Minimal residual space-time discretizations of parabolic equations

Asymmetric spatial operators

Stevenson, R. ; Westerdiep, J.

DOI
10.1016/j.camwa.2021.09.014

Publication date
2021

Document Version
Final published version

Published in
Computers & Mathematics with Applications

Computers and Mathematics with Applications 101 (2021) 107–118

Contents lists available at ScienceDirect

Computers and Mathematics with Applications

www.elsevier.com/locate/camwa

Minimal residual space-time discretizations of parabolic equations: Asymmetric spatial operators ☆

Rob Stevenson *, Jan Westerdiep

Korteweg–de Vries (KdV) Institute for Mathematics, University of Amsterdam, P.O. Box 94248, 1090 GE Amsterdam, the Netherlands

ARTICLE INFO

Keywords:
Parabolic PDEs
Space-time variational formulations
Quasi-best approximations
Stability
Robustness

ABSTRACT

We consider a minimal residual discretization of a simultaneous space-time variational formulation of parabolic evolution equations. Under the usual 'LBB' stability condition on pairs of trial- and test spaces we show quasi-optimality of the numerical approximations without assuming symmetry of the spatial part of the differential operator. Under a stronger LBB condition we show error estimates in an energy-norm that are independent of this spatial differential operator.

1. Introduction

This paper is about the numerical solution of parabolic evolution equations in a simultaneous space-time variational formulation. Compared to classical time-stepping schemes, simultaneous space-time methods are much better suited for a massively parallel implementation (e.g. [32,41]), allow for local refinements in space and time (e.g. [38,26,36,42]), and produce numerical approximations from the employed trial spaces that are quasi-best.

The standard bilinear form that results from a space-time variational formulation is non-coercive, which makes it difficult to construct pairs of discrete trial and test spaces that inherit the stability of the continuous formulation. For this reason, in [2] R. Andreev proposed to use minimal residual discretizations. They have an equivalent interpretation as Galerkin discretizations of an extended self-adjoint, but indefinite, mixed system having as secondary variable the Riesz lift of the residual of the primal variable w.r.t. the PDE.

For pairs of trial spaces that satisfy a Ladyzhenskaja–Babuška–Brezzi (LBB) condition, it was shown that w.r.t. the norm on the natural solution space, being an intersection of Bochner spaces, the Galerkin solutions are quasi-best approximations from the selected trial spaces. This LBB condition was verified in [2] for 'full' and 'sparse' tensor products of various finite element spaces in space and time. The sparse tensor product setting was then generalized in [36, Proposition 5.1] to allow for local refinements in space and time whilst retaining (uniform) LBB stability.

A different minimal residual formulation of first order system type was introduced in [22], see also [27]. Here the various residuals are all measured in L_2-norms, meaning that they do not have to be introduced as separate variables, and the resulting bilinear form is coercive.

Closer in spirit to [2] are the space-time methods presented in [35, 30,6], in which error bounds are presented w.r.t. mesh-dependent norms. In [12,40] space-time variational methods are presented that lead to coercive bilinear forms based on fractional Sobolev norms of order $\frac{1}{2}$. A first order space-time DPG formulation of the heat equation is presented in [16].

A restriction imposed in [2], as well as in the other mentioned references apart from [6,27], is that the spatial part of the PDO is not only coercive but also symmetric. In [37] we could remove the symmetry condition for the analysis of a related Brézis–Ekeland–Nayroles (BEN) ([4,31]) formulation of the parabolic PDE. In the current work, we prove that also for the minimal residual (MR) method the symmetry condition can be dropped. So for both MR and BEN we show that under the aforementioned LBB condition the Galerkin approximations are quasi-optimal, where the bound on the error in the numerical approximation for BEN improves upon the one from [37].

The error bounds for both MR and BEN degrade for increasing asymmetry. This is not an artefact of the theory but is confirmed by numerical experiments. Under a stronger LBB condition on the pair of trial spaces, however, we will prove that the MR and BEN approximations are quasi-best w.r.t. a continuous, i.e., 'mesh-*in*dependent', energy-norm, uniformly in the spatial PDO.

☆ The second author has been supported by the Netherlands Organization for Scientific Research (NWO) under contract. no. 613.001.652.
* Corresponding author.
E-mail addresses: r.p.stevenson@uva.nl (R. Stevenson), j.h.westerdiep@uva.nl (J. Westerdiep).

https://doi.org/10.1016/j.camwa.2021.09.014
Received 2 June 2021; Received in revised form 16 September 2021; Accepted 29 September 2021
Available online 6 October 2021

We present numerical tests for the evolution problem governed by the simple PDE $\partial_t - \varepsilon\partial_x^2 + \partial_x + e\mathrm{Id}$ on $(0,1)^2$ with initial and boundary conditions, where e is either 0 or 1. For the case that homogeneous Dirichlet boundary conditions are prescribed at the outflow boundary $x=1$, the results for very small ε illustrate that quasi-optimal approximations do no necessarily mean accurate approximations. Indeed the error in the computed solution is large because of the unresolved boundary layer. The minimization of the error in the energy-norm of least squares type causes a global spread of the error along the streamlines. We tackled this problem by imposing these boundary conditions only weakly.

1.1. Organization

In Sect. 2 we recall the well-posed space-time variational formulation of the parabolic problem and study its conditioning. Under the usual LBB condition, in Sect. 3 we show quasi-optimality of the MR method without assuming symmetry of the spatial differential operator. A similar result is shown for BEN in Sect. 4. Known results concerning the verification of this LBB condition are summarized in Sect. 5, together with results about optimal preconditioning.

In Sect. 6 we equip the solution space with an energy-norm, and, under a stronger LBB condition, show error estimates for MR and BEN that are independent of the spatial differential operator. We present an a posteriori error estimator that, under an even stronger LBB condition, is efficient and, modulo a date-oscillation term, is reliable.

In Sect. 7 we apply the general theory to the example of the convection-diffusion problem. We give pairs of trial- and test spaces that satisfy the 2nd and 3rd mentioned LBB conditions. Finally, in Sect. 8 we present numerical results for the MR method in the simple case of having a one-dimensional spatial domain. To solve the problems caused by an unresolved boundary layer, we modify the method by imposing a boundary condition weakly.

1.2. Notations

In this work, by $C \lesssim D$ we will mean that C can be bounded by a multiple of D, independently of parameters that C and D may depend on. 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, by $\mathcal{L}(E,F)$ we will denote 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 $\mathbb{R}$.

2. Well-posed variational formulation

Let V, H be separable Hilbert spaces of functions on some "spatial domain" such that $V \hookrightarrow H$ with dense embedding. Identifying H with its dual, we obtain the Gelfand triple $V \hookrightarrow H \simeq H' \hookrightarrow V'$. We use $\langle\cdot,\cdot\rangle$ to denote both the scalar product on $H\times H$ as well as its unique extension to the duality pairing on $V'\times V$ or $V\times V'$, and denote the norm on H by $\|\cdot\|$.

For a.e.

$$t \in I := (0,T),$$

let $a(t;\cdot,\cdot)$ denote a bilinear form on $V\times V$ such that for any $\eta,\zeta\in V$, $t\mapsto a(t;\eta,\zeta)$ is measurable on I, and such that for some $\varrho\in\mathbb{R}$, for a.e. $t\in I$,

$$|a(t;\eta,\zeta)| \lesssim \|\eta\|_V\|\zeta\|_V \quad (\eta,\zeta\in V) \quad (\textit{boundedness}), \tag{2.1}$$

$$a(t;\eta,\eta) + \varrho\langle\eta,\eta\rangle \gtrsim \|\eta\|_V^2 \quad (\eta\in V) \quad (\textit{Gårding inequality}). \tag{2.2}$$

With $A(t)\in\mathcal{L}\mathrm{is}(V,V')$ being defined by $(A(t)\eta)(\zeta) := a(t;\eta,\zeta)$, given a forcing function g and an initial value u_0, we are interested in solving the *parabolic initial value problem* to finding u such that

$$\begin{cases} \frac{\mathrm{d}u}{\mathrm{d}t}(t) + A(t)u(t) = g(t) & (t\in I),\\ u(0) = u_0. \end{cases} \tag{2.3}$$

In a simultaneous space-time variational formulation, the parabolic PDE reads as finding u from a suitable space of functions X of time and space such that

$$(Bw)(v) := \int_I \langle \tfrac{\mathrm{d}w}{\mathrm{d}t}(t), v(t)\rangle + a(t;w(t),v(t))\mathrm{d}t = \int_I \langle g(t),v(t)\rangle \mathrm{d}t =: g(v)$$

for all v from another suitable space of functions Y of time and space. One possibility to enforce the initial condition is by testing it against additional test functions. A proof of the following result can be found in [34], cf. [29, Ch. 3, Thm. 4.1], [43, Ch. IV, §26], [15, Ch.XVIII, §3], and [19, Thm. 6.6] for similar statements.

Theorem 2.1. *With $X := L_2(I;V)\cap H^1(I;V')$, $Y := L_2(I;V)$, under conditions (2.1) and (2.2) it holds that*

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

where for $t\in\bar I$, $\gamma_t : u\mapsto u(t,\cdot)$ denotes the trace map. That is, assuming $g\in Y'$ and $u_0\in H$, finding $u\in X$ such that

$$(B,\gamma_0) = (g,u_0) \tag{2.4}$$

is a well-posed simultaneous space-time variational formulation of (2.3).

With $\tilde u(t) := u(t)e^{-\varrho t}$, (2.3) is equivalent to $\frac{\mathrm{d}\tilde u}{\mathrm{d}t}(t) + (A(t)+\varrho\mathrm{Id})\tilde u(t) = g(t)e^{-\varrho t}$ $(t\in I)$, $\tilde u(0) = u_0$. Since $((A(t)+\varrho\mathrm{Id})\eta)(\eta) \gtrsim \|\eta\|_V^2$, w.l.o.g. we will always assume that, besides (2.1), (2.2) is valid for $\varrho = 0$, i.e., for a.e. $t\in I$,

$$a(t;\eta,\eta) \gtrsim \|\eta\|_V^2 \quad (\eta\in V) \quad (\textit{coercivity}). \tag{2.5}$$

We define $A, A_s\in\mathcal{L}\mathrm{is}(Y,Y')$, $A_a\in\mathcal{L}(Y,Y')$, and $C,\partial_t\in\mathcal{L}(X,Y')$ by

$$(Aw)(v) := \int_I a(t;w(t),v(t))\mathrm{d}t, \quad A_s := \tfrac12(A+A'), \quad A_a := \tfrac12(A-A'),$$

$$\partial_t := B - A, \quad C := B - A_s = \partial_t + A_a,$$

and equip Y with 'energy'-scalar product $\langle\cdot,\cdot\rangle_Y := (A_s\cdot)(\cdot)$, and norm

$$\|v\|_Y := \sqrt{(A_s v)(v)}$$

being, thanks to (2.1) and (2.5), equivalent to the standard norm on Y. Equipping Y' with the resulting dual norm, $A_s\in\mathcal{L}\mathrm{is}(Y,Y')$ is an isometric isomorphism, and so for $f\in Y'$ we have

$$f(A_s^{-1}f) = (A_sA_s^{-1}f)(A_s^{-1}f) = \|A_s^{-1}f\|_Y^2 = \|f\|_{Y'}^2.$$

For some constant $\beta\ge 1$, we equip X with norm

$$\|\cdot\|_X := \sqrt{\|\cdot\|_Y^2 + \|\partial_t\cdot\|_{Y'}^2 + \|\gamma_T\cdot\|^2 + (\beta-1)\|\gamma_0\cdot\|^2},$$

being, thanks to $X\hookrightarrow C(\bar I;H)$, equivalent to the standard norm on X. In addition, we define the energy-norm on X by

$$|||\cdot|||_X := \sqrt{\|B\cdot\|_{Y'}^2 + \beta\|\gamma_0\cdot\|^2},$$

which, thanks to Theorem 2.1, is indeed a norm on X.

Proposition 2.2. *With $\alpha := \|A_a\|_{\mathcal{L}(Y,Y')}$, for $0\neq w\in X$ it holds that*

$$\Big(1+\frac{\alpha}{2}\big(\alpha+\sqrt{\alpha^2+4}\big)\Big)^{-1} \le \frac{|||w|||_X^2}{\|w\|_X^2} \le 1+\frac{\alpha}{2}\big(\alpha+\sqrt{\alpha^2+4}\big),$$

so that, in particular, both norms are equal when $A_a = 0$.

Proof. Using that for $w, v \in X$,

$$((\partial_t + \partial_t' + \gamma_0'\gamma_0)w)(v) = \int_I \langle \tfrac{\mathrm{d}w}{\mathrm{d}t}(t), v(t)\rangle + \langle w(t), \tfrac{\mathrm{d}v}{\mathrm{d}t}(t)\rangle \mathrm{d}t + \langle w(0), v(0)\rangle$$
$$= \int_I \tfrac{\mathrm{d}}{\mathrm{d}t}\langle w(t), v(t)\rangle \mathrm{d}t + \langle w(0), v(0)\rangle = (\gamma_T'\gamma_T w)(v),$$

we find that

$$\begin{aligned} B' A_s^{-1} B + \beta\gamma_0'\gamma_0 &= (C' + A_s)A_s^{-1}(C + A_s) + \beta\gamma_0'\gamma_0 \\ &= C' A_s^{-1} C + A_s + C' + C + \beta\gamma_0'\gamma_0 \\ &= C' A_s^{-1} C + A_s + \partial_t' + \partial_t + \beta\gamma_0'\gamma_0 \\ &= C' A_s^{-1} C + A_s + \gamma_T'\gamma_T + (\beta - 1)\gamma_0'\gamma_0. \end{aligned} \tag{2.6}$$

For $w \in X$,

$$(C' A_s^{-1} C w)(w) = (Cw)(A_s^{-1} Cw) = \|(\partial_t + A_a)w\|_{Y'}^2 \le (\|\partial_t w\|_{Y'} + \alpha\|w\|_Y)^2,$$

and so, for any $\eta \neq 0$, Young's inequality shows that

$$\begin{aligned} &\|Bw\|_{Y'}^2 + \beta\|\gamma_0 w\|^2 \\ &= \big((C' A_s^{-1} C + A_s + \gamma_T'\gamma_T + (\beta - 1)\gamma_0'\gamma_0)(w)\big)(w) \\ &\le (1 + \eta^2)\|\partial_t w\|_{Y'}^2 + ((1 + \eta^{-2})\alpha^2 + 1)\|w\|_Y^2 + \|\gamma_T w\|^2 + (\beta - 1)\|\gamma_0 w\|^2. \end{aligned}$$

Solving $(1 + \eta^2) = (1 + \eta^{-2})\alpha^2 + 1$ gives $1 + \eta^2 = 1 + \frac{\alpha}{2}\big(\alpha + \sqrt{\alpha^2 + 4}\big)$, showing one of the bounds of the statement.

From

$$\|(\partial_t + A_a)w\|_{Y'}^2 \ge (\|\partial_t w\|_{Y'} - \alpha\|w\|_Y)^2 \ge (1 - \eta^2)\|\partial_t w\|_{Y'}^2 + (1 - \eta^{-2})\alpha^2\|w\|_Y^2$$

again by Young's inequality, by solving η^2 from $1 - \eta^2 = (1 - \eta^{-2})\alpha^2 + 1$ the other bound follows. □

Remark 2.3. Because $\|\cdot\|_Y$ is defined in terms of the symmetric part A_s of the spatial differential operator A, $\alpha = \|A_a\|_{\mathcal{L}(Y,Y')}$ is a measure for the *relative* asymmetry of the operator A. Indeed $\|A_a\|_{\mathcal{L}(Y,Y')} = \|A_s^{-\frac{1}{2}} A_a A_s^{-\frac{1}{2}}\|_{\mathcal{L}(L_2(I;H),L_2(I;H))} = \rho(A_s^{-\frac{1}{2}} A_a' A_s^{-1} A_a A_s^{-\frac{1}{2}})^{\frac{1}{2}} = \rho(A_s^{-1} A_a A_s^{-1} A_a)^{\frac{1}{2}}$, where we used that $A_a' = -A_a$.

A result on the conditioning of $(B, \gamma_0) \in \mathcal{L}\mathrm{is}(X, Y' \times H)$ similar to Proposition 2.2 but w.r.t. different norms on X and Y can be found in [21, Lemmas 71.1 & 71.2].

3. Minimal residual (MR) method

Let $(X^\delta, Y^\delta)_{\delta\in\Delta}$ a family of closed, non-zero subspaces of X and Y, respectively. For $\delta \in \Delta$, let E_X^δ and E_Y^δ denote the trivial embeddings $X^\delta \to X$ and $Y^\delta \to Y$, which we sometimes write for clarity, but that we mainly introduce because of their duals. We assume that

$$X^\delta \subseteq Y^\delta \quad (\delta \in \Delta), \tag{3.1}$$

$$\gamma_\Delta^{\partial_t} := \inf_{\delta\in\Delta} \inf_{\{w\in X^\delta :\, \partial_t E_X^\delta w \neq 0\}} \frac{\|{E_Y^\delta}' \partial_t E_X^\delta w\|_{Y^{\delta\prime}}}{\|\partial_t E_X^\delta w\|_{Y'}} > 0. \tag{3.2}$$

Furthermore, for efficiency reasons we assume to have available a $K_Y^\delta = {K_Y^\delta}' \in \mathcal{L}\mathrm{is}(Y^{\delta\prime}, Y^\delta)$ (a 'preconditioner'), such that for some constants $0 < r_\Delta \le R_\Delta < \infty$,

$$\frac{((K_Y^\delta)^{-1} v)(v)}{({E_Y^\delta}' A_s E_Y^\delta v)(v)} \in [r_\Delta, R_\Delta] \quad (\delta \in \Delta, v \in Y^\delta), \tag{3.3}$$

or, equivalently, $\frac{f(K_Y^\delta f)}{f(({E_Y^\delta}' A_s E_Y^\delta)^{-1} f)} \in [R_\Delta^{-1}, r_\Delta^{-1}]$ $(\delta \in \Delta, f \in Y^{\delta\prime})$.

Noticing that $\|f\|_{Y^{\delta\prime}}^2 = f(({E_Y^\delta}' A_s E_Y^\delta)^{-1} f)$, the expression

$$\|\cdot\|_{K_Y^\delta} := \sqrt{(\cdot)(K_Y^\delta \cdot)}$$

defines an equivalent norm on $Y^{\delta\prime}$, and our Minimal Residual approximation $u^\delta \in X^\delta$ of the solution $u \in X$ of (2.4) is defined as

$$u^\delta := \operatorname*{argmin}_{w\in X^\delta} \|{E_Y^\delta}'(B E_X^\delta w - g)\|_{K_Y^\delta}^2 + \beta\|\gamma_0 E_X^\delta w - u_0\|^2, \tag{3.4}$$

for some constant $\beta \ge 1$. Later we will see that, thanks to (3.2) and (3.3),

$$\inf_{0\neq w\in X^\delta} \sup_{(v_1,v_2)\in Y^\delta\times H} \frac{(B E_X^\delta w)(E_Y^\delta v_1) + \beta\langle\gamma_0 E_X^\delta w, v_2\rangle}{\sqrt{((K_Y^\delta)^{-1} v_1)(v_1) + \beta\|v_2\|^2}} > 0 \tag{3.5}$$

(even uniformly in $\delta \in \Delta$)[1] which implies that (3.4) has a unique solution. The numerical approximation (3.4) was proposed in [2],[2] and further investigated in [37]. In both these references the analysis of the MR method was restricted to the case that $A_a = 0$. The introduction of the parameter $\beta \ge 1$ allows to appropriately weight both terms in the least squares minimization.

The solution u^δ of the MR problem is the solution of the resulting Euler–Lagrange equations, which read as

$$({E_X^\delta}' B' E_Y^\delta K_Y^\delta E_Y^{\delta\prime} B E_X^\delta + {E_X^\delta}' \beta\gamma_0'\gamma_0 E_X^\delta) u^\delta = {E_X^\delta}' B' E_Y^\delta K_Y^\delta E_Y^{\delta\prime} g + {E_X^\delta}' \beta\gamma_0' u_0, \tag{3.6}$$

as also the second component of the solution $(\mu^\delta, u^\delta) \in Y^\delta \times X^\delta$ of

$$\begin{bmatrix} (K_Y^\delta)^{-1} & {E_Y^\delta}' B E_X^\delta \\ {E_X^\delta}' B' E_Y^\delta & -{E_X^\delta}' \beta\gamma_0'\gamma_0 E_X^\delta \end{bmatrix} \begin{bmatrix} \mu^\delta \\ u^\delta \end{bmatrix} = \begin{bmatrix} {E_Y^\delta}' g \\ -{E_X^\delta}' \beta\gamma_0' u_0 \end{bmatrix}, \tag{3.7}$$

being a useful representation when no efficient preconditioner is available and one has to resort to $(K_Y^\delta)^{-1} = {E_Y^\delta}' A_s E_Y^\delta$.

With the "projected" or "approximate" (because generally $Y^\delta \neq Y$) *trial-to-test operator* $T^\delta = (T_1^\delta, T_2^\delta) \in \mathcal{L}(X, Y^\delta \times H)$ defined by

$$\begin{aligned} &((K_Y^\delta)^{-1} T_1^\delta w)(v_1) + \beta\langle T_2^\delta w, v_2\rangle = (Bw)(E_Y^\delta v_1) + \beta\langle\gamma_0 u, v_2\rangle \\ &\quad ((v_1, v_2) \in Y^\delta \times H), \end{aligned} \tag{3.8}$$

and the "projected" or "approximate" *optimal test space* $Z^\delta := \operatorname{ran} T^\delta|_{X^\delta}$, a third equivalent formulation of (3.4) (see e.g. [13], [8, Prop. 2.2], [14]) is finding $u^\delta \in X^\delta$ that solves the Petrov–Galerkin system

$$(B E_X^\delta u^\delta)(E_Y^\delta v_1) + \beta\langle\gamma_0 E_X^\delta u^\delta, v_2\rangle = g(E_Y^\delta v_1) + \beta\langle u_0, v_2\rangle \quad ((v_1, v_2) \in Z^\delta). \tag{3.9}$$

Note that (3.9) avoids the 'normal equations' (3.6). It will allow us to derive a quantitatively sharp estimate for the error in u^δ. From (3.3) and (3.5), one infers that $\sup_{0\neq w\in X^\delta} \frac{\|T^\delta w\|_{Y\times H}}{\|w\|_X} > 0$, so that, thanks to X^δ being closed, Z^δ is a closed subspace of $Y^\delta \times H$. We orthogonally decompose $Y^\delta \times H$ into Z^δ and $(Z^\delta)^\perp$, where here we equip Y^δ with inner product $((K_Y^\delta)^{-1}\cdot)(\cdot)$. From (3.8) one infers that for $w \in X^\delta$ and $(v_1, v_2) \in (Z^\delta)^\perp$, it holds that $(Bw)(v_1) + \beta\langle\gamma_0 u, v_2\rangle = 0$, and so

$$\begin{aligned} &\sup_{(v_1,v_2)\in Y^\delta\times H} \frac{(B E_X^\delta w)(E_Y^\delta v_1) + \beta\langle\gamma_0 E_X^\delta w, v_2\rangle}{\sqrt{((K_Y^\delta)^{-1} v_1)(v_1) + \beta\|v_2\|^2}} \\ &= \sup_{(v_1,v_2)\in Z^\delta} \frac{(B E_X^\delta w)(E_Y^\delta v_1) + \beta\langle\gamma_0 E_X^\delta w, v_2\rangle}{\sqrt{((K_Y^\delta)^{-1} v_1)(v_1) + \beta\|v_2\|^2}}. \end{aligned} \tag{3.10}$$

[1] This follows by combining (3.13), (3.15), and (3.16).

[2] In [2], the norm $\|\gamma_0 E_X^\delta w - u_0\|$ reads as $\sup_{0\neq z\in Z^\delta} \frac{\langle\gamma_0 E_X^\delta w - u_0, z\rangle}{\|z\|}$ for some $H \supseteq Z^\delta \supseteq \operatorname{ran}\gamma_0|_{X^\delta}$ which generalization seems not very helpful.

Theorem 3.1. *Under conditions (3.1), (3.2), and (3.3), the solution $u^\delta \in X^\delta$ of (3.6) exists uniquely, and satisfies*

$$\|u-u^\delta\|_X \le \sqrt{\frac{\max(R_\Delta,1)\Big(1+\frac12\big(\alpha^2+\alpha\sqrt{\alpha^2+4}\big)\Big)}{\min(r_\Delta,1)\frac12\Big((\gamma_\Delta^{\partial_t})^2+\alpha^2+1-\sqrt{((\gamma_\Delta^{\partial_t})^2+\alpha^2+1)^2-4(\gamma_\Delta^{\partial_t})^2}\Big)}}\inf_{w\in X^\delta}\|u-w\|_X.$$

Before we give its proof, we make a few comments on this error bound. First, it shows that for $\gamma_\Delta^{\partial_t}=r_\Delta=R_\Delta=1$ and $\alpha=0$, u^δ is the best approximation to u from X^δ. Secondly, for $\alpha=0$ (and $\beta=1$), the bound equals the one found in [37, Thm. 3.7 & Rem. 3.8]. Thirdly, using Mathematica® [44] we find that[3]

$$\sqrt{\frac{\Big(1+\frac12\big(\alpha^2+\alpha\sqrt{\alpha^2+4}\big)\Big)}{\frac12\Big((\gamma_\Delta^{\partial_t})^2+\alpha^2+1-\sqrt{((\gamma_\Delta^{\partial_t})^2+\alpha^2+1)^2-4(\gamma_\Delta^{\partial_t})^2}\Big)}}\Big/\frac{1+\frac12\big(\alpha^2+\alpha\sqrt{\alpha^2+4}\big)}{\gamma_\Delta^{\partial_t}}\in\big[\tfrac12\sqrt{3},1\big]$$

for $\alpha\ge0$, $\gamma_\Delta^{\partial_t}\in(0,1]$, clarifying the behavior of the bound in terms of α and $\gamma_\Delta^{\partial_t}$.

Proof. Let u be the solution of (2.4), i.e., $g=Bu$ and $u_0=\gamma_0 u$. The mapping $P^\delta\in\mathcal{L}(X,X)$ from u to the solution $u^\delta\in X^\delta$ of (3.4) or, equivalently, (3.6) or (3.9), is a projector onto X^δ that, by our assumption $X^\delta\notin\{0,X\}$, is unequal to 0 or Id. Consequently $\|P^\delta\|_{\mathcal{L}(X,X)}=\|\mathrm{Id}-P^\delta\|_{\mathcal{L}(X,X)}$ ([28,45]), and

$$\begin{aligned}\|u-u^\delta\|_X&=\|(\mathrm{Id}-P^\delta)u\|_X=\inf_{w\in X^\delta}\|(\mathrm{Id}-P^\delta)(u-w)\|_X\\&\le\|P^\delta\|_{\mathcal{L}(X,X)}\inf_{w\in X^\delta}\|u-w\|_X.\end{aligned}\tag{3.11}$$

To bound $\|P^\delta\|_{\mathcal{L}(X,X)}=\sup_{0\neq w\in X}\frac{\|P^\delta w\|_X}{\|w\|_X}$, given $w\in X$, let $E_X^\delta w^\delta:=P^\delta w$. Using (3.3), (3.10), (3.9), and Proposition 2.2 we estimate

$$\begin{aligned}&\sup_{(v_1,v_2)\in Y^\delta\times H}\frac{\big((BE_X^\delta w^\delta)(E_Y^\delta v_1)+\beta\langle\gamma_0E_X^\delta w^\delta,v_2\rangle\big)^2}{\|E_Y^\delta v_1\|_Y^2+\beta\|v_2\|^2}\\&\le\frac{1}{\min(r_\Delta,1)}\sup_{(v_1,v_2)\in Y^\delta\times H}\frac{\big((BE_X^\delta w^\delta)(E_Y^\delta v_1)+\beta\langle\gamma_0E_X^\delta w^\delta,v_2\rangle\big)^2}{((K_Y^\delta)^{-1}v_1)(v_1)+\beta\|v_2\|^2}\\&=\frac{1}{\min(r_\Delta,1)}\sup_{(v_1,v_2)\in Z^\delta}\frac{\big((BE_X^\delta w^\delta)(E_Y^\delta v_1)+\beta\langle\gamma_0E_X^\delta w^\delta,v_2\rangle\big)^2}{((K_Y^\delta)^{-1}v_1)(v_1)+\beta\|v_2\|^2}\\&=\frac{1}{\min(r_\Delta,1)}\sup_{(v_1,v_2)\in Z^\delta}\frac{\big((Bw)(E_Y^\delta v_1)+\beta\langle\gamma_0w,v_2\rangle\big)^2}{((K_Y^\delta)^{-1}v_1)(v_1)+\beta\|v_2\|^2}\\&\le\frac{\max(R_\Delta,1)}{\min(r_\Delta,1)}\sup_{(v_1,v_2)\in Y\times H}\frac{\big((Bw)(v_1)+\beta\langle\gamma_0w,v_2\rangle\big)^2}{\|v_1\|_Y^2+\beta\|v_2\|^2}\\&=\frac{\max(R_\Delta,1)}{\min(r_\Delta,1)}|||w|||_X^2\le\frac{\max(R_\Delta,1)}{\min(r_\Delta,1)}\Big(1+\frac12\big(\alpha^2+\alpha\sqrt{\alpha^2+4}\big)\Big)\|w\|_X^2.\end{aligned}\tag{3.12}$$

On the other hand,

$$\begin{aligned}&\sup_{(v_1,v_2)\in Y^\delta\times H}\frac{\big((BE_X^\delta w^\delta)(E_Y^\delta v_1)+\beta\langle\gamma_0E_X^\delta w^\delta,v_2\rangle\big)^2}{\|E_Y^\delta v_1\|_Y^2+\beta\|v_2\|^2}\\&=\sup_{(v_1,v_2)\in Y^\delta\times H}\frac{\big((A_sE_Y^\delta(E_Y^{\delta\prime}A_sE_Y^\delta)^{-1}E_Y^{\delta\prime}BE_X^\delta w^\delta)(E_Y^\delta v_1)+\beta\langle\gamma_0E_X^\delta w^\delta,v_2\rangle\big)^2}{\|E_Y^\delta v_1\|_Y^2+\beta\|v_2\|^2}\\&=\sup_{(v_1,v_2)\in Y^\delta\times H}\frac{\big(\langle E_Y^\delta(E_Y^{\delta\prime}A_sE_Y^\delta)^{-1}E_Y^{\delta\prime}BE_X^\delta w^\delta,E_Y^\delta v_1\rangle_Y+\beta\langle\gamma_0E_X^\delta w^\delta,v_2\rangle\big)^2}{\|E_Y^\delta v_1\|_Y^2+\beta\|v_2\|^2}\\&=\|E_Y^\delta(E_Y^{\delta\prime}A_sE_Y^\delta)^{-1}E_Y^{\delta\prime}BE_X^\delta w^\delta\|_Y^2+\beta\|\gamma_0E_X^\delta w^\delta\|^2\\&=(A_sE_Y^\delta(E_Y^{\delta\prime}A_sE_Y^\delta)^{-1}E_Y^{\delta\prime}BE_X^\delta w^\delta)(E_Y^\delta(E_Y^{\delta\prime}A_sE_Y^\delta)^{-1}E_Y^{\delta\prime}BE_X^\delta w^\delta)\\&\qquad+\beta(E_X^{\delta\prime}\gamma_0'\gamma_0E_X^\delta w^\delta)(w^\delta)\\&=\big((E_X^{\delta\prime}B'E_Y^\delta(E_Y^{\delta\prime}A_sE_Y^\delta)^{-1}E_Y^{\delta\prime}BE_X^\delta+\beta E_X^{\delta\prime}\gamma_0'\gamma_0E_X^\delta)w^\delta\big)(w^\delta).\end{aligned}\tag{3.13}$$

Using (3.1), we write $E_X^\delta=E_Y^\delta F^\delta$ with F^δ denoting the trivial embedding $X^\delta\to Y^\delta$. Using $B=C+A_s$ and $C+C'+\gamma_0'\gamma_0=\gamma_T'\gamma_T$, similar to (2.6) we infer that

$$\begin{aligned}&E_X^{\delta\prime}B'E_Y^\delta(E_Y^{\delta\prime}A_sE_Y^\delta)^{-1}E_Y^{\delta\prime}BE_X^\delta+E_X^{\delta\prime}\beta\gamma_0'\gamma_0E_X^\delta\\&=F^{\delta\prime}\Big(E_Y^{\delta\prime}B'E_Y^\delta(E_Y^{\delta\prime}A_sE_Y^\delta)^{-1}E_Y^{\delta\prime}BE_Y^\delta+E_Y^{\delta\prime}\beta\gamma_0'\gamma_0E_Y^\delta\Big)F^\delta\\&=F^{\delta\prime}\Big(E_Y^{\delta\prime}C'E_Y^\delta(E_Y^{\delta\prime}A_sE_Y^\delta)^{-1}E_Y^{\delta\prime}CE_Y^\delta+E_Y^{\delta\prime}A_sE_Y^\delta\\&\qquad+E_Y^{\delta\prime}(\gamma_T'\gamma_T+(\beta-1)\gamma_0'\gamma_0)E_Y^\delta\Big)F^\delta\\&=E_X^{\delta\prime}C'E_Y^\delta(E_Y^{\delta\prime}A_sE_Y^\delta)^{-1}E_Y^{\delta\prime}CE_X^\delta+E_X^{\delta\prime}A_sE_X^\delta\\&\qquad+E_X^{\delta\prime}(\gamma_T'\gamma_T+(\beta-1)\gamma_0'\gamma_0)E_X^\delta.\end{aligned}\tag{3.14}$$

We conclude that for any $\eta\in(0,1]$,

$$\begin{aligned}&\big((E_X^{\delta\prime}B'E_Y^\delta(E_Y^{\delta\prime}A_sE_Y^\delta)^{-1}E_Y^{\delta\prime}BE_X^\delta+E_X^{\delta\prime}\beta\gamma_0'\gamma_0E_X^\delta)w^\delta\big)(w^\delta)\\&=\|E_Y^{\delta\prime}CE_X^\delta w^\delta\|_{Y^{\delta\prime}}^2+\|E_X^\delta w^\delta\|_Y^2+\|\gamma_TE_X^\delta w^\delta\|^2+(\beta-1)\|\gamma_0E_X^\delta w^\delta\|^2\\&\ge(\|E_Y^{\delta\prime}\partial_tE_X^\delta w^\delta\|_{Y^{\delta\prime}}-\alpha\|E_X^\delta w^\delta\|_Y)^2+\|E_X^\delta w^\delta\|_Y^2+\|\gamma_TE_X^\delta w^\delta\|^2\\&\qquad+(\beta-1)\|\gamma_0E_X^\delta w^\delta\|^2\\&\ge(1-\eta^2)\|E_Y^{\delta\prime}\partial_tE_X^\delta w^\delta\|_{Y^{\delta\prime}}^2+\big((1-\eta^{-2})\alpha^2+1\big)\|E_X^\delta w^\delta\|_Y^2+\|\gamma_TE_X^\delta w^\delta\|^2\\&\qquad+(\beta-1)\|\gamma_0E_X^\delta w^\delta\|^2\\&\overset{(3.2)}{\ge}(1-\eta^2)(\gamma_\Delta^{\partial_t})^2\|\partial_tE_X^\delta w^\delta\|_{Y'}^2+\big((1-\eta^{-2})\alpha^2+1\big)\|E_X^\delta w^\delta\|_Y^2+\|\gamma_TE_X^\delta w^\delta\|^2\\&\qquad+(\beta-1)\|\gamma_0E_X^\delta w^\delta\|^2\\&\ge\min\Big((1-\eta^2)(\gamma_\Delta^{\partial_t})^2,\big((1-\eta^{-2})\alpha^2+1\big)\Big)\|E_X^\delta w^\delta\|_X^2,\end{aligned}\tag{3.15}$$

where we applied Young's inequality. Solving $(1-\eta^2)(\gamma_\Delta^{\partial_t})^2=\big((1-\eta^{-2})\alpha^2+1\big)$ for η yields

$$(1-\eta^2)(\gamma_\Delta^{\partial_t})^2=\tfrac12\Big((\gamma_\Delta^{\partial_t})^2+\alpha^2+1-\sqrt{((\gamma_\Delta^{\partial_t})^2+\alpha^2+1)^2-4(\gamma_\Delta^{\partial_t})^2}\Big)>0.\tag{3.16}$$

Recalling (3.11) and $\|P^\delta\|_{\mathcal{L}(X,X)}=\sup_{0\neq w\in X}\frac{\|w^\delta\|_X}{\|w\|_X}$, the proof is completed by combining (3.12), (3.13), and (3.15). □

4. Brézis–Ekeland–Nayroles (BEN) formulation

The minimizer $u\in X$ of $\left\|\begin{bmatrix}B\\\sqrt{\beta}\gamma_0\end{bmatrix}w-\begin{bmatrix}g\\\sqrt{\beta}u_0\end{bmatrix}\right\|_{Y'\times H}^2$, that is equal to the unique solution of (2.4), is the unique solution of

$$(B'A_s^{-1}B+\beta\gamma_0'\gamma_0)u=B'A_s^{-1}g+\beta\gamma_0'u_0.\tag{4.1}$$

As we have seen in (2.6), this system is equivalent to

$$(C'A_s^{-1}C+A_s+\gamma_T'\gamma_T+(\beta-1)\gamma_0'\gamma_0)u=(\mathrm{Id}+C'A_s^{-1})g+\beta\gamma_0'u_0,\tag{4.2}$$

showing that u is the second component of the pair $(\lambda,u)\in Y\times X$ that solves

$$\begin{bmatrix}A_s&C\\C'&-(A_s+\gamma_T'\gamma_T+(\beta-1)\gamma_0'\gamma_0)\end{bmatrix}\begin{bmatrix}\lambda\\u\end{bmatrix}=\begin{bmatrix}g\\-(g+\beta\gamma_0'u_0)\end{bmatrix}.\tag{4.3}$$

Notice that $\lambda=u$.

The formulation (4.2) of the parabolic equation can alternatively be derived from the application of the Brézis–Ekeland–Nayroles variational principle ([4,31], cf. also [1, §3.2.4]), which generalizes beyond the linear, Hilbert space setting.

[3] `Reduce[{Sqrt[3]/2 <= Sqrt[(1 + 1/2*(a^2 + a*Sqrt[a^2 + 4]))/(1/2*(g^2 + a^2 + 1 - Sqrt[(g^2 + a^2 + 1)^2 - 4g^2]))] / ((1 + 1/2*(a^2 + a*Sqrt[a^2 + 4]))/g) <= 1}, {a, g}]` returns `a >= 0 && 0 < g <= 1`.

Given $\delta \in \Delta$, we consider the Galerkin discretization of (4.3), i.e.,

$$\begin{bmatrix} {E_Y^\delta}' A_s E_Y^\delta & {E_Y^\delta}' C E_X^\delta \\ ({E_Y^\delta}' C E_X^\delta)' & -{E_X^\delta}'(A_s + \gamma_T'\gamma_T + (\beta-1)\gamma_0'\gamma_0)E_X^\delta \end{bmatrix} \begin{bmatrix} \lambda^\delta \\ \bar{u}^\delta \end{bmatrix} = \begin{bmatrix} {E_Y^\delta}' g \\ -{E_X^\delta}'(g + \beta\gamma_0' u_0) \end{bmatrix} \tag{4.4}$$

or, equivalently

$$\begin{aligned} &{E_X^\delta}'\big(C' E_Y^\delta ({E_Y^\delta}' A_s E_Y^\delta)^{-1} {E_Y^\delta}' C + A_s + \gamma_T'\gamma_T + (\beta-1)\gamma_0'\gamma_0\big) E_X^\delta \bar{u}^\delta \\ &\quad = {E_X^\delta}'\big(C' E_Y^\delta ({E_Y^\delta}' A_s E_Y^\delta)^{-1} {E_Y^\delta}' g + g + \beta\gamma_0' u_0\big). \end{aligned} \tag{4.5}$$

Remark 4.1. Assuming $X^\delta \subseteq Y^\delta$ ((3.1)) and $K_Y^\delta = ({E_Y^\delta}' A_s E_Y^\delta)^{-1}$, it holds that $\bar{u}^\delta = u^\delta$, i.e., the solutions of BEN and MR *are equal*. Indeed, (3.14) shows that in this case the operator at the left-hand side of (4.5) equals the operator in (3.6), and from ${E_X^\delta}' A_s E_Y^\delta ({E_Y^\delta}' A_s E_Y^\delta)^{-1} {E_Y^\delta}' = {E_X^\delta}'$ when $X^\delta \subseteq Y^\delta$ one deduces that also the right-hand sides agree.

In contrast to MR, with BEN, however, it is *not* possible to replace $({E_Y^\delta}' A_s E_Y^\delta)^{-1}$ by a general preconditioner as in (3.7)-(3.6) and still obtain a quasi-best approximation to (λ, u) from $Y^\delta \times X^\delta$. This can be understood by noticing that replacing A_s^{-1} in (4.2) by another operator changes the solution, whereas this is not the case in (4.1). So for the iterative solution of BEN one has to operate on the saddle point system (4.4) instead of on a symmetric positive definite system as with MR, see (3.6).

On the other hand, with BEN it is not needed that $X^\delta \subseteq Y^\delta$, as we will see below.

The applicability of BEN for the case that $A_a \neq 0$ was already demonstrated in [37]. The following result gives a quantitatively better error bound.

Theorem 4.2. *Under the sole condition (3.2), the solution $\bar{u}^\delta \in X^\delta$ of (4.5) exists uniquely, and satisfies*

$$\|u - \bar{u}^\delta\|_X \leq \frac{\big(1 + \frac{1}{2}\big(\alpha^2 + \alpha\sqrt{\alpha^2+4}\big)\big) \inf_{w \in X^\delta} \|u - w\|_X + \sqrt{1+\alpha^2} \inf_{v \in Y^\delta} \|u - v\|_Y}{\frac{1}{2}\big((\gamma_\Delta^{\partial_t})^2 + \alpha^2 + 1 - \sqrt{((\gamma_\Delta^{\partial_t})^2 + \alpha^2 + 1)^2 - 4(\gamma_\Delta^{\partial_t})^2}\big)}.$$

Proof. With $g = Bu$ and $u_0 = \gamma_0 u$, using $B = C + A_s$ and $\gamma_0'\gamma_0 = \gamma_T'\gamma_T - (C' + C)$, the right-hand side of (4.5) reads as

$$\begin{aligned} &{E_X^\delta}'\big(C' E_Y^\delta ({E_Y^\delta}' A_s E_Y^\delta)^{-1} {E_Y^\delta}' (C + A_s) + A_s + \gamma_T'\gamma_T + (\beta-1)\gamma_0'\gamma_0 - C'\big) u = \\ &{E_X^\delta}'\big(C' E_Y^\delta ({E_Y^\delta}' A_s E_Y^\delta)^{-1} {E_Y^\delta}' C + A_s + \gamma_T'\gamma_T + (\beta-1)\gamma_0'\gamma_0 \\ &\qquad + C'\big[E_Y^\delta ({E_Y^\delta}' A_s E_Y^\delta)^{-1} {E_Y^\delta}' A_s - \mathrm{Id}\big]\big) u. \end{aligned}$$

So with $G(\delta) := C' E_Y^\delta ({E_Y^\delta}' A_s E_Y^\delta)^{-1} {E_Y^\delta}' C + A_s + \gamma_T'\gamma_T + (\beta-1)\gamma_0'\gamma_0$, it holds that

$$\begin{aligned} &u \mapsto E_X^\delta \bar{u}^\delta \\ &= E_X^\delta ({E_X^\delta}' G(\delta) E_X^\delta)^{-1} {E_X^\delta}'\big(G(\delta) + C'\big[E_Y^\delta ({E_Y^\delta}' A_s E_Y^\delta)^{-1} {E_Y^\delta}' A_s - \mathrm{Id}\big]\big) u, \end{aligned}$$

where we already used that ${E_X^\delta}' G(\delta) E_X^\delta$ is invertible, which will be verified below. Since $E_X^\delta ({E_X^\delta}' G(\delta) E_X^\delta)^{-1} {E_X^\delta}' G(\delta) \in \mathcal{L}(X, X)$ and $E_Y^\delta ({E_Y^\delta}' A_s E_Y^\delta)^{-1} {E_Y^\delta}' A_s \in \mathcal{L}(Y, Y)$ are projectors onto X^δ and Y^δ, respectively, the latter being orthogonal, for any $v \in Y^\delta$ and $w \in X^\delta$ it holds that

$$\begin{aligned} &u - \bar{u}^\delta = (\mathrm{Id} - E_X^\delta ({E_X^\delta}' G(\delta) E_X^\delta)^{-1} {E_X^\delta}' G(\delta))(u - E_X^\delta w) \\ &+ E_X^\delta ({E_X^\delta}' G(\delta) E_X^\delta)^{-1} {E_X^\delta}' C'\big[\mathrm{Id} - E_Y^\delta ({E_Y^\delta}' A_s E_Y^\delta)^{-1} {E_Y^\delta}' A_s\big](u - E_Y^\delta v) \end{aligned}$$

and so, also using $Y^\delta \notin \{0, Y\}$,

$$\|u - \bar{u}^\delta\|_X \leq \|({E_X^\delta}' G(\delta) E_X^\delta)^{-1}\|_{\mathcal{L}(X^{\delta\prime}, X^\delta)} \Big\{ \|G(\delta)\|_{\mathcal{L}(X,X')} \inf_{w \in X^\delta} \|u - w\|_X + \|C\|_{\mathcal{L}(X,Y')} \inf_{v \in Y^\delta} \|u - v\|_Y \Big\}.$$

For $w \in X$, we have

$$\begin{aligned} (G(\delta)w)(w) &= \|{E_Y^\delta}' C w\|_{Y^{\delta\prime}}^2 + \|w\|_Y^2 + \|\gamma_T w\|^2 + (\beta-1)\|\gamma_0 w\|^2 \\ &\leq \|Cw\|_{Y'}^2 + \|w\|_Y^2 + \|\gamma_T w\|^2 + (\beta-1)\|\gamma_0 w\|^2 \\ &= ((C' A_s^{-1} C + A_s + \gamma_T'\gamma_T + (\beta-1)\gamma_0'\gamma_0)w)(w) \\ &= \|Bw\|_{Y'}^2 + \beta\|\gamma_0 w\|^2 \\ &\leq \Big(1 + \tfrac{1}{2}\Big(\alpha^2 + \alpha\sqrt{\alpha^2+4}\Big)\Big)\|w\|_X^2 \end{aligned}$$

by Proposition 2.2. Since $(G(\delta)\cdot)(\cdot)$ is symmetric semi-positive-definite, we conclude that $\|G(\delta)\|_{\mathcal{L}(X,X')} \leq 1 + \frac{1}{2}\Big(\alpha^2 + \alpha\sqrt{\alpha^2+4}\Big)$.

For $w \in X^\delta$, one deduces

$$\begin{aligned} &(G(\delta)E_X^\delta w)(E_X^\delta w) \\ &= \|{E_Y^\delta}' C E_X^\delta w\|_{Y^{\delta\prime}}^2 + \|E_X^\delta w\|_Y^2 + \|\gamma_T E_X^\delta w\|^2 + (\beta-1)\|\gamma_0 E_X^\delta w\|^2 \\ &\geq \tfrac{1}{2}\Big((\gamma_\Delta^{\partial_t})^2 + \alpha^2 + 1 - \sqrt{((\gamma_\Delta^{\partial_t})^2 + \alpha^2 + 1)^2 - 4(\gamma_\Delta^{\partial_t})^2}\Big)\|E_X^\delta w\|_X^2 \end{aligned}$$

by following the lines starting at the second line of (3.15), in particular showing that ${E_X^\delta}' G(\delta) E_X^\delta$ is invertible.

Finally, for $w \in X$, $\|Cw\|_{Y'} \leq \|\partial_t w\|_{Y'} + \alpha\|w\|_Y \leq \sqrt{1+\alpha^2}\,\|w\|_X$. The theorem follows by combining the above estimates. □

5. Inf-sup condition (3.2), i.e., $\gamma_\Delta^{\partial_t} > 0$, and condition (3.3)

By the boundedness and coercivity assumptions (2.1) and (2.5), it holds that $\|\cdot\|_Y \eqsim \|\cdot\|_{L_2(I;V)}$. Since with

$$\gamma^\delta := \gamma^\delta(X^\delta, Y^\delta) := \inf_{\{w \in X^\delta :\, \partial_t w \neq 0\}} \sup_{0 \neq v \in Y^\delta} \frac{\int_I \langle \partial_t w, v\rangle \mathrm{d}t}{\|\partial_t w\|_{L_2(I;V')}\|v\|_{L_2(I;V)}}, \tag{5.1}$$

consequently it holds that $\gamma_\Delta^{\partial_t} \eqsim \inf_{\delta \in \Delta} \gamma^\delta$, we will summarize some known results about settings for which $\inf_{\delta\in\Delta}\gamma^\delta > 0$ has been demonstrated.

In the final subsection of this section we will briefly comment on the construction of preconditioners at the Y-side, i.e. condition (3.3), and the X-side. The preconditioner K_Y^δ has its application for the reduction of the saddle-point system (3.7) (reading $(K_Y^\delta)^{-1}$ as ${E_Y^\delta}' A_s E_Y^\delta$) to the elliptic system (3.6), and as an ingredient for building a preconditioner for the saddle-point system (4.4), whereas K_X^δ can be applied for preconditioning (3.6), and as the other ingredient to construct a preconditioner for (4.4).

Since inf-sup or Ladyzhenskaya–Babuška–Brezzi (LBB) conditions of type $\gamma^\delta > 0$ will be encountered often, in an abstract framework in the following Proposition 5.1 we establish their relation to existence of a Fortin operator, denoted by Q. Since the work of Fortin ([23]), it is well-known that existence of such an operator implies the LBB condition. We show that also the converse is true, and present a quantitatively optimal statement. Moreover, in contrast to the common presentation (although not in [23]), in view of applications the operator F in Proposition 5.1 is not required to be injective. The estimates from [20, Lemma 26.9], which apply under the 'continuous' inf-sup condition $\inf_{0\neq a\in\mathscr{A}} \frac{\|Fa\|_{\mathscr{B}'}}{\|a\|_{\mathscr{A}}} > 0$, are in that case similar to those from Proposition 5.1, and can easily be derived from this result.

Proposition 5.1. *For Hilbert spaces $\mathscr{A}$ and $\mathscr{B}$, let $F \in \mathcal{L}(\mathscr{A}, \mathscr{B}')$. Let $\mathfrak{A} \subset \mathscr{A}$ and $\mathfrak{B} \subset \mathscr{B}$ be closed subspaces with $F\mathfrak{A} \neq \{0\}$ and $\mathfrak{B} \neq \{0\}$. Let $E_{\mathfrak{A}} : \mathfrak{A} \to \mathscr{A}$ and $E_{\mathfrak{B}} : \mathfrak{B} \to \mathscr{B}$ denote the trivial embeddings, which we sometimes write for clarity, but that we mainly introduce for their duals. If there exists a*

$$Q \in \mathcal{L}(\mathscr{B},\mathscr{B}) \text{ with } \operatorname{ran} Q \subset \mathfrak{B} \text{ and } (F\mathfrak{A})((\mathrm{Id}-Q)\mathscr{B}) = 0, \tag{5.2}$$

then $\mathfrak{G} := \inf_{\{\mathfrak{a}\in\mathfrak{A}:\ F\mathfrak{a}\neq 0\}} \frac{\|E'_{\mathfrak{B}}FE_{\mathfrak{A}}\mathfrak{a}\|_{\mathfrak{B}'}}{\|F\mathfrak{a}\|_{\mathscr{B}'}} \geq \|Q\|^{-1}_{\mathcal{L}(\mathscr{B},\mathscr{B})}$. *Conversely, if* $\mathfrak{G} > 0$, *and* $\operatorname{ran} E'_{\mathfrak{B}}FE_{\mathfrak{A}}$ *is closed, then a* Q *as in* (5.2) *exists, which moreover is a projector, with* $\|Q\|_{\mathcal{L}(\mathscr{B},\mathscr{B})} = 1/\mathfrak{G}$. *The condition of the closedness of* $\operatorname{ran} E'_{\mathfrak{B}}FE_{\mathfrak{A}}$ *can be replaced by* $\dim\mathfrak{A} < \infty$, *or by the closedness of* $\operatorname{ran} F$.

Proof. This proof resembles that of [18, Thm. 3.11], but yields quantitatively optimal bounds.

If a Q as in (5.2) exists, then for $\mathfrak{a}\in\mathfrak{A}$ it holds that

$$\|F\mathfrak{a}\|_{\mathscr{B}'} = \sup_{0\neq\beta\in\mathscr{B}} \frac{(F\mathfrak{a})(\beta)}{\|\beta\|_{\mathscr{B}}} = \sup_{0\neq\beta\in\mathscr{B}} \frac{(F\mathfrak{a})(Q\beta)}{\|\beta\|_{\mathscr{B}}} \leq \|Q\|_{\mathcal{L}(\mathscr{B},\mathscr{B})} \sup_{0\neq\mathfrak{b}\in\mathfrak{B}} \frac{(F\mathfrak{a})(\mathfrak{b})}{\|\mathfrak{b}\|_{\mathscr{B}}},$$

or $\mathfrak{G} \geq \|Q\|^{-1}_{\mathcal{L}(\mathscr{B},\mathscr{B})}$.

Now let $\mathfrak{G} > 0$. By the open mapping, the closedness of $\operatorname{ran} F$ is equivalent to $\|F[\alpha]\|_{\mathscr{B}'} \eqsim \|[\alpha]\|_{\mathscr{A}/\ker F}$ ($[\alpha]\in\mathscr{A}/\ker F$). Thanks to $\mathfrak{G} > 0$, the latter implies

$$\|E'_{\mathfrak{B}}FE_{\mathfrak{A}}[\mathfrak{a}]\|_{\mathfrak{B}'} \eqsim \|[\mathfrak{a}]\|_{\mathscr{A}/\ker F} \quad ([\mathfrak{a}]\in\mathfrak{A}/\ker F), \tag{5.3}$$

which in turn is equivalent to the closedness of $\operatorname{ran} E'_{\mathfrak{B}}FE_{\mathfrak{A}}$. Obviously, the latter holds also true when $\dim\mathfrak{A} < \infty$.

With the Riesz map $R\colon \mathscr{B}\to\mathscr{B}'$, we define $Q\colon \mathscr{B}\to\mathfrak{B}\colon \beta\mapsto\mathfrak{b}$ with the latter being the first component[4] of $(\mathfrak{b},[\mathfrak{a}])\in\mathfrak{B}\times\mathfrak{A}/\ker F$ that solves

$$\begin{bmatrix} E'_{\mathfrak{B}}RE_{\mathfrak{B}} & E'_{\mathfrak{B}}FE_{\mathfrak{A}} \\ E'_{\mathfrak{A}}F'E_{\mathfrak{B}} & 0 \end{bmatrix} \begin{bmatrix} \mathfrak{b} \\ [\mathfrak{a}] \end{bmatrix} = \begin{bmatrix} 0 \\ E'_{\mathfrak{A}}F'\beta \end{bmatrix}.$$

We will see that this system is uniquely solvable.

We equip $\mathfrak{A}/\ker F$ with norm $\|E'_{\mathfrak{B}}FE_{\mathfrak{A}}\cdot\|_{\mathfrak{B}'}$. Thanks to (5.3), with this norm and corresponding scalar product, $\mathfrak{A}/\ker F$ is a Hilbert space, which implies the surjectivity of the corresponding Riesz map.

One verifies that both $E'_{\mathfrak{B}}RE_{\mathfrak{B}}\colon \mathfrak{B}\to\mathfrak{B}'$ and the Schur complement $S := E'_{\mathfrak{A}}F'E_{\mathfrak{B}}(E'_{\mathfrak{B}}RE_{\mathfrak{B}})^{-1}E'_{\mathfrak{B}}FE_{\mathfrak{A}}\colon \mathfrak{A}/\ker F\to(\mathfrak{A}/\ker F)'$ are Riesz maps. Using $S[\mathfrak{a}] = E'_{\mathfrak{A}}F'\beta$, we infer that

$$\|\mathfrak{b}\|_{\mathscr{B}} = \|E'_{\mathfrak{B}}FE_{\mathfrak{A}}[\mathfrak{a}]\|_{\mathfrak{B}'} = \|[\mathfrak{a}]\|_{\mathfrak{A}/\ker F} = \|E'_{\mathfrak{A}}F'\beta\|_{(\mathfrak{A}/\ker F)'}.$$

From

$$\begin{aligned}\|E'_{\mathfrak{A}}F'\|_{\mathcal{L}(\mathscr{B},(\mathfrak{A}/\ker F)')} &= \|FE_{\mathfrak{A}}\|_{\mathcal{L}(\mathfrak{A}/\ker F,\mathscr{B}')} \\ &= \sup_{\{\mathfrak{a}\in\mathfrak{A}:\ F\mathfrak{a}\neq 0\}} \inf_{0\neq\mathfrak{b}\in\mathfrak{B}} \frac{\|F\mathfrak{a}\|_{\mathscr{B}'}\|\mathfrak{b}\|_{\mathscr{B}}}{(F\mathfrak{a})(\mathfrak{b})} = 1/\mathfrak{G},\end{aligned}$$

we conclude that $\|Q\|_{\mathcal{L}(\mathscr{B},\mathscr{B})} = 1/\mathfrak{G}$, which completes the proof. □

5.1. 'Full' tensor product case

Concerning the verification of $\inf_{\delta\in\Delta}\gamma^\delta > 0$, we start with the easy case of X^δ and Y^δ being *'full' tensor products* of approximation spaces in time and space (as opposed to sparse tensor products, see below). With $Y_t := L_2(I)$ and $X_t := H^1(I)$, for $Z\in\{X,Y\}$ let $(Z_t^\delta)_{\delta\in\Delta}$ and $(Z_{\mathbf{x}}^\delta)_{\delta\in\Delta}$ be families of closed subspaces of Z_t and V, respectively, and let $Z^\delta := Z_t^\delta\otimes Z_{\mathbf{x}}^\delta$. Assuming that

$$\gamma_t^\delta := \inf_{\{w\in X_t^\delta:\ w'\neq 0\}} \sup_{0\neq v\in Y_t^\delta} \frac{\int_I w'v\,\mathrm{d}t}{\|w'\|_{L_2(I)}\|v\|_{L_2(I)}} \gtrsim 1, \tag{5.4}$$

$$\gamma_{\mathbf{x}}^\delta := \inf_{0\neq w\in X_{\mathbf{x}}^\delta} \sup_{0\neq v\in Y_{\mathbf{x}}^\delta} \frac{\langle w,v\rangle}{\|w\|_{V'}\|v\|_V} \gtrsim 1, \tag{5.5}$$

a tensor product argument shows that

$$\gamma^\delta = \gamma_t^\delta\gamma_{\mathbf{x}}^\delta \gtrsim 1.$$

[4] One may verify that $\mathfrak{b} = \operatorname{argmin}_{\{\tilde{\mathfrak{b}}:\ (F\mathfrak{A})(\beta-\tilde{\mathfrak{b}})=0\}}\|\tilde{\mathfrak{b}}\|_{\mathscr{B}}$.

Obviously, (5.4) is true when $\frac{\mathrm{d}}{\mathrm{d}t}X_t^\delta \subseteq Y_t^\delta$, which however is not a necessary condition. For example, when X_t^δ is the space of continuous piecewise linears w.r.t. some partition of I, and Y_t^δ is the space of continuous piecewise linears w.r.t. a once dyadically refined partition, an easy computation ([2, Prop. 6.1]) shows that $\gamma_t^\delta \geq \sqrt{3/4}$.

Considering, for a domain $\Omega\subset\mathbb{R}^d$ and $\Gamma\subset\partial\Omega$, $H = L_2(\Omega)$ and $V = H^1_{0,\Gamma}(\Omega) := \{v\in H^1(\Omega):\ v|_\Gamma = 0\}$, $H^1(\Omega)$-stability of the $L_2(\Omega)$-orthogonal projector onto Lagrange finite element spaces $X_{\mathbf{x}}^\delta = Y_{\mathbf{x}}^\delta$ is an extensively studied subject. In view of Proposition 5.1, taking F to be the Riesz map $H\to H'$ viewed as a mapping $V\to V'$, this stability implies (5.5). For finite element spaces w.r.t. shape regular quasi-uniform partitions into, say, d-simplices, where Γ is the union of faces of $T\in\mathcal{T}$, this stability follows easily from direct and inverse estimates. It is known that this stability holds also true for (shape regular) locally refined partitions when they are sufficiently mildly graded. In [24], it is shown that in two space dimensions the meshes generated by newest vertex bisection satisfy this requirement, see also [17] for extensions.

5.2. Sparse tensor product case

As shown in [2, Prop. 4.2], these results for full tensor products extend to *sparse tensor products*. When $(Z_t^\delta)_{\delta\in\Delta}$ and $(Z_{\mathbf{x}}^\delta)_{\delta\in\Delta}$ are nested sequences of closed subspaces $Z_t^{\delta_0}\subset Z_t^{\delta_1}\subset\cdots\subset Z_t$, $Z_{\mathbf{x}}^{\delta_0}\subset Z_{\mathbf{x}}^{\delta_1}\subset\cdots\subset V$ that satisfy (5.4)–(5.5), then for $Z^{\delta_n} := \sum_{\{0\leq n_t+n_{\mathbf{x}}\leq n\}} Z_t^{\delta_{n_t}}\otimes Z_{\mathbf{x}}^{\delta_{n_{\mathbf{x}}}}$ it holds that

$$\gamma^{\delta_n} \geq \min_{0\leq n_t\leq n}\gamma_t^{\delta_{n_t}} \min_{0\leq n_{\mathbf{x}}\leq n}\gamma_t^{\delta_{n_{\mathbf{x}}}} \gtrsim 1.$$

5.3. Time-slab partition case

Another extension of the full tensor product case is given by the following. Let $(\bar{X}^\delta,\bar{Y}^\delta)_{\delta\in\bar{\Delta}}$ be a family of pairs of closed subspaces of X and Y for which

$$\gamma_{\bar{\Delta}} := \inf_{\delta\in\bar{\Delta}} \inf_{\{w\in\bar{X}^\delta:\ \partial_t w\neq 0\}} \sup_{0\neq v\in\bar{Y}^\delta} \frac{\int_I\langle\partial_t w,v\rangle\mathrm{d}t}{\|w\|_{L_2(I;V')}\|v\|_{L_2(I;V)}} > 0.$$

Then if, for $\delta\in\Delta$, X^δ and Y^δ are such that for some finite partition $I^\delta = ([t_{i-1}^\delta,t_i^\delta])_i$ of I, with $G_i^\delta(t) := t_{i-1}^\delta + \frac{t}{T}(t_i^\delta - t_{i-1}^\delta)$ and arbitrary $\delta_i\in\bar{\Delta}$ it holds that

$$\begin{aligned} X^\delta &\subseteq \{u\in X:\ u|_{(t_{i-1}^\delta,t_i^\delta)}\circ G_i^\delta \in \bar{X}^{\delta_i}\}, \\ Y^\delta &\supseteq \{v\in L_2(I;V):\ v|_{(t_{i-1}^\delta,t_i^\delta)}\circ G_i^\delta \in \bar{Y}^{\delta_i}\}, \end{aligned}$$

then $\gamma^\delta \geq \gamma_{\bar{\Delta}} > 0$ as one easily verifies by writing $\int_I\langle\frac{\mathrm{d}u}{\mathrm{d}t},v\rangle\mathrm{d}t = \sum_i\int_{t_{i-1}}^{t_i}\langle\frac{\mathrm{d}u}{\mathrm{d}t},v\rangle\mathrm{d}t$. An example of this *'time-slab partition'* setting will be given in Sect. 7. Thinking of the $\bar{X}^\delta$ as being finite element spaces, notice that the condition $X^\delta\subset X$ will require that possible 'hanging nodes' on the interface between different time slabs do not carry degrees of freedom.

5.4. Generalized sparse tensor product case

Finally, we informally describe a *'generalized' sparse tensor product* setting that allows for *local refinements* driven by an a posteriori error estimator. For $Z\in\{X,Y\}$, let the nested sequences of closed subspaces $Z_t^{\delta_0}\subset Z_t^{\delta_1}\subset\cdots\subset Z_t$, $Z_{\mathbf{x}}^{\delta_0}\subset Z_{\mathbf{x}}^{\delta_1}\subset\cdots\subset V$ be equipped with hierarchical bases, meaning that the basis for $Z_t^{\delta_i}$ (analogously $Z_{\mathbf{x}}^{\delta_i}$) is inductively defined as the basis for $Z_t^{\delta_{i-1}}$ plus a basis for a complement space of $Z_t^{\delta_{i-1}}$ in $Z_t^{\delta_i}$. The *level* of the functions in the latter basis is defined as i.

Let us consider the usual case that the diameter of the support of a hierarchical basis function with level i is $\eqsim 2^{-i}$, and let us assign to each basis function ϕ on level $i>0$ one (or a few) parents with level $i-1$ whose supports intersect the support of ϕ. We now let $(Z^\delta)_{\delta\in\Delta}$

be the collection of all spaces that are spanned by sets of product hierarchical basis functions, which sets are *downward closed* (or *lower*) in the sense that if a product of basis functions is in the set, then so are all their parents in both directions. Note that the sparse tensor product spaces $\sum_{\{0\le n_t+n_{\mathbf{x}}\le n\}} Z_t^{\delta_{n_t}}\otimes Z_{\mathbf{x}}^{\delta_{n_{\mathbf{x}}}}$ are included in this collection, but that it contains many more spaces.

Under conditions on the hierarchical bases for $Z_t^{\delta_0}\subset Z_t^{\delta_1}\subset\cdots\subset Z_t$ for $Z\in\{X,Y\}$, that should be of *wavelet-type*, in [36] it is shown that to any X^δ one can assign a Y^δ with $\dim Y^\delta\lesssim\dim X^\delta$, such that $\gamma^\delta\gtrsim 1$ holds.

5.5. Preconditioners

Moving to condition (3.3), obviously we would like to construct K_Y^δ such that it is not only a uniform preconditioner, i.e., it satisfies (3.3), but also that its application can be performed in $\mathcal{O}(\dim Y^\delta)$ operations. In the full-tensor product case, after selecting bases for Y_t^δ and $Y_{\mathbf{x}}^\delta$, the construction of K_Y^δ boils down to tensorizing approximate inverses of the 'mass matrix' in time, which does not pose any problems, and the 'stiffness matrix' in space. For $V=H^1(\Omega)$ (or a subspace of aforementioned type), it is well-known that by taking a multi-grid preconditioner as the approximate inverse of the stiffness matrix the resulting K_Y^δ satisfies our needs. A straightforward generalization of this construction of K_Y^δ applies to spaces Y^δ that correspond to the time-slab partitioning approach.

Finally, for the efficient iterative solution of (3.6) or (4.4), one needs a $K_X^\delta=K_X^{\delta\prime}\in\mathcal{L}\mathrm{is}(X^{\delta\prime},X^\delta)$ whose norm and norm of its inverse are uniformly bounded, and whose application can be performed in $\mathcal{O}(\dim X^\delta)$ operations. For the full/sparse and generalized sparse tensor product setting such preconditioners have been constructed in [3] and [36], respectively.

6. Robustness

The quasi-optimality results presented in Theorems 3.1 and 4.2 for MR and BEN degenerate when $\alpha=\|A_a\|_{\mathcal{L}(Y,Y')}\to\infty$. Aiming at results that are robust for $\alpha\to\infty$, we now study convergence w.r.t. the energy-norm $|||\cdot|||_X$ on X. On its own this change of norms turns out not to be helpful. By replacing $\|\cdot\|_X$ by $|||\cdot|||_X$ in Theorems 3.1 and 4.2, and adapting their proofs in an obvious way yields for MR the same upper bound for $\frac{|||u-u^\delta|||_X}{\inf_{w\in X^\delta}|||u-w|||_X}$ as we found for $\frac{\|u-u^\delta\|_X}{\inf_{w\in X^\delta}\|u-w\|_X}$ (for $u\notin X^\delta$), whereas instead of Theorem 4.2 we arrive at the only slightly more favorable bound

$$|||u-\bar{u}^\delta|||_X \le \frac{2+\alpha^2+\alpha\sqrt{\alpha^2+4}}{(\gamma_\Delta^{\partial_t})^2+\alpha^2+1-\sqrt{((\gamma_\Delta^{\partial_t})^2+\alpha^2+1)^2-4(\gamma_\Delta^{\partial_t})^2}}\inf_{w\in X^\delta,v\in Y^\delta}|||u-w|||_X+\|u-v\|_Y,$$

which is, however, still far from being robust.

In order to obtain robust bounds, instead of the condition $\gamma_\Delta^{\partial_t}>0$ ((3.2)) we now impose

$$\gamma_\Delta^C:=\inf_{\delta\in\Delta}\inf_{\{0\neq w\in X^\delta:\ CE_X^\delta w\neq 0\}}\frac{\|E_Y^{\delta\prime}CE_X^\delta w\|_{Y^{\delta\prime}}}{\|CE_X^\delta w\|_{Y'}}>0, \tag{6.1}$$

which, when considering a family of operators A, we would like to hold uniformly for $\alpha\to\infty$.

Theorem 6.1. *Under conditions (3.1), (6.1), and (3.3), the solution $u^\delta\in X^\delta$ of (3.6) satisfies*

$$|||u-u^\delta|||_X\le\sqrt{\tfrac{\max(R_\Delta,1)}{\min(r_\Delta,1)}}(\gamma_\Delta^C)^{-1}\inf_{w\in X^\delta}|||u-w|||_X; \tag{6.2}$$

and under condition (6.1), the solution $\bar{u}^\delta\in X^\delta$ of (4.5) satisfies

$$|||u-\bar{u}^\delta|||_X\le(\gamma_\Delta^C)^{-2}\big\{\inf_{w\in X^\delta}|||u-w|||_X+\inf_{v\in Y^\delta}\|u-v\|_Y\big\}. \tag{6.3}$$

Proof. The first estimate follows from ignoring the last inequality in (3.12), and by replacing the first inequality in (3.15) by

$$\begin{aligned}&\|E_Y^{\delta\prime}CE_X^\delta w^\delta\|_{Y^{\delta\prime}}^2+\|E_X^\delta w^\delta\|_Y^2+\|\gamma_T E_X^\delta w^\delta\|^2+(\beta-1)\|\gamma_0E_X^\delta w^\delta\|^2\\ &\ge(\gamma_\Delta^C)^2\Big(\|CE_X^\delta w^\delta\|_{Y'}^2+\|E_X^\delta w^\delta\|_Y^2+\|\gamma_T E_X^\delta w^\delta\|^2+(\beta-1)\|\gamma_0E_X^\delta w^\delta\|^2\Big)\\ &=(\gamma_\Delta^C)^2\big((E_X^{\delta\prime}B'A_s^{-1}BE_X^\delta+E_X^{\delta\prime}\beta\gamma_0'\gamma_0E_X^\delta)w^\delta\big)(w^\delta)=(\gamma_\Delta^C)^2|||w^\delta|||_X^2.\end{aligned}$$

By following the proof of Theorem 4.2, recalling that now X is equipped with $|||\cdot|||_X$, from $\|C\|_{\mathcal{L}(X,Y')}\le 1$, $\|G(\delta)\|_{\mathcal{L}(X,X')}\le 1$, and $\|(E_X^{\delta\prime}G(\delta)E_X^\delta)^{-1}\|_{\mathcal{L}(X^{\delta\prime},X^\delta)}\le(\gamma_\Delta^C)^{-2}$, one infers the estimate for BEN. □

We conclude that for a family of (asymmetric) operators A robustness w.r.t. $|||\cdot|||_X$ is obtained when $(\gamma_\Delta^C)^{-1}$ is uniformly bounded for $\alpha=\|A_a\|_{\mathcal{L}(Y,Y')}\to\infty$. A family for which this will be realized is presented in Sect. 7.

6.1. A posteriori error estimation

In particular because for $\alpha=\|A_a\|_{\mathcal{L}(Y,Y')}\to\infty$ meaningful a priori error bounds for $\inf_{w\in X^\delta}|||u-w|||_X$ will be hard to derive, it is important to have (robust) a posteriori error bounds.

Let $Q_B^\delta\in\mathcal{L}(Y,Y)$ be such that $\operatorname{ran}Q_B^\delta\subset Y^\delta$ and $(\mathrm{Id}-Q_B^{\delta\prime})BX^\delta=0$. Then, with $e_{\mathrm{osc}}^\delta(g):=\|(\mathrm{Id}-Q_B^{\delta\prime})g\|_{Y'}$, for $w\in X^\delta$ and u the solution of (2.4) it holds that

$$\begin{aligned}r_\Delta\|E_Y^{\delta\prime}(g-Bw)\|_{K_Y^\delta}^2+\beta\|u_0-\gamma_0w\|^2\le|||u-w|||_X^2\le\\ \big(\|Q_B^\delta\|_{\mathcal{L}(Y,Y)}\sqrt{R_\Delta}\|E_Y^{\delta\prime}(g-Bw)\|_{K_Y^\delta}+e_{\mathrm{osc}}^\delta(g)\big)^2+\beta\|u_0-\gamma_0w\|^2,\end{aligned}$$

which follows from $\|g-Bw\|_{Y^{\delta\prime}}\le\|g-Bw\|_{Y'}\le\|Q_B^{\delta\prime}(g-Bw)\|_{Y'}+e_{\mathrm{osc}}^\delta(g)$.

We infer that if $\sup_{\delta\in\Delta}\|Q_B^\delta\|_{\mathcal{L}(Y,Y)}<\infty$, then the a posteriori error estimator

$$\mathcal{E}^\delta(w;g,u_0,\beta):=\sqrt{\|E_Y^{\delta\prime}(g-Bw)\|_{K_Y^\delta}^2+\beta\|u_0-\gamma_0w\|^2} \tag{6.4}$$

is an efficient and, modulo the *data oscillation term* $e_{\mathrm{osc}}^\delta(g)$, reliable estimator of the error $|||u-w|||_X$. If $\sup_{\delta\in\Delta}\|Q_B^\delta\|_{\mathcal{L}(Y,Y)}$ and $\frac{\max(R_\Delta,1)}{\min(r_\Delta,1)}$ are bounded uniformly in $\alpha\to\infty$, then this estimator is even robust.

Remark 6.2. In view of Proposition 5.1, the aforementioned assumptions $\operatorname{ran}Q_B^\delta\subset Y^\delta$, $(\mathrm{Id}-Q_B^{\delta\prime})BX^\delta=0$, and $\sup_{\delta\in\Delta}\|Q_B^\delta\|_{\mathcal{L}(Y,Y)}<\infty$ are equivalent to

$$\gamma_\Delta^B:=\inf_{\delta\in\Delta}\inf_{\{0\neq w\in X^\delta:\ BE_X^\delta w\neq 0\}}\frac{\|E_Y^{\delta\prime}BE_X^\delta w\|_{Y^{\delta\prime}}}{\|BE_X^\delta w\|_{Y'}}>0.$$

In applications the conditions $\gamma_\Delta^{\partial_t}>0$, $\gamma_\Delta^C>0$, and $\gamma_\Delta^B>0$ are increasingly more difficult to fulfill.

To have a meaningful reliability result, in addition we would like to find above Q_B^δ such that, for sufficiently smooth g, the term $e_{\mathrm{osc}}^\delta(g)$ is asymptotically, i.e. for the 'mesh-size' tending to zero, of equal or higher order than the approximation error $\inf_{w\in X^\delta}|||u-w|||_X$. We will realize this in the setting that will be discussed in Sect. 7.2.

7. Spatial differential operators with dominating asymmetric part

For some domain $\Omega\subset\mathbb{R}^d$, and $\Gamma\subset\partial\Omega$, let

$$\begin{aligned}&H:=L_2(\Omega),\ V:=H^1_{0,\Gamma}(\Omega):=\{v\in H^1(\Omega):\ v|_\Gamma=0\},\\ &a(t;\eta,\zeta):=\int_\Omega\varepsilon\nabla\eta\cdot\nabla\zeta+(\mathbf{b}\cdot\nabla\eta+e\eta)\zeta\,\mathrm{d}\mathbf{x},\quad\varepsilon>0,\\ &\mathbf{b}\in L_\infty(I;L_\infty(\mathrm{div};\Omega)),\ e\in L_\infty(I\times\Omega),\ \operatorname{ess\,inf}(e-\tfrac12\operatorname{div}_{\mathbf{x}}\mathbf{b})\ge 0,\end{aligned} \tag{7.1}$$

and $|\Gamma|>0$ when the latter ess inf is zero, so that (2.1) and (2.5) are valid. In this setting, the operators A_a, $A_s=A_s(\varepsilon)$, and so $A=A(\varepsilon)=A_s(\varepsilon)+A_a$, are given by

$$(A_a w)(v)=\int_I\int_\Omega (\mathbf{b}\cdot\nabla_{\mathbf{x}} w+\tfrac{1}{2}w\,\mathrm{div}_{\mathbf{x}}\,\mathbf{b})v\,\mathrm{d}\mathbf{x}\mathrm{d}t,$$
$$(A_s(\varepsilon)w)(v)=\int_I\int_\Omega \varepsilon\nabla_{\mathbf{x}} w\cdot\nabla_{\mathbf{x}} v+(e-\tfrac{1}{2}\,\mathrm{div}_{\mathbf{x}}\,\mathbf{b})wv\,\mathrm{d}\mathbf{x}\mathrm{d}t.$$

Thinking of $\mathbf{b}$ and e fixed, and variable $\varepsilon>0$, one infers that $\alpha=\alpha(\varepsilon)\to\infty$ when $\varepsilon\downarrow 0$ (cf. Remark 2.3).

In the next subsection we will construct $(X^\delta)_{\delta\in\Delta}\subset X$ and $(Y^\delta)_{\delta\in\Delta}\subset Y$ that (essentially) satisfy $\inf_{\varepsilon>0}\gamma_\Delta^C(\varepsilon)>0$ as families of finite element spaces w.r.t. subdivisions of $I\times\Omega$ into time-slabs with prismatic elements in each slab w.r.t. generally different partitions of Ω. Notice that although $C=\partial_t+A_a$ is independent of ε, $\gamma_\Delta^C(\varepsilon)$ depends on ε because it is defined in terms of the ε-dependent energy-norm $\|\cdot\|_Y=\sqrt{(A_s(\varepsilon)\cdot)(\cdot)}$.

As a consequence of $\gamma_\Delta^C(\varepsilon)$ being uniformly positive, for $K_Y^\delta\eqsim(E_Y^{\delta\,\prime}A_sE_Y^\delta)^{-1}$ uniformly in ε and δ, i.e., $\sup_{\varepsilon>0}\frac{\max(R_\Delta,1)}{\min(r_\Delta,1)}<\infty$, Theorem 6.1 gives ε-robust quasi-optimality results for MR and BEN w.r.t. the ε- (and β-) dependent norm $|||\cdot|||_X$.

7.1. Realization of $\inf_\varepsilon\gamma_\Delta^C(\varepsilon)>0$

Given a conforming partition $\mathcal{T}$ of a polytopal $\overline{\Omega}$ into (essentially disjoint) closed d-simplices, we define $S_{\mathcal{T}}^{-1,q}$ as the space of all (discontinuous) piecewise polynomials of degree q w.r.t. $\mathcal{T}$, and, for $q\geq 1$, set

$$S_{\mathcal{T},0}^{0,q}:=S_{\mathcal{T}}^{-1,q}\cap H^1_{0,\Gamma}(\Omega),$$

where we assume that Γ is the union of faces of $T\in\mathcal{T}$.

Let $(\mathcal{T}^\delta)_{\delta\in\bar\Delta}$, $(\mathcal{T}_S^\delta)_{\delta\in\bar\Delta}$ be families of such partitions of $\overline{\Omega}$ that are uniformly shape regular (which for $d=1$ should be read as to satisfy a uniform K-mesh property), and where $\mathcal{T}_S^\delta$ is a refinement of $\mathcal{T}^\delta$ of some *fixed* maximal depth in the sense that $|T|\gtrsim|T'|$ for $\mathcal{T}_S^\delta\ni T\subset T'\in\mathcal{T}^\delta$, so that $\dim\mathcal{T}_S^\delta\lesssim\dim\mathcal{T}^\delta$. On the other hand, fixing a $q\geq 1$, we require that the refinement from $\mathcal{T}^\delta$ to $\mathcal{T}_S^\delta$ is sufficiently deep that it permits the construction of a projector P_q^δ for which

$$\mathrm{ran}\,P_q^\delta\subseteq S_{\mathcal{T}_S^\delta,0}^{0,q},\quad \mathrm{ran}(\mathrm{Id}-P_q^\delta)\perp_{L_2(\Omega)}\big(S_{\mathcal{T}^\delta,0}^{0,q}+S_{\mathcal{T}^\delta}^{-1,q-1}\big),\tag{7.2}$$
$$\|P_q^\delta w\|_{L_2(T)}\lesssim\|w\|_{L_2(T)}\quad(T\in\mathcal{T}^\delta,\,w\in L_2(\Omega)).\tag{7.3}$$

As shown in [18, Lemma 5.1 and Rem. 5.2], regardless of the refinement rule (e.g. red-refinement or newest vertex bisection) that is (recursively) applied to create $(\mathcal{T}_S^\delta)_{\delta\in\bar\Delta}$ from $(\mathcal{T}^\delta)_{\delta\in\bar\Delta}$, there is a refinement of some fixed depth that suffices to satisfy (7.3) as well as

$$\mathrm{ran}\,P_q^\delta\subseteq\{w\in S_{\mathcal{T}_S^\delta,0}^{0,q}:\,w|_{\cup_{T\in\mathcal{T}}\partial T}=0\},\quad \mathrm{ran}(\mathrm{Id}-P_q^\delta)\perp_{L_2(\Omega)}S_{\mathcal{T}^\delta,0}^{-1,q}.\tag{7.4}$$

Condition (7.4) is stronger than (7.2), and will be relevant in Sect. 7.2 on robust a posteriori error estimation.

For $d\in\{1,2,3\}$ and $q\in\{1,2,3\}$, and both newest vertex bisection and red-refinement it was verified that it is sufficient that the aforementioned depth creates in the space $S_{\mathcal{T}_S^\delta,0}^{0,q}$ an additional number of degrees of freedom interior to any $T\in\mathcal{T}^\delta$ that is greater or equal to $\binom{q+d}{q}$.

Remark 7.1. To satisfy condition (7.2)–(7.3) generally a smaller number of degrees of freedom interior to any $T\in\mathcal{T}^\delta$ suffices. For $d=2=q$, in [18, Appendix A] it was shown that in order to satisfy (7.2)–(7.3) it is sufficient to create $\mathcal{T}_s^\delta$ from $\mathcal{T}^\delta$ by one red-refinement, which creates only three of such degrees of freedom, whereas to satisfy (7.3)–(7.4) six additional interior degrees of freedom are needed.

We show robustness of MR and BEN in a time-slab partition setting.

Theorem 7.2. *Let H, V, and $a(\cdot;\cdot,\cdot)$ be as in (7.1), with* constant $\mathbf{b}$ *and* constant $e\geq 0$, *and let $(\mathcal{T}^\delta)_{\delta\in\bar\Delta}$ and $(\mathcal{T}_S^\delta)_{\delta\in\bar\Delta}$ be as specified above. Then if, for $\delta\in\Delta$, X^δ and Y^δ are such that for some finite partition $I^\delta=([t_{i-1}^\delta,t_i^\delta])_i$ of I, and arbitrary $\delta_i\in\bar\Delta$,*

$$X^\delta\subseteq\{w\in C(I;H^1_{0,\Gamma}(\Omega)):\,w|_{(t_{i-1}^\delta,t_i^\delta)}\in\mathcal{P}_q(t_{i-1}^\delta,t_i^\delta)\otimes S_{\mathcal{T}^{\delta_i},0}^{0,q}\},$$
$$Y^\delta\supseteq\{v\in L_2(I;H^1_{0,\Gamma}(\Omega)):\,v|_{(t_{i-1}^\delta,t_i^\delta)}\in\mathcal{P}_q(t_{i-1}^\delta,t_i^\delta)\otimes S_{\mathcal{T}_S^{\delta_i},0}^{0,q}\},\tag{7.5}$$

then $\inf_{\varepsilon>0}\gamma_\Delta^C(\varepsilon)>0$. Consequently the bounds (6.2) and (6.3) show quasi-optimality of MR and BEN w.r.t. the (ε- and β-dependent) norm $|||\cdot|||_X$, uniformly in $\varepsilon>0$ and $\beta\geq 1$.

Proof. As follows from Proposition 5.1 the statement $\inf_{\varepsilon>0}\gamma_\Delta^C(\varepsilon)>0$ is equivalent to existence of $Q_C^\delta\in\mathcal{L}(Y,Y)$ with

$$\sup_{\varepsilon>0,\,\delta\in\Delta}\|Q_C^\delta\|_{\mathcal{L}(Y,Y)}<\infty,\quad \mathrm{ran}\,Q_C^\delta\subset Y^\delta,$$
$$\int_I\int_\Omega((\partial_t+\mathbf{b}\cdot\nabla_{\mathbf{x}})X^\delta)(\mathrm{Id}-Q_C^\delta)Y\,\mathrm{d}\mathbf{x}\mathrm{d}t=0,\tag{7.6}$$

where we recall that, thanks to constant $\mathbf{b}$, $Y=L_2(I;H^1_{0,\Gamma}(\Omega))$ is equipped with norm

$$\sqrt{(A_s(\varepsilon)v)(v)}=\sqrt{\int_I\varepsilon\|\nabla_{\mathbf{x}}v\|^2_{L_2(\Omega)^d}+e\|v\|^2_{L_2(\Omega)}\mathrm{d}t}$$
$$\eqsim\sqrt{\varepsilon}\|\nabla_{\mathbf{x}}v\|_{L_2(I\times\Omega)^d}+\sqrt{e}\|v\|_{L_2(I;L_2(\Omega))}.$$

It holds that

$$(\partial_t+\mathbf{b}\cdot\nabla_{\mathbf{x}})X^\delta\subseteq\big\{v\in L_2(I\times\Omega):\,v|_{(t_{i-1}^\delta,t_i^\delta)}\in\mathcal{P}_q(t_{i-1}^\delta,t_i^\delta)\otimes(S_{\mathcal{T}^{\delta_i},0}^{0,q}+S_{\mathcal{T}^{\delta_i}}^{-1,q-1})\big\}.\tag{7.7}$$

Let $(Q_{\mathbf{x}}^\delta)_{\delta\in\bar\Delta}$ denote a family of projectors such that

$$\sup_{\delta\in\Delta}\max\big(\|Q_{\mathbf{x}}^\delta\|_{\mathcal{L}(L_2(\Omega),L_2(\Omega))},\|Q_{\mathbf{x}}^\delta\|_{\mathcal{L}(H^1_{0,\Gamma}(\Omega),H^1_{0,\Gamma}(\Omega))}\big)<\infty,\tag{7.8}$$
$$\mathrm{ran}\,Q_{\mathbf{x}}^\delta\subset S_{\mathcal{T}_S^\delta,0}^{0,q},\quad \mathrm{ran}(\mathrm{Id}-Q_{\mathbf{x}}^\delta)\perp_{L_2(\Omega)}\big(S_{\mathcal{T}^\delta,0}^{0,q}+S_{\mathcal{T}^\delta}^{-1,q-1}\big),\tag{7.9}$$

and let $Q^{\delta,i}$ be the $L_2(t_{i-1}^\delta,t_i^\delta)$-orthogonal projector onto $\mathcal{P}_q(t_{i-1}^\delta,t_i^\delta)$. Then, the operator Q_C^δ, defined by

$$(Q_C^\delta v)|_{(t_{i-1}^\delta,t_i^\delta)\times\Omega}=(Q^{\delta,i}\otimes Q_{\mathbf{x}}^{\delta_i})v|_{(t_{i-1}^\delta,t_i^\delta)\times\Omega},$$

satisfies (7.6). Indeed its uniform boundedness w.r.t. the energy-norm on Y follows by the boundedness of $Q_{\mathbf{x}}^\delta$ w.r.t. both the $L_2(\Omega)$- and $H^1(\Omega)$-norms. By writing $\mathrm{Id}-Q^{\delta,i}\otimes Q_{\mathbf{x}}^{\delta_i}=(\mathrm{Id}-Q^{\delta,i})\otimes\mathrm{Id}+Q^{\delta,i}\otimes(\mathrm{Id}-Q_{\mathbf{x}}^{\delta_i})$, and using (7.7) one verifies the third condition in (7.6).

We seek $Q_{\mathbf{x}}^\delta$ of the form $Q_{\mathbf{x}}^\delta=\check{Q}_{\mathbf{x}}^\delta+\hat{Q}_{\mathbf{x}}^\delta+\hat{Q}_{\mathbf{x}}^\delta\check{Q}_{\mathbf{x}}^\delta$ where

$$\mathrm{ran}\,\check{Q}_{\mathbf{x}}^\delta,\mathrm{ran}\,\hat{Q}_{\mathbf{x}}^\delta\subset S_{\mathcal{T}_S^\delta,0}^{0,q},\quad \mathrm{ran}(\mathrm{Id}-\hat{Q}_{\mathbf{x}}^\delta)\perp_{L_2(\Omega)}(S_{\mathcal{T}^\delta,0}^{0,q}+S_{\mathcal{T}^\delta}^{-1,q-1}).\tag{7.10}$$

Then from $\mathrm{Id}-Q_{\mathbf{x}}^\delta=(\mathrm{Id}-\hat{Q}_{\mathbf{x}}^\delta)(\mathrm{Id}-\check{Q}_{\mathbf{x}}^\delta)$, we infer that (7.9) is satisfied.

We take $\hat{Q}_{\mathbf{x}}^\delta=P_q^\delta$ from (7.2)–(7.3). It satisfies the properties required in (7.10). With $\hbar_\delta$ being the piecewise constant function defined by $\hbar_\delta|_T=\mathrm{diam}\,T$ $(T\in\mathcal{T}^\delta)$, thanks to the uniform K-mesh property of $\mathcal{T}\in(\mathcal{T}^\delta)_{\delta\in\bar\Delta}$, (7.3) implies that $\|\hbar_\delta^{-1}P_q^\delta\hbar_\delta\|_{\mathcal{L}(L_2(\Omega),L_2(\Omega))}\lesssim 1$, as well as $\|P_q^\delta\|_{\mathcal{L}(L_2(\Omega),L_2(\Omega))}\lesssim 1$.

We take $\check{Q}_{\mathbf{x}}^\delta$ as a modified Scott-Zhang quasi-interpolator onto $S_{\mathcal{T}_S^\delta,0}^{0,q}$ ([25, Appendix]). The modification consists in setting the degrees of freedom on Γ to zero. When applied to a function from $H^1_{0,\Gamma}(\Omega)$ it equals the original Scott–Zhang interpolator ([39]), but thanks to the modification it is uniformly bounded w.r.t. $L_2(\Omega)$, and so $\|Q_{\mathbf{x}}^\delta\|_{\mathcal{L}(L_2(\Omega),L_2(\Omega))}$ is uniformly bounded.

Writing $Q_{\mathbf{x}}^{\delta}=\breve{Q}_{\mathbf{x}}^{\delta}+P_q^{\delta}(\mathrm{Id}-\breve{Q}_{\mathbf{x}}^{\delta})$, from $\hbar_{\delta}^{-1}(\mathrm{Id}-\breve{Q}_{\mathbf{x}}^{\delta})\in\mathcal{L}(H^1_{0,\Gamma}(\Omega),L_2(\Omega))$, $\hbar_{\delta}^{-1}P_q^{\delta}\hbar_{\delta}\in\mathcal{L}(L_2(\Omega),L_2(\Omega))$, and $\breve{Q}_{\mathbf{x}}^{\delta}\in\mathcal{L}(H^1_{0,\Gamma}(\Omega),H^1_{0,\Gamma}(\Omega))$ all being uniformly bounded, and $\|\cdot\|_{H^1(\Omega)}\lesssim\|\hbar_{\delta}^{-1}\cdot\|_{L_2(\Omega)}$ on $S^{0,q}_{\mathcal{T}_S^{\delta},0}$, we infer the uniform boundedness of $\|Q_{\mathbf{x}}^{\delta}\|_{\mathcal{L}(H^1_{0,\Gamma}(\Omega),H^1_{0,\Gamma}(\Omega))}$. □

Next under the condition that $\operatorname{ess\,inf}(e-\frac{1}{2}\operatorname{div}_{\mathbf{x}}\mathbf{b})>0$, we consider the case of *variable* $\mathbf{b}$ and e. The scaling argument that was applied directly below Theorem 2.1 shows that it is no real restriction to assume that $\operatorname{ess\,inf}(e-\frac{1}{2}\operatorname{div}_{\mathbf{x}}\mathbf{b})>0$. Although we will not be able to show $\inf_{\varepsilon>0}\gamma_{\Delta}^{C}(\varepsilon)>0$, this inf-sup condition will be valid modulo a perturbation which can be dealt with using Young's inequality similarly as in the proofs of Theorems 3.1 and 4.2. It will result in ε- (and β-) robust quasi-optimality results for MR and BEN similar as for constant $\mathbf{b}$ and constant $e\geq 0$.

Theorem 7.3. *Consider the situation of Theorem 7.2, but now without the assumption of* $\mathbf{b}$ *and* e *being constants. Assume* $\mathbf{b}\in W^1_{\infty}(I\times\Omega)^d$, $\operatorname{ess\,inf}(e-\frac{1}{2}\operatorname{div}_{\mathbf{x}}\mathbf{b})>0$, *and, only for the case that* $\mathbf{b}$ *is time-dependent,*

$$|t_{i-1}^{\delta}-t_i^{\delta}|\lesssim\max_{T\in\mathcal{T}^{\delta_i}}\operatorname{diam}(T). \tag{7.11}$$

Then for MR and BEN it holds

$$|||u-u^{\delta}|||_X\lesssim\tfrac{\max(R_{\Delta},1)}{\min(r_{\Delta},1)}\inf_{w\in X^{\delta}}|||u-w|||_X,$$

$$|||u-\bar{u}^{\delta}|||_X\lesssim\inf_{w\in X^{\delta}}|||u-w|||_X+\inf_{v\in Y^{\delta}}\|u-v\|_Y,$$

uniformly *in* $\varepsilon>0$ *and* $\beta\geq 1$.

Proof. As in the proof of Theorem 6.1, we follow the proofs of Theorems 3.1 (MR) and 4.2 (BEN). We only need to adapt the derivation of a lower bound for the expression in the second line of (3.15).

With $\xi=\operatorname{ess\,inf}(e-\frac{1}{2}\operatorname{div}_{\mathbf{x}}\mathbf{b})$, it holds that

$$\sqrt{\xi}\|\cdot\|_{Y'}\leq\|\cdot\|_{L_2(I\times\Omega)}\leq\tfrac{1}{\sqrt{\xi}}\|\cdot\|_Y.$$

Let $\mathbf{b}_{\delta}$ be the piecewise constant vector field defined by taking the average of $\mathbf{b}$ over each prismatic element $(t_{i-1}^{\delta},t_i^{\delta})\times T$ for $T\in\mathcal{T}^{\delta_i}$. We use $w\mapsto\mathbf{b}_{\delta}\cdot\nabla_{\mathbf{x}}w$ to approximate A_a. We have $\|\mathbf{b}-\mathbf{b}_{\delta}\|_{L_{\infty}((t_{i-1}^{\delta},t_i^{\delta})\times T)^d}\lesssim\operatorname{diam}(T)\|\mathbf{b}\|_{W^1_{\infty}((t_{i-1}^{\delta},t_i^{\delta})\times T)^d}$ by (7.11). An application of the inverse inequality on the family of spaces $(S^{0,q}_{\mathcal{T},0})_{\mathcal{T}\in\bar{\Delta}}$ shows that for some constant $L>0$, for $w\in X^{\delta}$ it holds that

$$\|(\mathbf{b}-\mathbf{b}_{\delta})\cdot\nabla_{\mathbf{x}}w+\tfrac{1}{2}w\operatorname{div}_{\mathbf{x}}\mathbf{b}\|_{L_2(I\times\Omega)}\leq L\|w\|_{L_2(I\times\Omega)}.$$

Because (7.7) is also valid for *piecewise* constant $\mathbf{b}$, and

$$\sqrt{(A_s(\varepsilon)v)(v)}\eqsim\sqrt{\varepsilon}\|\nabla_{\mathbf{x}}v\|_{L_2(I\times\Omega)^d}+\sqrt{\xi}\|v\|_{L_2(I;L_2(\Omega))},$$

only dependent on $\|e-\frac{1}{2}\operatorname{div}_{\mathbf{x}}\mathbf{b}\|_{L_{\infty}(I\times\Omega)}/\xi$, the proof of Theorem 7.2 shows that for some constant $\gamma>0$, for $w\in X^{\delta}$ it holds that

$$\|{E_Y^{\delta}}'(\partial_t+\mathbf{b}_{\delta}\cdot\nabla_{\mathbf{x}})E_Y^{\delta}w\|_{Y^{\delta}{}'}\geq\gamma\|(\partial_t+\mathbf{b}_{\delta}\cdot\nabla_{\mathbf{x}})E_Y^{\delta}w\|_{Y'}.$$

By combining these estimates, we find that for $w\in X^{\delta}$ it holds that

$$\begin{aligned}\|{E_Y^{\delta}}'CE_Y^{\delta}w\|_{Y^{\delta}{}'}&\geq\gamma\|(\partial_t+\mathbf{b}_{\delta}\cdot\nabla_{\mathbf{x}})E_Y^{\delta}w\|_{Y'}-\tfrac{L}{\sqrt{\xi}}\|E_Y^{\delta}w\|_{L_2(I\times\Omega)}\\&\geq\gamma\|CE_Y^{\delta}w\|_{Y'}-(\gamma+1)\tfrac{L}{\sqrt{\xi}}\|E_Y^{\delta}w\|_{L_2(I\times\Omega)}\\&\geq\gamma\|CE_Y^{\delta}w\|_{Y'}-(\gamma+1)\tfrac{L}{\xi}\|E_Y^{\delta}w\|_Y,\end{aligned}$$

and so

$$\begin{aligned}&\|{E_Y^{\delta}}'CE_Y^{\delta}w\|^2_{Y^{\delta}{}'}+\|E_X^{\delta}w\|_Y^2+\|\gamma_TE_X^{\delta}w\|^2+(\beta-1)\|\gamma_0E_X^{\delta}w\|^2\\&\geq\big(\gamma\|CE_Y^{\delta}w\|_{Y'}-(\gamma+1)\tfrac{L}{\xi}\|E_Y^{\delta}w\|_Y\big)^2+\|E_X^{\delta}w\|_Y^2+\|\gamma_TE_X^{\delta}w\|^2\\&\quad+(\beta-1)\|\gamma_0E_X^{\delta}w\|^2\\&\geq(1-\eta^2)\gamma^2\|CE_Y^{\delta}w\|^2_{Y'}+\big\{(1-\eta^{-2})(\gamma+1)^2\tfrac{L^2}{\xi^2}+1\big\}\|E_X^{\delta}w\|_Y^2\\&\quad+\|\gamma_TE_X^{\delta}w\|^2+(\beta-1)\|\gamma_0E_X^{\delta}w\|^2.\end{aligned}$$

Minimizing over η shows that, with $\alpha^2:=(\gamma+1)^2\frac{L^2}{\xi^2}$, the last expression is greater than or equal to

$$\tfrac{1}{2}\Big(\gamma^2+\alpha^2+1-\sqrt{(\gamma^2+\alpha^2+1)^2-4\gamma^2}\Big)|||E_X^{\delta}w|||_X^2,$$

which completes the proof. □

The undesirable condition (7.11) for time-dependent $\mathbf{b}$ might be pessimistic in practice, which however we have not tested so far.

7.2. Robust a posteriori error estimation

A robust error estimator will be realized in the following limited setting.

Consider the spaces and bilinear form a as in (7.1), where $\mathbf{b}$ is constant, $e=0$, and the polytope $\Omega\subset\mathbb{R}^d$ is convex. For families of quasi-uniform partitions $(I^{\delta})_{\delta\in\Delta}$ of $\overline{I}$, and $(\mathcal{T}^{\delta})_{\delta\in\Delta}$ and $(\mathcal{T}_S^{\delta})_{\delta\in\Delta}$ of $\overline{\Omega}$ as before, where $\mathcal{T}_S^{\delta}$ is a sufficiently deep refinement of $\mathcal{T}^{\delta}$ that permits the construction of a projector P_1^{δ} that satisfies (7.3)–(7.4), and for some $h_{\delta}>0$, $\operatorname{diam}T\eqsim h_{\delta}\eqsim\operatorname{diam}J$ $(T\in\mathcal{T}^{\delta},J\in I^{\delta})$, let $X^{\delta}:=S^{0,1}_{I^{\delta}}\otimes S^{0,1}_{\mathcal{T}^{\delta},0}$ and $Y^{\delta}:=S^{-1,1}_{I^{\delta}}\otimes S^{0,1}_{\mathcal{T}_S^{\delta},0}$. For completeness, $S^{-1,1}_{I^{\delta}}$ denotes the space of piecewise linears w.r.t. I^{δ}, and $S^{0,1}_{I^{\delta}}$ the space of continuous piecewise linears w.r.t. I^{δ}.

In this setting, in [18, Thm. 5.6] projectors $Q_B^{\delta}\in\mathcal{L}(Y,Y)$ have been constructed with $\operatorname{ran}Q_B^{\delta}\subset Y^{\delta}$ and $(\mathrm{Id}-{Q_B^{\delta}}')BX^{\delta}=0$. Moreover, these Q_B^{δ} are uniformly bounded in $Y=L_2(I;H^1_{0,\Gamma}(\Omega))$ equipped with the standard Bochner norm, with $H^1_{0,\Gamma}(\Omega)$ being equipped with $\|\nabla\cdot\|_{L_2(\Omega)^d}$. Since for the current bilinear form a, the energy-norm $\|\cdot\|_Y$ is equal to $\sqrt{\varepsilon}\|\cdot\|_{L_2(I;H^1_{0,\Gamma}(\Omega))}$, it holds that $\sup_{\delta\in\Delta,\varepsilon>0}\|Q_B^{\delta}\|_{\mathcal{L}(Y,Y)}<\infty$, and so

$$\inf_{\varepsilon>0}\gamma_{\Delta}^{B}(\varepsilon)>0.$$

Let $((\hat{K}_Y^{\delta})^{-1}v)(v)\eqsim\int_I\int_{\Omega}|\nabla_{\mathbf{x}}v|^2\mathrm{d}\mathbf{x}\mathrm{d}t$ $(\delta\in\Delta,\ v\in Y^{\delta})$, then $(\varepsilon^{-1}\hat{K}_Y^{\delta})^{-1}\eqsim{E_Y^{\delta}}'A_sE_Y^{\delta}$, i.e., using preconditioner $K_Y^{\delta}:=\varepsilon^{-1}\hat{K}_Y^{\delta}$ it holds that $\sup_{\varepsilon>0}\frac{\max(R_{\Delta},1)}{\min(r_{\Delta},1)}<\infty$.

What remains is to show that *data-oscillation* is asymptotically of higher or equal order as the approximation error in $|||\cdot|||_X=\sqrt{\|B\cdot\|^2_{Y'}+\beta\|\gamma_0\cdot\|^2}$. Noting that $\|\cdot\|_{Y'}=\frac{1}{\sqrt{\varepsilon}}\|\cdot\|_{L_2(I;H^1_{0,\Gamma}(\Omega)')}$, it is natural to select

$$\beta=\varepsilon^{-1}.$$

Then $\sqrt{\varepsilon}|||\cdot|||_X$ equals

$$\sqrt{\|(\partial_t+\mathbf{b}\cdot\nabla_{\mathbf{x}})\cdot\|^2_{L_2(I;H^1_{0,\Gamma}(\Omega)')}+\varepsilon^2\|\cdot\|^2_{L_2(I;H^1_{0,\Gamma}(\Omega))}+\varepsilon\|\gamma_T\cdot\|^2+(1-\varepsilon)\|\gamma_0\cdot\|^2},$$

and so even for a general smooth u, $\sqrt{\varepsilon}$ times the approximation error cannot be expected to be smaller than $\eqsim h_{\delta}^2$. Since for $g\in L_2(I;H^1(\Omega))\cap H^2(I;H^{-1}(\Omega))$ it holds that $\sqrt{\varepsilon}\|(\mathrm{Id}-{Q_B^{\delta}}')g\|_{Y'}=\|(\mathrm{Id}-{Q_B^{\delta}}')g\|_{L_2(I;H^1_{0,\Gamma}(\Omega)')}\lesssim h_{\delta}^2$ ([18, Thm. 5.6]), we conclude that $\mathcal{E}^{\delta}(w;g,u_0,\beta)$ from (6.4) is an efficient and, modulo above satisfactory data-oscillation term, reliable a posteriori estimator of the error in w in $|||\cdot|||_X$-norm.

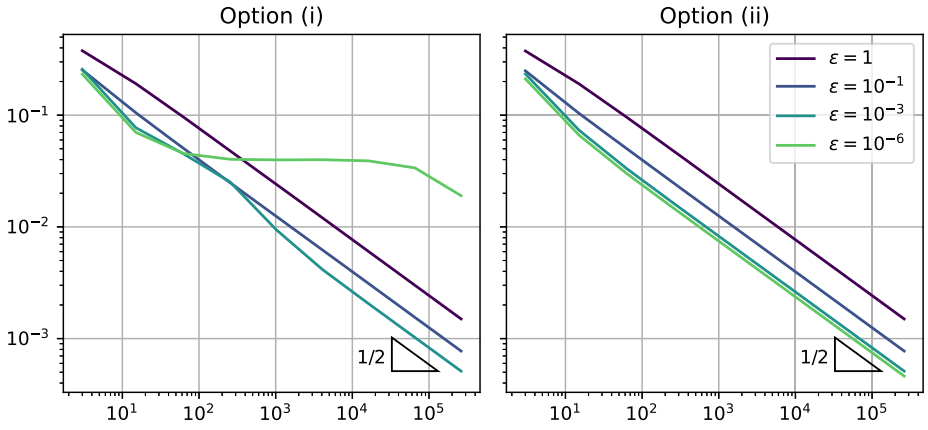

Fig. 1. Relative estimated error progression for the *smooth problem* as function of $\dim X^\delta$ for different diffusion rates ε. Left: test space Y^δ as in Option (i); right: Y^δ as in (ii).

8. Numerical test

We tested the minimal residual (MR) method applied to the parabolic initial value problem with the singularly perturbed 'spatial component' as given in (7.1). We considered the simplest case where $I = \Omega = (0,1)$, $\mathbf{b} = 1$, and e is either 0 or 1, and $X^\delta = S^{0,1}_{I^\delta} \otimes S^{0,1}_{\mathcal{T}^\delta,0}$, where $I^\delta = \mathcal{T}^\delta$ is a uniform partition of the unit interval with mesh size h_δ. Taking always $(K^\delta_Y)^{-1} = E^{\delta\prime}_Y A_s E^\delta_Y$, we took either

(i) $Y^\delta = S^{-1,1}_{I^\delta} \otimes S^{0,1}_{\mathcal{T}^\delta,0} (\supseteq X^\delta \cup \partial_t X^\delta)$ which for any *fixed* $\varepsilon > 0$ gives $\gamma^{\partial_t}_\Delta > 0$ (Sect. 5.1), so that the MR approximations are quasi-optimal approximations from the trial space w.r.t. $\|\cdot\|_X$ (Theorem 3.1), or

(ii) $Y^\delta := S^{-1,1}_{I^\delta} \otimes S^{0,1}_{\mathcal{T}^\delta_s,0}$ where $\mathcal{T}^\delta_s$ is a uniform partition with mesh-size $h_\delta/3$ which even gives $\inf_{\varepsilon>0} \gamma^C_\Delta(\varepsilon) > 0$ (Theorem 7.2), so that the MR approximations are quasi-optimal approximations from the trial space w.r.t. the energy-norm $|||\cdot|||_X$ also *uniformly in* $\varepsilon > 0$ (Theorem 6.1).

Remark 4.1 shows that in these cases the BEN and MR methods give the same solution.

As discussed in Sect. 7.2, for the case that $e = 0$ it is natural to take the weight $\beta = \varepsilon^{-1}$. Unlike with $e = 0$, for $e = 1$ and $0 \neq v \in Y$ the energy-norm $\sqrt{(A_s v)(v)}$ does not tend to zero for $\varepsilon \downarrow 0$ but converges to $\|v\|_{L_2(I\times\Omega)}$. In view of this there is no reason to let β tend to infinity for $\varepsilon \downarrow 0$, and we took $\beta = 1$.

For Y^δ as in (ii), in Sect. 7.2 it was shown that for $(e,\beta) = (0,\varepsilon^{-1})$ it holds that $\inf_{\varepsilon>0} \gamma^B_\Delta(\varepsilon) > 0$, and more specifically that the a posteriori error estimator $\mathcal{E}^\delta(w; g, u_0, \beta)$ from (6.4) is an efficient and, modulo a data-oscillation term that is at least of equal order, reliable estimator of the error $|||u - w|||_X$. Therefore to assess our numerical results, we used Y^δ as in Option (ii) for error estimation, even when solving with Y^δ as in (i).

For $(e,\beta) = (1,1)$, we numerically *observed* that for our model problems the a posteriori error estimator $\mathcal{E}^\delta(w; g, u_0, \beta)$ computed with Y^δ as in (ii) is efficient and reliable as, knowing that the estimator *equals* $|||u - w|||_X$ for $Y^\delta = Y$, we saw that further overrefinement of the test space Y^δ never increased the estimated error by more than a percent. So again, regardless of whether we took Y^δ as in Option (i) or (ii), we used Y^δ as in (ii) to compute $\mathcal{E}^\delta(w; g, u_0, \beta)$.

In experiments below, we choose $\varepsilon = 1, 10^{-1}, 10^{-3}, 10^{-6}$; to compare different values of ε, we show the estimated error divided by an accurate approximation for $\sqrt{\|g\|^2_{Y'} + \beta\|u_0\|^2}$, which is equal to the $|||\cdot|||_X$-norm of the exact solution.

8.1. Smooth problem

We take (homogeneous) Dirichlet boundary conditions at left- and right boundary, i.e. $\Gamma = \partial\Omega$, select $(e,\beta) = (0,\varepsilon^{-1})$, and prescribe the exact solution $u(t,x) := (t^2+1)\sin(\pi x)$ with derived data u_0 and g. For this problem, the best possible error in $|||\cdot|||_X$-norm, divided by $|||u|||_X$, decays proportionally to $(\dim X^\delta)^{-1/2}$.

Fig. 1 shows this relative estimated error as a function of $\dim X^\delta$. In accordance with Theorem 3.1, for this parabolic problem with non-symmetric spatial part, both Option (i) and Option (ii) give solutions that converge at the expected rate. For Option (i), however, this convergence is not uniform in ε, but in accordance with Theorem 6.1, for Option (ii) it is.

8.2. Internal layer problem

We choose $u_0 := 0$ and $g(t,x) := \mathbb{1}_{\{x>t\}}$, select $(e,\beta) = (0,\varepsilon^{-1})$, and prescribe a homogeneous Dirichlet boundary condition only at the left boundary $x = 0$, i.e. $\Gamma := \{0\}$, and so have a Neumann boundary condition at the 'outflow' boundary $x = 1$. Due to the jump in the forcing data, in the limit $\varepsilon \downarrow 0$, the solution $t \cdot \mathbb{1}_{\{x>t\}}$ is discontinuous along the diagonal $x = t$.

The left of Fig. 2 shows the relative estimated error progression of Option (ii) as a function of $\dim X^\delta$; as Option (i) again suffers from degradation for small ε (with results very similar to the left of Fig. 1), we omit a graph of its error progression. Its right shows the discrete solution at $h_\delta = \frac{1}{512}$ and $\varepsilon = 10^{-6}$. The solution resembles the *pure transport* solution quite well, with the exception of a small artefact near $x = t = 0$.

8.3. Boundary layer problem

We choose $u_0(x) := \sin(\pi x)$ and $g = 0$, select $(e,\beta) = (1,1)$, and set homogeneous Dirichlet boundary conditions on $\partial\Omega$, i.e. $\Gamma = \{0,1\}$. Due to the condition on the outflow boundary, the problem is ill-posed in the limit $\varepsilon = 0$, hence for ε small, the solution has a boundary layer at $x = 1$.

Fig. 3 shows that the method fails to make progress until the boundary layer is resolved at $h_\delta \lesssim \varepsilon$. Fig. 4 shows two discrete solutions at $h_\delta = \frac{1}{512}$ computed for Option (ii). We see that for $\varepsilon = 10^{-3}$, the boundary layer is resolved and the solution resembles the *pure transport* solution quite well, with the exception of a small artefact near $x = t = 1$. For $\varepsilon = 10^{-6}$ though, the boundary layer cannot be resolved with the current (uniform) mesh, and the solution is completely wrong. For $\varepsilon \downarrow 0$, the energy-norm of the error in an approximation w approaches $\sqrt{\|(\partial_t + \mathbf{b}\cdot\nabla_{\mathbf{x}})w\|^2_{L_2(I\times\Omega)} + \|u_0 - \gamma_0 w\|^2_{L_2(\Omega)}}$. As a result, for streamlines that hit the outflow boundary, the method 'chooses' to smear the unavoidably large error as a consequence of the layer along the whole

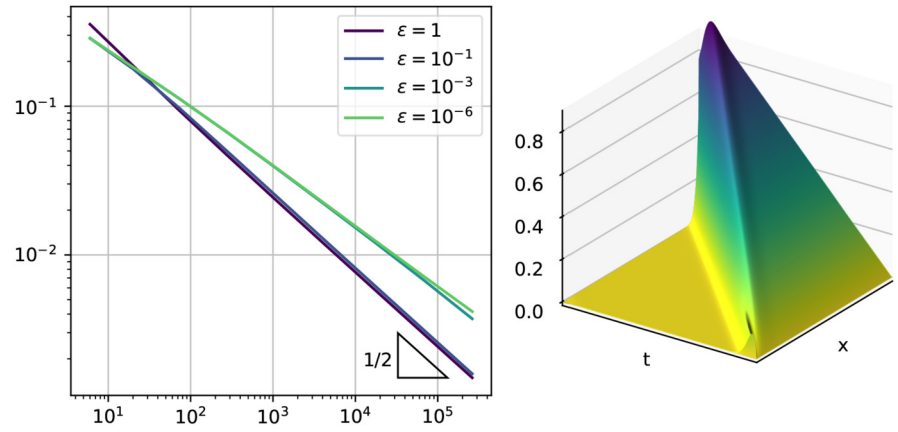

Fig. 2. Solving the *internal layer problem* with Option (ii). Left: relative estimated error progression as function of $\dim X^\delta$ for different diffusion rates ε. Right: solution at $h_\delta = \frac{1}{512}$ and $\varepsilon = 10^{-6}$.

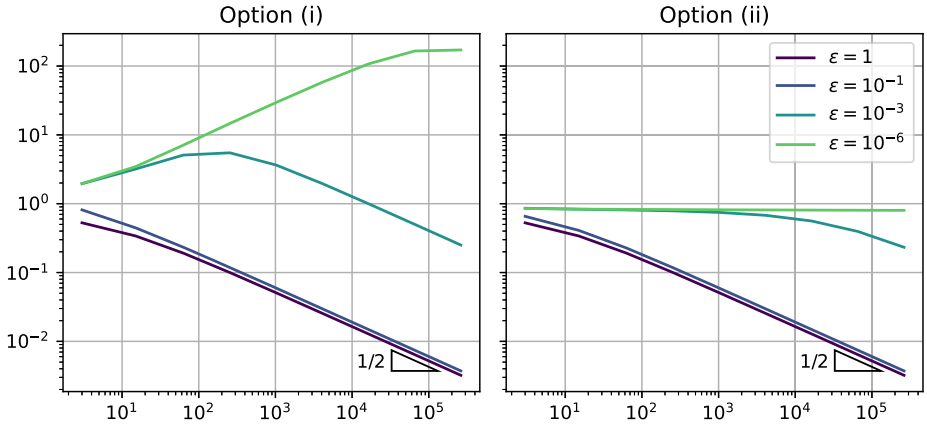

Fig. 3. Relative estimated error progression for the *boundary layer problem* as function of $\dim X^\delta$ for different diffusion rates ε. Left: test space Y^δ as in Option (i); right: Y^δ as in (ii).

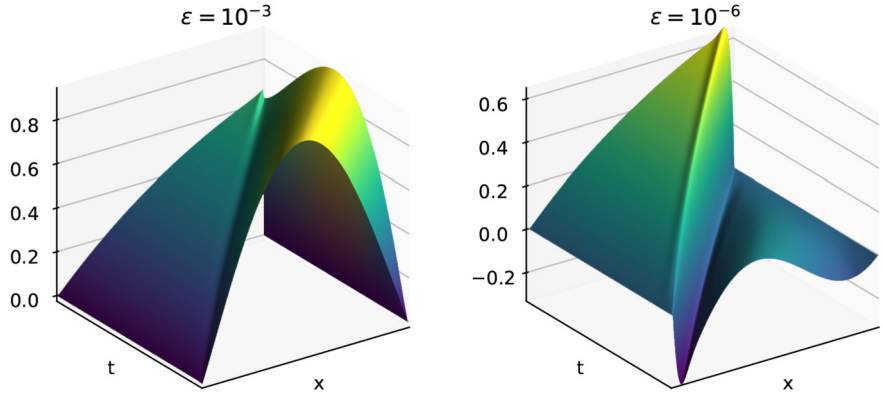

Fig. 4. Solutions of the *boundary layer problem* with Option (ii) at $h_\delta = \frac{1}{512}$. Left: diffusion $\varepsilon = 10^{-3}$; right: $\varepsilon = 10^{-6}$.

streamline resulting in a globally bad approximation. This is a well-known phenomenon when using a least squares method to approximate a solution that has a sharp layer or a shock.

8.4. Imposing outflow boundary conditions weakly

One common work-around to the problem caused by the boundary layer is to refine the mesh strongly towards this layer. An alternative is to impose at the outflow boundary the Dirichlet boundary condition only weakly, see e.g. the references [9,8,10,11] where this approach has been applied with least squares methods for stationary convection dominated convection-diffusion methods. This approach is also known from other contexts, as in [5,7,33]. Without having a rigorous analysis we tried this weak imposition of the Dirichlet boundary condition by computing, with Y^δ as in Option (ii),

$$u^\delta := \operatorname*{argmin}_{w \in \hat{X}^\delta} \|E_Y^{\delta\prime}(BE_X^\delta w - g)\|^2_{Y^{\delta\prime}} + \beta\|\gamma_0 E_X^\delta w - u_0\|^2 + \varepsilon\|w(\cdot,1)\|^2_{L_2(I)}.$$

Here, $\hat{X}^\delta$ denotes the space X^δ after removing the Dirichlet boundary condition at $x = 1$. Fig. 5 shows the resulting error progression, which is robust in ε, as well as the minimal residual solution at $h_\delta = \frac{1}{512}$ and

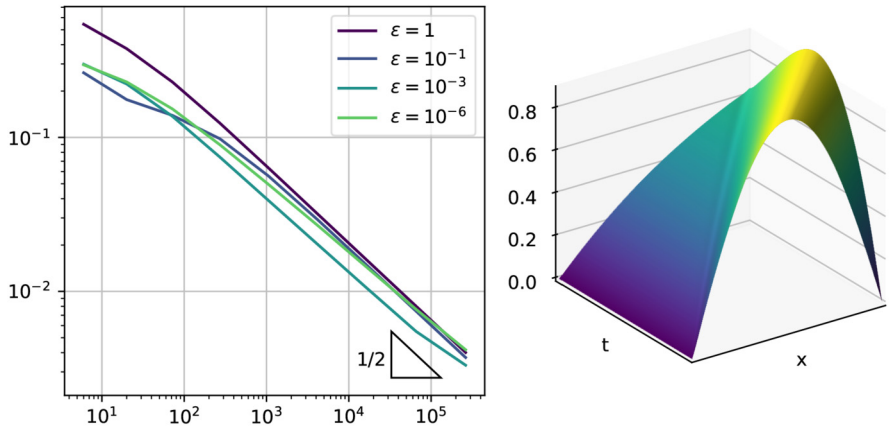

Fig. 5. Solving the *boundary layer problem* with Option (ii), and imposing the outflow boundary condition *weakly*. Left: relative estimated error progression as function of $\dim \hat{X}^\delta$ for different diffusion rates ε. Right: solution at $h_\delta = \frac{1}{512}$ and $\varepsilon = 10^{-6}$.

$\varepsilon = 10^{-6}$; it resembles the *pure transport* solution quite well, and does not suffer from the artifact present at the right of Fig. 4.

References

[1] R. Andreev, Stability of space-time Petrov-Galerkin discretizations for parabolic evolution equations, PhD thesis, ETH, Zürich, 2012.

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

[3] R. Andreev, Wavelet-in-time multigrid-in-space preconditioning of parabolic evolution equations, SIAM J. Sci. Comput. 38 (1) (2016) A216–A242.

[4] H. Brézis, I. Ekeland, Un principe variationnel associé à certaines équations paraboliques. Le cas dépendant du temps, C. R. Acad. Sci. Paris Sér. A-B 282 (20) (1976) A1197–A1198.

[5] Y. Bazilevs, T.J.R. Hughes, Weak imposition of Dirichlet boundary conditions in fluid mechanics, Comput. Fluids 36 (1) (2007) 12–26.

[6] T. Boiveau, V. Ehrlacher, A. Ern, A. Nouy, Low-rank approximation of linear parabolic equations by space-time tensor Galerkin methods, ESAIM: Math. Model. Numer. Anal. 53 (2) (2019) 635–658.

[7] E. Burman, M.A. Fernández, P. Hansbo, Continuous interior penalty finite element method for Oseen's equations, SIAM J. Numer. Anal. 44 (3) (2006) 1248–1274.

[8] D. Broersen, R.P. Stevenson, A robust Petrov-Galerkin discretisation of convection-diffusion equations, Comput. Math. Appl. 68 (11) (2014) 1605–1618.

[9] A. Cohen, W. Dahmen, G. Welper, Adaptivity and variational stabilization for convection-diffusion equations, ESAIM: Math. Model. Numer. Anal. 46 (2012) 1247–1273.

[10] J. Chan, J.A. Evans, W. Qiu, A dual Petrov-Galerkin finite element method for the convection-diffusion equation, Comput. Math. Appl. 68 (11) (2014) 1513–1529.

[11] H. Chen, G. Fu, J. Li, W. Qiu, First order least squares method with weakly imposed boundary condition for convection dominated diffusion problems, Comput. Math. Appl. 68 (12, part A) (2014) 1635–1652.

[12] D. Devaud, Petrov-Galerkin space-time hp-approximation of parabolic equations in $H^{1/2}$, IMA J. Numer. Anal. 40 (4) (2020) 2717–2745.

[13] L. Demkowicz, J. Gopalakrishnan, A class of discontinuous Petrov-Galerkin methods. II. Optimal test functions, Numer. Methods Partial Differ. Equ. 27 (1) (2011) 70–105.

[14] L. Demkowicz, J. Gopalakrishnan, An overview of the discontinuous Petrov Galerkin method, in: Recent Developments in Discontinuous Galerkin Finite Element Methods for Partial Differential Equations, in: IMA Vol. Math. Appl., vol. 157, Springer, Cham, 2014, pp. 149–180.

[15] R. Dautray, J.-L. Lions, Mathematical Analysis and Numerical Methods for Science and Technology, Evolution Problems I, vol. 5, Springer-Verlag, Berlin, 1992.

[16] L. Diening, J. Storn, A space-time DPG method for the heat equation, arXiv:2012.13229, 2020.

[17] L. Diening, J. Storn, T. Tscherpel, On the Sobolev and L^p-stability of the L^2-projection, arXiv:2008.01801, 2020.

[18] W. Dahmen, R.P. Stevenson, J. Westerdiep, Accuracy controlled data assimilation for parabolic problems, Math. Comput. (2021), https://doi.org/10.1090/mcom/3680, in press arXiv:2105.05836.

[19] A. Ern, J.-L. Guermond, Theory and Practice of Finite Elements, Applied Mathematical Sciences., vol. 159, Springer, New York, 2004.

[20] A. Ern, J.-L. Guermond, Finite Elements. II. Galerkin Approximation, Elliptic and Mixed PDEs, Texts in Applied Mathematics, vol. 73, Springer, Cham, 2021.

[21] A. Ern, J.-L. Guermond, Finite Elements. III. First-Order and Time-Dependent PDEs, Texts in Applied Mathematics, vol. 74, Springer, Cham, 2021.

[22] T. Führer, M. Karkulik, Space-time least-squares finite elements for parabolic equations, Comput. Math. Appl. 92 (2021) 27–36.

[23] M. Fortin, An analysis of the convergence of mixed finite element methods. III, RAIRO. Anal. Numér. 11 (4) (1977) 341–354.

[24] F.D. Gaspoz, C.-J. Heine, K.G. Siebert, Optimal grading of the newest vertex bisection and H^1-stability of the L_2-projection, IMA J. Numer. Anal. 36 (3) (2016) 1217–1241.

[25] V. Girault, J.-L. Lions, Two-grid finite-element schemes for the transient Navier-Stokes problem, ESAIM: Math. Model. Numer. Anal. 35 (5) (2001) 945–980.

[26] H. Gimperlein, J. Stocek, Space-time adaptive finite elements for nonlocal parabolic variational inequalities, Comput. Methods Appl. Mech. Eng. 352 (2019) 137–171.

[27] G. Gantner, R. Stevenson, Further results on a space-time FOSLS formulation of parabolic PDEs, ESAIM: Math. Model. Numer. Anal. 55 (1) (2021) 283–299.

[28] T. Kato, Estimation of iterated matrices, with application to the von Neumann condition, Numer. Math. 2 (1960) 22–29.

[29] J.-L. Lions, E. Magenes, Non-homogeneous Boundary Value Problems and Applications. Vol. I, Die Grundlehren der Mathematischen Wissenschaften, vol. 181, Springer-Verlag, New York-Heidelberg, 1972, Translated from the French by P. Kenneth.

[30] U. Langer, S.E. Moore, M. Neumüller, Space-time isogeometric analysis of parabolic evolution problems, Comput. Methods Appl. Mech. Eng. 306 (2016) 342–363.

[31] B. Nayroles, Deux théorèmes de minimum pour certains systèmes dissipatifs, C. R. Acad. Sci. Paris Sér. A-B 282 (17) (1976) A1035–A1038.

[32] M. Neumüller, I. Smears, Time-parallel iterative solvers for parabolic evolution equations, SIAM J. Sci. Comput. 41 (1) (2019) C28–C51.

[33] F. Schieweck, On the role of boundary conditions for CIP stabilization of higher order finite elements, Electron. Trans. Numer. Anal. 32 (2008) 1–16.

[34] Ch. Schwab, R.P. Stevenson, A space-time adaptive wavelet method for parabolic evolution problems, Math. Comput. 78 (2009) 1293–1318.

[35] O. Steinbach, Space-time finite element methods for parabolic problems, Comput. Methods Appl. Math. 15 (4) (2015) 551–566.

[36] R.P. Stevenson, R. van Venetië, J. Westerdiep, A wavelet-in-time, finite element-in-space adaptive method for parabolic evolution equations, arXiv:2101.03956, 2021.

[37] R.P. Stevenson, J. Westerdiep, Stability of Galerkin discretizations of a mixed space-time variational formulation of parabolic evolution equations, IMA J. Numer. Anal. 41 (1) (2021) 28–47.

[38] O. Steinbach, H. Yang, Comparison of algebraic multigrid methods for an adaptive space-time finite-element discretization of the heat equation in 3D and 4D, Numer. Linear Algebra Appl. 25 (3) (2018) e2143.

[39] L.R. Scott, S. Zhang, Finite element interpolation of nonsmooth functions satisfying boundary conditions, Math. Comput. 54 (190) (1990) 483–493.

[40] O. Steinbach, M. Zank, Coercive space-time finite element methods for initial boundary value problems, Electron. Trans. Numer. Anal. 52 (2020) 154–194.

[41] R. van Venetië, J. Westerdiep, A parallel algorithm for solving linear parabolic evolution equations, arXiv:2009.08875, 2021.

[42] R. van Venetië, J. Westerdiep, Efficient space-time adaptivity for parabolic evolution equations using wavelets in time and finite elements in space, arXiv:2104.08143, 2021.

[43] J. Wloka, Partielle Differentialgleichungen. Sobolevräume und Randwertaufgaben, B.G. Teubner, Stuttgart, 1982.

[44] Wolfram Research, Inc. Mathematica, Version 12.3.1, Champaign, IL, 2021.

[45] J. Xu, L. Zikatanov, Some observations on Babuška and Brezzi theories, Numer. Math. 94 (1) (2003) 195–202.