Omega)\times L_2(\Omega)}$, and thus in particular that G is injective.

Given $(\mathbf{q}, h, u_0) \in L_2(I\times\Omega)^d \times L_2(I\times\Omega) \times L_2(\Omega)$, let $u_1 \in X$ be the solution of

$$\begin{bmatrix} B \\ \gamma_0 \end{bmatrix} u_1 = \begin{bmatrix} v \mapsto \int_{I\times\Omega} (h + \mathbf{b}\cdot\mathbf{A}^{-1}\mathbf{q}) v + \mathbf{q}\cdot\nabla_{\mathbf{x}} v \,\mathrm{d}\mathbf{x}\,\mathrm{d}t \\ u_0 \end{bmatrix} \in L_2(I; H^1_D(\Omega)') \times L_2(\Omega),$$

so that for $v \in L_2(I; H^1_D(\Omega))$

$$\begin{aligned}
\int_{I\times\Omega} \partial_t u_1\, v + \mathbf{A}\nabla_{\mathbf{x}} u_1 \cdot \nabla_{\mathbf{x}} v + \mathbf{b}\cdot\nabla_{\mathbf{x}} u_1\, v + c\, u_1\, v \,\mathrm{d}\mathbf{x}\,\mathrm{d}t \\
= \int_{I\times\Omega} (h + \mathbf{b}\cdot\mathbf{A}^{-1}\mathbf{q}) v + \mathbf{q}\cdot\nabla_{\mathbf{x}} v \,\mathrm{d}\mathbf{x}\,\mathrm{d}t,
\end{aligned}$$

and thus for $\mathbf{u}_2 := \mathbf{q} - \mathbf{A}\nabla_{\mathbf{x}} u_1 \in L_2(I\times\Omega)^d$

$$\int_{I\times\Omega} \partial_t u_1\, v - \mathbf{u}_2\cdot\nabla_{\mathbf{x}} v \,\mathrm{d}\mathbf{x}\,\mathrm{d}t = \int_{I\times\Omega} \underbrace{(h + \mathbf{b}\cdot\mathbf{A}^{-1}\mathbf{u}_2 - c u_1)}_{=:\tilde{h}\in L_2(I\times\Omega)} v \,\mathrm{d}\mathbf{x}\,\mathrm{d}t.$$

For $v \in H^1(I\times\Omega)$ that vanish at $\partial(I\times\Omega)\backslash\overline{I\times\Gamma_N}$, one has $\int_{I\times\Omega} \partial_t u_1\, v \,\mathrm{d}\mathbf{x}\,\mathrm{d}t = -\int_{I\times\Omega} u_1 \partial_t v \,\mathrm{d}\mathbf{x}\,\mathrm{d}t$, and therefore $\int_{I\times\Omega} \partial_t u_1\, v - \mathbf{u}_2\cdot\nabla_{\mathbf{x}} v \,\mathrm{d}\mathbf{x}\,\mathrm{d}t = -\int_{I\times\Omega} \mathbf{u}\cdot\nabla v \,\mathrm{d}\mathbf{x}\,\mathrm{d}t$, which shows $\operatorname{div}\mathbf{u} = \tilde{h}$. Moreover, for such v, it holds that

$$\begin{aligned}
\int_{I\times\Gamma_N} \mathbf{u}\cdot\mathbf{n}\, v \,\mathrm{d}s = \int_{I\times\Omega} v \operatorname{div}\mathbf{u} + \mathbf{u}\cdot\nabla v \,\mathrm{d}\mathbf{x}\,\mathrm{d}t = \int_{I\times\Omega} v\tilde{h} + u_1\partial_t v + \mathbf{u}_2\cdot\nabla_{\mathbf{x}} v \,\mathrm{d}\mathbf{x}\,\mathrm{d}t \\
= \int_{I\times\Omega} v\tilde{h} - (\partial_t u_1\, v - \mathbf{u}_2\cdot\nabla_{\mathbf{x}} v) \,\mathrm{d}\mathbf{x}\,\mathrm{d}t = 0,
\end{aligned}$$

which proves that $\mathbf{u}|_{I\times\Gamma_N}\cdot\mathbf{n} = 0$, and so $\mathbf{u} \in U_0$. We conclude that $G\mathbf{u} = (\mathbf{q}, h, u_0)$, *i.e.*, G is surjective, which completes the proof. □

Next, using Theorem 2.3, we show that the well-posed standard variational formulation of the parabolic problem discussed in Sections 1.2 and 1.3, thus with homogeneous Dirichlet datum $u_D = 0$, has an equivalent formulation as a well-posed first-order system. As a preparation, we note that any forcing term $g \in L_2(I; H^1_D(\Omega)')$ can (non-uniquely) be written in the form

$$g(v) = \int_{I\times\Omega} g_1 v + \mathbf{g}_2 \cdot \nabla_{\mathbf{x}} v \,\mathrm{d}\mathbf{x}\,\mathrm{d}t \quad \text{for all } v \in L_2(I; H^1_D(\Omega)), \tag{2.2}$$

for some $g_1 \in L_2(I; L_2(\Omega))$ and $\mathbf{g}_2 \in L_2(I; L_2(\Omega)^d)$. Take, *e.g.*, $g_1 = w$ and $\mathbf{g}_2 = \nabla_{\mathbf{x}} w$ with $w \in L_2(I; H^1_D(\Omega))$ being the Riesz lift of g defined by

$$\int_{I\times\Omega} wv + \nabla_{\mathbf{x}} w \cdot \nabla_{\mathbf{x}} v \,\mathrm{d}\mathbf{x}\,\mathrm{d}t = g(v) \quad \text{for all } v \in L_2(I; H^1_D(\Omega)). \tag{2.3}$$

Proposition 2.5. *With a splitting of $g \in L_2(I; H^1_D(\Omega)')$ as in (2.2), where $(g_1, \mathbf{g}_2) \in L_2(I; L_2(\Omega)) \times L_2(I; L_2(\Omega)^d)$, and $u_0 \in L_2(\Omega)$, it holds that $u_1 \in X = L_2(I; H^1_D(\Omega)) \cap H^1(I; H^1_D(\Omega)')$ solves $(Bu_1, \gamma_0 u_1) = (g, u_0)$ and $\mathbf{u}_2 = -\mathbf{A}\nabla_{\mathbf{x}} u_1 + \mathbf{g}_2$ if and only if $\mathbf{u} = (u_1, \mathbf{u}_2) \in U_0$ solves*

$$G\mathbf{u} = \left(\mathbf{g}_2, g_1 - \mathbf{b}\cdot\mathbf{A}^{-1}\mathbf{g}_2, u_0\right).$$

Proof. With $\mathbf{u}_2 = -\mathbf{A}\nabla_{\mathbf{x}} u_1 + \mathbf{g}_2$, *i.e.*, $(G\mathbf{u})_1 = \mathbf{g}_2$, the equation $Bu_1 = g$, *i.e.*,

$$\int_{I\times\Omega} (\partial_t u_1 + \mathbf{b}\cdot\nabla_{\mathbf{x}} u_1 + cu_1)v + \mathbf{A}\nabla_{\mathbf{x}} u_1 \cdot \nabla_{\mathbf{x}} v \,\mathrm{d}\mathbf{x}\,\mathrm{d}t = g(v) \quad \left(v \in L_2(I; H^1_D(\Omega))\right),$$

is equivalent to

$$\int_{I\times\Omega} \partial_t u_1 v - \mathbf{u}_2 \cdot \nabla_{\mathbf{x}} v \,\mathrm{d}\mathbf{x}\,\mathrm{d}t = \int_{I\times\Omega} \underbrace{(\mathbf{b}\cdot\mathbf{A}^{-1}(\mathbf{u}_2 - \mathbf{g}_2) - cu_1 + g_1)}_{=:\tilde{g}\in L_2(I\times\Omega)} v \,\mathrm{d}\mathbf{x}\,\mathrm{d}t \,(v \in L_2(I; H^1_D(\Omega))). \tag{2.4}$$

As we have seen in the last paragraph of the proof of Theorem 2.3, (2.4) implies $\operatorname{div}\mathbf{u} = \tilde{g}$, *i.e.*, $(G\mathbf{u})_2 = g_1 - \mathbf{b}\cdot\mathbf{A}^{-1}\mathbf{g}_2$, and $\mathbf{u}|_{I\times\Gamma_N} = 0$.

Conversely, let $\mathbf{u} \in U_0$ satisfy $G\mathbf{u} = (\mathbf{g}_2, g_1 - \mathbf{b}\cdot\mathbf{A}^{-1}\mathbf{g}_2, u_0)$. Then, Proposition 2.1 shows that $u_1 \in X$. Since $\operatorname{div}\mathbf{u} = \tilde{g}$, it remains to show that

$$\int_{I\times\Omega} \partial_t u_1 v - \mathbf{u}_2 \cdot \nabla_{\mathbf{x}} v \,\mathrm{d}\mathbf{x}\,\mathrm{d}t = \int_{I\times\Omega} v \operatorname{div}\mathbf{u} \,\mathrm{d}\mathbf{x}\,\mathrm{d}t \quad (v \in L_2(I; H^1_D(\Omega))).$$

The latter relation is already valid for arbitrary $\mathbf{u} \in H_{0,I\times\Gamma_N}(\operatorname{div}; I\times\Omega)$ and $v \in L_2(I; H^1_D(\Omega))$. Indeed, for smooth $\mathbf{u}$ and v in these spaces, it follows by integration by parts, and so by using Lemma 2.2, it follows by the density of the sets of those functions in these spaces. □

When $f \in L_2(I\times\Omega)$ and $\phi = 0$ in (1.1), one has $g = f \in L_2(I\times\Omega)$ and one obviously takes $(g_1, \mathbf{g}_2) = (g, 0)$ in the previous proposition. For $g \in L_2(I; H^1_D(\Omega)') \setminus L_2(I\times\Omega)$ (*i.e.*, $f \notin L_2(I\times\Omega)$ and/or $\phi \neq 0$) generally the splitting of g requires solving (2.3). For the case that $\Gamma_N = \partial\Omega$, an alternative approach for inhomogeneous Neumann datum $\phi \neq 0$ will be presented in Theorem 2.9.

Remark 2.6. Theorem 2.3 extends the crucial result from [18]. For the case that $\mathbf{A} = \mathrm{Id}$, $\mathbf{b} = 0 = c$, and $\Gamma_D = \partial\Omega$, there it was shown that the harmlessly different operator $\widetilde{G}\colon \mathbf{u} \mapsto G(u, -\mathbf{u}_2)\colon \widetilde{U}_0 \mapsto L_2(I; L_2(\Omega)^d) \times L_2(I; L_2(\Omega)) \times L_2(\Omega)$ is in $\mathcal{L}\mathrm{is}(\widetilde{U}_0, \operatorname{ran}\widetilde{G})$, and that $\operatorname{ran}\widetilde{G} \supseteq \{\mathbf{0}\} \times L_2(I; L_2(\Omega)) \times L_2(\Omega)$. We showed that G, and thus $\widetilde{G}$, is also surjective. Notice that for well-posedness of a least-squares formulation, this surjectivity is not required. Indeed, bounded invertibility of the operator between its domain and its range is equivalent to

boundedness and coercivity of the bilinear form corresponding to the Euler–Lagrange equations resulting from the least-squares functional.

Our motivation to replace $\mathbf{u}_2$ by $-\mathbf{u}_2$ is that $\partial_t u_1 + \mathrm{div}_{\mathbf{x}}\, \mathbf{u}_2$ is the divergence of the vector field $\mathbf{u}\colon I\times\Omega \to \mathbb{R}^{d+1}$. When imposing, as we do, that the latter divergence is in $L_2(I\times\Omega)$, we know that $\mathbf{u}$ has a normal trace at $\partial(I\times\Omega)$, which allowed an easy extension to homogeneous Neumann boundary conditions. Furthermore, in Proposition 2.1, we made the observation that $\widetilde{U}_0 \eqsim U_0$, which freed ourselves from the dual norm which is part of the definition of $\widetilde{U}_0$. This will also play an essential role in the proofs of Theorem 2.8 and 2.9 dealing with inhomogeneous boundary conditions, and that of Theorem 3.3 concerning plain convergence of a standard adaptive algorithm.

2.2. Inhomogeneous boundary conditions

We extend the first-order formulation to cover both inhomogeneous (pure) Dirichlet boundary conditions and inhomogeneous (pure) Neumann boundary conditions, the latter now without the need to compute a Riesz lift of the boundary datum.

The following lemma is essentially a slight generalization of Theorem 2.1 from [36]. Thinking of S as being a trace operator, it shows how to append (essential) inhomogeneous boundary conditions to an equation that is well-posed for the corresponding homogeneous boundary conditions.

Lemma 2.7. *Let $\mathcal{X}$ and $\mathcal{Y}_2$ be Banach spaces, and $\mathcal{Y}_1$ be a normed linear space. Let $S\in\mathcal{L}(\mathcal{X},\mathcal{Y}_2)$ be surjective, let $F\in\mathcal{L}(\mathcal{X},\mathcal{Y}_1)$ be such that with $\mathcal{X}_0 := \{x\in\mathcal{X}\colon Sx=0\}$, $F|_{\mathcal{X}_0}\in\mathcal{L}\mathrm{is}(\mathcal{X}_0,\mathcal{Y}_1)$. Then, $\begin{bmatrix}F\\S\end{bmatrix}\in\mathcal{L}\mathrm{is}\big(\mathcal{X},\mathcal{Y}_1\times\mathcal{Y}_2\big)$.*

Proof. Knowing that S maps the open unit ball of $\mathcal{X}$ onto an open neighborhood of $0\in\mathcal{Y}_2$ (according to the open mapping theorem), there exists a constant $r>0$ such that for any $y\in\mathcal{Y}_2$ there exists an $x\in\mathcal{X}$ with $Sx=y$ and $\|x\|_{\mathcal{X}}\le r\|y\|_{\mathcal{Y}_2}$. Denoting this mapping $y\mapsto x$ by E, from $\mathrm{ran}(\mathrm{Id}-ES)\subseteq\mathcal{X}_0$ we have for $x\in\mathcal{X}$

$$\begin{aligned}\|x\|_{\mathcal{X}} &\le \|ESx\|_{\mathcal{X}} + \|(\mathrm{Id}-ES)x\|_{\mathcal{X}} \lesssim \|Sx\|_{\mathcal{Y}_2} + \|F(\mathrm{Id}-ES)x\|_{\mathcal{Y}_1}\\ &\le \|Sx\|_{\mathcal{Y}_2} + \|Fx\|_{\mathcal{Y}_1} + \|FESx\|_{\mathcal{Y}_1} \lesssim \|Sx\|_{\mathcal{Y}_2} + \|Fx\|_{\mathcal{Y}_1} \lesssim \|x\|_{\mathcal{X}}.\end{aligned}$$

Given $(y_1,y_2)\in\mathcal{Y}_1\times\mathcal{Y}_2$, let $x_2\in\mathcal{X}$ be such that $Sx_2=y_2$, and $x_0\in\mathcal{X}_0$ be such that $Fx_0 = y_1 - Fx_2$. Then, $\begin{bmatrix}F\\S\end{bmatrix}(x_0+x_2) = \begin{bmatrix}y_1\\y_2\end{bmatrix}$ showing that $\begin{bmatrix}F\\S\end{bmatrix}$ is surjective, which completes the proof. □

In combination with Theorem 2.3, Lemma 2.7 allows to prove the following theorem for inhomogeneous pure Dirichlet boundary conditions.

Theorem 2.8 (Inhomogeneous (pure) Dirichlet)**.** *It holds that*

$$\begin{aligned}G_D\colon \mathbf{u}=(u_1,\mathbf{u}_2) &\mapsto (\mathbf{u}_2 + \mathbf{A}\nabla_{\mathbf{x}}u_1, \mathrm{div}\,\mathbf{u} - \mathbf{b}\cdot\mathbf{A}^{-1}\mathbf{u}_2 + cu_1, u_1(0,\cdot), u_1|_{I\times\partial\Omega})\\ &\in \mathcal{L}\mathrm{is}\left(U, L_2(I\times\Omega)^d \times L_2(I\times\Omega)\times L_2(\Omega)\times (L_2(I;H^{\frac12}(\partial\Omega))\cap H^{\frac14}(I;L_2(\partial\Omega)))\right).\end{aligned}$$

Proof. An application of Lemma 2.2 for $\Gamma_N=\emptyset$ shows that for $\mathbf{u}=(u_1,\mathbf{u}_2)\in U$,

$$\|u_1\|_{L_2(I;H^1(\Omega))\cap H^1(I;H^{-1}(\Omega))} \lesssim \|\mathbf{u}\|_U. \tag{2.5}$$

We will combine this observation with the fact that

$$L_2\left(I;H^1(\Omega)\right)\cap H^1\left(I;H^{-1}(\Omega)\right) \hookrightarrow C\left(\bar I;L_2(\Omega)\right)\cap H^{\frac12}\left(I;L_2(\Omega)\right), \tag{2.6}$$

which follows from $[H^{-1}(\Omega),H^1(\Omega)]_{\frac12} = L_2(\Omega)$, see, *e.g.*, [15], pages 480 and 494. As shown in Chapter 4, Theorem 2.1 of [25],

$$u_1\mapsto u_1|_{I\times\partial\Omega}\in\mathcal{L}\left(L_2(I;H^1(\Omega))\cap H^{\frac12}\left(I;L_2(\Omega)\right), L_2(I;H^{\frac12}(\partial\Omega))\cap H^{\frac14}\left(I;L_2(\partial\Omega)\right)\right).$$

Together with (2.5) and (2.6), it shows that G_D is bounded.

Since in the current case of $\Gamma_D = \partial\Omega$, we have $\{\mathbf{u} \in U : u_1|_{I\times\partial\Omega} = 0\} = U_0$, knowing the result of Theorem 2.3, Lemma 2.7 shows that the proof will be completed once we have shown that

$$U \to L_2(I; H^{\frac{1}{2}}(\partial\Omega)) \cap H^{\frac{1}{4}}(I; L_2(\partial\Omega)) : \mathbf{u} \mapsto u_D := u_1|_{I\times\partial\Omega} \text{ is surjective.} \tag{2.7}$$

As shown in Theorem 2.9 from [14], the mapping

$$\begin{aligned} u_1 \mapsto (h, u_D) &:= \left(v \mapsto \int_{I\times\Omega} \partial_t u_1\, v + \nabla_{\mathbf{x}} u_1 \cdot \nabla_{\mathbf{x}} v \,\mathrm{d}\mathbf{x}\,\mathrm{d}t, u_1|_{I\times\partial\Omega} \right) \\ &\in \mathcal{L}\mathrm{is}\Big(L_2(I; H^1(\Omega)) \cap H^{\frac{1}{2}}_{00,\{0\}}(I; L_2(\Omega)), \quad \big(L_2(I; H^1_0(\Omega)) \cap H^{\frac{1}{2}}_{00,\{T\}}(I; L_2(\Omega))\big)' \\ &\quad \times L_2(I; H^{\frac{1}{2}}(\partial\Omega)) \cap H^{\frac{1}{4}}(I; L_2(\partial\Omega))\Big), \end{aligned}$$

where, with $H^1_{0,\{0\}}(I) := \{w \in H^1(I) : w(0) = 0\}$, $H^{\frac{1}{2}}_{00,\{0\}}(I) := [L_2(I), H^1_{0,\{0\}}(I)]_{\frac{1}{2}}$, with a similar definition of $H^{\frac{1}{2}}_{00,\{T\}}(I)$. For given h and u_D, the corresponding u_1 is in $L_2(I; H^1(\Omega))$. Taking $h \in L_2(I; L_2(\Omega))$ (*e.g.*, $h = 0$) and $\mathbf{u}_2 = -\nabla_{\mathbf{x}} u_1 \in L_2(I\times\Omega)^d$, from $\int_{I\times\Omega} \partial_t u_1\, v - \mathbf{u}_2 \cdot \nabla_{\mathbf{x}} v \,\mathrm{d}\mathbf{x}\,\mathrm{d}t = \int_{I\times\Omega} hv \,\mathrm{d}\mathbf{x}\,\mathrm{d}t$ for $v \in \mathcal{D}(I\times\Omega) \subset L_2(I; H^1_0(\Omega)) \cap H^{\frac{1}{2}}_{00,\{T\}}(I; L_2(\Omega))$, it follows that $\operatorname{div}\mathbf{u} = h \in L_2(I\times\Omega)$, *i.e.*, (2.7) is valid. □

Using Theorem 2.8, we formulate the parabolic problem with inhomogeneous pure Dirichlet boundary conditions as a well-posed first-order system. Let $(g_1, \mathbf{g}_2) \in L_2(I; L_2(\Omega)) \times L_2(I; L_2(\Omega)^d)$, $u_0 \in L_2(\Omega)$, and $u_D \in L_2(I; H^{\frac{1}{2}}(\partial\Omega)) \cap H^{\frac{1}{4}}(I; L_2(\partial\Omega))$, and set $g := v \mapsto \int_{I\times\Omega} g_1 v + \mathbf{g}_2 \cdot \nabla_{\mathbf{x}} v \,\mathrm{d}\mathbf{x}\,\mathrm{d}t \in L_2(I; H^{-1}(\Omega))$. Then the solution $\mathbf{u} = (u_1, \mathbf{u}_2) \in U$ of

$$G_D\mathbf{u} = \big(\mathbf{g}_2, g_1 - \mathbf{b}\cdot\mathbf{A}^{-1}\mathbf{g}_2, u_0, u_D\big), \tag{2.8}$$

satisfies

$$(Bu_1, \gamma_0 u_1, u_1|_{I\times\partial\Omega}) = (g, u_0, u_D),$$

i.e., u_1 satisfies the parabolic PDE in standard variational form and both the initial and Dirichlet boundary condition. Indeed, knowing $\mathbf{u}_2 + \mathbf{A}\nabla_{\mathbf{x}} u_1 = \mathbf{g}_2$, the second equation in (2.8) is equivalent to $\int_{I\times\Omega}(\partial_t u_1 + \mathbf{b}\cdot\nabla_{\mathbf{x}} u_1 + cu_1)v + \mathbf{A}\nabla_{\mathbf{x}} u_1 \cdot \nabla_{\mathbf{x}} v \,\mathrm{d}\mathbf{x}\,\mathrm{d}t = g(v)$ for all $v \in L_2(I; H^1_0(\Omega))$.

Analogously to the case of inhomogeneous pure Dirichlet boundary conditions, the combination of Theorem 2.3 and Lemma 2.7 allows to prove the following theorem for inhomogeneous pure Neumann boundary conditions.

Theorem 2.9 (Inhomogeneous (pure) Neumann)**.** *It holds that*

$$\begin{aligned} G_N : \mathbf{u} = (u_1, \mathbf{u}_2) &\mapsto \big(\mathbf{u}_2 + \mathbf{A}\nabla_{\mathbf{x}} u_1, \operatorname{div}\mathbf{u} - \mathbf{b}\cdot\mathbf{A}^{-1}\mathbf{u}_2 + cu_1, u_1(0,\cdot), \mathbf{u}|_{I\times\partial\Omega}\cdot\mathbf{n}\big) \\ &\in \mathcal{L}\mathrm{is}\left(U, L_2(I\times\Omega)^d \times L_2(I\times\Omega) \times L_2(\Omega) \times \Big(L_2(I; H^{\frac{1}{2}}(\partial\Omega)) \cap H^{\frac{1}{4}}(I; L_2(\partial\Omega))\Big)'\right). \end{aligned}$$

Proof. Clearly, the first two components of G_N are continuous. Recall from (2.5) and (2.6) that also the third one is bounded, and that $\|u_1\|_{H^{\frac{1}{2}}(I;L_2(\Omega))} \lesssim \|\mathbf{u}\|_U$. To see boundedness of the fourth one, we first remark that for smooth $\mathbf{u}$ and v on $I\times\Omega$, integration by parts shows that

$$\int_{I\times\partial\Omega} \mathbf{u}\cdot\mathbf{n}\, v \,\mathrm{d}\mathbf{s} = \int_{I\times\Omega} \mathbf{u}_2 \cdot \nabla_{\mathbf{x}} v + \operatorname{div}\mathbf{u}\, v - \partial_t u_1 v \,\mathrm{d}\mathbf{x}\,\mathrm{d}t. \tag{2.9}$$

As we have seen in the proof of Theorem 2.8, $v \in L_2(I;H^{\frac12}(\partial\Omega)) \cap H^{\frac14}(I;L_2(\partial\Omega))$ has a bounded extension to a $v_1 \in L_2\big(I;H^1(\Omega)\big) \cap H^{\frac12}_{00,\{0\}}(I;L_2(\Omega))$. Equally well it has a bounded extension to a $v_2 \in L_2\big(I;H^1(\Omega)\big) \cap H^{\frac12}_{00,\{T\}}(I;L_2(\Omega))$. Taking a smooth $\chi\colon I \to [0,1]$ with $\chi \equiv 1$ in a neighborhood of 0 and $\chi \equiv 0$ in a neighborhood of T, and $v_3(t,x) := \chi(t)v_1(t,x) + (1-\chi(t))v_2(t,x)$, we obtain a bounded extension to a $v_3 \in L_2(I;H^1(\Omega)) \cap H^{\frac12}_{00}(I;L_2(\Omega))$, where $H^{\frac12}_{00}(I) := [L_2(I), H^1_0(I)]_{\frac12}$. Given such an extension of $v \in L_2(I;H^{\frac12}(\partial\Omega)) \cap H^{\frac14}(I;L_2(\partial\Omega))$, for $\mathbf{u} \in U$ the right-hand side of (2.9) can be bounded by a multiple of $\|\mathbf{u}\|_U \, \|v\|_{L_2(I;H^{\frac12}(\partial\Omega))\cap H^{\frac14}(I;L_2(\partial\Omega))}$, where the term $\int_{I\times\Omega} \partial_t u_1 v_3 \mathrm{d}\mathbf{x}\,\mathrm{d}t$ is bounded *via* interpolation as follows

$$\begin{aligned}\Big|\int_{I\times\Omega} \partial_t u_1 v_3 \,\mathrm{d}\mathbf{x}\,\mathrm{d}t\Big| &\lesssim \|u_1\|_{[L_2(I;L_2(\Omega)),H^1(I;L_2(\Omega))]_{\frac12}} \|v_3\|_{[H^1_0(I;L_2(\Omega)),L_2(I;L_2(\Omega))]_{\frac12}} \\ &\eqsim \|u_1\|_{H^{\frac12}(I;L_2(\Omega))} \|v_3\|_{H^{\frac12}_{00}(I;L_2(\Omega))} \lesssim \|\mathbf{u}\|_U \, \|v\|_{L_2(I;H^{\frac12}(\partial\Omega))\cap H^{\frac14}(I;L_2(\partial\Omega))}.\end{aligned}$$

By a standard mollification argument as in the original proof of Meyers–Serrin, one sees that the set of smooth $\mathbf{u} \in U$ is dense in U. This yields that G_N is bounded.

Since in the current case of $\Gamma_N = \partial\Omega$, we have $\{\mathbf{u} \in U\colon \mathbf{u}|_{I\times\partial\Omega} \cdot \mathbf{n} = 0\} = U_0$, knowing the result of Theorem 2.3, Lemma 2.7 shows that the proof will be completed once we have shown that

$$U \to \Big(L_2(I;H^{\frac12}(\partial\Omega)) \cap H^{\frac14}(I;L_2(\partial\Omega))\Big)' : \mathbf{u} \mapsto \mathbf{u}|_{I\times\partial\Omega} \cdot \mathbf{n} \text{ is surjective.} \tag{2.10}$$

In Corollary 3.17 of [14], it has been shown that for any $\psi \in (L_2(I;H^{\frac12}(\partial\Omega)) \cap H^{\frac14}(I;L_2(\partial\Omega)))'$ there exists a $u_1 \in L_2(I;H^1(\Omega)) \cap H^{\frac12}_{00,\{0\}}(I;L_2(\Omega))$ with $\partial_t u_1 - \Delta_{\mathbf{x}} u_1 = 0$ on $I\times\Omega$, and $(\nabla_{\mathbf{x}} u_1)|_{I\times\partial\Omega} \cdot \mathbf{n}_{\mathbf{x}} = -\psi$. Taking $\mathbf{u}_2 = -\nabla_{\mathbf{x}} u_1$, it means $\operatorname{div}\mathbf{u} = 0$ and $\mathbf{u}|_{I\times\partial\Omega} \cdot \mathbf{n} = \psi$, so that $\mathbf{u} \in U$ and (2.10) is valid. □

Using Theorem 2.9, we formulate the parabolic problem with inhomogeneous pure Neumann boundary conditions as a well-posed first-order system. Let $(g_1, \mathbf{g}_2) \in L_2(I;L_2(\Omega)) \times L_2(I;L_2(\Omega)^d)$ with $\mathbf{g}_2|_{I\times\partial\Omega} \cdot \mathbf{n}_{\mathbf{x}} \in \big(L_2(I;H^{\frac12}(\partial\Omega)) \cap H^{\frac14}(I;L_2(\partial\Omega))\big)'$, $u_0 \in L_2(\Omega)$, and $\phi \in (L_2(I;H^{\frac12}(\partial\Omega)) \cap H^{\frac14}(I;L_2(\partial\Omega)))'$, and set $g := v \mapsto \int_{I\times\Omega} g_1 v + \mathbf{g}_2 \cdot \nabla_{\mathbf{x}} v \,\mathrm{d}\mathbf{x}\,\mathrm{d}t \in L_2(I;H^{-1}(\Omega))$. Then, the solution $\mathbf{u} = (u_1, \mathbf{u}_2) \in U$ of

$$G_N \mathbf{u} = \big(\mathbf{g}_2, g_1 - \mathbf{b}\cdot\mathbf{A}^{-1}\mathbf{g}_2, u_0, \mathbf{g}_2|_{I\times\partial\Omega} \cdot \mathbf{n}_{\mathbf{x}} - \phi\big), \tag{2.11}$$

satisfies

$$(Bu_1, \gamma_0 u_1, \mathbf{A}\nabla_{\mathbf{x}} u_1|_{I\times\partial\Omega} \cdot \mathbf{n}_{\mathbf{x}}) = (g, u_0, \phi),$$

i.e., u_1 satisfies the parabolic PDE in standard variational form and both the initial and Neumann boundary condition. Indeed, knowing $\mathbf{u}_2 + \mathbf{A}\nabla_{\mathbf{x}} u_1 = \mathbf{g}_2$, it holds that $\mathbf{A}\nabla_{\mathbf{x}} u_1|_{I\times\partial\Omega} \cdot \mathbf{n}_{\mathbf{x}} = (\mathbf{g}_2 - \mathbf{u}_2)|_{I\times\partial\Omega} \cdot \mathbf{n}_{\mathbf{x}} = \phi$, and the second equation in (2.11) is equivalent to $\int_{I\times\Omega} (\partial_t u_1 + \mathbf{b}\cdot\nabla_{\mathbf{x}} u_1 + c u_1) v + \mathbf{A}\nabla_{\mathbf{x}} u_1 \cdot \nabla_{\mathbf{x}} v \,\mathrm{d}\mathbf{x}\,\mathrm{d}t = g(v)$ for all $v \in L_2(I;H^1_0(\Omega))$.

3. Plain convergence of adaptive algorithm for homogeneous pure Dirichlet boundary conditions

Consider the setting of Section 1.3 with $\Gamma_D = \partial\Omega$ and homogeneous Dirichlet datum $u_D = 0$. Let $(f_1, \mathbf{f}_2) \in L_2(I;L_2(\Omega)) \times L_2(I;L_2(\Omega)^d)$, $f := v \mapsto \int_{I\times\Omega} f_1 v + \mathbf{f}_2 \cdot \nabla_{\mathbf{x}} v \,\mathrm{d}\mathbf{x}\,\mathrm{d}t \in L_2(I;H^{-1}(\Omega))$ and $u_0 \in L_2(\Omega)$. Since no Neumann boundary conditions are present, g from (1.2) coincides with f. Then, with u being the solution u of (1.1), Proposition 2.5 states that $\mathbf{u} = (u, -\mathbf{A}\nabla_{\mathbf{x}} u) \in U_0$ is the unique solution of

$$G\mathbf{u} = \mathbf{f},$$

where

$$\mathbf{f} := (\mathbf{f}_2, f_1 - \mathbf{b}\cdot\mathbf{A}^{-1}\mathbf{f}_2, u_0) \in L := L_2(I\times\Omega)^d \times L_2(I\times\Omega)\times L_2(\Omega).$$

For an arbitrary discrete subspace $U_0^\delta \subset U_0$, the corresponding least-squares approximation $\mathbf{u}^\delta \in U_0^\delta$ of $\mathbf{u}$ is given by

$$\mathbf{u}^\delta := \operatorname*{argmin}_{\mathbf{v}\in U_0^\delta} \|\mathbf{f} - G\mathbf{v}\|_L^2. \tag{3.1}$$

The resulting Euler–Lagrange equation reads as

$$\langle G\mathbf{u}^\delta, G\mathbf{v}\rangle_L = \langle \mathbf{f}, G\mathbf{v}\rangle_L \quad \text{for all } \mathbf{v}\in U_0^\delta. \tag{3.2}$$

As G is a linear isomorphism, the left-hand side defines an elliptic bilinear form and the Lax–Milgram lemma indeed guarantees unique solvability of (3.1) and (3.2).

Throughout the remainder of this section, for $p\in\mathbb{N}$ some fixed polynomial degree we consider discrete spaces of the form

$$U_0^\delta := S_0^p(\mathcal{T}^\delta)\times S^p(\mathcal{T}^\delta)^d \subset U_0$$

for conforming simplicial meshes $\mathcal{T}^\delta$ of $I\times\Omega$, where

$$S^p(\mathcal{T}^\delta) := \{u\in C(I\times\Omega) : u|_K \text{ polynomial of degree } p \text{ for all } K\in\mathcal{T}^\delta\},$$
$$S_0^p(\mathcal{T}^\delta) := \{u\in S^p(\mathcal{T}^\delta) : u|_{I\times\partial\Omega} = 0\}.$$

In particular, we consider such meshes that can be created by newest vertex bisection [35] starting from a given initial partition $\mathcal{T}^0$.

Finally, we define the reliable and efficient *a posteriori* error estimator

$$\eta(\mathbf{f},\mathbf{u}^\delta) := \|\mathbf{f} - G\mathbf{u}^\delta\|_L \eqsim \|\mathbf{u}-\mathbf{u}^\delta\|_U, \tag{3.3}$$

with corresponding error indicators

$$\eta(K;\mathbf{f},\mathbf{u}^\delta) := \|\mathbf{f} - G\mathbf{u}^\delta\|_{L(K)} \quad \text{for all } K\in\mathcal{T}^\delta, \tag{3.4}$$

where

$$L(\omega) := L_2(\omega)^d \times L_2(\omega)\times L_2(\partial_0\omega) \quad \text{for all measurable } \omega\subseteq I\times\Omega.$$

Here and throughout the remainder of this section, we use the notation $\partial_0\omega := \partial\omega\cap(\{0\}\times\Omega)$.

We consider the following adaptive algorithm.

Algorithm 3.1. Input: Right-hand side $\mathbf{f}\in L$, initial mesh $\mathcal{T}^0 = \mathcal{T}^{\delta_0}$, marking function $M:[0,\infty)\to[0,\infty)$ that is continuous at 0 with $M(0)=0$.
Loop: For each $\ell = 0,1,2,\dots$, iterate the following steps (i)–(iv):

(i) Compute least-squares approximation $\mathbf{u}^\ell = \mathbf{u}^{\delta_\ell}$ of $\mathbf{u}$.
(ii) Compute error indicators $\eta(K;\mathbf{f},\mathbf{u}^\ell)$ for all elements $K\in\mathcal{T}^\ell = \mathcal{T}^{\delta_\ell}$.
(iii) Determine a set of marked elements $\mathcal{M}^\ell\subseteq\mathcal{T}^\ell$ with the following marking property

$$\max_{K\in\mathcal{T}^\ell\setminus\mathcal{M}^\ell} \eta\left(K;\mathbf{f},\mathbf{u}^\ell\right) \le M\left(\max_{K\in\mathcal{M}^\ell}\eta\left(K;\mathbf{f},\mathbf{u}^\ell\right)\right).$$

(iv) Generate refined conforming simplicial mesh $\mathcal{T}^{\ell+1}$ by refining at least all marked elements $\mathcal{M}^\ell$ *via* newest vertex bisection.

Output: Refined meshes $\mathcal{T}^\ell$, corresponding exact discrete solutions $\mathbf{u}^\ell$, and error estimators $\eta(\mathbf{f},\mathbf{u}^\ell)$ for all $\ell\in\mathbb{N}_0$.

Remark 3.2. The criterion (iii) is satisfied for standard marking strategies:

– Suppose that the Dörfler criterion is used for fixed $0<\theta\leq 1$, *i.e.*,

$$\theta\,\eta(\mathbf{f},\mathbf{u}^\ell)^2\leq\sum_{K\in\mathcal{M}^\ell}\eta(K;\mathbf{f},\mathbf{u}^\ell)^2.$$

While this does not directly imply (iii), with the aim to realize optimal rates, the set $\mathcal{M}^\ell$ is constructed in practice *via* sorting of the indicators such that also

$$\max_{K\in\mathcal{T}^\ell\setminus\mathcal{M}^\ell}\eta\left(K;\mathbf{f},\mathbf{u}^\ell\right)\leq\min_{K\in\mathcal{M}^\ell}\eta\left(K;\mathbf{f},\mathbf{u}^\ell\right),$$

see [27]. Then, (iii) holds with $M(t):=t$.

– Suppose the maximum criterion is used for fixed $0\leq\theta\leq 1$, *i.e.*,

$$\mathcal{M}^\ell:=\left\{K\in\mathcal{T}^\ell:\eta\left(K;\mathbf{f},\mathbf{u}^\ell\right)\geq(1-\theta)\max_{K'\in\mathcal{T}^\ell}\eta\left(K';\mathbf{f},\mathbf{u}^\ell\right)\right\}.$$

Then, (iii) holds with $M(t):=t$. To see this, let $K\in\mathcal{T}^\ell\setminus\mathcal{M}^\ell$ and note that

$$\eta(K;\mathbf{f},\mathbf{u}^\ell)<(1-\theta)\max_{K'\in\mathcal{T}^\ell}\eta(K';\mathbf{f},\mathbf{u}^\ell)\leq\min_{K'\in\mathcal{M}^\ell}\eta(K';\mathbf{f},\mathbf{u}^\ell).$$

The following theorem states convergence of Algorithm 3.1. For the heat equation with $\mathbf{A}=\mathrm{Id}$, $\mathbf{b}=0$, and $c=0$, the performance of the algorithm has been numerically investigated in [18].

Theorem 3.3 (Convergence for homogeneous (pure) Dirichlet)**.** *There holds plain convergence of the error*

$$\|\mathbf{u}-\mathbf{u}^\ell\|_U\to 0\quad\text{as }\ell\to\infty.\tag{3.5}$$

As the considered estimator (3.3) *is equivalent to the error, convergence to zero also transfers to the estimator.*

Proof. It suffices to verify that the considered problem fits into the abstract framework of [32], which gives sufficient conditions for error convergence. This will be done in the following three steps.

Step 1. Define another equivalent norm on U

$$\|\mathbf{v}\|^2_{U(I\times\Omega)}:=\|v_1\|^2_{L_2(I;H^1(\Omega))}+\|\mathbf{v}_2\|^2_{L_2(I;L_2(\Omega)^d)}+\|\operatorname{div}\mathbf{v}\|^2_{L_2(I\times\Omega)}+\|v_1(0,\cdot)\|^2_{L_2(\Omega)}$$

for all $\mathbf{v}=(v_1,\mathbf{v}_2)\in U$. Moreover, define the following semi-norms for all measurable subsets $\omega\subseteq I\times\Omega$

$$\|\mathbf{v}\|^2_{U(\omega)}:=\|v_1\|^2_{L_2(\omega)}+\|\nabla_{\mathbf{x}}v_1\|^2_{L_2(\omega)}+\|\mathbf{v}_2\|^2_{L_2(\omega)}+\|\operatorname{div}\mathbf{v}\|^2_{L_2(\omega)}+\|v_1|_{\partial_0\omega}\|^2_{L_2(\partial_0\omega)}.$$

The additional term $\|v_1|_{\partial_0\omega}\|^2_{L_2(\partial_0\omega)}$ will be required to prove local stability (3.11). The semi-norms are *additive* as well as *absolutely continuous* in the sense of Section 2.1 from [32], *i.e.*,

$$\|\mathbf{v}\|^2_{U(\omega_1\cup\omega_2)}=\|\mathbf{v}\|^2_{U(\omega_1)}+\|\mathbf{v}\|^2_{U(\omega_2)}\quad\text{for all }\mathbf{v}\in U,\omega_1,\omega_2\subseteq I\times\Omega\text{ with }\omega_1\cap\omega_2=\emptyset;\tag{3.6}$$

as well as

$$\lim_{|\omega|\to 0}\|\mathbf{v}\|^2_{U(\omega)}=0\quad\text{for all }\mathbf{v}\in U.\tag{3.7}$$

Remark 3.4. For this proof step it was essential that we got rid of the dual norm in Proposition 2.1, see also Remark 2.6.

Step 2. We next show a *local approximation property* in the sense of Section 2.2.2 from [32], *i.e.*, existence of a dense subspace $W \subseteq U_0$ equipped with additive semi-norms $\|\cdot\|_{W(\omega)}$, $\omega \subseteq I \times \Omega$, such that $\|\cdot\|_{W(I\times\Omega)} = \|\cdot\|_W$, and a corresponding $\Pi^\delta \in \mathcal{L}(W, U_0^\delta)$ with

$$\|\mathbf{v} - \Pi^\delta \mathbf{v}\|_{U(K)} \lesssim |K|^{\frac{q}{d+1}} \|\mathbf{v}\|_{W(K)} \quad \text{for all } \mathbf{v} \in W, K \in \mathcal{T}^\delta, \tag{3.8}$$

where $q > 0$ is some fixed exponent. For $k := \min\{k' \in \mathbb{N} \colon k' \geq p+1, k' > \frac{d+1}{2}\}$, let

$$W := \big\{\mathbf{v} = (v_1, \mathbf{v}_2) \in H^k(I\times\Omega) \times H^k(I\times\Omega)^d : \, v_1|_{I\times\partial\Omega} = 0\big\} \subset U_0,$$

and let $I^\delta \in \mathcal{L}(H^k(I\times\Omega), S^p(\mathcal{T}^\delta))$ be the standard point-wise interpolation operator, which is well-defined because of $k > \frac{d+1}{2}$. Then, the operator $\Pi^\delta := \mathbf{I}_{d+1}^\delta := (I^\delta, \dots, I^\delta)$ (of length $d+1$) is in $\mathcal{L}(W, U_0^\delta)$, and with $\mathbf{I}_d^\delta$ defined analogously, it holds that

$$\begin{aligned}
\|\mathbf{v} - \Pi^\delta \mathbf{v}\|_{U(K)}^2 &= \|v_1 - I^\delta v_1\|_{L_2(K)}^2 + \|\nabla_{\mathbf{x}}(v_1 - I^\delta v_1)\|_{L_2(K)}^2 + \|(v_1 - I^\delta v_1)|_{\partial_0 K}\|_{L_2(\partial_0 K)}^2 \\
&\quad + \|\mathbf{v}_2 - \mathbf{I}_d^\delta \mathbf{v}_2\|_{L_2(K)}^2 + \|\operatorname{div}(\mathbf{v} - \mathbf{I}_{d+1}^\delta \mathbf{v})\|_{L_2(K)}^2 \\
&\lesssim \|(v_1 - I^\delta v_1)|_{\partial_0 K}\|_{L_2(\partial_0 K)}^2 + \|v_1 - I^\delta v_1\|_{H^1(K)}^2 + \|\mathbf{v}_2 - \mathbf{I}_d^\delta \mathbf{v}_2\|_{H^1(K)}^2.
\end{aligned}$$

A standard trace inequality ([5], Eq. (10.3.8)) further shows that

$$\begin{aligned}
\|(v_1 - I^\delta v_1)|_{\partial_0 K}\|_{L_2(\partial_0 K)}^2 &\leq \|(v_1 - I^\delta v_1)|_{\partial K}\|_{L_2(\partial K)}^2 \\
&\lesssim |K|^{-\frac{1}{d+1}} \|v_1 - I^\delta v_1\|_{L_2(K)}^2 + |K|^{\frac{1}{d+1}} |v_1 - I^\delta v_1|_{H^1(K)}^2.
\end{aligned}$$

To finish the proof, we show for $m \in \{0,1\}$ and $v \in H^k(I\times\Omega)$ that

$$\|v - I^\delta v\|_{H^m(K)} \lesssim |K|^{\frac{p+1-m}{d+1}} \|v\|_{H^k(K)}.$$

While this is standard if $p+1 > (d+1)/2$, *i.e.*, $k = p+1$, it is not evident if $p+1 \leq (d+1)/2$, and we thus provide a short proof. We first assume that K is the reference simplex, *i.e.*, the convex hull of the canonical basis vectors in $\mathbb{R}^{d+1}$. Let $\widetilde{v} \in P^{k-1}$ be the best approximation of v with respect to $\|\cdot\|_{H^k}$ in the space of polynomials of degree $k-1$, and let $\widehat{v} \in P^p$ be the best approximation of $\widetilde{v}$ with respect to $\|\cdot\|_{H^k}$ in the space of polynomials of degree p. The projection property as well as continuity of I^δ on $H^k(K)$ show that

$$\begin{aligned}
\|v - I^\delta v\|_{H^m(K)} &= \|(\mathrm{Id} - I^\delta)(v - \widehat{v})\|_{H^m(K)} \\
&\lesssim \|v - \widehat{v}\|_{H^k(K)} \leq \|v - \widetilde{v}\|_{H^k(K)} + \|\widetilde{v} - \widehat{v}\|_{H^k(K)}.
\end{aligned}$$

Equivalence of norms on finite-dimensional spaces and two applications of the Bramble–Hilbert lemma further yield that

$$\begin{aligned}
\|v - \widetilde{v}\|_{H^k(K)} + \|\widetilde{v} - \widehat{v}\|_{H^k(K)} &\eqsim \|v - \widetilde{v}\|_{H^k(K)} + \|\widetilde{v} - \widehat{v}\|_{H^{p+1}(K)} \\
&\lesssim \|v - \widetilde{v}\|_{H^k(K)} + |\widetilde{v}|_{H^{p+1}(K)} \leq \|v - \widetilde{v}\|_{H^k(K)} + |v - \widetilde{v}|_{H^{p+1}(K)} + |v|_{H^{p+1}(K)} \\
&\leq 2\|v - \widetilde{v}\|_{H^k(K)} + |v|_{H^{p+1}(K)} \lesssim |v|_{H^k(K)} + |v|_{H^{p+1}(K)}.
\end{aligned}$$

If K is arbitrary, the fact that we use newest vertex bisection allows to apply a standard scaling argument, which yields that

$$\|v - I^\delta v\|_{H^m(K)} \lesssim |K|^{\frac{k-m}{d+1}} |v|_{H^k(K)} + |K|^{\frac{p+1-m}{d+1}} |v|_{H^{p+1}(K)} \lesssim |K|^{\frac{p+1-m}{d+1}} \|v\|_{H^k(K)}.$$

Overall, we thus conclude (3.8) with $q = p$.

Step 3. With the patch $\omega^\delta(K) := \bigcup\{K' \in \mathcal{T}^\delta : K \cap K' \neq \emptyset\}$ of an element $K \in \mathcal{T}^\delta$, we finally show that the employed error estimator is *locally stable* as in Section 2.2.3 from [32], *i.e.*,

$$\eta(K;\mathbf{f},\mathbf{u}^\delta) \lesssim \|\mathbf{u}^\delta\|_{U(\omega^\delta(K))} + \|D\|_{\widetilde{W}(\omega^\delta(K))} \quad \text{for all } K \in \mathcal{T}^\delta \tag{3.9}$$

for a suitable D depending only on the data in a normed space $\widetilde{W}$ equipped with additive and absolutely continuous semi-norms $\|\cdot\|_{\widetilde{W}(\omega)}$, $\omega \subseteq I \times \Omega$, such that $\|\cdot\|_{\widetilde{W}(I\times\Omega)} = \|\cdot\|_{\widetilde{W}}$; as well as *strongly reliable* as in Section 2.2.3 from [32]

$$\langle \mathbf{f} - G\mathbf{u}^\delta, G\mathbf{v}\rangle_L \lesssim \sum_{K\in\mathcal{T}^\delta} \eta(K;\mathbf{f},\mathbf{u}^\delta)\|\mathbf{v}\|_{U(\omega^\delta(K))} \quad \text{for all } \mathbf{v} \in U_0. \tag{3.10}$$

Remark 3.5. Actually, [32] assumes that $\widetilde{W} = L_2(I\times\Omega)$. It is, however, straightforward to see that our mildly relaxed assumption is already sufficient for the convergence proof. Indeed, local stability is only employed in the elementary ([32], Lem. 3.5).

Local stability (3.9) follows from the triangle inequality

$$\eta(K) = \eta(K;\mathbf{f},\mathbf{u}^\delta) = \|\mathbf{f} - G\mathbf{u}^\delta\|_{L(K)} \le \|\mathbf{f}\|_{L(K)} + \|G\mathbf{u}^\delta\|_{L(K)}$$

and the following local stability of G

$$\begin{aligned}\|G\mathbf{u}^\delta\|^2_{L(K)} \lesssim \|\mathbf{u}^\delta_2\|^2_{L_2(K)} + \|\nabla_{\mathbf{x}} u^\delta_1\|^2_{L_2(K)} + \|\operatorname{div}\mathbf{u}^\delta\|^2_{L_2(K)} + \|u^\delta_1\|^2_{L_2(K)} \\ + \|u^\delta_1(0,\cdot)\|^2_{L_2(\partial_0 K)} = \|\mathbf{u}^\delta\|^2_{U(K)}.\end{aligned} \tag{3.11}$$

Strong reliability (3.10) follows from the Cauchy–Schwarz inequality together with the previous local stability of G

$$\langle \mathbf{f} - G\mathbf{u}^\delta, G\mathbf{v}\rangle_L \le \sum_{K\in\mathcal{T}^\delta} \eta(K;\mathbf{f},\mathbf{u}^\delta)\,\|G\mathbf{v}\|_{L(K)} \lesssim \sum_{K\in\mathcal{T}^\delta} \eta(K;\mathbf{f},\mathbf{u}^\delta)\|\mathbf{v}\|_{U(K)},$$

which concludes the proof. □

Remark 3.6. Together with the Céa lemma and with $h^\delta_{\max} := \max\{|K|^{1/(d+1)} : K \in \mathcal{T}^\delta\}$, Step 2 from the proof particularly yields the *a priori* estimate

$$\|\mathbf{u} - \mathbf{u}^\delta\|_U \lesssim \inf_{\mathbf{v}\in U^\delta_0} \|\mathbf{u} - \mathbf{v}\|_U \le \|\mathbf{u} - \Pi^\delta\mathbf{u}\|_U \lesssim (h^\delta_{\max})^p \|\mathbf{u}\|_{H^k(I\times\Omega)\times H^k(I\times\Omega)^d}$$

whenever the solution $\mathbf{u}$ satisfies the additional regularity $\mathbf{u} \in H^k(I\times\Omega)\times H^k(I\times\Omega)^d$, where $k = \min\{k' \in \mathbb{N}\colon k' \ge p+1,\ k' > \frac{d+1}{2}\}$. Instead of the standard interpolation operator I^δ, one can also consider the Scott–Zhang operator $\widetilde{I}^\delta$ from [31] which preserves homogeneous Dirichlet boundary conditions. Then, Equation (4.3) of [31] gives an alternative local bound for the resulting operator $\widetilde{\Pi}^\delta$

$$\|\mathbf{v} - \widetilde{\Pi}^\delta\mathbf{v}\|_{U(K)} \lesssim |K|^{\frac{p}{d+1}} \|\mathbf{v}\|_{H^{p+1}(\omega^\delta(K))\times H^{p+1}(\omega^\delta(K))^d}$$

for all $\mathbf{v} \in H^{p+1}(I\times\Omega)\times H^{p+1}(I\times\Omega)^d$ with $\mathbf{v}|_{I\times\partial\Omega} = 0$ and all $K \in \mathcal{T}^\delta$. In particular this yields the *a priori* estimate

$$\|\mathbf{u} - \mathbf{u}^\delta\|_U \lesssim (h^\delta_{\max})^p \|\mathbf{u}\|_{H^{p+1}(I\times\Omega)\times H^{p+1}(I\times\Omega)^d} \tag{3.12}$$

under the milder assumption that $\mathbf{u} \in H^{p+1}(I\times\Omega)\times H^{p+1}(I\times\Omega)^d$. We mention that Theorem 14 of [18] already proved the latter inequality in the lowest-order case $p = 1$ under even weaker assumptions on $\mathbf{u}$. However, their proof is restricted to simplicial meshes that directly result from a tensor-product mesh ([18], Sect. 4.1.2).

Remark 3.7. (a) We stress that the proof of Theorem 3.3 is relatively abstract in the sense that it generalizes to a large class of least-squares formulations: Suppose that U (instead of U_0) and L are arbitrary Hilbert spaces. Consider the equation

$$Gu = f \quad \text{for given } G \in \mathcal{L}\mathrm{is}(U, L) \text{ and } f \in L.$$

Moreover, suppose that U as well as L are equipped with additive and absolutely continuous (see (3.6) and (3.7)) semi-norms $\|\cdot\|_{U(\omega)}$, $\|\cdot\|_{L(\omega)}$ for all measurable subsets ω of some set $\Omega \subseteq \mathbb{R}^n$ being the union of an initial conforming simplicial mesh $\mathcal{T}^0$. To any conforming simplicial mesh $\mathcal{T}^\delta$ of Ω, we associate a finite-dimensional subspace $U^\delta \subseteq U$ such that $U^\delta \subseteq U^{\widehat{\delta}}$ for all refinements $\mathcal{T}^{\widehat{\delta}}$ of $\mathcal{T}^\delta$. We define the least-squares approximation u^δ as in (3.1) and (3.2) and the error estimator $\eta(f, u^\delta)$ with indicators $\eta(K; f, u^\delta)$ as in (3.3) and (3.4). In this setting, Algorithm 3.1 can be applied. Then, the (analogous) local approximation property of Step 2 (where one could also allow for $W((\omega^\delta)^m(K))$ for fixed $m \in \mathbb{N}$ instead of $W(K)$ in (3.8)) and local stability of G as in (3.11) (where again $U(K)$ could be replaced by $U((\omega^\delta)^m(K))$) yield error and estimator convergence

$$\|u - u^\ell\|_U \eqsim \eta(f, u^\ell) \to 0 \quad \text{as } \ell \to \infty. \tag{3.13}$$

Independently, it has also been recently observed in [19] that the given abstract assumptions yield (3.13) for least-squares methods. However, we stress that Theorem 3.3 is not available in [19].
(b) The setting of (a) is for instance satisfied for a standard least-squares formulation of the Poisson model problem ([4], page 56), the Helmholtz problem [8], the linear elasticity problem [10], and the Stokes problem [9], see also Chapter 3 of [39] for a brief overview of all these formulations. The involved spaces $H^1(\Omega)$ and $H(\mathrm{div}; \Omega)$ can be discretized by usual finite element spaces, *i.e.*, continuous piecewise polynomials and Raviart–Thomas functions, respectively. The required corresponding approximation properties (3.8) are well-known, see, *e.g.*, [17], Section 1.5.

Only for the Stokes problem [9], one requires a special interpolation operator on (a dense subspace of) $\{\mathbf{v} \in H(\mathrm{div}; \Omega)^d : \int_\Omega \mathrm{tr}(\mathbf{v})\, d\mathbf{x} = 0\}$, where tr denotes the trace of square matrices. Since $S^1(\mathcal{T}^\delta)^d$ is contained in the Raviart–Thomas space of order ≥ 1 (excluding the lowest-order case), such an operator can be defined component-wise as an integral-preserving $J^\delta \in \mathcal{L}(H^2(\Omega), S^1(\mathcal{T}^\delta))$ with a local approximation property, *i.e.*, $\int_\Omega v\, d\mathbf{x} = \int_\Omega J^\delta v\, d\mathbf{x}$ and

$$\|v - J^\delta v\|_{H^1(K)} \lesssim |K|^{\frac{1}{d}} \|v\|_{H^2((\omega^\delta)^m(K))} \tag{3.14}$$

for all $v \in H^2(\Omega)$, $K \in \mathcal{T}^\delta$, and some fixed $m \in \mathbb{N}_0$. The operator J^δ is for instance constructed as follows: Inspired by Section 4.1 of [37] and given the nodal Lagrange basis $\{\phi_i : i \in \{1, \dots, N\}\}$ with corresponding local dual basis $\{\psi_i : i \in \{1, \dots, N\}\}$ as in [31], one first defines

$$\widetilde{\psi}_i := \frac{\phi_i + \int_\Omega (1 - \phi_i)\phi_i\, d\mathbf{x}\, \psi_i - \sum_{j \neq i} \left(\int_\Omega \phi_i \phi_j\, d\mathbf{x}\, \psi_j \right)}{\int_\Omega \phi_i\, d\mathbf{x}}$$

for all $i \in \{1, \dots, N\}$. This provides a second local dual basis in the sense that $\mathrm{supp}(\widetilde{\psi}_i) \subset \mathrm{supp}(\phi_i)$ and $\int_{\mathrm{supp}(\widetilde{\psi}_j)} \phi_i \widetilde{\psi}_j\, d\mathbf{x} = \delta_{ij}$ for all $i, j \in \{1, \dots, N\}$. Moreover, from $\sum_i \phi_i = 1$, one verifies that $\sum_i (\int_\Omega \phi_i d\mathbf{x}) \widetilde{\psi}_i = 1$ meaning that this dual basis has (lowest-order) approximation properties. Defining

$$J^\delta : H^1(\Omega) \to S^1\left(\mathcal{T}^\delta\right), \quad v \mapsto \sum_{i=1}^N \int_{\mathrm{supp}\, \widetilde{\psi}_i} v \widetilde{\psi}_i\, d\mathbf{x}\, \phi_i,$$

the latter property implies that this biorthogonal projector is integral-preserving, and the desired approximation property (3.14) with $m = 2$ follows as in [31].

Moreover, Führer and Praetorius [19] verify the setting of (a) for another least-squares formulation of the Stokes problem as well as the Maxwell problem.
(c) Optimal convergence of adaptive least-square finite element methods driven by an equivalent weighted error estimator has been already proved for the Poisson problem in [11, 12], the linear elasticity problem [7], and the Stokes problem [6]. However, apart from the very recent and independent work [19], convergence for adaptive algorithms driven by the natural estimator is only known for the Poisson problem if Dörfler marking with a sufficiently large bulk parameter is used, see [13], where Q-linear convergence has been demonstrated.

Acknowledgements. The first author has been supported by the Austrian Science Fund (FWF) under grant J4379-N. The second author has been supported by NSF Grant DMS 172029.

References

[1] R. Andreev, Stability of sparse space-time finite element discretizations of linear parabolic evolution equations. *IMA J. Numer. Anal.* **33** (2013) 242–260.
[2] I. Babuška and T. Janik, The *h-p* version of the finite element method for parabolic equations. I. The *p*-version in time. *Numer. Methods Part. Differ. Equ.* **5** (1989) 363–399.
[3] I. Babuška and T. Janik, The *h-p* version of the finite element method for parabolic equations. II. The *h-p* version in time. *Numer. Methods Part. Differ. Equ.* **6** (1990) 343–369.
[4] P.B. Bochev and M.D. Gunzburger, *Least-squares finite element methods*. In: Vol. 166 of *Applied Mathematical Sciences*. Springer, New York (2009).
[5] S.C. Brenner and L.R. Scott, The mathematical theory of finite element methods, 3rd edition. In: Vol. 15 of *Texts in Applied Mathematics*. Springer, New York (2008).
[6] P. Bringmann and C. Carstensen, h-adaptive least-squares finite element methods for the 2D Stokes equations of any order with optimal convergence rates. *Comput. Math. Appl.* **74** (2017) 1923–1939.
[7] P. Bringmann, C. Carstensen and G. Starke, An adaptive least-squares FEM for linear elasticity with optimal convergence rates. *SIAM J. Numer. Anal.* **56** (2018) 428–447.
[8] Z. Cai, R. Lazarov, T.A. Manteuffel and S.F. McCormick, First-order system least squares for second-order partial differential equations. I. *SIAM J. Numer. Anal.* **31** (1994) 1785–1799.
[9] Z. Cai, B. Lee and P. Wang, Least-squares methods for incompressible Newtonian fluid flow: linear stationary problems. *SIAM J. Numer. Anal.* **42** (2004) 843–859.
[10] Z. Cai, J. Korsawe and G. Starke, An adaptive least squares mixed finite element method for the stress-displacement formulation of linear elasticity. *Numer. Methods Part. Differ. Equ.* **21** (2005) 132–148.
[11] C. Carstensen, Collective marking for adaptive least-squares finite element methods with optimal rates. *Math. Comput.* **89** (2020) 89–103.
[12] C. Carstensen and E.-J. Park, Convergence and optimality of adaptive least squares finite element methods. *SIAM J. Numer. Anal.* **53** (2015) 43–62.
[13] C. Carstensen, E.-J. Park and P. Bringmann, Convergence of natural adaptive least squares finite element methods. *Numer. Math.* **136** (2017) 1097–1115.
[14] M. Costabel, Boundary integral operators for the heat equation. *Integral Equ. Oper. Theory* **13** (1990) 498–552.
[15] R. Dautray and J.-L. Lions, Mathematical analysis and numerical methods for science and technology. In: Vol. 5 of Evolution Problems I. Springer-Verlag, Berlin (1992).
[16] D. Devaud and Ch. Schwab, Space-time *hp*-approximation of parabolic equations. *Calcolo* **55** (2018) 23.
[17] A. Ern and J.-L. Guermond, Theory and practice of finite elements. In: Vol. 159 of *Applied Mathematical Sciences*. Springer, New York (2004).
[18] T. Führer and M. Karkulik, Space-time least-squares finite elements for parabolic equations. Preprint arXiv:1911.01942 (2019).
[19] T. Führer and D. Praetorius, A short note on plain convergence of adaptive least-squares finite element methods. *Comput. Math. Appl.* **80** (2020) 1619–1632.
[20] M.J. Gander and M. Neumüller, Analysis of a new space-time parallel multigrid algorithm for parabolic problems. *SIAM J. Sci. Comput.* **38** (2016) A2173–A2208.
[21] V. Girault and P.A. Raviart, Finite Element Methods for Navier–Stokes Equations, Theory and Algorithms. Springer-Verlag, Berlin (1986).
[22] M.D. Gunzburger and A. Kunoth, Space-time adaptive wavelet methods for control problems constrained by parabolic evolution equations. *SIAM J. Contr. Optim.* **49** (2011) 1150–1170.
[23] U. Langer, S.E. Moore and M. Neumüller, Space-time isogeometric analysis of parabolic evolution problems. *Comput. Methods Appl. Mech. Eng.* **306** (2016) 342–363.
[24] J.-L. Lions and E. Magenes, Non-Homogeneous Boundary Value Problems and Applications. Vol. I. Translated from the French by P. Kenneth, *Die Grundlehren der mathematischen Wissenschaften*, Band 181. Springer-Verlag, New York-Heidelberg (1972).

[25] J.-L. Lions and E. Magenes, Non-Homogeneous Boundary Value Problems and Applications. Vol. II. Translated from the French by P. Kenneth, *Die Grundlehren der mathematischen Wissenschaften*, Band 182. Springer-Verlag, New York-Heidelberg (1972).
[26] M. Neumüller and I. Smears, Time-parallel iterative solvers for parabolic evolution equations. *SIAM J. Sci. Comput.* **41** (2019) C28–C51.
[27] C.-M. Pfeiler and D. Praetorius, Dörfler marking with minimal cardinality is a linear complexity problem. *Math. Comput.* **89** (2020) 2735–2752.
[28] N. Rekatsinas and R. Stevenson, An optimal adaptive tensor product wavelet solver of a space-time fosls formulation of parabolic evolution problems. *Adv. Comput. Math.* **45** (2018) 1031–1066.
[29] Ch. Schwab and R.P. Stevenson, A space-time adaptive wavelet method for parabolic evolution problems. *Math. Comput.* **78** (2009) 1293–1318.
[30] Ch. Schwab and R.P. Stevenson, Fractional space-time variational formulations of (Navier)–Stokes equations. *SIAM J. Math. Anal.* **49** (2017) 2442–2467.
[31] L.R. Scott and S. Zhang, Finite element interpolation of nonsmooth functions satisfying boundary conditions. *Math. Comput.* **54** (1990) 483–493.
[32] K.G. Siebert, A convergence proof for adaptive finite elements without lower bound. *IMA J. Numer. Anal.* **31** (2011) 947–970.
[33] O. Steinbach, Space-time finite element methods for parabolic problems. *Comput. Methods Appl. Math.* **15** (2015) 551–566.
[34] O. Steinbach and M. Zank, Coercive space-time finite element methods for initial boundary value problems. Berichte aus dem Institut für Angewandte Mathematik, Bericht 2018/7, Technische Universität Graz (2018).
[35] R.P. Stevenson, The completion of locally refined simplicial partitions created by bisection. *Math. Comput.* **77** (2008) 227–241.
[36] R.P. Stevenson, First-order system least squares with inhomogeneous boundary conditions. *IMA J. Numer. Anal.* **34** (2014) 863–878.
[37] R.P. Stevenson and R. van Venetië, Uniform preconditioners for problems of negative order. *Math. Comput.* **89** (2020) 645–674.
[38] R.P. Stevenson and J. Westerdiep, Stability of Galerkin discretizations of a mixed space-time variational formulation of parabolic evolution equations. *IMA J. Numer. Anal.* (2020).
[39] J. Storn, *Topics in least-squares and discontinuous Petrov-Galerkin finite element analysis*. Ph.D. thesis, Humboldt-Universität zu Berlin (2019).
[40] K. Urban and A.T. Patera, An improved error bound for reduced basis approximation of linear parabolic problems. *Math. Comput.* **83** (2014) 1599–1615.
[41] I. Voulis and A. Reusken, A time dependent Stokes interface problem: Well-posedness and space-time finite element discretization. *ESAIM:M2AN* **52** (2018) 2187–2213.
[42] J. Wloka, Partielle Differentialgleichungen: Sobolevräume und Randwertaufgaben. B.G. Teubner, Stuttgart (1982).