Università degli Studi di Milano
Graduate School in Mathematical Sciences
Dipartimento di Matematica
Ph. D. Program in Mathematics and Statistics for
Computational Sciences, Ciclo XXVI

Quasi-optimality in the backward Euler-Galerkin method for linear parabolic problems

MAT/08 Numerical Analysis

Thesis of
Francesca Tantardini

Advisor:
Prof. Andreas Veeser

Ph. D. Program Coordinator:
Prof. Giovanni Naldi

Academic Year 2012-13

Contents

Introduction

Galerkin finite element methods are widely used for the numerical solution of linear parabolic problems. There is a vast literature of corresponding a priori error bounds, most of which are derived without invoking a quasi-optimality result like Ceá's Lemma; see the monograph of Thomée [41]. This differs from the analysis of linear elliptic problems, as Douglas and Dupont [21] recognized already in 1970.

A quasi-optimality result states the equivalence between the error of the method and the best error within the underlying discrete space. The interest for such a result is motivated by the fact that it is stronger than error bounds of optimal order. In fact, given a suitable discrete space, quasi-optimality implies error bounds of optimal order, but not vice versa: for example, optimal order error bounds may require more regularity than the minimal one indicated by approximation theory or, more subtle, the bound may not vanish whenever the error does so. Moreover optimal order error bounds provide information of asymptotic nature, while quasi-optimality goes beyond and in particular covers the computational range.

There are some known results of quasi-optimality in the framework of parabolic problems. Concerning fixed spatial semi-discretizations, Douglas and Dupont [21] derived a quasi-optimality result in a norm involving a time derivative of order $1/2$, assuming that the initial error vanishes. This approach has been generalized by Baiocchi and Brezzi [3] to fractions around $1/2$ and by Tomarelli [42] also to general initial values. Other three results concern quasi-optimality for the approximation in space in simpler norms, which are related to the standard weak formulation of parabolic problems. However, none of these results perfectly mimics the elliptic case. Dupont [22] derived quasi-optimality in a norm close to the one of $H^1(H^{-1}) \cap L^2(H^1)$, but that depends on the discretization. This approach has been generalized in [4, 24, 26] in the context of moving mesh methods and in [25] also for Navier-Stokes equations. The other two results, [13, 30] require that the L^2-projection is H^1-stable. With this assumption, Hackbusch [30] in particular established stability in $L^2(H^1)$, while Chrysafinos and Hou [13] showed quasi-

optimality in $H^1(H^{-1}) \cap L^2(H^1)$.

Concerning more general spatial discretizations, Dupont in [22] analyses the case where the spatial mesh is allowed to change with time and proposes a remarkable counterexample. The discretization is based on one dimensional finite elements in space, and backward Euler in time. The best approximation error converges to zero as the mesh-size h and the time step τ independently converge to zero. However, the spatial mesh changes every time-step in a way that, if h, $\tau \to 0$ such that $h^4/\tau \to \infty$, then the discrete solution does not converge to the exact solution. This reveals that a quasi-optimality result does not hold, at least when the spatial discretization changes. For more general problems, Chrysafinos and Walkington [14] prove that the error in the $L^2(L^\infty) \cap L^2(H^1)$-norm can be bounded by the error given by a suitable local projection, and an extra term, that vanishes if the spatial discretization remains the same.

This thesis concerns linear parabolic equations, with a uniformly elliptic operator that may depend on time. The approach to the analysis of the backward Euler-Galerkin method is based on the framework given by the inf-sup condition. In order to shed light on various aspects, we analyse separately the spatial discretization, the time discretization and the issue of varying the spatial discretization, and combine them in an analysis of the backward Euler- Galerkin method.

Concerning approximation in space we prove that the H^1-stability of the L^2-projection is a necessary condition for quasi-optimality, both in the $H^1(H^{-1}) \cap L^2(H^1)$-norm and in the $L^2(H^1)$-norm.

Furthermore we investigate the discretization only in time by the backward Euler method. Under the assumption that the time partition is locally quasi-uniform, we prove that the error in a norm that mimics the $H^1(H^{-1}) \cap L^2(H^1)$-norm is equivalent to the sum of the best errors with piecewise constants for the exact solution and its time derivative. Concerning the $L^2(H^1)$-norm, we observe lack of stability, and therefore of quasi-optimality. Nevertheless, the $L^2(H^1)$-norm of the discrete solution is bounded in terms of a stronger norm of the exact solution, and we derive an abstract error estimate where the right-hand side is equivalent to the error.

Moreover we address the topic of varying the spatial discretization. Given a partition of the time interval, we allow for modifications of the spatial discretization every time step. Assuming the L^2-projection to be H^1-stable, we prove that the error is bounded, up to a constant, by the best error and an extra term that arises from the modifications of the spatial discretization, and vanishes if these do not occur. This result is valid for both the broken $H^1(H^{-1}) \cap L^2(H^1)$-norm and the $L^2(H^1)$-norm. This extra term is consistent with Dupont's counterexample, in that it converges to zero if $h^4/\tau \to 0$.

Collecting all the previous results, we analyse the backward Euler-Galerkin method.

Finally we derive to error estimates, in the case the spatial discretization is based on finite elements. We provide error bounds in terms of the local mesh-size, the local time step and the regularity of the exact solution. The latter is measured with Sobolev spaces of possible fractional order.

Organization

The thesis is organized as follows. In Chapter 1 we recall Petrov-Galerkin approximations and we derive lower bounds for the quasi-optimality constant. In Chapter 2 we cast abstract parabolic problems into the setting given by the inf-sup condition. We consider two formulations: in the "standard" one the solution is sought in $L^2(H^1) \cap H^1(H^{-1})$, while in the "natural" formulation the solution belongs to $L^2(H^1)$. Chapters 3–6 are dedicated to the discretization. Each one is divided into two parts: the first part is associated to the standard formulation and concerns approximation in $L^2(H^1) \cap H^1(H^{-1})$, while the second part is related to the natural formulation and deals with approximation in $L^2(H^1)$. More specifically, in Chapter 3 we study the spatial semidiscretization, in Chapter 4 the semidiscretization in time, in Chapter 5 the variable spatial discretization and in Chapter 6 the backward-Euler Galerkin method. We conclude in Chapter 7 with the error bounds. The approximation of the time derivative involves the H^{-1}-norm in space. In order to deal with this term, we define a suitable interpolation operator, which allows for duality arguments.

Theorems, propositions, remarks, etc. share the same numbering and are numbered per chapter.

Acknowledgements

I am deeply thankful to Professor Andreas Veeser for his precious advice, the fruitful discussions and his constant kindness.

Chapter 1

The Quasi-Optimality Constant for Petrov-Galerkin Approximations

Petrov-Galerkin approximations that are inf-sup stable are known to be near-best [2], in the sense that there exists a constant $q > 0$ such that

$$\|u - U\| \leq q \inf_{V} \|u - V\| . \tag{1.1}$$

This means that the error of the Galerkin approximation U is equivalent to the best error with respect to the discrete space. This is also called a symmetric error estimate [21, 22], to stress that the norm on the right-hand side coincides with the one on the left-hand side. The result in (1.1) provides information of non-asymptotic nature about the quality of the approximation, and it is useful for deriving a priori error estimates. Of course this is significant for practical computation only if q is of moderate size. Upper bounds for q are given in [2, 7, 12, 48].

In this chapter we provide a formula for q in terms of the bilinear form of the variational problem and in terms of the discrete spaces. We regain the known upper bound and we also furnish lower bounds for q, which are useful in Chapter 3.

The chapter is organized as follows. In Section 1.1, we recall the framework of Petrov-Galerkin approximations and the inf-sup theory, which is a key instrument in establishing our results. In Section 1.2 we derive the aforementioned formula for q. Finally in Section 1.3 we analyse the non-conforming case.

1.1 Petrov-Galerkin approximation

We start by briefly reviewing Petrov-Galerkin approximation in Banach reflexive spaces. See also the original work of Babuška [2] and, for example, the textbook [27].

1.1.1 The abstract problem

We introduce the abstract problem to be approximated. Let $(X_1, \|\cdot\|_1)$ and $(X_2, \|\cdot\|_2)$ be two real Banach spaces, and let X_2 be also reflexive. The dual space X_2^* of X_2 is equipped with the usual dual norm $\|\ell\|_2^* = \sup_{\|\varphi\|_2=1} \ell(\varphi)$ for $\ell \in X_2^*$. Moreover, let b be a real-valued bounded bilinear form on $X_1 \times X_2$ and let C_b be the continuity constant of b:

$$C_b := \sup_{v \in X_1} \sup_{\varphi \in X_2} \frac{b(v,\varphi)}{\|v\|_1 \, \|\varphi\|_2}. \tag{1.2}$$

We consider the problem

$$\text{given } \ell \in X_2^*, \text{ find } u \in X_1 \text{ such that } \forall \varphi \in X_2 \quad b(u,\varphi) = \ell(\varphi) \tag{1.3}$$

and say that it is well-posed if, for any $\ell \in X_2^*$, there exists a unique solution that continuously depends on ℓ. The spaces X_1 and X_2 are called trial and test space, respectively. Problem (1.3) is well-posed if and only if there hold the following two conditions:

$$c_b := \inf_{v \in X_1} \sup_{\varphi \in X_2} \frac{b(v,\varphi)}{\|v\|_1 \, \|\varphi\|_2} > 0 \qquad \text{(uniqueness)}, \tag{1.4a}$$

$$(b(v,\varphi) = 0 \quad \forall v \in X_1) \Rightarrow \varphi = 0 \qquad \text{(existence)}. \tag{1.4b}$$

The quantity c_b is the so-called inf-sup constant. An equivalent condition to (1.4) is

$$\inf_{v \in X_1} \sup_{\varphi \in X_2} \frac{b(v,\varphi)}{\|v\|_1 \, \|\varphi\|_2} = \inf_{\varphi \in X_2} \sup_{v \in X_1} \frac{b(v,\varphi)}{\|v\|_1 \, \|\varphi\|_2} > 0. \tag{1.5}$$

This equality allows to exchange the spaces where infimum and supremum are taken and it is a consequence of

$$\frac{1}{c_b} = \left\|B^{-1}\right\|_{\mathcal{L}(X_2^*, X_1)} = \left\|B^{-*}\right\|_{\mathcal{L}(X_1^*, X_2)}, \tag{1.6}$$

where the linear operator $B \in \mathcal{L}(X_1, X_2^*)$ is given by $B(v)(\varphi) = b(v,\varphi)$, and $B^* \in \mathcal{L}(X_2, X_1^*)$ is its adjoint $B^*(\varphi)(v) = B(v)(\varphi)$. Equation (1.6) tells that

$\|u\|_1 \leq c_b^{-1} \|\ell\|_2^*$, so c_b^{-1} may be viewed as an absolute condition number for solving (1.3) with respect to $\|\cdot\|_1$ and $\|\cdot\|_2$. Consequently, problem (1.3) is well-conditioned if the inf-sup constant c_b is not too small.

One could also consider a problem with right-hand side in X_1^*, so that X_1 becomes the test space and X_2 the trial space:

$$\text{given } g \in X_1^*, \text{ find } \phi \in X_2 \text{ such that } \forall v \in X_1 \quad b(v, \phi) = g(v). \tag{1.7}$$

We call (1.7) the dual problem of the primal problem (1.3). Thanks to (1.5) the well-posedness of primal and dual problem are equivalent.

1.1.2 Petrov-Galerkin method

We next review Petrov-Galerkin methods for problem (1.3). For notational simplicity, we take the viewpoint that a Petrov-Galerkin method is characterized by one pair of subspaces, instead of a family of pairs. Given two nontrivial and proper subspaces $M_i \subset X_i$, $i = 1, 2$, the Petrov-Galerkin method $M = M_1 \times M_2$ reads

$$\text{given } \ell \in X_2^*, \text{ find } U_M \in M_1 \text{ such that } \forall \varphi \in M_2 \quad b(U_M, \varphi) = \ell(\varphi). \tag{1.8}$$

Replacing X_i by M_i, $i = 1, 2$, we see that method M is well-defined, or problem (1.8) is well-posed, if and only if there hold

$$c_M := \inf_{v \in M_1} \sup_{\varphi \in M_2} \frac{b(v, \varphi)}{\|v\|_1 \|\varphi\|_2} > 0, \tag{1.9a}$$

$$(b(v, \varphi) = 0 \quad \forall v \in M_1) \Rightarrow \varphi = 0 \tag{1.9b}$$

The quantity c_M is the so-called discrete inf-sup constant. If M_1 and M_2 have finite dimension, it is necessary for (1.9) that $\dim(M_1) = \dim(M_2)$. In this case, (1.9a) and (1.9b) are equivalent.

We say that the method M is stable if there exists a constant k such that, for every $\ell \in H_2^*$,

$$\|U_M\|_1 \leq k \|u\|_1 . \tag{1.10}$$

It is easy to see that the best constant k in (1.10) is the norm of the Ritz projection $R_M : X_1 \to M_1$, which maps any exact solution on its corresponding approximate solution. More precisely, R_M is defined by

$$\forall \varphi \in M_2 \qquad b(R_M u, \varphi) = b(u, \varphi). \tag{1.11}$$

For a well-defined method M, the map R_M is also well-defined and it is actually a projection onto the nontrivial proper subspace M_1.

One may also consider the dual Petrov-Galerkin method $M_2 \times M_1$ for the dual problem (1.7):

$$\text{given } g \in X_1^*, \text{ find } \Phi_M \in M_2 \text{ such that } \forall v \in M_1 \quad b(v, \Phi_M) = g(v). \quad (1.12)$$

Thanks again to (1.5) the well-posedness of (1.12) is equivalent to the one of (1.8). We denote by $R_M^* : X_2 \to M_2$ the dual Ritz projection defined, similarly to (1.11), by

$$\forall v \in M_1 \qquad b(v, R_M^* \phi) = b(v, \phi).$$

1.2 The quasi-optimality constant

The quasi-optimality constant q_M of a method M is the smallest constant $q \geq 0$ such that, for any $\ell \in X_2^*$, there holds

$$\|u - U_M\|_1 \leq q \inf_{v \in M_1} \|u - v\|_1 . \quad (1.13)$$

In view of $U_M \in M_1$, there holds $q_M \geq 1$.

We briefly provide an overview of the history of (1.13), in case X_1 is a Hilbert space. The first result of type (1.13) is due to Céa, [12, Prop. 3.1], who proved it in 1964 for a symmetric bilinear form. Denoted by α_b the coercivity constant of the bilinear form, the upper bound for q is given by

$$q_M \leq \sqrt{\frac{C_b}{\alpha_b}}.$$

Birkhoff, Schultz and Varga in 1968 [7, Thm. 13] extended the result to the non-symmetric case, but still with identical trial and test space, with

$$q_M \leq \frac{C_b}{\alpha_b}.$$

In 1970, Babuška [2, Thm. 2.2] proved (1.13) for the more general setting described in Section 1.1, with

$$q_M \leq 1 + \frac{C_b}{c_M}.$$

Finally Xu and Zikatanov [48, Thm. 2] in 2003 improved the bound by Babuška:

$$q_M \leq \frac{C_b}{c_M}.$$

Theorem 1.2 provides a formula for q_M in terms of b and the discrete spaces. It allows to derive also lower bounds for q_M and regain the upper bound by Xu and Zikatanov. We remark that the bound by Babuška is valid in the more generic case when X_1 is a Banach space.

We first show the equivalence between quasi-optimality and stability, with $q_M = \|R_M\|_{\mathcal{L}(X_1)}$. In fact, since b is a bounded bilinear form, R_M is also linear and bounded. Therefore,

$$\forall v \in M_1 \qquad u - R_M u = (I - R_M)(u - v),$$

which implies $q_M = \|I - R_M\|_{\mathcal{L}(X_1)}$. In order to link this to $\|R_M\|_{\mathcal{L}(X_1)}$ we exploit [48, Lemma 5], that we recall for convenience.

Lemma 1.1. *Let H be a Hilbert space, and $P : H \to H$ a nontrivial idempotent operator, that is, $0 \neq P^2 = P \neq I$. Then the following identity holds*

$$\|P\|_{\mathcal{L}(H)} = \|I - P\|_{\mathcal{L}(H)}.$$

Applying Lemma 1.1 with $P = R_M$ yields

$$q_M = \|R_M\|_{\mathcal{L}(X_1)}. \tag{1.14}$$

This equality allows to derive the following theorem.

Theorem 1.2 (Quasi-optimality)**.** *Assume that X_1 is a Hilbert space,* (1.2) *is finite, problem* (1.3) *is well-posed, and method M is well-defined. Then the quasi-optimality constant of M satisfies:*

$$q_M = \sup_{\varphi\in M_2}\inf_{v\in M_1}\frac{\|v\|_1}{b(v,\varphi)}\sup_{x\in X_1}\frac{b(x,\varphi)}{\|x\|_1} = \sup_{v\in M_1}\inf_{\varphi\in M_2}\frac{\|v\|_1}{b(v,\varphi)}\sup_{x\in X_1}\frac{b(x,\varphi)}{\|x\|_1}. \tag{1.15}$$

Proof. We introduce the following norm on X_2

$$\|\varphi\|_b := \sup_{v\in X_1}\frac{b(v,\varphi)}{\|v\|_1},$$

which is equivalent to $\|\cdot\|_2$, with $c_b\|\varphi\|_2 \le \|\varphi\|_b \le C_b\|\varphi\|_2$, for every $\varphi \in M_2$. Moreover the inf-sup and continuity constants of b with respect to $\|\cdot\|_b$ are equal:

$$\inf_{\varphi\in X_2}\sup_{v\in X_1}\frac{b(v,\varphi)}{\|\varphi\|_b\|v\|_1} = \sup_{\varphi\in X_2}\sup_{v\in X_1}\frac{b(v,\varphi)}{\|\varphi\|_b\|v\|_1} = 1.$$

Method M is well-defined with respect to $\|\cdot\|_b$ too, and the discrete inf-sup constant, denoted by β_M, enjoys the symmetry (1.5):

$$\beta_M := \inf_{\varphi\in M_2}\sup_{v\in M_1}\frac{b(v,\varphi)}{\|\varphi\|_b\|v\|_1} = \inf_{v\in M_1}\sup_{\varphi\in M_2}\frac{b(v,\varphi)}{\|\varphi\|_b\|v\|_1}.$$

Using formula (1.14), we derive that $q_M = \beta_M^{-1}$. Indeed, given $v \in X_1$, there holds

$$\beta_M \|R_M v\|_1 \leq \sup_{\varphi \in M_2} \frac{b(R_M v, \varphi)}{\|\varphi\|_b} = \sup_{\varphi \in M_2} \frac{b(v, \varphi)}{\|\varphi\|_b} \leq \|v\|_1 \tag{1.16}$$

whence

$$q_M = \|R_M\|_{\mathcal{L}(X_1)} \leq \beta_M^{-1}.$$

The other direction follows by similar arguments and the symmetry (1.5). For every $\varphi \in M_2$, we have

$$\|\varphi\|_b \leq \sup_{v \in X_1} \frac{b(v, \varphi)}{\|v\|_1} = \sup_{v \in X_1} \frac{b(R_M v, \varphi)}{\|v\|_1} \leq \|R_M\|_{\mathcal{L}(X_1)} \sup_{v \in M_1} \frac{b(v, \varphi)}{\|v\|_1},$$

that is, $\beta_M \geq \|R_M\|_{\mathcal{L}(X_1)}^{-1} = q_M^{-1}$. □

We notice that (1.15) involves only discrete test functions. We consider a norm $\|\cdot\|_\sharp$ on M_2, that can differ from the inherited norm $\|\cdot\|_2$, and can depend on the discrete space M_2, but such that

$$c_M^\sharp := \inf_{\varphi \in M_2} \sup_{v \in M_1} \frac{b(v, \varphi)}{\|\varphi\|_\sharp \|v\|_1} > 0 \quad \text{and} \tag{1.17a}$$

$$C_M^\sharp := \sup_{\varphi \in M_2} \sup_{v \in M_1} \frac{b(v, \varphi)}{\|\varphi\|_\sharp \|v\|_1} \leq C_{X_1 \times M_2}^\sharp := \sup_{v \in X_1} \sup_{\varphi \in M_1} \frac{b(v, \varphi)}{\|\varphi\|_\sharp \|v\|_1} < \infty. \tag{1.17b}$$

The following bounds follow from (1.15) and elementary inequalities.

Corollary 1.3 (Upper and Lower bounds)**.** *Under the hypothesis of Theorem 1.2 the quasi-optimality constant of M satisfies:*

$$\max\left\{\frac{c_b}{c_M}, \frac{C_{X_1 \times M_2}^\sharp}{C_M^\sharp}\right\} \leq q_M \leq \min\left\{\frac{C_b}{c_M}, \frac{C_{X_1 \times M_2}^\sharp}{c_M^\sharp}\right\}. \tag{1.18}$$

In the upper bound we recognize the constant in [48].

Remark 1.4 (Another error norm)**.** One could think to measure the error in another norm $\|\cdot\|_\sim$. Of course $\|\cdot\|_\sim$ has to be well-defined on X_1, but it may happen that the bilinear form b is not continuous on $X_1 \times X_2$ equipped with $\|\cdot\|_\sim$ and $\|\cdot\|_2$, respectively. However, if $(X_1, \|\cdot\|_\sim)$ is still a Hilbert space, we get $q_M = \|R_M\|_{\mathcal{L}((X_1, \|\cdot\|_\sim))}$. If b is continuous on $X_1 \times M_2$ equipped

respectively with $\|\cdot\|_\sim$ and $\|\cdot\|_2$, and the corresponding inf-sup constant does not degenerate, we deduce from Corollary 1.3 that

$$\frac{\tilde{C}_{X_1\times M_2}}{\tilde{C}_M} \le q_M \le \frac{\tilde{C}_{X_1\times M_2}}{\tilde{c}_M},$$

where $\tilde{C}_{X_1\times M_2}$, $\tilde{C}_M$ and $\tilde{c}_M$ are respectively defined as $C^\sharp_{X_1\times M_2}$, $C^\sharp_M$ and $c^\sharp_M$ of (1.17) with $\|\cdot\|_\sim$ in place of $\|\cdot\|_1$ and $\|\cdot\|_\sharp = \|\cdot\|_2$.

Remark 1.5 (Quasi-optimality of dual method)**.**
We can exchange the spaces X_1 and M_1 with X_2 and M_2 respectively, and consider the quasi-optimality constant q^*_M related to the dual Galerkin solution and defined as the smallest constant $q^* \ge 0$ such that

$$\|\phi - \Phi_M\|_2 \le q^* \inf_{\eta\in M_2} \|\phi - \eta\|_2 .$$

Assuming X_2 to be a Hilbert space, we derive as above that $q^*_M = \|R^*_M\|_{\mathcal{L}(X_2)}$ and recalling (1.5) we have that

$$\max\left\{\frac{c_b}{c_M}, \frac{C^\sharp_{M_1\times X_2}}{C^\sharp_M}\right\} \le q^*_M \le \min\left\{\frac{C_b}{c_M}, \frac{C^\sharp_{M_1\times X_2}}{c^\sharp_M}\right\},$$

where $C^\sharp_{M_1\times X_2}$ is defined as $C^\sharp_{X_1\times M_2}$ of (1.17) and $C^\sharp_M$ and $c^\sharp_M$ are as in (1.17) with $(M_1, \|\cdot\|_\sharp)$ in place of $(M_2, \|\cdot\|_\sharp)$, and $(X_2, \|\cdot\|_2)$ in place of $(X_1, \|\cdot\|_1)$. We notice that q^*_M and q_M share the bounds

$$\frac{c_b}{c_M} \le q_M, q^*_M \le \frac{C_b}{c_M}.$$

1.3 The non–conforming case

In the previous section $M_1 \subset X_1$, $M_2 \subset X_2$ and the exact solution satisfies

$$b(u,\varphi) = \ell(\varphi), \qquad \forall\varphi \in M_2. \tag{1.19}$$

In this section we analyse the case where $M_1 \nsubseteq X_1$ or $M_2 \nsubseteq X_2$ or the discrete solution is defined via a bilinear form b_M and a linear functional ℓ_M possibly different from b and ℓ. Inserting the exact solution u in the discrete problem may not give an equality as in (1.19).

We endow M_1 and M_2 with $\|\cdot\|_{1,\sim}$ and $\|\cdot\|_{2,\sim}$ respectively, which may differ from $\|\cdot\|_1$ and $\|\cdot\|_2$. We assume that $b_M : X_1 + M_1 \times M_2 \to \mathbb{R}$ is a continuous bilinear form with respect to $\|\cdot\|_{1,\sim}$ and $\|\cdot\|_{2,\sim}$, with constant $C_\sim$

and that it satisfies the inf-sup condition on $(M_1, \|\cdot\|_{1,\sim}) \times (M_2, \|\cdot\|_{2,\sim})$, with constant c_M. Given $\ell_M \in M_2^*$, the discrete problem reads

$$\text{find } U_M \in M_1 \text{ such that, } \forall \varphi \in M_2, \quad b_M(U_M, \varphi) = \ell_M(\varphi).$$

In order to estimate the error $\|u - U_M\|_{1,\sim}$ we follow the strategy in [11, Sect. 10.1]. We consider the Ritz-projection $R_M : X_1 + M_1 \to M_1$ such that

$$\forall \varphi \in M_2 \qquad b_M(R_M v, \varphi) = b_M(v, \varphi),$$

and we bound the error in terms of the deviation of U_M from $R_M u$:

$$\|u - U_M\|_{1,\sim} \leq \|u - R_M u\|_{1,\sim} + \|R_M u - U_M\|_{1,\sim}.$$

We assume that $(X_1 + M_1, \|\cdot\|_{1,\sim})$ is a Hilbert space, applying the results of the previous section we get

$$\|u - R_M u\|_{1,\sim} \leq \frac{C_\sim}{c_M} \inf_{v \in M_1} \|u - v\|_{1,\sim}.$$

Regarding $\|R_M u - U_M\|_{1,\sim}$, we have

$$c_M \|R_M u - U_M\|_{1,\sim} \leq \sup_{\varphi \in M_2} \frac{b_M(R_M u - U_M, \varphi)}{\|\varphi\|_{2,\sim}} \leq C_\sim \|R_M u - U_M\|_{1,\sim}.$$

The term

$$\sup_{\varphi \in M_2} \frac{b_M(R_M u - U_M, \varphi)}{\|\varphi\|_{2,\sim}} = \sup_{\varphi \in M_2} \frac{b_M(u, \varphi) - \ell_M(\varphi)}{\|\varphi\|_{2,\sim}}$$

can be seen as a consistency error, due to the fact that the exact solution does not satisfy the discrete problem. It measures implicitly the discrepancy between b and b_M and ℓ and ℓ_M. Therefore

$$\|u - U_M\|_{1,\sim} \leq \frac{C_\sim}{c_M} \inf_{v \in M_1} \|u - v\|_{1,\sim} + c_M^{-1} \sup_{\varphi \in M_2} \frac{b_M(u, \varphi) - \ell_M(\varphi)}{\|\varphi\|_{2,\sim}}.$$

We notice that the right-hand side is equivalent to the error, in fact

$$\begin{aligned}
&\frac{C_\sim}{c_M} \inf_{v \in M_1} \|u - v\|_{1,\sim} + c_M^{-1} \sup_{\varphi \in M_2} \frac{b_M(u, \varphi) - \ell_M(\varphi)}{\|\varphi\|_{2,\sim}} \\
&\qquad \leq \frac{C_\sim}{c_M} \left(\|u - U_M\|_{1,\sim} + \|R_M u - U_M\|_{1,\sim} \right) \\
&\qquad \leq \frac{C_\sim}{c_M} \left(\frac{C_\sim}{c_M} + 2 \right) \|u - U_M\|_{1,\sim}.
\end{aligned}$$

Chapter 2

Abstract Linear Parabolic Problem

The purpose of this chapter is to recall the setting of abstract linear parabolic problems, whose Petrov-Galerkin approximations we are interested in. In order to apply the results of Chapter 1 we reformulate the problem by means of a bilinear form in two different ways. The first one, called "standard", involves the time derivative of the exact solution, while in the second one, called "natural", the time derivative is shifted to the test function.

The chapter is organized as follows. In Section 2.1 we specify the notations and the common assumptions on the two bilinear formulations of the problem. In Section 2.2 and 2.3 we present the standard and natural weak formulation respectively. We recall the proof of well-posedness by means of the inf-sup theory, whose structure is reproduced in the discrete framework, and we provide bounds for the constants associated to the bilinear forms. In Section 2.4 we recall some regularity results to motivate the assumptions for the error estimates in Chapter 7.

2.1 Abstract parabolic problem

Parabolic initial boundary value problems are defined on a space-time cylinder $Q = \Omega \times I$, where $\Omega \subset \mathbb{R}^n$ and $I = (0, T)$, $T > 0$ is a time interval. The unknown function u can be interpreted as a time-dependent function with values in a functional space. The problem can thus be rewritten as an initial value problem that formally reads

$$u' + Au = f \text{ in } I, \qquad u(0) = w, \tag{2.1}$$

where A is an elliptic operator acting on a Hilbert space V which also takes into account the boundary conditions, f is a forcing term, and w is an initial

value from a Hilbert space W. We specify below the assumptions on the spaces V and W and the elliptic operator A.

We assume that $(V, \|\cdot\|_V)$ and $(W, \|\cdot\|_W)$ are two separable Hilbert spaces such that

$$V \subset W \subset V^*$$

forms a Hilbert triplet. More precisely, we assume that the embedding $V \subset W$ is continuous and dense, view W^* as a subspace of V^*, and identify W and its dual W^* with the help of Riesz representation theorem. The scalar product in W as well as the duality pairing of $V^* \times V$ is denoted by $\langle \cdot, \cdot \rangle$. The norm on V^* is indicated by $\|\cdot\|_{V^*} = \sup_{\|v\|_V=1} \langle \cdot, v \rangle$.

We will use H-valued functions depending on time, where H is a Hilbert space, e.g. $H = V, W, V^*$. For the corresponding function spaces, see e.g. [28, Sect. 5.9.2] for a brief review. In particular, we use the following ones over a proper time interval J. Let $C^r(J; H)$, $r \in \mathbb{N}$, denote the space of all functions from J to H that are continuous, together with their classical derivatives up to order r and let $C_0^\infty(J; H)$ denote the space of all functions in $\cap_{r\in\mathbb{N}} C^r(J; H)$ with compact support in J. Furthermore $L^2(J; H)$ denotes the space of functions of the form $J \to H$ that are measurable and square-integrable with respect to the Bochner integral. With $H^1(J; H)$ we denote the space of all functions in $L^2(J; H)$ whose distributional derivative is square-integrable. Finally, we set $H^1(J; V, V^*) := \{v \in L^2(J; V) \mid v' \in L^2(J; V^*)\}$. If $J = I$, we suppress the time interval and write, e.g., $H^1(V, V^*) := H^1(I; V, V^*)$ for short.

We assume that the elliptic operator A arises from a bilinear form a that depends on time and is bounded and coercive in the following sense:

$$a(\cdot; v, \varphi) \text{ is measurable in } I \text{ for any } v, \varphi \in V, \tag{2.2a}$$

$$\nu_a := \inf_{t\in I} \inf_{\|v\|_V=1} a(t; v, v) > 0, \tag{2.2b}$$

$$C_a := \sup_{t\in I} \sup_{\|v\|_V=\|\varphi\|_V=1} a(t; v, \varphi) < \infty, \tag{2.2c}$$

where inf and sup with respect to time are essential ones. Note that (2.2b) and (2.2c) are equivalent to requiring that $\int_I a(t, \cdot, \cdot)\, \mathrm{d}t$ is a bilinear form on $L^2(V)$ with optimal coercivity and continuity constants ν_a and C_a, respectively.

The operator A in (2.1) is of the form $I \to \mathcal{L}(V, V^*)$ and defined by the requirement

$$\langle A(t)v, \varphi \rangle = a(t; v, \varphi),$$

for any $t \in I$. In the following lemma about the adjoint inverse $A^{-*}(t) := [A(t)^*]^{-1}$ we suppress t for notational simplicity.

Lemma 2.1 (Adjoint inverse of elliptic operator)**.** *If A arises from a bilinear form a that is coercive and continuous on V with constants ν_a and C_a, then the adjoint inverse exists and satisfies*

$$C_a^{-1} \|\ell\|_{V^*} \le \left\|A^{-*}\ell\right\|_V \le \nu_a^{-1} \|\ell\|_{V^*} \text{ and } \left\langle \ell, A^{-*}\ell \right\rangle \ge \nu_a \left\|A^{-*}\ell\right\|_V^2 .$$

for any $\ell \in V^*$.

Proof. The Lax-Milgram Theorem implies that A is invertible and so its (adjoint) inverse exists. Taking $\psi = A^{-*}\ell$, we obtain the first inequality

$$\|\ell\|_{V^*} = \|A^*\psi\|_{V^*} = \sup_{\|\varphi\|_V=1} a(\varphi,\psi) \le C_a \left\|A^{-*}\ell\right\|_V$$

from the continuity of a. Its coercivity and the definition of A^* give the third inequality,

$$\nu_a \left\|A^{-*}\ell\right\|_V^2 \le a(A^{-*}\ell, A^{-*}\ell) = \left\langle \ell, A^{-*}\ell \right\rangle ,$$

which in turn implies the second one. □

A typical example of a parabolic problem is

$$\begin{aligned} \partial_t u - \operatorname{div}(\mathbf{A}\nabla u) &= f && \text{in } \Omega \times (0,T) \\ u &= 0 && \text{on } \partial\Omega \times (0,T) \\ u(0) &= u_0 && \text{in } \Omega, \end{aligned} \tag{2.3}$$

where $\mathbf{A}$ satisfies, for every $x \in \Omega$, $t \in (0,T)$, and $\xi \in \mathbb{R}^n$

$$\lambda|\xi|^2 \le \xi \cdot \mathbf{A}(x,t)\xi \le \Lambda|\xi|^2,$$

with $\Lambda \ge \lambda > 0$. Problem (2.3) fits into the above framework with $V = H_0^1(\Omega)$, $W = L^2(\Omega)$, and $V^* = H^{-1}(\Omega)$.

2.2 Standard weak formulation

We recall the setting of the standard weak formulation, rewrite it in the form (1.3), and provide bounds for the constants associated to the bilinear form.

Assume that $V \subset W \subset V^*$, a and A are as in §2.1. The standard weak formulation of the abstract initial value problem (2.1) reads

$$\begin{gathered} \text{given } f \in L^2(V^*) \text{ and } w \in W, \text{ find } u \in H^1(V,V^*) \text{ such that} \\ u' + Au = f \text{ in } I, \quad u(0) = w. \end{gathered} \tag{2.4}$$

The differential equation should be interpreted in the sense of V^*-valued distributions. The initial condition, which is formulated in an essential manner,

is meaningful thanks to the embedding $H^1(V,V^*) \subset C^0(W)$; see [28, Sect. 5.9.2, Thm. 3]. Problem (2.4) is well-posed; see, e.g. [28, Sect. 7.1.2] for a proof by means of the Faedo-Galerkin method.

Throughout this section we set

$$X_1 := H^1(V,V^*) \quad \text{with} \quad \|v\|_1^2 := \|v(0)\|_W^2 + \int_I \|v\|_V^2 + \|v'\|_{V^*}^2 \,, \tag{2.5a}$$

$$\begin{aligned} X_2 &:= \{\varphi = (\varphi_0,\varphi_1) \mid \varphi_0 \in W, \varphi_1 \in L^2(V)\} \\ &\text{with} \quad \|\varphi\|_2^2 := \|\varphi_0\|_W^2 + \int_I \|\varphi_1\|_V^2 \,, \end{aligned} \tag{2.5b}$$

$$b(v,\varphi) := \langle v(0),\varphi_0\rangle + \int_I \langle v',\varphi_1\rangle + a(\cdot;v,\varphi_1) \tag{2.5c}$$

and

$$\ell(\varphi) := \langle w,\varphi_0\rangle + \int_I \langle f,\varphi_1\rangle \,.$$

We refer to b as the standard bilinear form. With these definitions, (2.4) is equivalent to (1.3). Note that the decoupling of differential equation and initial condition in the test space reflects the essential nature of the latter. The term $\|v(0)\|_W$ in the definition of $\|\cdot\|_1$ allows to avoid the use of the embedding $H^1(V,V^*) \subset C^0(W)$ when bounding the continuity constant C_b. If it is omitted, there appears a dependence on T for small T in the following results.

The following proposition investigates the properties of the bilinear form b. Its proof of the inf-sup condition for b contains elements from Ern and Guermond [27, Thm. 6.6] and Stevenson and Schwab [38, Thm. 5.1]. Together with the abstract theory of Chapter 1, Proposition 2.2 provides not only an alternative approach to existence and uniqueness for (2.4) but also serves as a guideline for analysing Galerkin approximation in space.

Proposition 2.2 (Standard bilinear form)**.** *The bilinear form b in* (2.5) *is continuous and satisfies the inf-sup condition with*

$$C_b \le \sqrt{2}\max\{1,C_a\}, \qquad c_b \ge \frac{\min\{\nu_a, C_a^{-1}, \nu_a C_a^{-1}\}}{2}.$$

Proof. In order to verify the first bound, let $v \in X_1$ and $\varphi = (\varphi_0,\varphi_1) \in X_2$ and derive

$$\begin{aligned} b(v,\varphi) &\le \|v(0)\|_W \|\varphi_0\|_W + \int_I \Big(\|v'\|_{V^*} + C_a \|v\|_V \Big) \|\varphi_1\|_V \\ &\le \left(\|v(0)\|_W^2 + 2\int_I \|v'\|_{V^*}^2 + C_a^2 \|v\|_V^2 \right)^{1/2} \left(\|\varphi_0\|_W^2 + \int_I \|\varphi_1\|_V^2 \right)^{1/2} \\ &\le \sqrt{2}\max\{1,C_a\} \|v\|_1 \|\varphi\|_2 \,. \end{aligned}$$

Next, we verify the second bound, which implies the inf-sup condition (1.4a). Given $v \in X_1 \setminus \{0\}$, we choose

$$\varphi_0 = 2v(0) \quad \text{and} \quad \varphi_1(t) = v(t) + A^{-*}(t)v'(t),\ t \in I.$$

Using the identities $a(\cdot; v, A^{-*}v') = \langle v', v\rangle$ and $2\int_I \langle v', v\rangle = \|v(T)\|_W^2 - \|v(0)\|_W^2$, coercivity (2.2b) of the elliptic bilinear form and Lemma 2.1, we derive

$$\begin{aligned} b(v,\varphi) &= \|v(0)\|_W^2 + \int_I \left\langle v', A^{-*}v' \right\rangle + a(\cdot; v, v) + \|v(T)\|_W^2 \\ &\geq \min\{1, \nu_a\} \left(\|v(0)\|_W^2 + \int_I \left\|A^{-*}v'\right\|_V^2 + \|v\|_V^2 \right) \end{aligned} \tag{2.6}$$

and

$$\|v(0)\|_W^2 + \int_I \left\|A^{-*}v'\right\|_V^2 + \|v\|_V^2 \geq \min\{1, C_a^{-2}\} \|v\|_1^2 . \tag{2.7}$$

On the other hand, using again Lemma 2.1, we obtain

$$\|\varphi\|_2^2 \leq 4 \left(\|v(0)\|_W^2 + \int_I \|v\|_V^2 + \left\|A^{-*}v'\right\|_V^2 \right). \tag{2.8}$$

Combining (2.6) and (2.7) yields $b(v,\varphi) > 0$ and so $\varphi \neq 0$. Using first (2.6) and (2.8) and then (2.7) we arrive at

$$\frac{b(v,\varphi)}{\|\varphi\|_2} \geq \frac{\min\{\nu_a, C_a^{-1}, \nu_a C_a^{-1}\}}{2} \|v\|_1$$

which implies the second claimed bound.

Finally, we verify the non-degeneracy condition (1.4b). To this end, we assume that φ satisfies

$$\forall v \in X_1 \quad b(v,\varphi) = 0 \tag{2.9}$$

and observe

$$\int_I \langle v', \varphi_1\rangle = -\int_I a(\cdot; v, \varphi_1) \leq C_a \|v\|_{L^2(V)} \|\varphi_1\|_{L^2(V)} .$$

for all $v \in C_0^\infty(V)$. Since $C_0^\infty(V)$ is dense in $L^2(V)$ and the spaces $L^2(V)^*$ and $L^2(V^*)$ are isomorphic, we obtain the additional regularity $\varphi_1 \in H^1(V^*)$. We therefore can integrate by parts in (2.9) and see that $\varphi_1 \in H^1(V, V^*)$ solves the backward-in-time problem

$$\forall v \in X_1 \quad \langle v(T), \varphi_1(T)\rangle - \langle v(0), \varphi_1(0) - \varphi_0\rangle + \int_I \langle -\varphi_1', v\rangle + a(\cdot; v, \varphi_1) = 0.$$

We derive $-\varphi_1' + A^*\varphi_1 = 0$, $\varphi_1(T) = 0$ and $\varphi_1(0) = \varphi_0$ by testing with appropriate functions $v \in X_1$. Using these facts for $v = \varphi_1$ yields

$$\frac{1}{2}\|\varphi_0\|_W^2 + \nu_a \|\varphi_1\|_{L^2(V)}^2 \leq 0$$

and we conclude $\varphi = 0$. □

2.3 Natural weak formulation

In order to obtain a solution notion that requires less regularity in time, one may integrate by parts the terms with the time derivative, assuming the test function to be more regular. There is essentially an exchange between trial and test space, which loses its two-components structure. This entails that the initial condition is formulated in a natural way.

Assume that $V \subset W \subset V^*$, a and A are as in Section 2.1. In their terms the natural weak formulation may be written as:

$$\begin{aligned} &\text{given } \ell \in \{\varphi \in H^1(V, V^*) : \varphi(T) = 0\}^* \text{ find } u \in L^2(V) \text{ such that} \\ &\forall \varphi \in H^1(V, V^*) \text{ with } \varphi(T) = 0, \quad \int_I -\langle \varphi', u\rangle + \langle Au, \varphi\rangle = \ell(\varphi). \end{aligned} \tag{2.10}$$

We could choose ℓ of the form

$$\ell(\varphi) := \langle w, \varphi(0)\rangle + \int_I \langle f, \varphi\rangle\,, \tag{2.11}$$

where $f \in L^2(V^*)$ and $w \in W$. In this case the solution of Problem (2.4) also verifies (2.10). Throughout this section we set

$$\begin{aligned} X_1 &:= L^2(V) \qquad \text{with} \quad \|v\|_1^2 := \int_I \|v\|_V^2\,, \\ X_2 &:= \{\varphi \in L^2(V) \mid \varphi' \in L^2(V'), \varphi(T) = 0\} \\ &\qquad \text{with} \quad \|\varphi\|_2^2 := \int_I \|\varphi\|_V^2 + \|\varphi'\|_{V^*}^2\,, \\ b(v, \varphi) &:= \int_I -\langle \varphi', v\rangle + a(\cdot; v, \varphi). \end{aligned} \tag{2.12}$$

With these definitions, (2.10) is equivalent to (1.3). The following proposition verifies the inf-sup condition for b, and shows that Problem (2.10) is well-posed. In the case ℓ is as in (2.11) its unique solution coincides with the one of Problem (2.4) and belongs to $H^1(V, V^*)$.

Proposition 2.3 (Natural bilinear form)**.** *The bilinear form b in* (2.12) *is continuous and satisfies the inf-sup condition* (1.4) *with*

$$C_b \leq \sqrt{2}\max\{1, C_a\}, \qquad c_b \geq \frac{\nu_a}{\sqrt{2}}\min\{1, C_a^{-1}\}. \tag{2.13}$$

Proof. The proof follows the same lines as the one of Proposition 2.2. The main difference concerns the second bound, where we prove the symmetric variant of (1.4a). For every $\varphi \in X_2$ we take as test function $v = \varphi - A^{-1}\varphi'$, and we exploit Lemma 2.1 with A^* in place of A. □

Remark 2.4 (Duality)**.** The similarities between Propositions 2.2 and 2.3 are not a coincidence. In fact, Problem (2.10) and the dual problem of (2.4) with equal right-hand-side $\ell \in H^1(V, V^*)^*$ are strictly related. If we apply an affine transformation in time, and define $\hat{u} := u(T - \cdot)$, from (2.10) we have that $\hat{u}$ satisfies, for every $v \in H^1(V, V^*)$ with $v(0) = 0$,

$$\int_I \langle v', \hat{u}\rangle + \langle A^*(T-\cdot)v, \hat{u}\rangle = \ell(v). \tag{2.14}$$

On the other hand, the solution (ϕ_0, ϕ_1) of the aforementioned dual problem satisfies, for every $v \in H^1(V, V^*)$ with $v(0) = 0$,

$$\int_I \langle v', \phi_1\rangle + \langle Av, \phi_1\rangle = \ell(v),$$

that is (2.14) with A in place of $A^*(T - \cdot)$.

2.4 Regularity results

We recall some regularity results for the standard formulation. For the rest of this section, let u be the solution of (2.4).

We start with temporal regularity. We set $\mathcal{D}(A) := \{v \in V,\ Av \in W\}$ and assume $u_0 \in \mathcal{D}(A(0))$, $f \in H^1(V^*) \cap L^2(V)$ and $A \in C^1(\mathcal{L}(V, V^*))$. Then, see [36, Sect. 11.1.4], it holds

$$u' \in H^1(V^*) \cap L^2(V).$$

Concerning spatial regularity, we assume that A is independent of time and symmetric, that is $\langle Au, v\rangle = \langle Av, u\rangle$ for every $u, v \in V$. Moreover assume that the injection of V in W is compact, $u_0 \in V$ and $f \in L^2(W)$. Then, see [40, Ch. II, Thm. 3.3], it holds

$$u \in L^2(D(A)) \cap C^0(V), \qquad \text{and} \qquad u' \in L^2(W).$$

For higher regularity, assume that $\Omega \subset \mathbb{R}^d$ is a bounded domain, $V = H_0^1(\Omega)$, $W = L^2(\Omega)$, $V^* = H^{-1}(\Omega)$ and A is of the form

$$Au = -\operatorname{div}(\mathbf{A}\nabla u)$$

with $\mathbf{A} = (a_{ij}(x))_{i,j=1}^d$. Assume that $\partial\Omega$ is C^{2m+2}, $a_{ij} \in C^{2m+1}(\bar{\Omega})$, $u_0 \in H^{2m+1}(\Omega) \cap H_0^1(\Omega)$, and

$$\frac{\mathrm{d}^k f}{\mathrm{d}t^k} \in L^2(H^{2m-2k}), \qquad k = 0, \dots, m.$$

Assume also that the following compatibility conditions hold

$$g_1 := f(0) - Au_0 \in H_0^1(\Omega), \dots, g_m := \frac{\mathrm{d}^{m-1} f}{\mathrm{d}t^{m-1}}(0) - Ag_{m-1} \in H_0^1(\Omega).$$

Then, see [28, Sect. 7.1, Thm. 6], it holds

$$\frac{\mathrm{d}^k u}{\mathrm{d}t^k} \in L^2(H^{2m+2-2k}), \qquad k = 0, \dots, m+1.$$

Chapter 3

The Role of the L^2-projection in the Spatial Semidiscretization

Galerkin finite element methods are widely used for the numerical solution of parabolic problems, see the monograph [41] of Thomée for an overview of the corresponding error bounds. Remarkably, the derivation of most a priori error bounds for linear parabolic problems differs from those for linear elliptic problems: a quasi-optimality (or near best) result like Céa's Lemma is not invoked.

Douglas and Dupont [21] recognized this difference in 1970, and derived a quasi-optimality result for the approximation in space in a norm involving a time derivative of order 1/2, assuming that the initial error vanishes. This approach has been generalized by Baiocchi and Brezzi [3] to fractions around 1/2 and by Tomarelli [42] also to general initial values.

Other three results, that we are aware of, concern quasi-optimality for the approximation in space in simpler norms, which are related to the standard and natural weak formulations of parabolic problems recalled in Chapter 2. However, none of these results perfectly mimics the elliptic case. Dupont [22] derived quasi-optimality in a norm close to the one of $H^1(H^{-1}) \cap L^2(H^1)$, but that depends on the discretization. The other two results, [13, 30] require that the L^2-projection is H^1-stable. With this assumption, Hackbusch [30] in particular established stability in $L^2(H^1)$, while Chrysafinos and Hou [13] showed quasi-optimality in $H^1(H^{-1}) \cap L^2(H^1)$.

The purpose of this chapter is to clarify the role of this hypothesis, providing an approach to quasi-optimality by means of the inf-sup theory, recalled in Chapter 1. In particular we re-establishes the last three results, show that they are interrelated, and that the H^1-stability of the L^2-projection in [30, 13] is necessary.

The chapter is organized as follows. Section 3.1 concerns conforming di-

cretizations of Hilbert triplets and the relationship between the L^2-projection and the norms on the dual discrete space. In Sections 3.2 and 3.3 we analyse Galerkin approximations in the $L^2(H^1) \cap H^1(H^{-1})$- and $L^2(H^1)$-norm respectively. In other words, we apply the Petrov-Galerkin method of Section 1.1 to the standard and natural weak formulation respectively. In both cases we prove that the H^1-stability of the L^2-projection is necessary for quasi-optimality.

3.1 Conforming discretization of Hilbert triplets

Let $V \subset W \subset V^*$ be a Hilbert triplet like in §2.1 and S a finite-dimensional, non-trivial, and proper subspace of V. Observe that S is also a subspace of W and thus, with the identification $S^* = S$, also of V^*.

As a subspace of V or W, we equip S with the norm $\|\cdot\|_V$ or $\|\cdot\|_W$, respectively. As a subspace of V^*, the situation is less clear. In fact, we may equip $S^* = S$ with

$$\|s\|_{V^*} = \sup_{\|\varphi\|_V=1} \langle s, \varphi \rangle \qquad \text{or} \qquad \|s\|_{S^*} := \sup_{\varphi \in S : \|\varphi\|_V=1} \langle s, \varphi \rangle \,. \tag{3.1}$$

The two alternatives give precedence to one of the following two properties of $S = S^*$: S is a subset of V^* and S^* is a dual space of $(S, \|\cdot\|_V)$. In what follows we show that parabolic quasi-optimality requires that the two norms in (3.1) are equivalent. In view of $S \subset V$, we immediately see that

$$\forall s \in S \qquad \|s\|_{S^*} \leq \|s\|_{V^*} \,.$$

Since S is finite-dimensional, also the other direction is true up to a constant, which may not be uniform for a family of subspaces.

In order to reveal the nature of the critical equivalence constant

$$c_S := \sup_{s \in S^*} \frac{\|s\|_{V^*}}{\|s\|_{S^*}},$$

we observe that the duality pairing arising in both norms of (3.1) is closely related with the scalar product of W in the given setting. We therefore associated with S its W-orthogonal projection and investigate its relationship with the spaces of the Hilbert triplet and dual of S.

The W-orthogonal projection onto S, or W-projection for short, is defined as follows:

$$\forall w \in W, \varphi \in S \qquad P_S w \in S \text{ and } \langle P_S w, \varphi \rangle = \langle w, \varphi \rangle \,. \tag{3.2}$$

This linear projection acting on W has the following properties: it is symmetric and there hold $\|P_S\|_{\mathcal{L}(W)} = 1 = \|I - P_S\|_{\mathcal{L}(W)}$, where I is the identity operator. It is also a linear projection acting on V thanks to $S \subset V$. The following lemma, which essentially can be found also in Chrysafinos and Hou [13], shows that P_S may be viewed also as a linear projection acting on V^*. For the sake of completeness, we provide its proof.

Lemma 3.1 (W-projection in V^*)**.** *The linear projection P_S extends to V^*, maintaining its symmetry in that $\langle \ell_1, P_S \ell_2 \rangle = \langle \ell_2, P_S \ell_1 \rangle$ for all $\ell_1, \ell_2 \in V^*$.*

Proof. First note that, thanks to $S \subset V$, the right-hand side of (3.2) is defined also if $w \in W$ is replaced by some functional $\ell \in V^*$. In this case, given some basis $\{e_j\}_{j=1}^n$ of S and assuming $P_S \ell = \sum_{j=1}^n \alpha_j e_j$, definition (3.2) is equivalent to the linear system

$$\sum_{j=1}^{n} \langle e_i, e_j \rangle \, \alpha_j = \langle \ell, e_i \rangle \,, \quad i = 1, \ldots, n.$$

Since its matrix is symmetric and positive definite, the latter admits a unique solution and $P_S \ell$ is well-defined for any $\ell \in V^*$.
If ℓ_1, $\ell_2 \in V^*$, the symmetry of the scalar product in W yields

$$\langle \ell_1, P_S \ell_2 \rangle = \langle P_S \ell_1, P_S \ell_2 \rangle = \langle P_S \ell_2, P_S \ell_1 \rangle = \langle \ell_2, P_S \ell_1 \rangle \,. \qquad \square$$

The next proposition shows that the critical equivalence constant for the norms in (3.1) is intimately related to P_S.

Proposition 3.2 (W-projection and norm equivalence on discrete dual)**.** *The equivalence constant c_S can be expressed in terms of the projection P_S as follows:*

$$c_S = \|P_S\|_{\mathcal{L}(V^*)} = \|I - P_S\|_{\mathcal{L}(V^*)} = \|I - P_S\|_{\mathcal{L}(V)} = \|P_S\|_{\mathcal{L}(V)} \,.$$

Proof. Since P_S is a linear projection on the nontrivial proper subspace S of V and V^*, Lemma 1.1 implies the second and last inequality. We conclude by showing

$$\|P_S\|_{\mathcal{L}(V^*)} = \|P_S\|_{\mathcal{L}(V)} \quad \text{and} \quad \|P_S\|_{\mathcal{L}(V^*)} \leq c_S \leq \|P_S\|_{\mathcal{L}(V)} \,.$$

The equality readily follows from

$$\forall \varphi \in V, \ell \in V^* \qquad \langle P_S \ell, \varphi \rangle = \langle \ell, P_S \varphi \rangle \,,$$

which is a consequence of the symmetry statement in Lemma 3.1. To show the first inequality, let $\ell \in V^*$ and observe

$$\|P_S \ell\|_{V^*} \le c_S \|P_S \ell\|_{S^*} = c_S \|\ell\|_{S^*} \le c_S \|\ell\|_{V^*} .$$

Finally, the second inequality follows from

$$\|s\|_{V^*} = \sup_{\|\varphi\|_V = 1} \langle s, \varphi \rangle = \sup_{\|\varphi\|_V = 1} \langle s, P_S \varphi \rangle \le \|P_S\|_{\mathcal{L}(V)} \|s\|_{S^*}$$

for every $s \in S$. □

3.2 Galerkin approximation in $L^2(H^1) \cap H^1(H^{-1})$

We recall the Galerkin approximation of the standard weak formulation and derive a discrete inf-sup condition by mimicking the proof of Proposition 2.2. This allows to establish that the Galerkin approximation is well-defined and that it satisfies a symmetric error estimate. Throughout this section we set $X_1 := H^1(V, V^*)$, $X_2 := W \times L^2(V)$ with norms

$$\|v\|_1^2 = \|v(0)\|_W^2 + \int_0^T \|v'\|_{V^*}^2 + \|v\|_V^2 , \qquad \|\varphi\|_2^2 = \|\varphi_0\|_W^2 + \int_0^T \|\varphi_1\|_V^2 .$$

Moreover $u_0 \in W$, $f \in L^2(V)$, and the bilinear form $b : X_1 \times X_2 \to \mathbb{R}$ and $\ell \in X_2^*$ are given by

$$\begin{aligned} b(v, \varphi) &:= \langle v(0), \varphi_0 \rangle + \int_I \langle v', \varphi_1 \rangle + a(\cdot; v, \varphi_1), \\ \ell(\varphi) &:= \langle w, \varphi_0 \rangle + \int_I \langle f, \varphi_1 \rangle , \end{aligned}$$

as in Section 2.2.

Let $S \subset V$ be a finite-dimensional subspace and recall that S is also a subspace of W and V^* and that it is identified with its dual. The Galerkin approximation of (2.4) is

$$\begin{gathered} \text{given } f \in L^2(V^*) \text{ and } s \in S, \text{ find } U_S \in H^1(S) \text{ such that} \\ \forall t \in I, \varphi \in S, \quad \langle U_S'(t), \varphi \rangle + a(t; U_S(t), \varphi) = \langle f(t), \varphi \rangle , \quad U(0) = s; \end{gathered} \tag{3.3}$$

see also Thomée [41, Ch. 1]. Of course $s \in S$ should be an approximation of the initial value $w \in W$.

We set

$$M := M_1 \times M_2 \quad \text{with} \quad M_1 := H^1(S), \qquad M_2 := S \times L^2(S) \tag{3.4}$$

and observe $M_i \subset X_i$ for $i = 1, 2$, where we use that S is also a subset of W and V^*. With these definitions, (3.3) is equivalent to (1.8) if and only if $s = P_S w$, where P_S is the W-orthogonal projection onto S; see also (3.2). For the rest of this section we assume (3.4) and $s = P_S w$.

In order to establish the crucial inf-sup condition for M, one may try to proceed as in the corresponding part of the proof of Proposition 2.2. However for $v \in M_1$, the function $A^{-*}v'$ may not be in M_2. To remedy, we propose to replace A by its discrete counterpart $A_S : I \to \mathcal{L}(S, S^*)$ defined by

$$\langle A_S(t)v, \varphi \rangle = a(t; v, \varphi).$$

Unfortunately, this replacement comes at the price that the set M_1 has to be equipped with the S-dependent norm

$$\|v\|_{1;S}^2 := \|v(0)\|_W^2 + \int_I \|v\|_V^2 + \|v'\|_{S^*}^2 \tag{3.5}$$

and that the argument provides a lower bound for

$$\tilde{c}_M := \inf_{v \in M_1} \sup_{\varphi \in M_2} \frac{b(v, \varphi)}{\|v\|_{1;S} \, \|\varphi\|_2},$$

a variant of the discrete inf-sup constant c_M.

Proposition 3.3 (Standard bilinear form and Galerkin approximation)**.** *If we equip, respectively, M_1 and M_2 with $\|\cdot\|_{1;S}$ and $\|\cdot\|_2$, the bilinear form b in* (2.5) *is continuous and satisfies the inf-sup condition on $M_1 \times M_2$ with*

$$\tilde{C}_M \le \sqrt{2} \max\{1, C_a\}, \qquad \tilde{c}_M \ge \frac{\min\{\nu_a, C_a^{-1}, \nu_a C_a^{-1}\}}{2}.$$

Proof. As the proof mimics the one of Proposition 2.2, we comment only on the differences after having replaced, respectively, X_1, X_2, A and $\|\cdot\|_1$ by M_1, M_2, A_S and $\|\cdot\|_{1;S}$.

When verifying the continuity, the replacement of the norm $\|\cdot\|_{V^*}$ by the weaker one $\|\cdot\|_{S^*}$ is compensated by the fact that the test function comes from the semidiscrete space M_2. In order to derive the lower bound for $\tilde{c}_M$, we choose

$$\varphi_0 = 2v(0) \quad \text{and} \quad \varphi_1(t) = v(t) + A_S^{-*}(t)v'(t), \ t \in I,$$

and exploit

$$C_a^{-1} \|\ell\|_{S^*} \le \left\|A_S^{-*}(t)\ell\right\|_V \le \nu_a^{-1} \|\ell\|_{S^*} \text{ and } \left\langle \ell, A_S^{-*}(t)\ell \right\rangle \ge \nu_a \left\|A_S^{-*}(t)\ell\right\|_V^2,$$

which follow by applying Lemma 2.1 with M_1, M_2 and A_S in place of X_1, X_2 and A. Verifying the non-degeneracy goes along the same lines. □

Remark 3.4 (Relation between A and A_S)**.** The replacement of A by A_S may be seen also in the following way: we replace $A^{-*}v' \in X_2$ by an approximation in M_2. In view of the norm of X_2, another natural approximation appears to be the pointwise Ritz projection R_S with respect to the elliptic bilinear form a. These two replacements actually coincide. In fact, there holds

$$\begin{aligned} \forall t \in I, \ell \in V^*, \varphi \in S \quad a(t; A_S(t)^{-1}\ell_{|S}, \varphi) &= \langle \ell, \varphi \rangle = a(t; A(t)^{-1}\ell, \varphi) \\ &= a(t; R_S A(t)^{-1}\ell, \varphi) \end{aligned}$$

and so $A_S^{-1}\ell = R_S A^{-1}\ell_{|S}$ for all $\ell \in V^*$.

Of course, Proposition 3.3, together with the abstract theory of Chapter 1, provides an alternative proof that the Galerkin approximation with $s = P_S w$ is well-defined for any $f \in L^2(V^*)$ and $w \in W$. It also provides an error estimate, with the help of Remark 1.4. In fact, the S-dependent norm $\|\cdot\|_{1;S}$ is defined also on X_1 and so can be used to measure the error. Noteworthy, the bilinear b is not continuous when X_1 and X_2 are equipped with $\|\cdot\|_{1;S}$ and $\|\cdot\|_2$: consider $v(t) = \alpha t v_0$, $t \in I$, where $v_0 \perp S$, for $\alpha \to \infty$. The resulting error estimate corresponds to [22, Thm. 2.1] by Dupont.

Corollary 3.5 (Quasi-optimality in the S-dependent norm)**.** *The Galerkin approximation* (3.3) *with $s = P_S w$ satisfies the following symmetric error estimate with respect to the S-dependent norm* (3.5)*:*

$$\|u - U_M\|_{1;S} \le 2\sqrt{2} \max\left\{\nu_a^{-1}, C_a^2, \nu_a^{-1} C_a^2\right\} \inf_{v \in M_1} \|u - v\|_{1;S} \tag{3.6}$$

Proof. In view of Proposition 3.3, it remains to observe

$$\tilde{C}_{X_1 \times M_2} = \sup_{v \in X_1} \sup_{\varphi \in M_2} \frac{b(v, \varphi)}{\|v\|_{1,S} \|\varphi\|_2} \le \sqrt{2} \max\{1, C_a\}$$

and to apply Corollary 1.3. □

3.2.1 Quasi-optimality and L^2-projection

The error notion in Corollary 3.5 depends on the discretization through the space S. This dependence may be troublesome, for example when comparing errors corresponding to different meshes. We now aim at a result without this disadvantage, by replacing the norm $\|\cdot\|_{1;S}$ by $\|\cdot\|_1$.

Proposition 3.6 (Discrete inf-sup constant)**.** *The discrete inf-sup constant c_M of the Galerkin method* (3.4) *is encased in terms of the V-stability of the W-projection:*

$$\frac{\tilde{c}_M}{\|P_S\|_{\mathcal{L}(V)}} \leq c_M \leq \frac{\tilde{C}_M}{\|P_S\|_{\mathcal{L}(V)}}.$$

Proof. Elementary inequalities yield

$$\inf_{v\in M_1} \frac{\|v\|_{1,S}}{\|v\|_1}\tilde{c}_M \leq c_M \leq \inf_{v\in M_1} \frac{\|v\|_{1,S}}{\|v\|_1}\tilde{C}_M.$$

We then prove

$$c_S = \sup_{v\in M_1} \frac{\|v\|_1}{\|v\|_{1,S}} = \left(\inf_{v\in M_1} \frac{\|v\|_{1,S}}{\|v\|_1}\right)^{-1}. \tag{3.7}$$

In fact, since $c_S \geq 1$, we easily derive $\|v\|_1 \leq c_S\|v\|_{1,S}$. On the other hand, choosing $v_n(t) = \phi\sin(\frac{2\pi nt}{T})$, with $\phi \in S$, we have

$$\frac{\|v_n\|_1^2}{\|v_n\|_{1,S}^2} = \frac{\frac{1}{2}T\Big(\|\phi\|_V^2 + \frac{4\pi^2n^2}{T^2}\|\phi\|_{V^*}^2\Big)}{\frac{1}{2}T\Big(\|\phi\|_V^2 + \frac{4\pi^2n^2}{T^2}\|\phi\|_{S^*}^2\Big)}.$$

Taking the supremum over $n \in \mathbb{N}$ we get

$$\sup_{v\in M_1} \frac{\|v\|_1}{\|v\|_{1,S}} \geq \sup_{\phi\in S} \frac{\|\phi\|_{V^*}}{\|\phi\|_{S^*}} = c_S.$$

Combining this with Proposition 3.2 gives the claimed bounds. □

Taking advantage that $\tilde{c}_M$ and $\tilde{C}_M$ can be bounded as in Proposition 3.3, we derive the following theorem.

Theorem 3.7 (Quasi-optimality)**.** *The Galerkin method* (3.4) *is quasi-optimal with*

$$\kappa_a^{-1}\|P_S\|_{\mathcal{L}(V)} \leq q_M \leq \kappa_a\|P_S\|_{\mathcal{L}(V)}$$

with $\kappa_a := 2\sqrt{2}\max\{\nu_a^{-1}, C_a^2, C_a^2\nu_a^{-1}\}$.

Proof. We insert the bounds of Propositions 2.2, 3.3 and 3.6 in Corollary 1.3

$$\frac{c_b}{c_M} \le q_M \le \frac{C_b}{c_M}. \qquad \square$$

The upper bound in Theorem 3.7, which corresponds to [13, Thm. 3.4] shows that the V-stability of the W-projection is sufficient for quasi-optimality. The lower bound reveals that this stability is not just a convenient assumption, but also necessary.

3.3 Galerkin approximation in $L^2(H^1)$

We consider the Galerkin approximation of the natural weak formulation, and derive, in a similar fashion as in Proposition 2.3, a discrete inf-sup constant. For the rest of this section, let $X_1 := L^2(V)$, $X_2 := \{\varphi \in H^1(V, V^*) : \varphi(T) = 0\}$, with norms

$$\|v\|_1^2 = \int_0^T \|v\|_V^2 \,, \qquad \|\varphi\|_2^2 = \int_0^T \|\varphi'\|_{V^*}^2 + \|\varphi\|_V^2 \,.$$

The bilinear form $b : X_1 \times X_2 \to \mathbb{R}$

$$b(v, \varphi) := \int_I - \langle \varphi', v \rangle + a(\cdot; v, \varphi),$$

and the linear functional $\ell \in X_2^*$ are as in Section 2.3.

Let $S \subset V$ be a finite-dimensional subspace. We set

$$M_1 := L^2(S) \subset X_1, \qquad M_2 := \{\varphi \in H^1(S) : \ \varphi(T) = 0\} \subset X_2. \tag{3.8}$$

The results in Section 3.2 suggest to endow M_2 with the S-dependent norm

$$\|\varphi\|_{2;S}^2 := \int_I \|\varphi\|_V^2 + \|\varphi'\|_{S^*}^2 \,,$$

while M_1 inherits the X_1-norm.

With these choices the Galerkin method (3.8) for the natural weak formulation is well-defined:

Proposition 3.8 (Natural bilinear form and Galerkin approximation)**.** *The bilinear form in* (2.12) *is continuous and satisfies the inf-sup condition on* $(M_1, \|\cdot\|_1) \times (M_2, \|\cdot\|_{2;S})$ *with*

$$C_M \le \sqrt{2} \max\{1, C_a\}, \qquad c_M \ge \frac{\nu_a}{\sqrt{2}} \min\{1, C_a^{-1}\}.$$

Proof. The proof mimics the one of Proposition 2.3, in the same way as the proof of Proposition 3.3 follows the one of Proposition 2.2. In particular the test function used to verify the bound for c_M is $v = \varphi - A_S^{-1}\varphi'$. $\square$

3.3.1 Quasi-optimality and L^2-projection

We recall from Chapter 1 that stability is equivalent to quasi-optimality with $q_M = \|R_M\|_{\mathcal{L}(X_1)}$. In [30, Thm. 3.4] Hackbusch proved some error estimates for the Galerkin approximation in space, assuming the L^2-projection to be H^1-stable. With a particular choice of the involved parameters, the result corresponds to the $L^2(H^1)$-stability of the Galerkin solution. For convenience we recall the proof in the case of interest.

Theorem 3.9. *The Galerkin method* (3.8) *is quasi-optimal, satisfying the following estimate*

$$\|u - U_M\|_1 \leq \|P_S\|_{\mathcal{L}(V)} \left(1 + \frac{C_a}{\nu_a}\right) \inf_{v \in X_1} \|u - v\|_1 .$$

Proof. In view of $q_M = \|R_M\|_{\mathcal{L}(X_1)}$, our aim is to find a stability estimate for $\|U_M\|_1$. We first exploit the triangle inequality to get

$$\|U_M\|_1 \leq \|U_M - P_S u\|_1 + \|P_S u\|_1 . \tag{3.9}$$

Secondly, we estimate $\|U_M - P_S u\|_1$ in terms of $\|u - P_S u\|_1$, which we finally bound thanks to the stability of P_S. Let us set $\eta := U_M - P_S u$. From the definition of P_S and U_M, we have, for every $\varphi \in M_2$,

$$\int_I - \langle \varphi', \eta \rangle = \int_I - \langle \varphi', U_M - u \rangle = \int_I \langle A(u - U_M), \varphi \rangle . \tag{3.10}$$

This implies that $\eta \in H^1(S)$. Integrating by parts in (3.10) gives, for every $\varphi \in C_0^\infty(S)$,

$$\int_I \langle \eta', \varphi \rangle + \langle A\eta, \varphi \rangle = \int_I \langle A(u - P_S u), \varphi \rangle , \tag{3.11}$$

which, by density, holds for every $\varphi \in L^2(S)$. Testing (3.10) with $\varphi = (T - t)\phi$, $\phi \in S$, integrating by parts and subtracting (3.11), gives $\eta(0) = 0$, and finally testing (3.11) with $\varphi = \eta$, we get

$$\frac{1}{2} \|\eta(T)\|_W^2 + \nu_a \|\eta\|_1^2 \leq C_a \|u - P_S u\|_1 \|\eta\|_1 ,$$

from which we deduce

$$\|\eta\|_1 \leq \frac{C_a}{\nu_a} \|u - P_S u\|_1 .$$

Inserting this result in (3.9) we get

$$\|U\|_1 \leq \frac{C_a}{\nu_a} \|I - P_S\|_{\mathcal{L}(V)} \|u\|_1 + \|P_S\|_{\mathcal{L}(V)} \|u\|_1 \leq \|P_S\|_{\mathcal{L}(V)} \left(1 + \frac{C_a}{\nu_a}\right) \|u\|_1 ,$$

which is equivalent to the thesis. □

The following theorem reveals that the stability of P_S is also necessary for the quasi-optimality of the Galerkin method (3.8).

Theorem 3.10. *The Galerkin method* (3.8) *is quasi-optimal with*

$$\frac{\nu_a \min\{1, C_a^{-2}\}}{2} \|P_S\|_{\mathcal{L}(V)} \le q_M \le \frac{2\max\{C_a^2, 1\}}{\nu_a} \|P_S\|_{\mathcal{L}(V)}. \tag{3.12}$$

Proof. We resort to Corollary 1.3, where we take $\|\cdot\|_\sharp = \|\cdot\|_{2;S}$. We start by bounding $C^\sharp_{X_1\times M_2}$ from below and from above. To derive a lower bound for $C^\sharp_{X_1\times M_2}$, we proceed in a similar way as in the proof of Proposition 2.3. Thanks also to (3.7) we obtain

$$\begin{aligned} C^\sharp_{X_1\times M_2} &= \sup_{\varphi\in M_2}\sup_{v\in X_1} \frac{b(v,\varphi)}{\|\varphi\|_{2;S}\|v\|_1} \ge \sup_{\varphi\in M_2} \frac{b(\varphi - A^{-1}\varphi', \varphi)}{\|\varphi\|_{2;S}\|\varphi - A^{-1}\varphi'\|_1} \\ &\ge \frac{\nu_a}{\sqrt{2}} \min\{1, C_a^{-1}\} \sup_{\varphi\in M_2} \frac{\|\varphi\|_2}{\|\varphi\|_{2;S}} = \frac{\nu_a}{\sqrt{2}} \min\{1, C_a^{-1}\} c_S. \end{aligned}$$

Concerning the upper bound we recall the bound for C_b in Proposition 2.3 and we get

$$\begin{aligned} C^\sharp_{X_1\times M_2} &= \sup_{\varphi\in M_2}\sup_{v\in X_1} \frac{b(v,\varphi)}{\|\varphi\|_{2;S}\|v\|_1} \le \left(\sup_{\varphi\in M_2} \frac{\|\varphi\|_2}{\|\varphi\|_{2;S}}\right)\left(\sup_{\varphi\in M_2}\sup_{v\in X_1} \frac{b(v,\varphi)}{\|\varphi\|_2\|v\|_1}\right) \\ &\le C_b c_S \le \sqrt{2}\max\{1, C_a\} c_S. \end{aligned}$$

To complete the proof we combine these bounds with Propositions 3.2 and 3.8. □

Remark 3.11 (Duality)**.** Theorem 3.10 can essentially be deduced from Theorem 3.7. To explain this, we consider the Galerkin solution (Φ_M^0, Φ_M^1) associated to the dual method of (3.4) for a problem with right-hand side $\ell \in H^1(V, V^*)^*$, defined by:

$$\forall v \in H^1(S) \quad \big\langle v(0), \Phi_M^0\big\rangle + \int_I \big\langle v', \Phi_M^1\big\rangle + \big\langle Av, \Phi_M^1\big\rangle = \ell(v).$$

The relationship between the Galerkin solution U_M of method (3.8) and (Φ_M^0, Φ_M^1) mirrors the one in Remark 2.4 between the corresponding exact solutions u and (ϕ_0, ϕ_1). If we set $\hat U_M := U_M(T - \cdot)$ and $\hat u := u(T - \cdot)$, we have, for every $v \in H^1(S)$ with $v(0) = 0$,

$$\int_I \langle v', \hat U_M - \hat u\rangle + \langle A^*(T-\cdot)v, \hat U_M - \hat u\rangle = 0 \quad \text{and} \tag{3.13a}$$

$$\int_I \big\langle v', \Phi_M^1 - \phi_1\big\rangle + \big\langle Av, \Phi_M^1 - \phi_1\big\rangle = 0. \tag{3.13b}$$

From Remark 1.5 and Theorem 3.7 we deduce that the quasi-optimality constant q_M^* of the dual method of (3.4) satisfies $q_M^* \approx \|P_S\|_{\mathcal{L}(V)}$. This link between q_M^* and $\|P_S\|_{\mathcal{L}(V)}$ arises from the second component $\phi_1 - \Phi_M^1$ of the error, the one related to the differential equation. In view of (3.13) also $\hat{u} - \hat{U}_M$ behaves in the same manner and therefore the quasi-optimality constant of method (3.8) also satisfies $q_M \approx \|P_S\|_{\mathcal{L}(V)}$. This also illustrates the link between the results by Hackbusch [30] and Chrysafinos and Hou [13].

Chapter 4

Discretization in time with the Backward Euler Method

In Chapter 3 we study one aspect of the approximation of parabolic problems, namely Galerkin approximations in space. In this chapter we analyse only the time discretization. We focus on one particular method, the implicit or backward Euler method. We adopt the viewpoint that the discrete trial and test spaces are subsets of the continuous ones, but the bilinear form that defines the discrete solution differs from the one that defines the exact solution. We assess quasi-optimality within the framework of the standard and natural weak formulations.

The chapter is organized as follows. Section 4.1 concerns the standard formulation. Under the assumption that the time partition is locally quasi-uniform, we prove that the error in a norm that mimics the $H^1(H^{-1}) \cap L^2(H^1)$-norm is equivalent to the sum of the best errors with piecewise constants for the exact solution and its time derivative. Section 4.2 concerns the natural formulation. We observe lack of stability in the $L^2(H^1)$-norm. Therefore a quasi-optimality result cannot hold, and we propose an abstract error estimate in the spirit of Section 1.3. Moreover we consider a modification of the right-hand side, which gives rise to a stable method. However, since the error does not vanish whenever the best error does, the method is still not quasi-optimal.

4.1 Standard formulation

Assume that $V \subset W \subset V^*$, a and A are as in Section 2.1, while, as in Section 2.2, $X_1 = H^1(V, V^*)$ and $X_2 = W \times L^2(V)$ with norms

$$\|v\|_1^2 = \|v(0)\|_W^2 + \int_0^T \|v'\|_{V^*}^2 + \|v\|_V^2\,, \qquad \|\varphi\|_2^2 = \|\varphi_0\|_W^2 + \int_0^T \|\varphi_1\|_V^2\,.$$

Moreover $u_0 \in W$, $f \in L^2(V)$, and the bilinear form $b : X_1 \times X_2 \to \mathbb{R}$ and $\ell \in X_2^*$ are given by

$$b(v,\varphi) = \langle v(0), \varphi_0\rangle + \int_0^T \langle v', \varphi_1\rangle + \langle Av, \varphi_1\rangle\,,$$
$$\ell(\varphi) = \langle u_0, \varphi_0\rangle + \int_0^T \langle f, \varphi_1\rangle\,.$$

Finally let $N \in \mathbb{N}$ and $\mathcal{P}$ be a partition

$$0 =: t_0 < t_1 < \ldots < t_N := T$$

of $I = (0,T)$ into N subintervals $I_n := (t_{n-1}, t_n]$, with size $\tau_n := |I_n| = t_n - t_{n-1}$, and let $\tau_{\mathcal{P}} := \max_n \tau_n$. The discrete solution provided by the backward Euler scheme is given by

$$\begin{aligned} &U_0 = u_0 \in W \\ &U_n \in V \ \text{ such that, for every } \phi \in V, \\ &\qquad \left\langle \frac{U_n - U_{n-1}}{\tau_n}, \phi \right\rangle + \langle A_n U_n, \phi\rangle = \langle F_n, \phi\rangle\,, \qquad n = 1, \ldots, N, \end{aligned} \tag{4.1}$$

where A_n and F_n are approximations of $A|_{I_n}$ and $f|_{I_n}$, respectively. In order to cast this scheme in the framework of Section 1.1, we consider the spaces

$$M_1 := \mathcal{S}^{1,0}(\mathcal{P}, V) := \{v \in C^0(V), v|_{I_n} \in \mathbb{P}^1(I_n, V),\ n = 1, \ldots, N\}, \tag{4.2a}$$
$$\widehat{M}_2 := \mathcal{S}^{0,-1}(\mathcal{P}, V) := \{\varphi \in L^2(V), \varphi|_{I_n} = \varphi_n \in V,\ n = 1, \ldots, N\}, \tag{4.2b}$$
$$M_2 := W \times \widehat{M}_2, \tag{4.2c}$$

where $\mathbb{P}^1(J, V)$ indicates the space of functions from the time interval J to V that are piecewise polynomials of degree at most one. With this choice $M_1 \subset X_1$ and $M_2 \subset X_2$. In order that it satisfies the inf-sup condition, we define the bilinear form $b_M : M_1 \times M_2 \to \mathbb{R}$ as follows

$$b_M(v,\varphi) := \langle v(0), \varphi_0\rangle + \sum_{n=1}^N \int_{I_n} \langle v', \varphi_n\rangle + \left\langle A\widehat{\Pi} v, \varphi_n \right\rangle, \tag{4.3}$$

where $\widehat{\Pi} : X_1 \to \widehat{M_2}$ is a suitable operator, such that, for every $v \in \mathbb{P}^1(I_n, V)$,

$$\widehat{\Pi} v|_{I_n} = v(t_n). \tag{4.4}$$

The presence of the operator $\widehat{\Pi}$ introduces a non-consistency, in that, for $\varphi \in M_2$, in general $b(u, \varphi) \neq b_M(u, \varphi)$. Given $\psi^n : I_n \to \mathbb{R}$

$$\psi^n(t) := \frac{6(t - t_{n-1})}{\tau_n^2} - \frac{2}{\tau_n},$$

the operator $\widehat{\Pi}$ can be defined by

$$\widehat{\Pi} v|_{I_n} := \Pi^n v := \int_{I_n} v\psi^n, \qquad v \in X_1. \tag{4.5}$$

Before examining the properties of $\widehat{\Pi}$, we recall the Poincaré and Friedrichs inequality in one dimension with optimal constants. Assume that $J = [a, b] \subset \mathbb{R}$ is an interval and that $v \in H^1(J)$. Then, see [8, p. 105],

$$\left\| v - \frac{1}{|J|} \int_J v \right\|_{L^2(J)} \leq \frac{b-a}{\pi} \left\| v' \right\|_{L^2(J)}. \tag{4.6}$$

Moreover, if $v(a) = v(b) = 0$,

$$\left\| v \right\|_{L^2(J)} \leq \frac{b-a}{\pi} \left\| v' \right\|_{L^2(J)}. \tag{4.7}$$

The optimal constant $1/\pi$ corresponds to the square root of the inverse of the first eigenvalue of the Laplacian over the interval J. If the function vanishes in only one of the endpoints, namely $v(a) = 0$ or $v(b) = 0$, then (4.7) is still valid with the price of a bigger constant. In fact, assume $v(a) = 0$ and consider the symmetric extension $\tilde{v}$ of v to the interval $\tilde{J} = [a, 2b - a]$. Since $\tilde{v} \in H^1(\tilde{J})$ and $v(a) = v(2b - a) = 0$, we can apply (4.7) and get

$$\left\| v \right\|_{L^2(J)} = \frac{1}{\sqrt{2}} \left\| \tilde{v} \right\|_{L^2(\tilde{J})} \leq \frac{2b - 2a}{\sqrt{2}\pi} \left\| \tilde{v}' \right\|_{L^2(\tilde{J})} = \frac{2}{\pi}(b - a) \left\| v' \right\|_{L^2(J)}. \tag{4.8}$$

To see that the constant $2/\pi$ is optimal consider the function $v \in H^1(J)$ such that $v(x) = \sin(\frac{\pi}{2(b-a)}(x - a))$, for $x \in J$. Similar considerations hold if $v(b) = 0$.

We collect the properties of $\widehat{\Pi}$ in the following remark.

Remark 4.1 (Properties of $\widehat{\Pi}$)**.** The operator $\widehat{\Pi}$ defined in (4.5) is linear and satisfies the following properties, for $n = 1, \ldots, N$:

(i) for every $v \in \mathbb{P}(I_n, V)$, $\Pi^n v = v(t_n)$,

(ii) for every $v \in L^2(I_n, V)$, $\|\Pi^n v\|_{L^2(I_n,V)} \leq 2 \|v\|_{L^2(I_n,V)}$, and the constant is optimal.

(iii) for every $v \in H^1(I_n, V^*)$,

$$\|\Pi^n v - v(t_n)\|_{V^*} \leq \frac{4}{\pi}\sqrt{\tau_n} \inf_{c \in V^*} \|v' - c\|_{L^2(I_n,V^*)} .$$

Proof. Property (i) follows from

$$\int_{I_n} \psi^n(t)\,\mathrm{d}t = \int_0^1 6s - 2\,\mathrm{d}s = 1 \quad \text{and}$$
$$\int_{I_n} \psi^n(t)\frac{t - t_{n-1}}{\tau_n}\,\mathrm{d}t = \int_0^1 (6s-2)s\,\mathrm{d}s = 1.$$

Property (ii) follows from

$$\|\Pi^n v\|_V \leq \int_{I_n} \|v\|_V\,|\psi^n| \leq \|v\|_{L^2(I_n,V)} \|\psi^n\|_{L^2(I_n)} = 2\tau_n^{-1/2} \|v\|_{L^2(I_n,V)} . \quad (4.9)$$

To see that the constant is optimal, take $\phi \in V$ and $v \in L^2(I_n, V)$ such that

$$\forall t \in I_n, \qquad v(t) = \frac{\phi}{\tau_n^{-3/8}}(t - t_{n-1})^{-3/8}.$$

We have that $\|\Pi^n v\|_{L^2(I_n,V)} = \sqrt{\tau_n}\,\|\phi\|_V$, and

$$\|v\|^2_{L^2(I_n,V)} = \int_{I_n} \|\phi\|_V^2 \frac{(t - t_{n-1})^{-3/4}}{\tau_n^{-3/4}}\,\mathrm{d}t = 4\tau_n \|\phi\|_V^2 .$$

Concerning Property (iii), we take $P \in \mathbb{P}^1(I_n, V^*)$ such that $P(t_{n-1}) = v(t_{n-1})$ and $P(t_n) = v(t_n)$. Exploiting Property (i), (4.9) and the Friedrichs inequality we get

$$\begin{aligned}\|\Pi^n v - v(t_n)\|_{V^*} = \|\Pi^n(v - P)\|_{V^*} &\leq 2\tau_n^{-1/2} \|v - P\|_{L^2(I_n,V^*)} \\ &\leq \frac{4}{\pi}\tau_n^{1/2} \|v' - P'\|_{L^2(I_n,V^*)} .\end{aligned}$$

Property (iii) follows from $P' = \frac{1}{\tau_n}\int_{\tau_n} v'$. □

The superscript n, as well as the hat over Π, reminds that Π^n applied to an affine function gives its value at the right endpoint of the interval I_n.

Taking $A_n : V \to V^*$ such that, for every $v \in V$,

$$A_n v := \frac{1}{\tau_n} \int_{I_n} A(t) v \, \mathrm{d}t, \tag{4.10}$$

and $F_n := \frac{1}{\tau_n} \int_{I_n} f(t) \, \mathrm{d}t$, the solution $U_M \in M_1$ such that, for every $\varphi \in M_2$,

$$b_M(U_M, \varphi) = \langle u_0, \varphi_0 \rangle + \sum_{n=1}^{N} \int_{I_n} \langle f, \varphi_n \rangle \,,$$

is such that $U_M(t_n)$ coincides with U_n given by (4.1).

We endow M_1 with

$$\|v\|_{1,\mathcal{P}}^2 := \|v(0)\|_W^2 + \sum_{n=1}^{N} \int_{I_n} \|v'\|_{V^*}^2 + \|\Pi^n v\|_V^2 \,,$$

while the space M_2 inherits the $\|\cdot\|_2$-norm. With these choices, the continuity and inf-sup constants of b_M are uniformly bounded.

Proposition 4.2. *The bilinear form* (4.3) *is continuous and fulfills the inf-sup condition* (1.9) *on* $(M_1, \|\cdot\|_{1,\mathcal{P}}) \times (M_2, \|\cdot\|_2)$ *with*

$$C_M \leq \sqrt{2} \max\{1, C_a\}, \qquad c_M \geq \frac{\min\{\nu_a, C_a^{-1}, \nu_a C_a^{-1}\}}{2}.$$

Moreover, it is also continuous on $(X_1, \|\cdot\|_{1,\mathcal{P}}) \times (M_2, \|\cdot\|_2)$ *and* $C_{X_1 \times M_2}$ *satisfies the same bound as* C_M*.*

Proof. The proof mimics the one of Proposition 2.2. Concerning the lower bound for c_M, we observe first that A_n satisfies the hypotheses of Lemma 2.1, so that

$$\varphi_0 = 2v(0), \qquad \varphi_n = \Pi^n v + A_n^{-*} v'|_{I_n}, \qquad n = 1, \dots, N$$

is a suitable test function. We remark that

$$\int_{I_n} \langle A \Pi^n v, \varphi_n \rangle = \int_{I_n} \langle A_n \Pi^n v, \varphi_n \rangle$$

and that

$$\begin{aligned}
&\|v(0)\|_W^2 + \sum_{n=1}^{N} \int_{I_n} 2\,\langle v', v(t_n)\rangle \\
&\qquad = \|v(0)\|_W^2 + \sum_{n=1}^{N} 2\,\|v(t_n)\|_W^2 - 2\,\langle v(t_{n-1}), v(t_n)\rangle \\
&\qquad = \sum_{n=1}^{N} \|v(t_n)\|_W^2 - 2\,\langle v(t_{n-1}), v(t_n)\rangle + \|v(t_{n-1})\|_W^2 \\
&\qquad = \sum_{n=1}^{N} \|v(t_n) - v(t_{n-1})\|_W^2 \geq 0.
\end{aligned}$$

To verify the non-degeneracy condition (1.9b), it suffices to take $v(t_n) = \varphi_n$, $n = 1, \ldots, N$, and $v(0) = 0$. We have that

$$\begin{aligned}
0 &= \|\varphi_1\|_W^2 + \sum_{n=2}^{N} \|\varphi_n\|_W^2 - \langle \varphi_{n-1}, \varphi_n\rangle + \sum_{n=1}^{N} \int_{I_n} \langle A\varphi_n, \varphi_n\rangle \\
&\geq \nu_a \sum_{n=1}^{N} \int_{I_n} \|\varphi_n\|_V^2
\end{aligned}$$

implies $\varphi_n = 0$, for $n = 1, \ldots, N$. By the density of V in W we also get that $\varphi_0 = 0$. □

We are in the situation described in Section 1.3. In order to derive an abstract error estimate, we need to bound the consistency error. We observe that, for every $\varphi \in M_2$,

$$\begin{aligned}
b_M(u,\varphi) - \ell(\varphi) = b_M(u,\varphi) - b(u,\varphi) &= \int_0^T \Big\langle A(\widehat{\Pi}u - u), \varphi \Big\rangle \\
&\leq C_a \left\| u - \widehat{\Pi}u \right\|_{L^2(V)} \|\varphi\|_{L^2(V)}\,.
\end{aligned}$$

We consider $Q : X_1 \to M_2$ defined by $Qv|_{I_n} := Q_n v := \frac{1}{\tau_n} \int_{I_n} v$. Because of stability of $\widehat{\Pi}$ and its invariance over piecewise constants, we have

$$\begin{aligned}
\left\| u - \widehat{\Pi}u \right\|_{L^2(V)}^2 &= \|u - Qu\|_{L^2(V)}^2 + \left\| Qu - \widehat{\Pi}u \right\|_{L^2(V)}^2 \\
&= \|u - Qu\|_{L^2(V)}^2 + \left\| \widehat{\Pi}(Qu - u) \right\|_{L^2(V)}^2 \leq 5\,\|u - Qu\|_{L^2(V)}^2\,.
\end{aligned} \tag{4.11}$$

This implies that the consistency error can be bounded in terms of a best error:

$$\sup_{\varphi\in M_2}\frac{b_M(u,\varphi)-\ell(\varphi)}{\|\varphi\|_2}\le\sqrt{5}C_a\,\|u-Qu\|_{L^2(V)}=\sqrt{5}C_a\inf_{z\in\widehat{M}_2}\|u-z\|_{L^2(V)}\,. \tag{4.12}$$

Theorem 4.3. *The Galerkin Method* (4.2) *with* b_M *as in* (4.3) *satisfies*

$$\begin{aligned}\Big(\|u'-U_M'\|^2_{L^2(V^*)}+\Big\|u-\widehat{\Pi}U_M\Big\|^2_{L^2(V)}+\|u(0)-U_M(0)\|^2_W\Big)^{1/2}\\ \le\sqrt{2}\kappa_a\inf_{v\in M_1}\Big(\|u'-v'\|^2_{L^2(V^*)}+\Big\|\widehat{\Pi}u-\widehat{\Pi}v\Big\|^2_{L^2(V)}+\|u(0)-v(0)\|^2_W\Big)^{1/2}\\ +(\sqrt{2}k_a+\sqrt{10})\inf_{z\in\widehat{M}_2}\|u-z\|_{L^2(V)}\,,\end{aligned} \tag{4.13}$$

where $\kappa_a=\sqrt{8}\max\{\nu_a^{-1},C_a^2,\nu_a^{-1}C_a^2\}$ *and* $k_a:=2\sqrt{5}C_a\max\{\nu_a^{-1},C_a,\nu_a^{-1}C_a\}$. *Moreover*

$$\|u'-U_M'\|^2_{L^2(V^*)}+\Big\|u-\widehat{\Pi}U_M\Big\|^2_{L^2(V)}+\|u(0)-U_M(0)\|^2_W\to 0\ \text{as}\ \tau_{\mathcal{P}}\to 0.$$

Proof. From Proposition 4.2, the results in Section 1.3 and (4.12) we get

$$\begin{aligned}\Big(\|u'-U_M'\|^2_{L^2(V^*)}+\Big\|\widehat{\Pi}u-\widehat{\Pi}U_M\Big\|^2_{L^2(V)}+\|u(0)-U_M(0)\|^2_W\Big)^{1/2}\\ \le\kappa_a\inf_{v\in M_1}\Big(\|u'-v'\|^2_{L^2(V^*)}+\Big\|\widehat{\Pi}u-\widehat{\Pi}v\Big\|^2_{L^2(V)}+\|u(0)-v(0)\|^2_W\Big)^{1/2}\\ +k_a\inf_{z\in\widehat{M}_2}\|u-z\|_{L^2(V)}\,,\end{aligned}$$

Combining with (4.11), we get (4.13). In order to prove convergence, we exploit the density of $C^\infty(V)$ in $H^1(V,V^*)$, see [47, Lemma 25.1]. Given $\epsilon>0$, we take $w\in C^\infty(V)$ such that

$$\|u'-w'\|^2_{L^2(V^*)}+\|u-w\|^2_{L^2(V)}\le\epsilon.$$

By the trace theorem we have $\|u(0)-w(0)\|^2_W\le C(T)\epsilon$. We use the triangle inequality in (4.13), and choose suitable functions $v\in M_1$ and $z\in\widehat{M}_2$. Concerning the infimum on M_1, we take $v\in M_1$ such that

$$v(t_n)=w(t_n),\qquad n=0,\dots,N.$$

Consider first $\|w' - v'\|^2_{L^2(I_n,V^*)}$. We can write

$$w'(t) - \frac{w(t_n) - w(t_{n-1})}{\tau_n} = \frac{1}{\tau_n}\int_{I_n} w'(t) - w'(s)\,\mathrm{d}s = \frac{1}{\tau_n}\int_{I_n}\int_s^t w''(\xi)\,\mathrm{d}\xi,$$

so that

$$\left\| w'(t) - \frac{w(t_n) - w(t_{n-1})}{\tau_n} \right\|_{V^*} \leq \tau_n^{1/2}\left(\int_{I_n} \|w''\|^2_{V^*}\right)^{1/2}.$$

Squaring, integrating over I_n, and summing over n gives

$$\|w' - v'\|^2_{L^2(V^*)} \leq \tau_{\mathcal{P}}^2 \|w''\|^2_{L^2(V^*)}. \tag{4.14}$$

Regarding the term $\left\|\widehat{\Pi}w - \widehat{\Pi}v\right\|^2_{L^2(I_n,V)}$, thanks to (4.4) we have

$$\begin{aligned}\widehat{\Pi}w - \widehat{\Pi}v &= \int_{I_n} w\psi^n - w(t_n) = \int_{I_n} (w(t) - w(t_n))\psi^n(t)\,\mathrm{d}t \\ &= \int_{I_n}\int_{t_n}^t w'(s)\,\mathrm{d}s\psi^n(t)\,\mathrm{d}t,\end{aligned}$$

so that

$$\left\|\widehat{\Pi}w - \widehat{\Pi}v\right\|_V \leq 2\tau_n^{1/2}\left(\|w'\|^2_V\right)^{1/2}.$$

Squaring, integrating over I_n and summing over n gives

$$\left\|\widehat{\Pi}w - \widehat{\Pi}v\right\|^2_{L^2(V)} \leq \tau_{\mathcal{P}}^2 \|w'\|^2_{L^2(V)}. \tag{4.15}$$

Concerning the infimum on $\widehat{M}_2$ we take $z \in \widehat{M}_2$ such that $z_n = w(t_n)$, $n = 1, \ldots, N$. Thanks to the Friedrichs inequality we get

$$\|w - z\|^2_{L^2(V)} \leq \tau_{\mathcal{P}} \|w'\|_{L^2(V)}. \tag{4.16}$$

Combining (4.14)–(4.16) it is possible to choose $\tau_{\mathcal{P}}$ such that

$$\|w' - v'\|^2_{L^2(V^*)} + \left\|\widehat{\Pi}w - \widehat{\Pi}v\right\|^2_{L^2(V)} + \|w - z\|^2_{L^2(V)} \leq \epsilon,$$

which completes the proof. □

In the right-hand side of (4.13) the distance between u and M_1 is measured in a norm which involves both the function itself and its time derivative. We simplify this coupled approximation problem and provide a bound in terms of the best errors for u' and u with piecewise constants. The price we pay is that the constant depends on

$$\mu_{\mathcal{P}} := \sup_n \frac{\tau_{n-1}}{\tau_n},$$

which measures how small the following step is compared to the previous one. In analogy to the definition of $\mathcal{S}^{0,-1}(\mathcal{P}, V)$, we set

$$\mathcal{S}^{0,-1}(\mathcal{P}, V^*) := \{\varphi \in L^2(V^*), \varphi|_{I_n} = \varphi_n \in V^*,\, n = 1, \dots, N\},$$

which appears in the following theorem.

Theorem 4.4. *The Galerkin method* (4.2) *with* b_M *as in* (4.3) *satisfies*

$$\begin{aligned}\|u' - U_M'\|^2_{L^2(V^*)} + \left\|u - \widehat{\Pi} U_M\right\|^2_{L^2(V)} + \|u(0) - U_M(0)\|^2_W \\ \le C_1 \inf_{v \in \mathcal{S}^{0,-1}(\mathcal{P}, V^*)} \|u' - v\|^2_{L^2(V^*)} + C_2 \inf_{z \in \mathcal{S}^{0,-1}(\mathcal{P}, V)} \|u - z\|^2_{L^2(V)},\end{aligned}$$

where $C_1 := 2\kappa_a^2(\pi^2 + 32 + 32\mu_{\mathcal{P}})/\pi^2$ *and* $C_2 := (\sqrt{2}k_a + \sqrt{10})^2$.

Proof. We prove the result for $u \in X_1 \cap C^0(V)$, the case $u \in X_1 \subset C^0(W)$ follows by density. We choose $v \in M_1$ such that

$$v(t_n) = \Pi^n u, \quad n = 1, \dots, N, \text{ and } v(0) = u(0).$$

This implies $\widehat{\Pi} u - \widehat{\Pi} v = 0$. Concerning $\|u' - v'\|_{L^2(I_n, V^*)}$ we observe, for $n = 2, \dots, N$,

$$u' - v' = u' - \frac{\Pi^n u - \Pi^{n-1} u}{\tau_n} = u' - \frac{1}{\tau_n}\int_{I_n} u' + \frac{u(t_n) - \Pi^n u - u(t_{n-1}) + \Pi^{n-1} u}{\tau_n}. \tag{4.17}$$

Thanks to Property (iii) of Remark 4.1, we get

$$\begin{aligned}\frac{1}{\tau_n}\left(\|\Pi^n u - u(t_n)\|^2_{V^*} + \left\|\Pi^{n-1} u - u(t_{n-1})\right\|^2_{V^*}\right) \\ \le \frac{16}{\pi^2} \inf_{c \in V^*} \|u' - c\|^2_{L^2(I_n, V^*)} + \frac{16}{\pi^2}\frac{\tau_{n-1}}{\tau_n} \inf_{c \in V^*} \|u' - c\|^2_{L^2(I_{n-1}, V^*)}.\end{aligned} \tag{4.18}$$

Combining (4.17)–(4.18) we get

$$\begin{aligned}\|u' - v'\|^2_{L^2(I_n, V^*)} \le \frac{\pi^2 + 32}{\pi^2} \inf_{c \in V^*} \|u' - c\|^2_{L^2(I_n, V^*)} \\ + \frac{32}{\pi^2}\mu_{\mathcal{P}} \inf_{c \in V^*} \|u' - c\|^2_{L^2(I_{n-1}, V^*)}.\end{aligned}$$

For $n = 1$, we have

$$u' - \frac{\Pi^1 u - u(0)}{\tau_1} = u' - \frac{1}{\tau_1}\int_{I_1} u' + \frac{u(t_1) - \Pi^1 u}{\tau_1},$$

so that

$$\|u' - v'\|^2_{L^2(I_1,V^*)} \le \frac{\pi^2 + 16}{\pi^2} \inf_{c\in V^*} \|u' - c\|^2_{L^2(I_1,V^*)}.$$

Summing over n we get

$$\|u' - v'\|_{L^2(V^*)} \le \frac{\pi^2 + 32 + 32\mu_{\mathcal{P}}}{\pi^2} \inf_{w\in\mathcal{S}^{0,-1}(\mathcal{P},V^*)} \|u' - w\|^2_{L^2(V^*)}.$$

The assertion follows combining this with Theorem 4.3. □

Assuming additional regularity

We provide an alternative to Theorem 4.4. We assume that the exact solution is more regular and we get a bound where the constant does not depend on $\mu_{\mathcal{P}}$.

We propose a different definition of $\widehat{\Pi}$. In place of (4.5), if $v \in C^0(V)$, we set

$$\widehat{\Pi} v|_{I_n} := v(t_n), \qquad n = 1, \dots, N, \tag{4.19}$$

so that (4.4) is trivially fulfilled. Proposition 4.7 remains valid, but the interpolation operator $\widehat{\Pi}$ is not stable in $L^2(V)$ any more, and therefore, for the consistency error, we only get that

$$\sup_{\varphi\in M_2} \frac{b_M(u,\varphi) - \ell(\varphi)}{\|\varphi\|_2} \le C_a \left\| u - \widehat{\Pi} u \right\|_{L^2(V)}. \tag{4.20}$$

Mimicking the reasoning in the proof of Theorem 4.3 we get the following theorem.

Theorem 4.5. *Assume $u \in C^0(V)$. The Galerkin Method (4.2) with b_M as in (4.3), and $\widehat{\Pi}$ as in (4.19) satisfies*

$$\begin{aligned}
&\left(\|u' - U_M'\|^2_{L^2(V^*)} + \left\| u - \widehat{\Pi} U_M \right\|^2_{L^2(V)} + \|u(0) - U_M(0)\|^2_W \right)^{1/2} \\
&\quad \le \sqrt{2}\kappa_a \inf_{v\in M_1} \left(\|u' - v'\|^2_{L^2(V^*)} + \left\| \widehat{\Pi} u - \widehat{\Pi} v \right\|^2_{L^2(V)} + \|u(0) - v(0)\|^2_W \right)^{1/2} \\
&\qquad + \left(\frac{k_a}{\sqrt{5}} + 1 \right) \left\| u - \widehat{\Pi} u \right\|_{L^2(V)},
\end{aligned}$$

where $\kappa_a = \sqrt{8}\max\{\nu_a^{-1}, C_a^2, \nu_a^{-1}C_a^2\}$ and $k_a = 2\sqrt{5}C_a \max\{\nu_a^{-1}, C_a, \nu_a^{-1}C_a\}$.

4.2 Natural formulation

In this section we look at the discretization of the natural formulation and analyse two different methods. To motivate them, observe first that, if $v(0) = 0$ and we set $\varphi_{N+1} = 0$, we can rewrite b_M in (4.3) as

$$\begin{aligned} b_M(v,\varphi) &= \sum_{n=1}^{N} \int_{I_n} \left\langle \frac{v(t_n) - v(t_{n-1})}{\tau_n}, \varphi_n \right\rangle + \left\langle A\widehat{\Pi} v, \varphi_n \right\rangle \\ &= \sum_{n=1}^{N} \langle v(t_n), \varphi_n \rangle - \langle v(t_{n-1}), \varphi_n \rangle + \int_{I_n} \langle Av(t_n), \varphi_n \rangle \\ &= \sum_{n=1}^{N} \langle v(t_n), \varphi_n \rangle - \langle v(t_n), \varphi_{n+1} \rangle + \int_{I_n} \langle Av(t_n), \varphi_n \rangle \\ &= \sum_{n=1}^{N} \int_{I_n} - \left\langle v(t_n), \frac{\varphi_{n+1} - \varphi_n}{\tau_n} \right\rangle + \langle Av(t_n), \varphi_n \rangle . \end{aligned}$$

We can thus interpret v as a piecewise constant function, whose value in each interval is given by $v(t_n)$. Correspondingly, φ can be seen as a continuous piecewise polynomial of first degree, with values in the left-endpoint of the intervals equal to φ_n. Assuming $\breve{\Pi}$ is an operator such that $\breve{\Pi}\varphi|_{I_n} = \varphi(t_{n-1})$ for an affine function in I_n, we can write

$$b_M(v,\varphi) = \sum_{n=1}^{N} \int_{I_n} - \langle v_n, \varphi' \rangle + \left\langle Av_n, \breve{\Pi}\varphi \right\rangle .$$

In this view, the right-hand side ℓ becomes

$$\ell(\varphi) = \langle u_0, \varphi(0) \rangle + \sum_{n=1}^{N} \int_{I_n} \left\langle f, \breve{\Pi}\varphi \right\rangle . \tag{4.21}$$

However, it is also possible to consider

$$\tilde{\ell}(\varphi) = \langle u_0, \varphi(0) \rangle + \sum_{n=1}^{N} \int_{I_n} \langle f, \varphi \rangle .$$

In the following, we analyse these two points of view, highlighting their advantages and drawbacks. For the rest of this section, assume that $V \subset W \subset V^*$, a and A are as in Section 2.1, while, as in Section 2.3, $X_1 = L^2(V)$, $X_2 = \{\varphi \in H^1(V, V^*), \varphi(T) = 0\}$ with norms

$$\|v\|_1^2 = \int_0^T \|v\|_V^2 , \qquad \|\varphi\|_2^2 = \int_0^T \|\varphi'\|_{V^*}^2 + \|\varphi\|_V^2 ,$$

and the bilinear form $b : X_1 \times X_2 \to \mathbb{R}$ is given by

$$b(v, \varphi) = \int_0^T - \langle \varphi', v \rangle + \langle Av, \varphi \rangle .$$

4.2.1 Backward Euler method

We first notice that $\ell \in X_2^*$ could be of the following form

$$\ell(\varphi) = \langle u_0, \varphi(0) \rangle + \int_0^T \langle \varphi', f_1 \rangle + \langle f_2, \varphi \rangle ,$$

where $u_0 \in W$, $f_1 \in L^2(V)$ and $f_2 \in L^2(V^*)$. Its representation in terms of u_0, f_1 and f_2 is not unique, since we can add, for example,

$$0 = \langle f(0), \varphi(0) \rangle + \int_0^T \langle f', \varphi \rangle + \langle \varphi', f \rangle , \qquad f \in H^1(V).$$

Therefore, given $\ell \in X_2^*$ it is not always possible to define univocally a right-hand side for the discrete problem mimicking (4.21), and thus, given an exact solution in X_1, the discrete solution is not well-defined.

To avoid this problem, we consider $\ell \in X_2^*$ of the following form

$$\ell(\varphi) := \langle g_0, \varphi(0) \rangle + \sum_{j=1}^{\jmath-1} \left\langle g_j, \varphi(\tilde{t}_j) \right\rangle + \int_0^T \langle f, \varphi \rangle , \tag{4.22}$$

where $f \in L^2(V^*)$, $(g_j)_{j=0}^{\jmath} \subset W$ and $0 =: \tilde{t}_0 < \tilde{t}_1 < \ldots < \tilde{t}_\jmath =: T$. Note that, thanks to the embedding $X_2 \subset C^0(W)$, there holds $\ell \in X_2^*$. With this choice, the solution u of

$$b(u, \varphi) = \ell(\varphi), \qquad \forall \varphi \in X_2, \tag{4.23}$$

belongs to $H^1((\tilde{t}_{j-1}, \tilde{t}_j); V, V^*) \subset C^0([\tilde{t}_{j-1}, \tilde{t}_j]; W)$, for $j = 1, \ldots, \jmath$. Integrating by parts piecewise shows that u also satisfies, for every $(\varphi, \phi) \in L^2(V) \times W^\jmath$,

$$\begin{aligned} \langle u(0), \phi_0 \rangle + \sum_{j=1}^{\jmath-1} \left\langle u(\tilde{t}_j^+) - u(\tilde{t}_j^-), \phi_j \right\rangle + \sum_{j=1}^{\jmath} \int_{\tilde{t}_{j-1}}^{\tilde{t}_j} \langle u' + Au, \varphi \rangle \\ = \sum_{j=0}^{\jmath-1} \langle g_j, \phi_j \rangle + \int_0^T \langle f, \varphi \rangle , \end{aligned} \tag{4.24}$$

where $u(t^-) := \lim_{s \nearrow t} u(s)$ and $u(t^+) := \lim_{s \searrow t} u(s)$ denote respectively the left and right limit of the function u to the time t. Equation (4.24) implies in particular that $u(\tilde{t}_j^+) - u(\tilde{t}_j^-) = g_j$, that is, u is allowed to jump at the points $\tilde{t}_j$, $j = 1, \dots, \mathcal{J} - 1$.

In order to discretize (4.23), as in Section 4.1 we consider a partition $\mathcal{P}$ of $(0, T)$, $0 =: t_0 < t_1 < \ldots < t_N := T$. We require that $\mathcal{P}$ is subordinate to $(\tilde{t}_j)_{j=1}^{\mathcal{J}-1}$, that is, for every $j = 1, \dots, \mathcal{J}-1$ there exists $n \in \{1, \dots, N-1\}$ such that $t_n = \tilde{t}_j$. Every subinterval $I_n := [t_{n-1}, t_n)$ is left-closed and right-open, and its size is denoted by $\tau_n := |I_n|$. Moreover let $\tau_{\mathcal{P}} := \max_n \tau_n$ be the biggest time-step. We consider the spaces

$$M_1 := \{v \in L^2(V), v|_{I_n} = v_n \in V,\, n = 1, \dots, N\}, \tag{4.25a}$$
$$M_2 := \{\varphi \in C^0(V), \varphi|_{I_n} \in \mathbb{P}^1(I_n, V),\, n = 1, \dots, N, \varphi(T) = 0\}. \tag{4.25b}$$

We notice that $M_1 \subset X_1$, and $M_2 \subset X_2$. We define the bilinear form $b_M : X_1 \times X_2 \to \mathbb{R}$ as follows

$$b_M(v, \varphi) := \sum_{n=1}^{N} \int_{I_n} -\langle \varphi', v_n \rangle + \left\langle A v_n, \breve{\Pi} \varphi \right\rangle, \tag{4.26}$$

where the operator $\breve{\Pi} : X_2 \to \widehat{M_2}$ is defined by

$$\breve{\Pi} v|_{I_n} := \Pi_n v := \int_{I_n} v \psi_n, \tag{4.27}$$

with

$$\psi_n(t) := \frac{-6(t - t_{n-1})}{\tau_n^2} + \frac{4}{\tau_n}.$$

Remark 4.6 (Properties of $\breve{\Pi}$)**.** The operator $\breve{\Pi}$ defined in (4.27) satisfies the following properties, for $n = 1, \dots, N$:

(i) for every $v \in \mathbb{P}(I_n, V)$, $\Pi_n v = v(t_{n-1})$,

(ii) for every $v \in L^2(I_n, V)$, $\|\Pi_n v\|_{L^2(I_n, V)} \leq 2 \|v\|_{L^2(I_n, V)}$,

Proof. Property (i) follows from

$$\int_{I_n} \psi_n(t) \, \mathrm{d}t = \int_0^1 -6s + 4 \, \mathrm{d}s = 1 \quad \text{and}$$
$$\int_{I_n} \psi^n(t) \frac{t - t_{n-1}}{\tau_n} \, \mathrm{d}t = \int_0^1 (-6s + 4) s \, \mathrm{d}s = 0.$$

Property (ii) follows from

$$\|\Pi_n v\|_V \le \int_{I_n} \|v\|_V \, |\psi_n| \le \|v\|_{L^2(I_n,V)} \, \|\psi_n\|_{L^2(I_n)} = 2\tau_n^{-1/2} \, \|v\|_{L^2(I_n,V)} \,. \tag{4.28}$$

□

The subscript n, as well as the overturned hat over Π, reminds that Π_n applied to an affine function gives its value in the left endpoint of the interval I_n. The right-hand side ℓ is replaced by

$$\ell_M(\varphi) := \langle g_0, \varphi(t_0) \rangle + \sum_{j=1}^{\jmath-1} \left\langle g_j, \varphi(\tilde{t}_j) \right\rangle + \int_0^T \left\langle f, \check{\Pi}\varphi \right\rangle . \tag{4.29}$$

We endow M_2 with

$$\|\varphi\|_{2,\mathcal{P}}^2 := \sum_{n=1}^{N} \int_{I_n} \|\varphi'\|_{V^*}^2 + \|\Pi_n \varphi\|_V^2 \,,$$

while the space M_1 inherits the $\|\cdot\|_1$-norm. We observe that, for every $\varphi \in M_2$,

$$\begin{aligned}
\left\| \varphi - \check{\Pi}\varphi \right\|_{L^2(V)}^2 &= \sum_{n=1}^{N} \int_{I_n} \left\| \frac{\varphi(t_n) - \varphi(t_{n-1})}{\tau_n} (t - t_{n-1}) \right\|_V^2 \mathrm{d}t \\
&= \sum_{n=1}^{N} \frac{\tau_n}{3} \, \|\varphi(t_n) - \varphi(t_{n-1})\|_V^2 \\
&\le \frac{2}{3} \sum_{n=1}^{N-1} \frac{\tau_n}{\tau_{n+1}} \int_{I_{n+1}} \|\varphi(t_n)\|_V^2 + \frac{2}{3} \sum_{n=1}^{N} \int_{I_n} \|\varphi(t_{n-1})\|_V^2 \\
&\le \frac{2}{3} (\mu_{\mathcal{P}} + 1) \left\| \check{\Pi}\varphi \right\|_{L^2(V)}^2 ,
\end{aligned} \tag{4.30}$$

so that

$$\|\varphi\|_2 \le \sqrt{\frac{4}{3}(2\mu_{\mathcal{P}} + 5)} \, \|\varphi\|_{2,\mathcal{P}} \,. \tag{4.31}$$

With an example, we show that we cannot avoid the dependence on $\mu_{\mathcal{P}}$ in (4.31). In fact, assume $V = H_0^1(\Omega)$, and $W = L^2(\Omega)$, with $\Omega \subset \mathbb{R}^d$. Fix $\bar{n} \in \{1, \dots, N-1\}$, and consider the function $\varphi \in M_2$ such that

$$\varphi(t_n) = \begin{cases} 0 & \text{if } n \neq \bar{n}, \\ \phi_m & \text{if } n = \bar{n}, \end{cases}$$

where ϕ_m is the m-th eigenfunction of the Laplacian, with corresponding eigenvalue λ_m. We have

$$\|\varphi'\|^2_{L^2(V^*)} = \int_{I_{\bar{n}}} \frac{\|\phi_m\|^2_{V^*}}{\tau_{\bar{n}}} + \int_{I_{\bar{n}+1}} \frac{\|\phi_m\|^2_{V^*}}{\tau_{\bar{n}+1}} = \lambda_m^{-1}\left(\tau_{\bar{n}}^{-1} + \tau_{\bar{n}+1}^{-1}\right)$$

and

$$\begin{aligned}\|\varphi\|^2_{L^2(V)} &= \int_{I_{\bar{n}}} \|\phi_m\|^2_V \frac{(t-t_{\bar{n}-1})^2}{\tau_{\bar{n}}^2}\,\mathrm{d}t + \int_{I_{\bar{n}+1}} \|\phi_m\|^2_V \frac{(t-t_{\bar{n}+1})^2}{\tau_{\bar{n}+1}^2}\,\mathrm{d}t \\ &= \frac{\lambda_m}{3}(\tau_{\bar{n}} + \tau_{\bar{n}+1}),\end{aligned}$$

while

$$\left\|\breve{\Pi}\varphi\right\|^2_{L^2(V)} = \int_{I_{\bar{n}+1}} \|\phi_m\|^2_V = \tau_{\bar{n}+1}\lambda_m.$$

Therefore we obtain

$$\begin{aligned}\frac{\|\varphi\|_2^2}{\|\varphi\|^2_{2,\mathcal{P}}} &= \frac{\tau_{\bar{n}}^2\tau_{\bar{n}+1}\lambda_m^2 + \tau_{\bar{n}}\tau_{\bar{n}+1}^2\lambda_m^2 + 3\tau_{\bar{n}} + 3\tau_{\bar{n}+1}}{3(\tau_{\bar{n}}\tau_{\bar{n}+1}^2\lambda_m^2 + \tau_{\bar{n}} + \tau_{\bar{n}+1})} \\ &\geq \frac{\tau_{\bar{n}}^2\tau_{\bar{n}+1}\lambda_m^2}{3(\tau_{\bar{n}}\tau_{\bar{n}+1}^2\lambda_m^2 + \tau_{\bar{n}} + \tau_{\bar{n}+1})} \overset{m\to\infty}{\longrightarrow} \frac{\tau_{\bar{n}}}{3\tau_{\bar{n}+1}}.\end{aligned}$$

We prove that the continuity and inf-sup constants of b_M are uniformly bounded.

Proposition 4.7. *The bilinear form* (4.26) *is continuous and fulfills the inf-sup condition* (1.9) *on* $(M_1, \|\cdot\|_1) \times (M_2, \|\cdot\|_{2,\mathcal{P}})$ *with*

$$C_M \leq \sqrt{2}\max\{1, C_a\} \qquad\qquad c_M \geq \frac{\nu_a}{\sqrt{2}}\min\{1, C_a^{-1}\}.$$

Moreover, it is also continuous on $(X_1, \|\cdot\|_{1,\mathcal{P}}) \times (M_2, \|\cdot\|_2)$ *and* $C_{X_1\times M_2}$ *satisfies the same bound as* C_M*.*

Proof. The proof mimics the one of Proposition 2.3. Concerning the lower bound for c_M we take as test function

$$v_n = \Pi_n\varphi - A_n^{-1}\varphi|'_{I_n}, \qquad n = 1, \dots, N$$

where A_n is defined in (4.10) and observe

$$\begin{aligned}\sum_{n=1}^{N}\int_{I_n} -2\left\langle\varphi',\varphi(t_{n-1})\right\rangle &= \sum_{n=1}^{N} 2\left\|\varphi(t_{n-1})\right\|_W^2 - 2\left\langle\varphi(t_n),\varphi(t_{n-1})\right\rangle \\ &= \left\|\varphi(t_0)\right\|_W^2 + \sum_{n=1}^{N}\left\|\varphi(t_n)\right\|_W^2 - 2\left\langle\varphi(t_n),\varphi(t_{n-1})\right\rangle + \left\|\varphi(t_{n-1})\right\|_W^2 \\ &\geq \sum_{n=1}^{N}\left\|\varphi(t_n)-\varphi(t_{n-1})\right\|_W^2 \geq 0. \end{aligned} \tag{4.32}$$

The non-degeneracy condition (1.9b) is proved taking $\varphi(t_n) = v_{n+1}$. □

From (4.32) we also get that

$$\sum_{n=1}^{N}\left\|\varphi(t_n)-\varphi(t_{n-1})\right\|_W^2 \leq \sum_{n=1}^{N}\int_{I_n} -2\left\langle\varphi',\varphi(t_{n-1})\right\rangle \leq \left\|\varphi\right\|_{2,\mathcal{P}}. \tag{4.33}$$

Remark 4.8. Assume $N = \jmath$, so that $\mathcal{P}$ is given by $0 = \tilde{t}_0 < \tilde{t}_1 < \ldots < \tilde{t}_\jmath = T$. If $u \in M_1$ then $f = Au$ and $b_M(u,\varphi) = \ell(\varphi)$ for every $\varphi \in M_2$. Therefore $U_M = u$.

We are in the situation described on Section 1.3. In order to derive an abstract error estimate, we write the consistency error as follows:

$$\begin{aligned} b_M(u,\varphi) - \ell_M(\varphi) &= b_M(u,\varphi) - b(u,\varphi) + \ell(\varphi) - \ell_M(\varphi) \\ &= \int_0^T \left\langle Au - f, \breve{\Pi}\varphi - \varphi \right\rangle \end{aligned}$$

Combining Proposition 4.7 with the results in Section 1.3, we obtain the following.

Proposition 4.9. *The Galerkin method* (4.25) *with* b_M *as in* (4.26) *and* ℓ_M *as in* (4.29) *satisfies*

$$\begin{aligned}\left\|u-U_M\right\|_1 \leq{}& 2\nu_a^{-1}\max\{1,C_a^2\}\inf_{v\in M_1}\left\|u-v\right\|_1 \\ &+ 2\nu_a^{-1}\max\{1,C_a\}\sup_{\varphi\in M_2}\frac{\displaystyle\int_0^T\left\langle Au-f,\breve{\Pi}\varphi-\varphi\right\rangle}{\left\|\varphi\right\|_{2,\mathcal{P}}}. \end{aligned} \tag{4.34}$$

Moreover, if $\mu_{\mathcal{P}} = \sup_n \frac{\tau_{n-1}}{\tau_n} < \infty$, *then*

$$\left\|u-U_M\right\|_1 \to 0 \qquad \textit{as} \qquad \tau_{\mathcal{P}} \to 0.$$

Proof. Given $\epsilon > 0$, the best error in $L^2(V)$ can be bounded in terms of ϵ as in the proof of Theorem 4.3. Concerning the consistency error, take $w \in C^0(W)$ such that $\|Au - f - w\|_{L^2(V^*)} < \epsilon$. This is possible thanks to the density of $C^0(W)$ in $L^2(V^*)$. We add and subtract w, so that

$$\sup_{\varphi \in M_2} \frac{\displaystyle\int_I \Big\langle Au - f, \varphi - \check{\Pi}\varphi \Big\rangle}{\|\varphi\|_{2,\mathcal{P}}} \leq \sup_{\varphi \in M_2} \frac{\displaystyle\int_I \Big\langle Au - f - w, \varphi - \check{\Pi}\varphi \Big\rangle}{\|\varphi\|_{2,\mathcal{P}}} + \sup_{\varphi \in M_2} \frac{\displaystyle\int_I \Big\langle w, \varphi - \check{\Pi}\varphi \Big\rangle}{\|\varphi\|_{2,\mathcal{P}}}.$$

Concerning the first term on the right-hand side, we recall (4.30) and get

$$\int_I \Big\langle Au - f - w, \varphi - \check{\Pi}\varphi \Big\rangle \leq \sqrt{\mu_{\mathcal{P}} + 1}\, \|Au - f - w\|_{L^2(V^*)}\, \|\varphi\|_{2,\mathcal{P}}\,.$$

Regarding the second term, we recall (4.33) and get

$$\Big\|\varphi - \check{\Pi}\varphi\Big\|^2_{L^2(W)} = \sum_{n=1}^{N} \frac{\tau_n}{3}\, \|\varphi(t_n) - \varphi(t_{n-1})\|^2_W \leq \frac{\tau_{\mathcal{P}}}{3}\, \|\varphi\|^2_{2,\mathcal{P}}\,,$$

so that

$$\int_I \Big\langle w, \varphi - \check{\Pi}\varphi \Big\rangle \leq \frac{\sqrt{3\tau_{\mathcal{P}}}}{3}\, \|w\|_{L^2(W)}\, \|\varphi\|_{2,\mathcal{P}}\,.$$

It is thus possible to choose $\tau_{\mathcal{P}}$ such that

$$\sup_{\varphi \in M_2} \frac{\displaystyle\int_I \Big\langle w, \varphi - \check{\Pi}\varphi \Big\rangle}{\|\varphi\|_{2,\mathcal{P}}} \leq \epsilon,$$

so that

$$\sup_{\varphi \in M_2} \frac{\displaystyle\int_I \Big\langle Au - f, \varphi - \check{\Pi}\varphi \Big\rangle}{\|\varphi\|_{2,\mathcal{P}}} \leq (1 + \sqrt{\mu_{\mathcal{P}} + 1})\epsilon. \qquad \square$$

The second term in the right-hand side of (4.34), due to the consistency error, is not expressed in terms of a best error. Proposition 4.10 below reveals that the method is not stable in $L^2(V)$, so it is not possible to bound the consistency error in terms of the best error in $L^2(V)$. However estimate (4.34) satisfies the property that the right-hand side is equivalent to the error

$\|u - U_M\|_1$ and can be used as a starting point for deriving error estimates. Moreover, recalling that $Au - f = -u'$ on every I_n and (4.30) we get

$$\|U_M\|_1 \lesssim \|u\|_1 + \Big(\sum_{n=1}^{N} \int_{I_n} \|u'\|_{V^*}^2 \Big)^{1/2},$$

where the hidden constant depends also on $\mu_{\mathcal{P}}$.

Proposition 4.10. *The Galerkin method* (4.25) *with* b_M *as in* (4.26) *and* ℓ_M *as in* (4.29) *is not stable in* $L^2(V)$.

Proof. To prove that the method is not stable, an example is sufficient. Assume $A = -\Delta$, the function ϕ_1 is the first eigenfunction of the Laplacian with eigenvalue λ_1, and $\tau := \tau_1 = \ldots = \tau_N$, with $\tau \leq \lambda_1$. The W-norm, V-norm and V^*-norm of ϕ_1 are given by

$$\|\phi_1\|_W^2 = 1, \qquad \|\phi_1\|_V^2 = \lambda_1, \qquad \|\phi_1\|_{V^*}^2 = \frac{1}{\lambda_1}.$$

We consider the function

$$u(t) = \begin{cases} 0 & 0 \leq t \leq t_{N-1} \\ (t - t_{N-1})^\rho \phi_1 & t_{N-1} < t \leq T \end{cases},$$

where $\rho > 0$. Its time derivative is given by

$$u'(t) = \begin{cases} 0 & 0 \leq t < t_{N-1} \\ \rho (t - t_{N-1})^{\rho-1} \phi_1 & t_{N-1} < t \leq T \end{cases}.$$

We prove that

$$\frac{\|u - U_M\|_1}{\|u\|_1} \to \infty \qquad \text{as} \qquad \rho \to \infty$$

by bounding the consistency error from below. To this end, we choose $\varphi_u \in M_2$ such that $\varphi_u(t_{N-1}) = -\phi_1$, $\varphi_u(t_0) = \ldots = \varphi_u(t_{N-2}) = 0$. Moreover, since $\varphi_u - \breve{\Pi}\varphi_u \in L_N := \operatorname{span}(t - t_{N-1})\xi$, $\xi \in V$, we project $u'|_{I_N}$ on L_N that is, we take $Q_N u'$ such that, for every $\xi \in V$,

$$\int_{I_N} \langle u'(t), (t - t_{N-1})\xi \rangle \, \mathrm{d}t = \int_{I_N} \langle Q_N u'(t), (t - t_{N-1})\xi \rangle \, \mathrm{d}t.$$

A calculation gives

$$Q_N u'(t) = \frac{3\rho}{\rho + 1} \tau_N^{\rho-2} (t - t_{N-1}) \phi_1, \qquad \forall t \in I_N.$$

We compute

$$\begin{aligned}\int_{I_N}\left\langle u',\varphi_u-\breve{\Pi}\varphi_u\right\rangle &= \int_{I_N}\left\langle Q_N u',-\frac{\varphi_u(t_{N-1})}{\tau}(t-t_{N-1})\right\rangle \,\mathrm{d}t\\ &= \langle\phi_1,\phi_1\rangle\int_{I_N}\frac{3\rho}{\rho+1}\tau^{\rho-3}(t-t_{N-1})^2\\ &= \frac{\rho}{\rho+1}\tau^{\rho},\end{aligned}$$

and

$$\begin{aligned}\|\varphi_u\|_{2,\mathcal{P}}^2 &= \int_{I_{N-1}}\left\|\frac{\phi_1}{\tau}\right\|_{V^*}^2\,\mathrm{d}t+\int_{I_N}\left\|\frac{\phi_1}{\tau}\right\|_{V^*}^2+\|\phi_1\|_V^2\,\mathrm{d}t\\ &= \frac{2+\lambda_1^{-2}\tau^2}{\tau\lambda_1}\\ &\leq \frac{3}{\tau\lambda_1},\end{aligned}$$

while the $L^2(V)$-norm of u is given by

$$\begin{aligned}\|u\|_{L^2(V)}^2 &= \int_{I_N}\|\phi_1\|_V^2\,(t-t_{N-2})^{2\rho}\,\mathrm{d}t\\ &= \frac{\tau^{2\rho+1}}{2\rho+1}\,\|\phi_1\|_V^2=\lambda_1\frac{\tau^{2\rho+1}}{2\rho+1}.\end{aligned}$$

We can bound the consistency error from below as follows

$$\begin{aligned}\sup_{\varphi\in M_2}\frac{\sum_n\int_{I_n}\left\langle f-Au,\varphi-\breve{\Pi}\varphi\right\rangle}{\|\varphi\|_{2,\mathcal{P}}} &\geq \frac{\int_{I_N}\left\langle u',\varphi_u-\breve{\Pi}\varphi_u\right\rangle}{\|\varphi_u\|_{2,\mathcal{P}}}\\ &= \frac{\rho}{\rho+1}\frac{\tau^{\rho+\frac{1}{2}}\sqrt{\lambda_1}}{\sqrt{3}},\end{aligned}$$

so that

$$\frac{\|u-U_M\|_1}{\|u\|_1}\geq\frac{1}{3\sqrt{2}}\sup_{\varphi\in M_2}\frac{b_M(u,\varphi)-\ell_M(\varphi)}{\|\varphi\|_{2,\mathcal{P}}}\geq\frac{\rho(2\rho+1)^{1/2}}{3(\rho+1)\sqrt{6}}\overset{\rho\to+\infty}{\longrightarrow}+\infty.\quad\square$$

4.2.2 A variant

The second method still involves the spaces M_1 and M_2 of (4.25), the bilinear form of (4.26) with $\breve{\Pi}$ as in (4.27), but it does not modify the right-hand

side of the continuous problem. More precisely, given $\ell \in X_2^*$, the discrete problem reads

$$\text{find } U_M \in M_1 \text{ such that, } \forall \varphi \in M_2,\ b_M(U_M, \varphi) = \ell(\varphi).$$

Every $\ell \in X_2^*$ also belongs to M_2^*, and (4.31) guarantees that $\|\ell\|_{M_2^*} \leq \sqrt{4(2\mu_{\mathcal{P}}+5)/3}\,\|\ell\|_{X_2^*}$. Proposition 4.7 ensures that the discrete problems are thus uniformly well-posed for every $\ell \in X_2^*$, and not only for those of the form (4.29) considered in Section 4.2.1.

Proposition 4.11. *Assume that the exact solution u belongs to M_1. If $b_M(u, \varphi) = \ell(\varphi)$ for every $\varphi \in M_2$ then $u = 0$.*

Proof. If $b_M(u, \varphi) = \ell(\varphi)$, then $b_M(u, \varphi) = b(u, \varphi)$ for every $\varphi \in M_2$. Therefore

$$\int_0^T \left\langle Au, \varphi - \breve{\Pi}\varphi \right\rangle = 0 \qquad \forall \varphi \in M_2.$$

Taking $\varphi(t_{k-1}) = u_n \delta_{nk}$, where $u_n := u|_{I_n}$ we get, for $n = 2, \ldots, N$,

$$\int_{I_{n-1}} \left\langle Au_{n-1}, \frac{u_n}{\tau_{n-1}}(t - t_{n-2}) \right\rangle \mathrm{d}t + \int_{I_n} \left\langle Au_n, -\frac{u_n}{\tau_n}(t - t_{n-1}) \right\rangle \mathrm{d}t = 0,$$

and, for $n = 1$,

$$\begin{aligned} 0 &= -\int_{I_1} \left\langle Au_1, -\frac{\varphi(t_0)}{\tau_1} t \right\rangle \mathrm{d}t = \int_{I_1} \left\langle Au_1, \frac{u_1}{\tau_1} t \right\rangle \mathrm{d}t \\ &= \int_{I_1} \left\langle A\frac{u_1}{\sqrt{\tau_1}}\sqrt{t}, \frac{u_1}{\sqrt{\tau_1}}\sqrt{t} \right\rangle \mathrm{d}t \geq \nu_a \int_{I_1} \frac{t}{\tau_1} \|u_1\|_V^2 \,\mathrm{d}t = \frac{\nu_a \tau_1}{2} \|u_1\|_V^2 . \end{aligned}$$

Therefore $u_1 = 0$, and by induction $u_n = 0$ for every $n = 1, \ldots, N$. □

Therefore the method cannot be quasi-optimal, because it occurs that the best error vanishes, whereas the error does not.

We write the consistency error as follows:

$$\begin{aligned} b_M(u, \varphi) - \ell_M(\varphi) &= b_M(u, \varphi) - b(u, \varphi) + \ell(\varphi) - \ell_M(\varphi) \\ &= \int_0^T \left\langle Au, \breve{\Pi}\varphi - \varphi \right\rangle . \end{aligned}$$

Combining Proposition 4.7 with the results in Section 1.3, we obtain the following.

Proposition 4.12. *The Galerkin method* (4.25) *with* b_M *as in* (4.26) *and without modifications of the right-hand side satisfies*

$$\|u - U_M\|_1 \leq 2\nu_a^{-1} \max\{1, C_a^2\} \inf_{v \in M_1} \|u - v\|_1 + 2\nu_a^{-1} \max\{1, C_a\} \sup_{\varphi \in M_2} \frac{\int_0^T \left\langle Au, \breve{\Pi}\varphi - \varphi \right\rangle}{\|\varphi\|_{2,\mathcal{P}}}. \tag{4.35}$$

Moreover, if $\mu_{\mathcal{P}} = \sup_n \frac{\tau_{n-1}}{\tau_n} < \infty$, *then*

$$\|u - U_M\|_1 \to 0 \qquad \text{as} \qquad \tau_{\mathcal{P}} \to 0.$$

Proof. It follows the same lines as the proof of Proposition 4.9. □

Remark 4.13. Assuming $\mu_{\mathcal{P}} < \infty$ the Galerkin method (4.25) with b_M as in (4.26) and without modifications of the right-hand side is stable, since from (4.35) follows

$$\|u - U_M\|_1 \leq C(\nu_a, C_a, \mu_{\mathcal{P}}) \|u\|_1 .$$

Chapter 5

Varying the Spatial Discretization

In Chapter 3 we show that a necessary and sufficient condition for the semidiscrete Galerkin approximation to be quasi-optimal is the H^1-stability of the L^2-projection. In Chapter 4 we prove that the error related to the standard formulation is equivalent to the sum of best errors for the exact solution and its time derivative. What if we discretize in both space and time? Can we expect an equivalence between the error and suitable best errors?

Todd Dupont in [22] presents a remarkable example. The discretization takes place in space with one-dimensional finite elements, and in time with backward Euler. The spatial mesh changes every time-step in such a way that, if the mesh-size h and the time step τ are such that $h^4/\tau \to \infty$, then the discrete solution does not converge to the exact solution as $h,\ \tau \to 0$. However, the best errors for the solution and its time derivative tend to zero as $h,\ \tau \to 0$, independently of their ratio. These best errors are intended with respect to the space of piecewise constants in time with values, in each subinterval, in the corresponding finite element space.

This reveals that, in addition to the best errors of above, some extra terms arise in the bound for the error, at least when the spatial discretizations are allowed to change. In a more general setting, Chrysafinos and Walkington [14] prove that the error in the $L^2(L^\infty)\cap L^2(H^1)$-norm can be bounded by the error given by a suitable local projection, and an extra term, that vanishes if the spatial discretization remains the same.

In order to better understand the situation, in this chapter we deal with the modification of the spatial discretizations and we add the time discretization only in Chapter 6.

The chapter is organized as follows. Section 5.1 concerns the standard formulation, while Section 5.2 the natural formulation. In both cases we

find the same term that disturbs the quasi-optimality. It vanishes if the spatial discretization does not change, and behaves like h^4/τ in the context of Dupont's example, as shown in Chapter 7.

The results of this chapter are the outcome of a collaboration with Christian Kreuzer.

5.1 Standard formulation

Assume that $V \subset W \subset V^*$, a and A are as in Section 2.1, while, as in Section 2.2, $X_1 = H^1(V, V^*)$ and $X_2 = W \times L^2(V)$ with norms

$$\|v\|_1^2 = \|v(0)\|_W^2 + \int_0^T \|v'\|_{V^*}^2 + \|v\|_V^2\,, \qquad \|\varphi\|_2^2 = \|\varphi_0\|_W^2 + \int_0^T \|\varphi_1\|_V^2\,.$$

Moreover $u_0 \in W$, $f \in L^2(V)$, and the bilinear form $b : X_1 \times X_2 \to \mathbb{R}$ and $\ell \in X_2^*$ are given by

$$b(v,\varphi) = \langle v(0), \varphi_0\rangle + \int_0^T \langle v', \varphi_1\rangle + \langle Av, \varphi_1\rangle\,,$$
$$\ell(\varphi) = \langle u_0, \varphi_0\rangle + \int_0^T \langle f, \varphi_1\rangle\,.$$

Furthermore, as in Section 4.1, let $N \in \mathbb{N}$ and $\mathcal{P}$ be a partition

$$0 = t_0 < t_1 < \ldots < t_N = T$$

of $I = (0,T)$ into N subintervals $I_n := (t_{n-1}, t_n]$. We consider a sequence of finite-dimensional subspaces $\{\mathbb{V}_n\}_{n=0}^N \subset V$. Approximation of the initial value takes place in $\mathbb{V}_0$, while for $n = 1, \ldots, N$, the approximation in the n-th interval I_n occurs in $\mathbb{V}_n$. The W-orthogonal projection on $\mathbb{V}_n$ is denoted by P_n, and, motivated by the results in Chapter 3, we assume that $\{P_n\}_{n=0}^N$ is uniformly stable in V, with

$$\sigma := \sup_{n=0,\ldots,N} \sup_{v\in V} \frac{\|P_n v\|_V}{\|v\|_V}.$$

In addition we denote by $\mathbb{A}_n : I \to \mathcal{L}(\mathbb{V}_n, \mathbb{V}_n^*)$ the discrete counterparts of A, that is,

$$\forall \varphi \in \mathbb{V}_n, \qquad \langle \mathbb{A}_n(t) v, \varphi\rangle = \langle A(t) v, \varphi\rangle\,.$$

Finally we set

$$\mathcal{S}^{L^2}(\mathcal{P}, \mathbb{V}) := \{\varphi \in L^2(V),\ \varphi|_{I_n} \in L^2(I_n, \mathbb{V}_n),\ n = 1, \ldots, N\},$$
$$\mathcal{S}^{H^1}(\mathcal{P}, \mathbb{V}) := \{v \in L^2(V),\ v|_{I_n} \in H^1(I_n, \mathbb{V}_n),\ n = 1, \ldots, N\},$$

and we consider the spaces

$$M_1 := \{v \in \mathcal{S}^{H^1}(\mathcal{P}, \mathbb{V}),\ v(0) \in \mathbb{V}_0, \qquad v(t_{n-1}^+) = P_n v(t_{n-1}),\ n = 1, \dots, N\}, \tag{5.1a}$$

$$M_2 := \mathbb{V}_0 \times \mathcal{S}^{L^2}(\mathcal{P}, \mathbb{V}), \tag{5.1b}$$

where $v(t^+) := \lim_{s \searrow t} v(s)$ denotes the right limit of the function v to the time t. We notice that $M_2 \subset X_2$, while in general $M_1 \nsubseteq X_1$. Thus we are in a non-conforming situation and we invoke the results of Section 1.3. If we choose $\mathbb{V}_n = S$ for every $n = 0, \dots, N$, then M_1 coincides with $H^1(S)$, M_2 coincides with $S \times L^2(S)$, and we are back in the situation of Section 3.2. We remark that the constraints $v(t_{n-1}^+) = P_n v(t_{n-1})$ in the definition of M_1 can be seen as a discrete replacement of the embedding $X_1 \subset C^0(W)$. The space M_2 inherits the $\|\cdot\|_2$-norm, while M_1 is endowed with the broken counterpart of $\|\cdot\|_1$:

$$\|v\|_{1,\mathcal{P}}^2 := \|v(0)\|_W^2 + \sum_{n=1}^N \int_{I_n} \|v'\|_{V^*}^2 + \|v\|_V^2 .$$

We replace b with its broken counterpart

$$b_M(v, \varphi) := \langle v(0), \varphi_0 \rangle + \sum_{n=1}^N \int_{I_n} \langle v', \varphi \rangle + \langle Av, \varphi \rangle , \tag{5.2}$$

so that b_M is well-defined also on $M_1 \times M_2$. We prove that the discrete problem is well-posed and derive a quasi-optimality result in M_1.

5.1.1 Quasi-optimality in a space with constraints

Proposition 5.1. *The bilinear form* (5.2) *is continuous and satisfies the inf-sup condition* (1.9) *on* $M_1 \times M_2$ *with*

$$C_M \le \sqrt{2} \max\{1, C_a\}, \qquad c_M \ge \frac{\min\{\nu_a, C_a^{-1}\sigma^{-1}, \nu_a C_a^{-1}\sigma^{-1}\}}{2}.$$

Moreover, it is also continuous on $(X_1, \|\cdot\|_{1,\mathcal{P}}) \times (M_2, \|\cdot\|_2)$ *and* $C_{X_1 \times M_2}$ *satisfies the same bound as* C_M.

Proof. The proof mimics those of Propositions 2.2 and 3.3. In order to derive the lower bound for c_M we choose, for every $v \in M_1$,

$$\varphi_0 = 2v(0), \qquad \varphi|_{I_n} = v|_{I_n} + \mathbb{A}_n^{-*}(v'|_{I_n}), \quad n = 1, \dots, N.$$

In addition we exploit $\|v'\|_{\mathbb{V}_n^*} \geq \sigma^{-1}\|v'\|_{V^*}$ and

$$\|v(0)\|_W^2 + \sum_{n=1}^{N}\|v(t_n)\|_W^2 - \left\|v(t_{n-1}^+)\right\|_W^2 \geq \sum_{n=0}^{N-1}\|v(t_n)\|_W^2 - \left\|v(t_n^+)\right\|_W^2$$
$$= \sum_{n=0}^{N-1}\|v(t_n)\|_W^2 - \|P_{n+1}v(t_n)\|_W^2 = \sum_{n=0}^{N-1}\|v(t_n) - P_{n+1}v(t_n)\|_W^2 \geq 0.$$

Concerning the non-degeneracy condition (1.9b), if $b_M(v,\varphi) = 0$ for every $v \in M_1$, the argument in the proof of Proposition 2.2 gives $\varphi|_{I_n} := \varphi_n \in H^1(I_n, \mathbb{V}_n)$ for every $n = 1, \ldots, N$. Integrating by parts in $b_M(v,\varphi) = 0$ we have

$$\langle v(0), \varphi_0\rangle + \sum_{n=1}^{N}\langle v(t_n), \varphi(t_n)\rangle - \left\langle v(t_{n-1}^+), \varphi(t_{n-1}^+)\right\rangle + \int_{I_n} -\langle\varphi', v\rangle + \langle Av, \varphi\rangle = 0.$$

Testing with suitable functions, we get, for $n = 1, \ldots, N$, $-\varphi_n' + \mathbb{A}_n^*\varphi_n = 0$, $\varphi(t_n) = P_n\varphi(t_n^+)$, $\varphi_0 = P_0\varphi(0^+)$ and $\varphi(T) = 0$. Proceeding with a backward induction we get that $\varphi_n = 0$, for every $n = N, \ldots, 1$ and $\varphi_0 = 0$. □

We notice that $b_M = b$ on $X_1 \times M_2$, so that $b_M(u,\varphi) = \ell(\varphi)$ for every $\varphi \in M_2$ and the consistency error vanishes. Applying the results of Section 1.3 we get the following proposition.

Proposition 5.2. *The Galerkin solution U_M of method* (5.1) *with b_M as in* (5.2) *satisfies the following estimate*

$$\|u - U_M\|_{1,\mathcal{P}} \leq \kappa_\sigma \inf_{v\in M_1}\|u - v\|_{1,\mathcal{P}}\,, \tag{5.3}$$

with $\kappa_\sigma := 2\sqrt{2}\max\{\nu_a^{-1}, C_a^2\sigma, \nu_a^{-1}C_a^2\sigma\}$.

5.1.2 Abstract error estimate

The infimum on the right-hand side of (5.3) is on functions v that belong to M_1. Therefore $v|_{I_n} \in \mathbb{V}_n$ and $v|_{I_{n+1}} \in \mathbb{V}_{n+1}$ are not independent but linked by $v(t_n^+) = P_{n+1}v(t_n)$. This puts some limitations in the approximation power of $\mathbb{V}_{n+1}$. We consider the unconstrained version of M_1:

$$\widehat{M_1} := \{v \in \mathcal{S}^{H^1}(\mathcal{P}, \mathbb{V}),\ v(0) \in \mathbb{V}_0\} \tag{5.4}$$

and aim at an error estimate that involves the best error on $\widehat{M_1}$. To this end we insert a particular choice of $v = \mathcal{I}u \in M_1$ in the right-hand side of (5.3).

In order to have a near-best approximation on I_n, one could think of taking $\mathcal{I}u|_{I_n} = P_n u$, $n = 1, \dots, N$. However, in general

$$\lim_{t \searrow t_n} P_{n+1} u(t) = P_{n+1} u(t_n) \neq P_{n+1} P_n u(t_n),$$

which violates the constraint in the definition of M_1. To overcome this problem we introduce a correction $z \in \widehat{M}_1$ defined iteratively starting with $z(0) := 0$ and such that

$$\begin{aligned} \lim_{t \searrow t_{n-1}} z(t) &= P_n P_{n-1} u(t_{n-1}) + P_n z(t_{n-1}) - P_n u(t_{n-1}) \\ &=: z_n^+ \in \mathbb{V}_n, \end{aligned} \tag{5.5a}$$

$n = 1, \dots, N$. The function $z_n^+ \in \mathbb{V}_n$ represents the deviation from the W-projection at t_{n-1}. We let evolve this defect in I_n by defining $z|_{I_n} =: z_n \in H^1(I_n; \mathbb{V}_n)$ to be the semidiscrete solution of the homogeneous parabolic problem

$$z_n' + \mathbb{A}_n z_n = 0 \quad \text{in } L^2(I_n; \mathbb{V}_n^*) \qquad \text{and} \qquad z_n(t_{n-1}) = z_n^+. \tag{5.5b}$$

We define $\mathcal{I} : X_1 + M_1 \to M_1$ as

$$(\mathcal{I}u)|_{I_n} := (P_n u)|_{I_n} + z_n, \quad n = 1, \dots, N, \qquad \text{and} \qquad (\mathcal{I}u)(0) := P_0 u(0). \tag{5.6}$$

We investigate the properties of $\mathcal{I}$ in the following

Proposition 5.3 (Properties of $\mathcal{I}$)**.** *The interpolation operator $\mathcal{I}$ defined in* (5.6) *with z_n as in* (5.5) *is a linear projection onto M_1, and it is stable with respect to $\|\cdot\|_{1,\mathcal{P}}$ with*

$$\|\mathcal{I}u\|_{1,\mathcal{P}} \leq \sigma \kappa_\sigma \|u\|_{1,\mathcal{P}}.$$

Proof. Thanks to (5.5a), $\mathcal{I}u \in M_1$ for every $u \in X_1 + M_1$. Linearity follows from linearity of P_n and of the equation in (5.5b). Invariance over M_1 is due to the fact that, for every $u \in M_1$, $P_n u = u$ and $z_n^+ = 0$, $n = 1, \dots, N-1$. Concerning stability, we exploit the fact that b_M satisfies the inf-sup condition on $M_1 \times M_2$. Therefore there exists $\varphi \in M_2$ such that

$$c_M \|\mathcal{I}u\| \|\varphi\| \leq b_M(\mathcal{I}u, \varphi). \tag{5.7}$$

Using the definition of z, the continuity of b_M and the V-stability of P_n we

get

$$\begin{aligned} b_M(\mathcal{I}u,\varphi) &= \langle \mathcal{I}u(0),\varphi_0\rangle + \sum_{n=1}^{N}\int_{I_n}\langle (\mathcal{I}u)',\varphi_n\rangle + \langle A\mathcal{I}u,\varphi_n\rangle \\ &= \langle P_0u(0),\varphi_0\rangle + \sum_{n=1}^{N}\int_{I_n}\langle (P_nu+z_n)',\varphi_n\rangle + \langle A(P_nu+z_n),\varphi_n\rangle \\ &= \langle P_0u(0),\varphi_0\rangle + \sum_{n=1}^{N}\int_{I_n}\langle (P_nu)',\varphi_n\rangle + \langle AP_nu,\varphi_n\rangle \\ &\le C_M\sigma \left\|u\right\|_{1,\mathcal{P}} \left\|\varphi\right\|_2 . \end{aligned}$$

Combining this with (5.7) gives the assertion. □

Thanks to Propositions 5.2–5.3, the choice of $\mathcal{I}$ ensures that

$$\left\|u-\mathcal{I}u\right\|_{1,\mathcal{P}} \approx \left\|u-U_M\right\|_{1,\mathcal{P}},$$

with hidden constants depending on ν_a, C_a and σ. In order to further estimate $\left\|u-\mathcal{I}u\right\|_{1,\mathcal{P}}$ we split it as

$$\left\|u-\mathcal{I}u\right\|_{1,\mathcal{P}} \le \left\|u-Pu\right\|_{1,\mathcal{P}} + \left\|z\right\|_{1,\mathcal{P}}, \tag{5.8}$$

where $(Pu)|_{I_n} := P_nu$. The stability of P_n allows to relate the first term on the right-hand side with the best-error in $\widehat{M}_1$. We bound $\left\|z\right\|_{1,\mathcal{P}}$ with the help of the following proposition.

Proposition 5.4. *The correction z defined in* (5.5)*, satisfies*

$$C_z^{-1}\left\|z\right\|_{1,\mathcal{P}}^2 \le \sum_{n=1}^{N}\left\|z_n(t_{n-1}^+)\right\|_W^2 - \left\|z_n(t_n)\right\|_W^2 \le c_z^{-1}\left\|z\right\|_{1,\mathcal{P}}^2, \tag{5.9a}$$

where

$$C_z := \nu_a^{-1}\max\{1,C_a^2\sigma^2\} \quad \textit{and} \quad c_z := \min\{C_a^{-1},\nu_a\}. \tag{5.9b}$$

Moreover,

$$\begin{aligned} &\sum_{n=1}^{N}\left\|z_n(t_{n-1}^+)\right\|_W^2 - \left\|z_n(t_n)\right\|_W^2 \\ &\qquad \le \left\|P_1(I-P_0)u(t_0)\right\|_W^2 + \sum_{n=1}^{N-1}\left\|P_n^+(I-P_n)u(t_n)\right\|_W^2, \end{aligned} \tag{5.9c}$$

where P_n^+ denotes the W-projection onto $\mathbb{V}_n\oplus\mathbb{V}_{n+1}$, $n=1,\dots,N-1$.

Proof. To proof (5.9a) we proceed as in the proof of Proposition 5.1 and test (5.5b) with $z_n + \mathbb{A}_n^{-*} z_n'$ and integrate over I_n. We get

$$0 = \|z_n(t_n)\|_W^2 - \left\|z_n(t_{n-1}^+)\right\|_W^2 + \int_{I_n} \langle \mathbb{A}_n z_n, z_n \rangle + \left\langle z_n', \mathbb{A}_n^{-*} z_n' \right\rangle .$$

The bounds follow from continuity and coercivity of $\mathbb{A}_n$, Lemma 2.1 and summing over n. Concerning (5.9c), we have

$$\begin{aligned} &\sum_{n=1}^{N} \left\|z_n(t_{n-1}^+)\right\|_W^2 - \|z_n(t_n)\|_W^2 \\ &\qquad \le \|P_1 P_0 u(t_0) - P_1 u(t_0)\|_W^2 \\ &\qquad\quad + \sum_{n=1}^{N-1} \|P_{n+1} P_n u(t_n) + P_{n+1} z_n(t_n) - P_{n+1} u(t_n)\|_W^2 - \|z_n(t_n)\|_W^2 . \end{aligned} \tag{5.10}$$

Since $\mathbb{V}_{n+1} \subset \mathbb{V}_n \oplus \mathbb{V}_{n+1}$ we have $P_{n+1} = P_n^+ P_{n+1} = P_{n+1} P_n^+$ and we can bound every term in the sum in the right-hand side in the following way:

$$\begin{aligned} &\|P_{n+1} P_n u(t_n) + P_{n+1} z_n(t_n) - P_{n+1} u(t_n)\|_W^2 \\ &\qquad = \left\|P_{n+1} P_n u(t_n) + P_{n+1} z_n(t_n) - P_{n+1} P_n^+ u(t_n)\right\|_W^2 \\ &\qquad \le \left\|P_n u(t_n) + z_n(t_n) - P_n^+ u(t_n)\right\|_W^2 \\ &\qquad = \left\|P_n u(t_n) - P_n^+ u(t_n)\right\|_W^2 + \|z_n(t_n)\|_W^2 . \end{aligned}$$

Inserting this in (5.10) gives (5.9c). □

Combining (5.3) with (5.8) and (5.9), and recalling that P_n is uniformly stable we get the following result.

Theorem 5.5. *The Galerkin solution U_M of method* (5.1) *with b_M as in* (5.2) *satisfies the following estimate*

$$\begin{aligned} \|u - U_M\|_{1,\mathcal{P}} &\le \kappa_\sigma \sigma \inf_{v \in \widehat{M}_1} \|u - v\|_{1,\mathcal{P}} \\ &\quad + \kappa_\sigma \sqrt{C_z} \left(\|P_1 (I - P_0) u(t_0)\|_W^2 + \sum_{n=1}^{N-1} \left\|P_n^+ (I - P_n) u(t_n)\right\|_W^2 \right)^{1/2} . \end{aligned} \tag{5.11}$$

Remark 5.6. If $\mathbb{V}_{n+1} \subset \mathbb{V}_n$ for every $n = 0, \dots, N-1$, then $P_1(I - P_0)$ and $P_n^+(I - P_n) = 0$, and we get

$$\|u - U_M\|_1 \le \kappa_\sigma \sigma \inf_{v \in H^1(S)} \|u - v\|_1 .$$

In particular, if $\mathbb{V}_n = S$ for every $n = 0, \dots, N$, we recover qualitatively one of the results of Theorem 3.7.

To give an idea of the sharpness and the limitations of estimate (5.11), we consider the following example.

Example 5.7 (Discretization with eigenfunctions)**.** Assume $\Omega \subset \mathbb{R}^d$, $V = H_0^1(\Omega)$, $W = L^2(\Omega)$, $V^* = H^{-1}(\Omega)$, $u_0 \in L^2(\Omega)$ and $A = -\Delta$. We consider the homogeneous equation

$$\begin{aligned} \partial_t u - \Delta u &= 0 && \text{in } \Omega \times (0,T), \\ u &= 0 && \text{on } \partial\Omega \times (0,T), \\ u(x,0) &= u_0 && \text{in } \Omega, \end{aligned}$$

whose exact solution is given by

$$u(x,t) = \sum_{j=1}^{\infty} \langle u_0, \phi_j \rangle \, e^{-\lambda_j t} \phi_j(x),$$

where $\{\phi_j\}_{j=1}^{\infty}$ are the eigenfunctions of the Laplacian, with corresponding eigenvalues $\{\lambda_j\}_{j=1}^{\infty}$.

For the discretization, we consider an even N and a uniform partition in time, with $\tau_1 = \ldots = \tau_N =: T/N$, and thus $t_n = nT/N$, $n = 0, \dots, N$. Moreover we set $\mathbb{V}_n = S_{\text{even}}$ if n is even, otherwise $\mathbb{V}_n = S_{\text{odd}}$, with

$$S_{\text{even}} := \operatorname{span}\{\phi_1, \dots, \phi_{m-1}, \phi_m\}, \qquad S_{\text{odd}} := \operatorname{span}\{\phi_1, \dots, \phi_{m-1}, \phi_{m+1}\}.$$

We want to compute the difference between the error and the bound in (5.11). The function U given by

$$U(x,0) = \sum_{j=1}^{m} \langle u_0, \phi_j \rangle \, \phi_j(x), \qquad U(x,t) = \sum_{j=1}^{m-1} \langle u_0, \phi_j \rangle \, e^{-\lambda_j t} \phi_j(x), \ t > 0,$$

belongs to M_1 and solves the discrete problem. The error $\|u - U\|_{1,\mathcal{P}}$ is given by

$$\begin{aligned} \|u - U\|_{1,\mathcal{P}}^2 &= \|u(0) - U(0)\|_W^2 + \sum_{n=1}^{N} \int_{I_n} \|u' - U'\|_{V^*}^2 + \|u - U\|_V^2 \\ &= \sum_{j=m+1}^{\infty} \langle u_0, \phi_j \rangle^2 + \int_I \sum_{j=m}^{\infty} \langle u_0, \phi_j \rangle^2 \, 2\lambda_j e^{-2\lambda_j t} \, \mathrm{d}t \\ &= \sum_{j=m+1}^{\infty} \langle u_0, \phi_j \rangle^2 + \sum_{j=m}^{\infty} \langle u_0, \phi_j \rangle^2 \, (1 - e^{-2\lambda_j T}). \end{aligned}$$

Since $S_{\text{even}} \oplus S_{\text{odd}} = \operatorname{span}\{\phi_1, \ldots, \phi_{m+1}\}$, we get

$$P_n^+(I - P_n)u(t_n) = \begin{cases} \langle u_0, \phi_{m+1}\rangle \, \phi_{m+1} e^{-\lambda_{m+1} t_n} & \text{if } n \text{ even} \\ \langle u_0, \phi_m\rangle \, \phi_m e^{-\lambda_m t_n} & \text{if } n \text{ odd} \end{cases},$$

while $P_1(I - P_0)u_0 = \langle u_0, \phi_{m+1}\rangle \, \phi_{m+1}$. We recall that N is even and obtain

$$\begin{aligned} &\|P_1(I - P_0)u_0\|_W^2 + \sum_{n=1}^{N-1} \left\|P_n^+(I - P_n)u(t_n)\right\|_W^2 \\ &\quad = \langle u_0, \phi_{m+1}\rangle^2 + \sum_{k=1}^{N/2-1} \langle u_0, \phi_{m+1}\rangle^2 \, e^{-2\lambda_{m+1} t_{2k}} + \sum_{k=0}^{N/2-1} \langle u_0, \phi_m\rangle^2 \, e^{-2\lambda_m t_{2k+1}} \\ &\quad = \langle u_0, \phi_{m+1}\rangle^2 \sum_{k=0}^{N/2-1} \left(e^{-\frac{4\lambda_{m+1} T}{N}}\right)^k + \langle u_0, \phi_m\rangle^2 \, e^{-\frac{2\lambda_m T}{N}} \sum_{k=0}^{N/2-1} \left(e^{-\frac{4\lambda_m T}{N}}\right)^k, \end{aligned}$$

and

$$\begin{aligned} \inf_{v \in \widehat{M}_1} \|u - v\|_{1,\mathcal{P}}^2 = &\sum_{j=m+1}^{\infty} \langle u_0, \phi_j\rangle^2 + \int_I \sum_{j=m+2}^{\infty} \langle u_0, \phi_j\rangle^2 \, 2\lambda_j e^{-2\lambda_j t} \, \mathrm{d}t \\ &+ \sum_{k=1}^{N/2} \int_{I_{2k}} \langle u_0, \phi_{m+1}\rangle^2 \, 2\lambda_{m+1} e^{-2\lambda_{m+1} t} \, \mathrm{d}t \\ &+ \sum_{k=0}^{N/2-1} \int_{I_{2k+1}} \langle u_0, \phi_m\rangle^2 \, 2\lambda_m e^{-2\lambda_m t} \, \mathrm{d}t. \end{aligned}$$

Therefore

$$\begin{aligned} &\inf_{v \in \widehat{M}_1} \|u - v\|_{1,\mathcal{P}}^2 + \|P_1(I - P_0)u_0\|_W^2 + \sum_{n=1}^{N-1} \left\|P_n^+(I - P_n)u(t_n)\right\|_W^2 \\ &\qquad - \|u - U\|_{1,\mathcal{P}}^2 \\ &\quad = \langle u_0, \phi_{m+1}\rangle^2 \, e^{-\frac{2\lambda_{m+1} T}{N}} \left(\frac{1 - e^{-2\lambda_{m+1} T}}{1 - e^{-\frac{4\lambda_{m+1} T}{N}}}\right) \\ &\qquad + \langle u_0, \phi_m\rangle^2 \, e^{\frac{-4\lambda_m T}{N}} \left(\frac{1 - e^{-2\lambda_m T}}{1 - e^{-\frac{4\lambda_m T}{N}}}\right). \end{aligned}$$

We notice that this difference may get big, if we take $N \gg \lambda_{m+1}$. On the other hand, if $N \ll \lambda_m$, it converges to 0 as $m \to \infty$.

5.2 Natural formulation

Assume that $V \subset W \subset V^*$, a and A are as in Section 2.1, while, as in Section 2.3, $X_1 = L^2(V)$, $X_2 = \{\varphi \in H^1(V, V^*), \varphi(T) = 0\}$ with norms

$$\|v\|_1^2 = \int_0^T \|v\|_V^2 \,, \qquad \|\varphi\|_2^2 = \int_0^T \|\varphi'\|_{V^*}^2 + \|\varphi\|_V^2 \,,$$

and the bilinear form $b : X_1 \times X_2 \to \mathbb{R}$ is given by

$$b(v, \varphi) = \int_0^T -\langle \varphi', v\rangle + \langle Av, \varphi\rangle \,.$$

We take $\ell \in X_2^*$ of the form

$$\ell(\varphi) = \langle g_0, \varphi(0)\rangle + \sum_{j=1}^{\jmath-1} \big\langle g_j, \varphi(\tilde{t}_j)\big\rangle + \int_0^T \langle f, \varphi\rangle \,,$$

with $f \in L^2(V^*)$, $(g_j)_{j=0}^{\jmath-1} \subset W$ and $0 = \tilde{t}_0 < \tilde{t}_1 < \ldots < \tilde{t}_\jmath = T$, as in Section 4.2.1. We recall that, with this choice, the solution u of

$$b(u, \varphi) = \ell(\varphi), \qquad \forall \varphi \in X_2, \tag{5.12}$$

belongs to $H^1((\tilde{t}_{j-1}, \tilde{t}_j); V, V^*) \subset C^0([\tilde{t}_{j-1}, \tilde{t}_j]; W)$, for $j = 1, \ldots, \jmath$. Moreover it also satisfies, for every $(\varphi, \phi) \in L^2(0, T; V) \times W^\jmath$,

$$\begin{aligned} \langle u(0), \phi_0\rangle + \sum_{j=1}^{\jmath-1} \big\langle u(\tilde{t}_j^+) - u(\tilde{t}_j^-), \phi_j\big\rangle + \sum_{j=1}^{\jmath} \int_{\tilde{t}_{j-1}}^{\tilde{t}_j} \langle u' + Au, \varphi\rangle \\ = \sum_{j=0}^{\jmath-1} \langle g_j, \phi_j\rangle + \int_0^T \langle f, \varphi\rangle \,, \end{aligned} \tag{5.13}$$

where $u(t^-) := \lim_{s \nearrow t} u(s)$ denotes the left limit of the function u to the time t.

Let $\mathcal{P}$ be a partition of $(0, T)$, $0 =: t_0 < t_1 < \ldots < t_N := T$, that is subordinate to $(\tilde{t}_j)_{j=1}^{\jmath-1}$, that is, for every $j = 1, \ldots, \jmath - 1$ there exists $n \in \{1, \ldots, N\}$ such that $t_n = \tilde{t}_j$. We indicate by $\mathfrak{J}$ the set of indices n which correspond to an index j. Similarly as Section 5.1 a finite-dimensional subspace $\mathbb{V}_n \subset V$ is related to every subinterval $I_n := [t_{n-1}, t_n)$. Note that I_n is left-closed and right-open. The W-orthogonal projections P_n onto $\mathbb{V}_n$, $n = 1, \ldots, N$ are assumed to be uniformly stable in V, with

$$\sigma := \sup_{n=1,\ldots,N} \sup_{v \in V} \frac{\|P_n v\|_V}{\|v\|_V}.$$

Moreover let $\mathbb{A}_n : I \to \mathcal{L}(\mathbb{V}_n, \mathbb{V}_n^*)$ be the discrete counterpart of A. We consider the spaces

$$M_1 := \{v \in L^2(V), v|_{I_n} \in L^2(I_n; \mathbb{V}_n),\ n = 1, \dots, N\}, \tag{5.14a}$$

$$M_2 := \{\varphi \in L^2(V), \varphi|_{I_n} \in H^1(I_n; \mathbb{V}_n),\ n = 1, \dots, N, \\ \varphi(t_n^-) = P_n\varphi(t_n),\ n = 1, \dots, N-1, \varphi(T) = 0\}. \tag{5.14b}$$

The space $M_1 \subset X_1$ and inherits its norm, while $M_2 \not\subseteq X_2$ in general and it is endowed with the broken counterpart of $\|\cdot\|_2$:

$$\|\varphi\|_{2,\mathcal{P}}^2 := \sum_{n=1}^{N} \int_{I_n} \|\varphi'\|_{V^*}^2 + \|\varphi\|_V^2 \,.$$

If we choose $\mathbb{V}_n = S$ for every $n = 1, \dots, N$, then M_1 and M_2 coincide respectively with $L^2(S)$ and $H^1(S)$ and we are in the situation of Section 3.3. We remark that the constraints $\varphi(t_n^-) = P_n\varphi(t_n)$ in the definition of M_2 can be seen as a discrete replacement of the embedding $X_2 \subset C^0(W)$.

We replace b with its broken counterpart

$$b_M(v, \varphi) := \sum_{n=1}^{N} \int_{I_n} -\langle \varphi', v\rangle + \langle Av, \varphi\rangle \,, \tag{5.15}$$

so that b_M is well-defined also on M_2. Moreover we replace ℓ with

$$\ell_M(\varphi) := \langle g_0, \varphi(0)\rangle + \sum_{j=1}^{\jmath-1} \big\langle g_j, \varphi(\tilde{t}_j^+)\big\rangle + \int_0^T \langle f, \varphi\rangle \,, \tag{5.16}$$

so that ℓ_M is well-defined also on M_2. We prove that the discrete problem is well-posed, and then invoke the results of Section 1.3, since we are in a non-conforming setting.

Proposition 5.8. *The bilinear form* (5.15) *is continuous and satisfies the inf-sup condition* (1.9) *on* $M_1 \times M_2$ *with*

$$C_M \le \sqrt{2}\max\{1, C_a\}, \qquad c_M \ge \frac{\nu_a}{\sqrt{2}}\min\{1, C_a^{-1}\sigma^{-1}\}.$$

Moreover it is also continuous on $(X_1, \|\cdot\|_1) \times (M_2, \|\cdot\|_{2,\mathcal{P}})$ *and* $C_{X_1\times M_2}$ *satisfies the same bound as* C_M*.*

Proof. The proof mimics the one of Propositions 2.3 and 5.1. In order to derive the lower bound for c_M we choose, for every $\varphi \in M_2$,

$$v|_{I_n} = \varphi|_{I_n} - \mathbb{A}_n^{-1}(\varphi'|_{I_n}), \qquad n = 1, \dots, N.$$

In addition we exploit $\|\varphi'\|_{\mathbb{V}_n^*} \geq \sigma^{-1} \|\varphi'\|_{V^*}$ and

$$\begin{aligned}\sum_{n=1}^{N} \|\varphi(t_{n-1})\|_W^2 - \left\|\varphi(t_n^-)\right\|_W^2 &\geq \|\varphi(0)\|_W^2 + \sum_{n=1}^{N-1} \|\varphi(t_n)\|_W^2 - \left\|\varphi(t_n^-)\right\|_W^2 \\ &= \|\varphi(0)\|_W^2 + \sum_{n=1}^{N-1} \|\varphi(t_n)\|_W^2 - \|P_n\varphi(t_n)\|_W^2 \\ &= \|\varphi(0)\|_W^2 + \sum_{n=1}^{N-1} \|\varphi(t_n) - P_n\varphi(t_n)\|_W^2 \geq 0. \end{aligned} \tag{5.17}$$

Concerning the non-degeneracy condition (1.9b), if $b_M(v,\varphi) = 0$ for every $\varphi \in M_2$, the argument in the proof of Proposition 2.2 gives, for every $n = 1,\dots,N$, $v \in H^1(I_n, \mathbb{V}_n)$. As in the proof of Proposition 5.1 we get, for $n = 1,\dots,N$, $v_n' + \mathbb{A}_n v_n = 0$, $v(t_n) = P_{n+1} v(t_n^-)$ and $v(0) = 0$. By induction on n we get $v|_{I_n} = 0$, for every $n = 1,\dots,N$. □

Applying the results of Section 1.3 we get

$$\begin{aligned}\|u - U_M\|_1 \leq 2\nu_a^{-1} \max\{1, C_a\sigma, C_a^2\sigma\} \inf_{v\in M_1} \|u - v\|_1 \\ + \sqrt{2}\nu_a^{-1} \max\{1, C_a\sigma\} \sup_{\varphi\in M_2} \frac{b_M(u,\varphi) - \ell_M(\varphi)}{\|\varphi\|_{2,\mathcal{P}}}.\end{aligned} \tag{5.18}$$

5.2.1 Abstract error estimate

Our purpose is to bound the consistency error that appears in the right-hand side of (5.18).

Proposition 5.9 (Bound for the consistency error)**.** *The consistency error in* (5.18) *satisfies*

$$\sup_{\varphi\in M_2} \frac{b_M(u,\varphi) - \ell_M(\varphi)}{\|\varphi\|_{2,\mathcal{P}}} \leq \left(\sum_{n=1}^{N-1} \left\|(I - P_n) P_n^+ u(t_n^-)\right\|_W^2 \right)^{\frac{1}{2}}, \tag{5.19}$$

where P_n^+ is the W-projection onto $\mathbb{V}_n \oplus \mathbb{V}_{n+1}$, $n = 1,\dots,N-1$.

Proof. We notice first that from (5.17) we deduce that

$$\begin{aligned}\sum_{n=1}^{N-1} \|\varphi(t_n) - P_n\varphi(t_n)\|_W^2 &\leq \sum_{n=1}^{N} \|\varphi(t_{n-1})\|_W^2 - \left\|\varphi(t_n^-)\right\|_W^2 \\ &= \sum_{n=1}^{N} \int_{I_n} -\langle \varphi', \varphi\rangle \leq \|\varphi\|_{2,\mathcal{P}}^2 .\end{aligned} \tag{5.20}$$

Our aim is then to bound $b_M(u,\varphi)-\ell_M(\varphi)$ in terms of the left-hand side of (5.20). Since the exact solution u of (5.12) is piecewise in $H^1(V^*)$ we can integrate by parts in time, and get, for every $\varphi\in M_2$,

$$\begin{aligned}
&b_M(u,\varphi)-\ell_M(\varphi)\\
&\quad=\sum_{n=1}^{N}\left(\int_{I_n}\langle-\varphi',u\rangle+\langle Au,\varphi\rangle\right)-\sum_{j=0}^{\jmath-1}\big\langle g_j,\varphi(\tilde t_j^+)\big\rangle-\int_0^T\langle f,\varphi\rangle\\
&\quad=\sum_{n=1}^{N}\left(\big\langle\varphi(t_{n-1}^+),u(t_{n-1}^+)\big\rangle-\big\langle\varphi(t_n^-),u(t_n^-)\big\rangle+\int_{I_n}\langle u',\varphi\rangle+\langle Au,\varphi\rangle\right)\\
&\qquad\qquad-\sum_{j=0}^{\jmath-1}\big\langle g_j,\varphi(\tilde t_j^+)\big\rangle-\int_0^T\langle f,\varphi\rangle\,.
\end{aligned}$$

Moreover u also solves (5.13) and $u(t_n^+)=u(t_n^-)$ for every $n\notin\mathfrak{J}$. Therefore,

$$\begin{aligned}
&b_M(u,\varphi)-\ell_M(\varphi)\\
&\quad=\sum_{n=1}^{N}\big\langle\varphi(t_{n-1}^+),u(t_{n-1}^+)\big\rangle-\big\langle\varphi(t_n^-),u(t_n^-)\big\rangle-\sum_{j=1}^{\jmath-1}\big\langle g_j,\varphi(\tilde t_j^+)\big\rangle\\
&\quad=\sum_{n=1,n\notin\mathfrak{J}}^{N-1}\big\langle\varphi(t_n^+),u(t_n^+)\big\rangle-\big\langle\varphi(t_n^-),u(t_n^-)\big\rangle\\
&\qquad\qquad+\sum_{j=1}^{\jmath-1}\big\langle\varphi(\tilde t_j^+),u(\tilde t_j^+)\big\rangle-\big\langle\varphi(\tilde t_j^-),u(\tilde t_j^-)\big\rangle-\big\langle u(\tilde t_j^+)-u(\tilde t_j^-),\varphi(\tilde t_j^+)\big\rangle\\
&\quad=\sum_{n=1}^{N-1}\big\langle\varphi(t_n^+),u(t_n^-)\big\rangle-\big\langle\varphi(t_n^-),u(t_n^-)\big\rangle\\
&\quad=\sum_{n=1}^{N-1}\big\langle\varphi(t_n^+),u(t_n^-)\big\rangle-\big\langle P_n\varphi(t_n^+),u(t_n^-)\big\rangle\,.
\end{aligned}\tag{5.21}$$

Every term in the sum on the right-hand side of (5.21) can be rewritten, introducing P_n^+, as

$$\begin{aligned}
\big\langle\varphi(t_n^+),u(t_n^-)\big\rangle-\big\langle P_n\varphi(t_n^+),u(t_n^-)\big\rangle&=\big\langle\varphi(t_n^+),P_n^+u(t_n^-)\big\rangle-\big\langle\varphi(t_n^+),P_n^+u(t_n^-)\big\rangle\\
&=\big\langle\varphi(t_n^+)-P_n\varphi(t_n^+),P_n^+u(t_n^-)\big\rangle\\
&=\big\langle\varphi(t_n^+)-P_n\varphi(t_n^+),(I-P_n)P_n^+u(t_n^-)\big\rangle\\
&\le\big\|\varphi(t_n^+)-P_n\varphi(t_n^+)\big\|_W\,\big\|(I-P_n)P_n^+u(t_n^-)\big\|_W\,.
\end{aligned}\tag{5.22}$$

Combining (5.21)–(5.22) with (5.20) we get (5.19). □

Combining (5.18) with Proposition 5.9 we get

Theorem 5.10. *The Galerkin solution U_M of method* (5.14) *with b_M as in* (5.15) *and ℓ_M as in* (5.16) *satisfies*

$$\|u - U_M\|_1 \leq 2\nu_a^{-1} \max\{1, C_a\sigma, C_a^2\sigma\} \inf_{v \in M_1} \|u - v\|_1 + \sqrt{2}\nu_a^{-1} \max\{1, C_a\sigma\} \left(\sum_{n=1}^{N-1} \left\| (I - P_n) P_n^+ u(t_n^-) \right\|_W^2 \right)^{\frac{1}{2}}.$$

Remark 5.11. The following statements are equivalent

(i) $\mathbb{V}_{n+1} \subseteq \mathbb{V}_n$, for every $n = 1, \ldots, N-1$;

(ii) $\sup_{\varphi \in M_2} \dfrac{b_M(u, \varphi) - \ell_M(\varphi)}{\|\varphi\|_{2,\mathcal{P}}} = 0.$

In this case,

$$\|u - U_M\|_1 \leq C(\nu_a, C_a, \sigma) \inf_{v \in M_1} \|u - v\|_1 ,$$

that is, we recover one of the results of Theorem 3.10.

Proof. (i)⇒(ii) follows from (5.19). Concerning (ii)⇒(i), from (5.21)–(5.22) and (ii) follows that

$$\left\langle \varphi, (I - P_n) P_n^+ u(t_n^-) \right\rangle = 0, \qquad \forall \varphi \in \mathbb{V}_{n+1}.$$

Therefore, it holds $0 = P_{n+1}(I - P_n)Pn^+ = P_{n+1}(I - P_n)$. Hence, for every $v_{n+1} \in \mathbb{V}_{n+1}$, we have

$$\begin{aligned} \|v_{n+1} - P_n v_{n+1}\|_W^2 &= \langle v_{n+1} - P_n v_{n+1}, v_{n+1} - P_n v_{n+1} \rangle \\ &= \langle v_{n+1}, v_{n+1} - P_n v_{n+1} \rangle \\ &= \langle v_{n+1}, P_{n+1} v_{n+1} - P_{n+1} P_n v_{n+1} \rangle = 0, \end{aligned}$$

that is $v_{n+1} = P_n v_{n+1}$, and $v_{n+1} \in \mathbb{V}_n$. □

Chapter 6

Full Discretization with the Backward Euler-Galerkin Method

In this chapter we analyse the backward Euler-Galerkin method, with the help of the results in Chapters 3–5. We discretize in both time and space and the spatial discretization may vary.

6.1 Standard formulation

Assume that $V \subset W \subset V^*$, a and A are as in Section 2.1, while, as in Section 2.2, $X_1 = H^1(V, V^*)$ and $X_2 = W \times L^2(V)$ with norms

$$\|v\|_1^2 = \|v(0)\|_W^2 + \int_0^T \|v'\|_{V^*}^2 + \|v\|_V^2 \,, \qquad \|\varphi\|_2^2 = \|\varphi_0\|_W^2 + \int_0^T \|\varphi_1\|_V^2 \,.$$

Moreover $u_0 \in W$, $f \in L^2(V)$, and the bilinear form $b : X_1 \times X_2 \to \mathbb{R}$ and $\ell \in X_2^*$ are given by

$$\begin{aligned} b(v, \varphi) &= \langle v(0), \varphi_0 \rangle + \int_0^T \langle v', \varphi_1 \rangle + \langle Av, \varphi_1 \rangle \,, \\ \ell(\varphi) &= \langle u_0, \varphi_0 \rangle + \int_0^T \langle f, \varphi_1 \rangle \,. \end{aligned}$$

Moreover, as in Section 4.1, let $N \in \mathbb{N}$ and $\mathcal{P}$ be a partition

$$0 = t_0 < t_1 < \ldots < t_N = T$$

of $I = (0, T)$ into N subintervals $I_n = (t_{n-1}, t_n]$, with

$$\mu_{\mathcal{P}} = \sup_n \frac{\tau_{n-1}}{\tau_n} < \infty.$$

As in Section 5.1 we consider a sequence of finite-dimensional subspaces $\{\mathbb{V}_n\}_{n=0}^N \subset V$. Approximation of the initial value takes place in $\mathbb{V}_0$, while for $n = 1, \dots, N$, the approximation in the n-th interval I_n occurs in $\mathbb{V}_n$. The W-orthogonal projection on $\mathbb{V}_n$ is denoted by P_n, and $\{P_n\}_{n=0}^N$ is assumed to be uniformly stable in V, with

$$\sigma = \sup_{n=0,\dots,N} \sup_{v \in V} \frac{\|P_n v\|_V}{\|v\|_V}.$$

In addition the operator $\mathcal{A}_n : \mathbb{V}_n \to \mathbb{V}_n^*$, which can be seen both as the discrete-in-time counterpart of $\mathbb{A}_n$ of Section 5.1, or the discrete-in-space counterpart of A_n of Section 4.1, is defined by

$$\mathcal{A}_n v := \frac{1}{\tau_n} \int_{I_n} \mathbb{A}_n(t) v \, \mathrm{d}t.$$

We set

$$\mathcal{S}^{0,-1}(\mathcal{P}, \mathbb{V}) := \{\varphi \in L^2(V), \varphi|_{I_n} \in \mathbb{V}_n, n = 1, \dots, N\}$$

and consider the spaces

$$M_1 := \{v \in L^2(V), v(0) \in \mathbb{V}_0, v|_{I_n} \in \mathbb{P}^1(I_n, \mathbb{V}_n), \quad v(t_{n-1}^+) = P_n v(t_{n-1}), n = 1, \dots, N\}, \tag{6.1a}$$

$$M_2 := \mathbb{V}_0 \times \mathcal{S}^{0,-1}(\mathcal{P}, \mathbb{V}), \tag{6.1b}$$

where $v(t^+) := \lim_{s \searrow t} v(s)$ denotes the right limit of the function v to the time t. We notice that $M_2 \subset X_2$, while in general $M_1 \nsubseteq X_1$. They are finite-dimensional spaces, with $\dim(M_1) = \dim(M_2)$. We remark that the constraints $v(t_{n-1}^+) = P_n v(t_{n-1})$ can be seen as a discrete replacement of the embedding $X_1 \subset C^0(W)$. The space M_2 inherits the $\|\cdot\|_2$-norm, while M_1 is endowed with

$$\|v\|_{1,\mathcal{P}}^2 := \|v(0)\|_W^2 + \sum_{n=1}^N \int_{I_n} \|v'\|_{V^*}^2 + \left\| \widehat{\Pi} v \right\|_V^2,$$

where $\widehat{\Pi}$ is defined, for $n = 1, \dots, N$, as in Section 4.1:

$$\widehat{\Pi} v|_{I_n} = \Pi^n v = \int_{I_n} v \psi^n, \quad \text{with} \quad \psi^n(t) = \frac{6(t - t_{n-1})}{\tau_n^2} - \frac{2}{\tau_n}.$$

Moreover we set $\Pi^0 u = u(0)$. We replace b with

$$b_M(v,\varphi) := \langle v(0), \varphi_0 \rangle + \sum_{n=1}^{N} \int_{I_n} \langle v', \varphi_n \rangle + \left\langle A\widehat{\Pi} v, \varphi_n \right\rangle, \tag{6.2}$$

so that b_M is well-defined also on $M_1 \times M_2$. We prove that the discrete problem is well-posed.

Proposition 6.1. *The bilinear form* (6.2) *is continuous and satisfies the inf-sup condition* (1.9) *on* $M_1 \times M_2$ *with*

$$C_M \le \sqrt{2}\max\{1, C_a\}, \qquad c_M \ge \frac{\min\{\nu_a, C_a^{-1}\sigma^{-1}, \nu_a C_a^{-1}\sigma^{-1}\}}{2}.$$

Moreover, it is also continuous on $(X_1, \|\cdot\|_{1,\mathcal{P}}) \times (M_2, \|\cdot\|_2)$ *and* $C_{X_1 \times M_2}$ *satisfies the same bound as* C_M*.*

Proof. The proof mimics those of Proposition 4.2 and 5.1. We derive the lower bound for c_M and the non-degeneracy condition (1.9b) follows since $\dim(M_1) = \dim(M_2) < \infty$. We choose, for every $v \in M_1$,

$$\varphi_0 = 2v(0), \qquad \varphi_n = \Pi^n v + \mathcal{A}_n^{-*}(v'|_{I_n}), \quad n = 1, \dots, N.$$

In addition, we exploit $\|v'\|_{\mathbb{V}_n^*} \ge \sigma^{-1} \|v'\|_{V^*}$ and

$$\begin{aligned}
&\|v(0)\|_W^2 + \sum_{n=1}^{N} \int_{I_n} 2\langle v', v(t_n)\rangle \\
&\quad = \|v(0)\|_W^2 + \sum_{n=1}^{N} 2\|v(t_n)\|_W^2 - 2\left\langle v(t_{n-1}^+), v(t_n)\right\rangle \\
&\quad \ge \sum_{n=1}^{N} \|v(t_n)\|_W^2 - 2\langle P_n v(t_{n-1}), v(t_n)\rangle + \|v(t_{n-1})\|_W^2 \\
&\quad = \sum_{n=1}^{N} \|v(t_n) - v(t_{n-1})\|_W^2 \ge 0.
\end{aligned}$$

□

We are in the situation described in Section 1.3. In order to derive an abstract error estimate, we need to bound the consistency error. We recall that

$$\mathcal{S}^{0,-1}(\mathcal{P}, V) = \{v \in L^2(V), v|_{I_n} \in V, n = 1, \dots, N\}$$

is the space of piecewise constants with values in V. We observe that, thanks to (4.11), we have, for every $\varphi \in M_2$,

$$b_M(u,\varphi) - \ell(\varphi) = b_M(u,\varphi) - b(u,\varphi) = \int_0^T \left\langle A(\widehat{\Pi}u - u), \varphi \right\rangle \\ \leq \sqrt{5} C_a \inf_{z \in \mathcal{S}^{0,-1}(\mathcal{P},V)} \|u - z\|_{L^2(V)} \|\varphi\|_{L^2(V)} . \tag{6.3}$$

Proposition 6.2. *The Galerkin solution U_M of method* (6.1) *with b_M as in* (6.2) *satisfies the following estimate*

$$\left(\|u' - U_M'\|^2_{L^2(V^*)} + \left\| u - \widehat{\Pi} U_M \right\|^2_{L^2(V)} + \|u(0) - U_M(0)\|^2_W \right)^{1/2} \\ \leq \sqrt{2}\kappa_\sigma \inf_{v \in M_1} \left(\|u' - v'\|^2_{L^2(V^*)} + \left\| \widehat{\Pi} u - \widehat{\Pi} v \right\|^2_{L^2(V)} + \|u(0) - v(0)\|^2_W \right)^{1/2} \\ + (\sqrt{2} k_\sigma + \sqrt{10}) \inf_{z \in \mathcal{S}^{0,-1}(\mathcal{P},V)} \|u - z\|_{L^2(V)} , \tag{6.4}$$

where we recall $\kappa_\sigma = 2\sqrt{2}\max\{\nu_a^{-1}, C_a^2\sigma, \nu_a^{-1}C_a^2\sigma\}$, while we set $k_\sigma := 2\sqrt{5}C_a \max\{\nu_a^{-1}, C_a\sigma, \nu_a^{-1}C_a\sigma\}$.

Proof. From Proposition 6.1, the results in Section 1.3 and (6.3) we get

$$\left(\|u' - U_M'\|^2_{L^2(V^*)} + \left\| \widehat{\Pi} u - \widehat{\Pi} U_M \right\|^2_{L^2(V)} + \|u(0) - U_M(0)\|^2_W \right)^{1/2} \\ \leq \kappa_\sigma \inf_{v \in M_1} \left(\|u' - v'\|^2_{L^2(V^*)} + \left\| \widehat{\Pi} u - \widehat{\Pi} v \right\|^2_{L^2(V)} + \|u(0) - v(0)\|^2_W \right)^{1/2} \\ + k_\sigma \inf_{z \in \mathcal{S}^{0,-1}(\mathcal{P},V)} \|u - z\|_{L^2(V)} .$$

Combining with (4.11), we get (6.4). □

The first infimum on the right-hand side of (6.4) is on functions v that belong to M_1. Therefore $v|_{I_n} \in \mathbb{V}_n$ and $v|_{I_{n+1}} \in \mathbb{V}_{n+1}$ are not independent but linked by $v(t_n^+) = P_{n+1}v(t_n)$. Moreover the norm involves both the function itself and its time derivative. We aim at an error estimate that involves the best errors for u and u' in $\mathcal{S}^{0,-1}(\mathcal{P},\mathbb{V})$. To this end we insert a particular choice of $v = \widehat{\mathcal{I}}u \in M_1$ in the first infimum in right-hand side of (6.4). We imitate the structure of the interpolation operator of Section 5.1.2 and define $\widehat{\mathcal{I}} : X_1 + M_1 \to M_1$ as

$$\widehat{\mathcal{I}}u := \Xi u + Z, \tag{6.5}$$

where $\Xi u(0) := P_0 u(0)$ and, for $n = 1, \dots, N$,

$$\forall t \in I_n, \qquad (\Xi u)(t) := (\Xi_n u)(t) := P_n \left(\frac{\Pi^{n-1} u - \Pi^n u}{\tau_n} (t_n - t) + \Pi^n u \right), \tag{6.6}$$

and Z is such that $Z(0) = 0$ and $Z|_{I_n} := Z_n \in \mathbb{P}^1(I_n, \mathbb{V}_n)$ satisfies

$$Z_n' + \mathcal{A}_n \Pi^n Z_n = 0 \qquad \text{and} \tag{6.7a}$$

$$Z_n(t_{n-1}^+) = P_n P_{n-1} \Pi^{n-1} u + P_n Z(t_{n-1}) - P_n \Pi^{n-1} u =: Z_n^+ \in \mathbb{V}_n. \tag{6.7b}$$

We investigate the properties of $\widehat{\mathcal{I}}$ in the following

Proposition 6.3 (Properties of $\widehat{\mathcal{I}}$)**.** *The interpolation operator $\widehat{\mathcal{I}}$ defined in* (6.5) *with Ξ as in* (6.6) *and Z as in* (6.7) *is a linear projection onto M_1, and it is stable with respect to $\|\cdot\|_{1,\mathcal{P}}$ with*

$$\left\| \widehat{\mathcal{I}} u \right\|_{1,\mathcal{P}} \leq C(\nu_a, C_a, \sigma, \mu_{\mathcal{P}}) \, \|u\|_{1,\mathcal{P}} \, .$$

Proof. Thanks to (6.7b), $\widehat{\mathcal{I}} u \in M_1$ for every $u \in X_1 + M_1$. Linearity follows from linearity of P_n and of the equation in (6.7a). Invariance over M_1 is due to the fact that, for every $u \in M_1$, $\Xi_n u = u|_{I_n}$ and $Z_n^+ = 0$, $n = 1, \dots, N-1$. Concerning stability, we proceed as in the proof of Proposition 5.3. We exploit the fact that b_M satisfies the inf-sup condition on $M_1 \times M_2$, and the definition of Z_n. We get that

$$c_M \left\| \widehat{\mathcal{I}} u \right\|_{1,\mathcal{P}} \leq C_M \, \|\Xi u\|_{1,\mathcal{P}} \, . \tag{6.8}$$

Moreover, thanks to the V-stability of P_n we have

$$\|\Xi u\|_{1,\mathcal{P}} \leq \sqrt{3} \sigma \sqrt{4\mu_{\mathcal{P}} + 5} \, \|u\|_{1,\mathcal{P}} \, . \tag{6.9}$$

In fact, for $n = 1, \dots, N$,

$$\|\Pi^n \Xi_n u\|_V = \|P_n \Pi^n u\|_V \leq \sigma \, \|\Pi^n u\|_V \, ,$$

while for $\|\Xi_n u'\|_{V^*}^2$ we have for $n = 1$

$$\begin{aligned} \|\Xi_1 u'\|_{V^*}^2 &\leq \sigma^2 \left\| \frac{\Pi^1 u - u(0)}{\tau_1} \right\|_{V^*}^2 \\ &\leq 2\sigma^2 \left(\left\| \frac{1}{\tau_1} \int_{I_1} u' \right\|_{V^*}^2 + \left\| \frac{\Pi^1 u - u(t_1)}{\tau_1} \right\|_{V^*}^2 \right) \\ &\leq \frac{2\pi^2 + 32}{\pi^2} \frac{\sigma^2}{\tau_1} \, \|u'\|_{L^2(I_n, V^*)}^2 \, , \end{aligned}$$

and for $n = 2, \dots, N$

$$\begin{aligned}\|\Xi_n u'\|_{V^*}^2 &\le \sigma^2 \left\| \frac{\Pi^n u - \Pi^{n-1} u}{\tau_n} \right\|_{V^*}^2 \\ &\le 3\sigma^2 \left(\left\| \frac{1}{\tau_n} \int_{I_n} u' \right\|_{V^*}^2 + \left\| \frac{\Pi^n u - u(t_n)}{\tau_n} \right\|_{V^*}^2 + \left\| \frac{\Pi^{n-1} u - u(t_{n-1})}{\tau_n} \right\|_{V^*}^2 \right) \\ &\le 3\sigma^2 \left(\frac{\pi^2 + 16}{\pi^2 \tau_n} \|u'\|_{L^2(I_n, V^*)}^2 + \frac{16}{\pi^2} \frac{\tau_{n-1}}{\tau_n^2} \|u'\|_{L^2(I_{n-1}, V^*)}^2 \right).\end{aligned}$$

Combining (6.8)–(6.9) gives the assertion. □

Thanks to Proposition (6.3), the choice of $\widehat{\mathcal{I}}$ ensures that

$$\left\| u - \widehat{\mathcal{I}} u \right\|_{1,\mathcal{P}} \approx \inf_{v \in M_1} \|u - v\|_{1,\mathcal{P}},$$

with hidden constants depending on ν_a, C_a, σ and $\mu_{\mathcal{P}}$. In order to further estimate $\|u - \mathcal{I} u\|_{1,\mathcal{P}}$, we split it as

$$\left\| u - \widehat{\mathcal{I}} u \right\|_{1,\mathcal{P}} \le \|u - \Xi u\|_{1,\mathcal{P}} + \|Z\|_{1,\mathcal{P}}, \tag{6.10}$$

and we bound the two terms on the right-hand side separately.

Proposition 6.4. *The operator* Ξ *defined in* (6.6) *satisfies the following bound:*

$$\begin{aligned}\|u - \Xi u\|_{1,\mathcal{P}}^2 \le 3\sigma^2 \frac{\pi^2 + 16 + 16\pi^2}{\pi^2} &\inf_{v \in \mathcal{S}^{0,-1}(\mathcal{P}, \mathbb{V})} \|u' - v\|_{L^2(V^*)}^2 \\ &+ 4\sigma^2 \inf_{w \in \mathcal{S}^{L^2}(\mathcal{P}, \mathbb{V})} \|u - w\|_{L^2(V)}^2 + \inf_{v_0 \in \mathbb{V}_0} \|u(0) - v_0\|_W^2,\end{aligned}$$

where $\mathcal{S}^{L^2}(\mathcal{P}, \mathbb{V}) := \{z \in L^2(V), z|_{I_n} \in L^2(I_n, \mathbb{V}_n), n = 1, \dots, N\}$.

Proof. We notice first that $\|u(0) - \Xi u(0)\|_W = \|u(0) - P_0 u(0)\|_W$ is the best error in $\mathbb{V}_0$ in the W-norm. Regarding $\left\| \widehat{\Pi}(u - \Xi u) \right\|_{L^2(V)}^2$ we have, for $n = 1, \dots, N$,

$$\left\| \widehat{\Pi}(u - \Xi u) \right\|_{L^2(I_n, V)}^2 = \int_{I_n} \|\Pi^n u - P_n \Pi^n u\|_V^2 \le 4 \|u - P_n u\|_{L^2(I_n, V)}^2.$$

By V-stability of P_n we get

$$\left\| \widehat{\Pi}(u - \Xi u) \right\|_{L^2(V)}^2 \le 4\sigma^2 \inf_{w \in \mathcal{S}^{L^2}(\mathcal{P}, \mathbb{V})} \|u - w\|_{L^2(V)}^2.$$

Concerning $\|u' - (\Xi u)'\|^2_{L^2(V^*)}$, we insert $P_n\left(\frac{1}{\tau_n}\int_{I_n} u'\right)$ for $n = 2, \dots, N$:

$$\begin{aligned} u' - \frac{P_n\Pi^n u - P_n\Pi^{n-1}u}{\tau_n} \\ = u' - P_n\left(\frac{1}{\tau_n}\int_{I_n} u'\right) + \frac{P_n u(t_n) - P_n\Pi^n u}{\tau_n} \\ + \frac{P_n\Pi^{n-1}u - P_n u(t_{n-1})}{\tau_n}. \end{aligned} \tag{6.11}$$

We bound separately

$$\begin{aligned} \left\|\frac{P_n u(t_n) - P_n\Pi^n u}{\tau_n}\right\|^2_{L^2(I_n,V^*)} &= \frac{1}{\tau_n}\|P_n u(t_n) - P_n\Pi^n u\|^2_{V^*} \\ &\le \frac{16\sigma^2}{\pi^2}\inf_{c\in V^*}\|u' - c\|^2_{L^2(I_n,V^*)}, \end{aligned} \tag{6.12}$$

and

$$\begin{aligned} \left\|\frac{P_n u(t_{n-1}) - P_n\Pi^{n-1} u}{\tau_n}\right\|^2_{L^2(I_n,V^*)} &= \frac{1}{\tau_n}\left\|P_n u(t_{n-1}) - P_n\Pi^{n-1}u\right\|^2_{V^*} \\ &\le \frac{16\sigma^2\mu_{\mathcal{P}}}{\pi^2}\inf_{c\in V^*}\|u' - c\|^2_{L^2(I_{n-1},V^*)}. \end{aligned} \tag{6.13}$$

Combining (6.11)–(6.13) we get, for $n = 2, \dots, N$,

$$\begin{aligned} \|u' - (\Xi u)'\|^2_{L^2(I_n,V^*)} \le 3\sigma^2\Bigg(&\frac{\pi^2+16}{\pi^2}\inf_{c\in\mathbb{V}_n}\|u' - c\|^2_{L^2(I_n,V^*)} \\ &\frac{16\mu_{\mathcal{P}}}{\pi^2}\inf_{c\in\mathbb{V}_{n-1}}\|u' - c\|^2_{L^2(I_n,V^*)}\Bigg). \end{aligned}$$

For $n = 1$, we have

$$\begin{aligned} u' - \frac{P_1\Pi^1 u - P_1 u(0)}{\tau_1} \\ = u' - P_1\left(\frac{1}{\tau_1}\int_{I_1} u'\right) + \frac{P_1 u(t_1) - P_1\Pi^1 u}{\tau_1} \end{aligned}$$

and thus

$$\|u' - (\Xi u)'\|^2_{L^2(I_1,V^*)} \le \frac{2\pi^2+32}{\pi^2}\sigma^2\inf_{c\in\mathbb{V}_1}\|u' - c\|^2_{L^2(I_1,V^*)}.$$

The thesis follows by summing $\|u' - (\Xi u)'\|^2_{L^2(I_n,V^*)}$ over n. □

Proposition 6.5. *The correction Z defined in* (6.7)*, satisfies*

$$\|Z\|_{1,\mathcal{P}}^2 \approx \sum_{n=1}^{N} \left\|Z_n(t_{n-1}^+)\right\|_W^2 - \|Z_n(t_n)\|_W^2 - \left\|Z_n(t_n) - Z_n(t_{n-1}^+)\right\|_W^2, \quad (6.14a)$$

where the hidden constants are given by C_z in the $\lesssim$-direction, and c_z in the $\gtrsim$-direction, being C_z and c_z as in (5.9b)*. Moreover,*

$$\sum_{n=1}^{N} \left\|Z_n(t_{n-1}^+)\right\|_W^2 - \|Z_n(t_n)\|_W^2 \leq \|P_1(I-P_0)u(t_0)\|_W^2 + \sum_{n=1}^{N-1} \left\|P_n^+(I-P_n)\Pi^n u\right\|_W^2, \quad (6.14b)$$

where P_n^+ denotes the W-projection onto $\mathbb{V}_n \oplus \mathbb{V}_{n+1}$, $n = 1, \dots, N-1$.

Proof. The proof follows the same lines as the proof of Proposition 5.4. The only difference is that, when testing (6.7a) with $Z_n + \mathcal{A}_n^{-*} Z_n'$ and integrating over I_n, we get

$$0 = 2\|Z_n(t_n)\|_W^2 - 2\left\langle Z_n(t_{n-1}^+), Z_n(t_n)\right\rangle_W + \int_{I_n} \langle \mathcal{A}_n Z_n, Z_n\rangle + \left\langle Z_n', \mathcal{A}_n^{-*} Z_n'\right\rangle,$$

and

$$\sum_{n=1}^{N} 2\left\langle Z_n(t_{n-1}^+), Z_n(t_n)\right\rangle_W - 2\|Z_n(t_n)\|_W^2 = \sum_{n=1}^{N} \left\|Z_n(t_{n-1}^+)\right\|_W^2 - \|Z_n(t_n)\|_W^2 - \left\|Z_n(t_n) - Z_n(t_{n-1}^+)\right\|_W^2. \qquad \square$$

Combining Propositions 6.2, 6.4 and 6.5 we get

Theorem 6.6. *The Galerkin solution U_M of method* (6.1) *with b_M as in* (6.2) *satisfies the following estimate*

$$\begin{aligned}
\|u' - U_M'\|_{L^2(V^*)}^2 + \left\|u - \widehat{\Pi} U_M\right\|_{L^2(V)}^2 + \|u(0) - U_M(0)\|_W^2 \\
\leq C_0 \inf_{v_0 \in \mathbb{V}_0} \|u(0) - v_0\|_W^2 + C_1 \inf_{v \in \mathcal{S}^{0,-1}(\mathcal{P},\mathbb{V})} \|u' - v\|_{L^2(V^*)}^2 \\
+ C_2 \inf_{w \in \mathcal{S}^{L^2}(\mathcal{P},\mathbb{V})} \|u - w\|_{L^2(V)}^2 + C_3 \inf_{z \in \mathcal{S}^{0,-1}(\mathcal{P},V)} \|u - z\|_{L^2(V)} \\
+ C_4 \left(\|P_1(I-P_0)u(t_0)\|_W^2 + \sum_{n=1}^{N-1} \left\|P_n^+(I-P_n)\Pi^n u\right\|_W^2 \right),
\end{aligned}$$

where $C_0 := 6\kappa_\sigma^2$, $C_1 := 18\kappa_\sigma^2\sigma^2(\pi^2 + 16 + 16\mu_{\mathcal{P}})/\pi^2$, $C_2 := 24\kappa_\sigma^2\sigma^2$, $C_3 := 3(\sqrt{2}k_\sigma + \sqrt{10})^2$, $C_4 := 6\nu_a^{-1}\kappa_\sigma \max\{1, C_a^2\sigma^2\}$.

We notice that there are two best errors involving u and the $L^2(V)$-norm. One regards only the time discretization, while the other just the spatial discretization. The best error for u' in the $L^2(V^*)$-norm, instead, couples the spatial and the time discretizations.

Assuming additional regularity

As in Section 4.1, we assume that the exact solution u belongs to $C^0(V)$, and provide a bound with constants independent of $\mu_{\mathcal{P}}$. For $v \in C^0(V)$ we set

$$\widehat{\Pi} v|_{I_n} := v(t_n). \tag{6.15}$$

Proposition 6.1 is still valid, but the consistency error can only be bounded, as in (4.20), by

$$\sup_{\varphi \in M_2} \frac{b_M(u,\varphi) - \ell(\varphi)}{\|\varphi\|_2} \le C_a \left\| u - \widehat{\Pi} u \right\|_{L^2(V)}.$$

The following proposition is the counterpart of Proposition 6.2.

Proposition 6.7. *Assume $u \in C^0(V)$. The Galerkin solution U_M of method (6.1) with b_M as in (6.2), and $\widehat{\Pi}$ as in (6.15) satisfies*

$$\begin{aligned}
&\left(\|u' - U_M'\|^2_{L^2(V^*)} + \left\| u - \widehat{\Pi} U_M \right\|^2_{L^2(V)} + \|u(0) - U_M(0)\|^2_W \right)^{1/2} \\
&\quad \le \sqrt{2}\kappa_\sigma \inf_{v \in M_1} \left(\|u' - v'\|^2_{L^2(V^*)} + \left\| \widehat{\Pi} u - \widehat{\Pi} v \right\|^2_{L^2(V)} + \|u(0) - v(0)\|^2_W \right)^{1/2} \\
&\qquad + \left(\frac{k_\sigma}{\sqrt{5}} + 1 \right) \left\| u - \widehat{\Pi} u \right\|_{L^2(V)},
\end{aligned} \tag{6.16}$$

where we recall $\kappa_\sigma = 2\sqrt{2}\max\{\nu_a^{-1}, C_a^2\sigma, \nu_a^{-1}C_a^2\sigma\}$, and $k_\sigma = 2\sqrt{5}C_a \cdot$ $\cdot \max\{\nu_a^{-1}, C_a\sigma, \nu_a^{-1}C_a\sigma\}$.

In place of Ξ defined in (6.6) we can set, for every $n = 1, \dots, N$,

$$\forall t \in I_n, \qquad (\Xi u)(t) := P_n \left(\frac{u(t_n) - u(t_{n-1})}{\tau_n}(t_n - t) + u(t_n) \right), \tag{6.17}$$

while $\Xi u(0)$ is still defined as $P_0 u(0)$.

Correspondingly, we change the definition of Z_n^+ in (6.7), setting

$$Z_n^+ := P_n P_{n-1} u(t_{n-1}) + P_n Z(t_{n-1}) - P_n u(t_{n-1}).$$

With this choices $\widehat{\mathcal{I}} u = \Xi u + Z$ belongs to M_1. In place of Theorem 6.6 we get the following result.

Theorem 6.8. *Assume $u \in C^0(V)$. The Galerkin solution U_M of method (6.1) with b_M as in (6.2) and $\widehat{\Pi}$ as in (6.15) satisfies*

$$\begin{aligned}
&\|u' - U_M'\|^2_{L^2(V^*)} + \left\|u - \widehat{\Pi} U_M\right\|^2_{L^2(V)} + \|u(0) - U_M(0)\|^2_W \\
&\quad \le \widetilde{C}_1 \inf_{v_0 \in \mathbb{V}_0} \|u(0) - v_0\|^2_W + \widetilde{C}_1 \sum_{n=1}^{N} \int_{I_n} \left\| u' - P_n \frac{1}{\tau_n} \int_{I_n} u' \right\|^2_{V^*} \\
&\qquad + \widetilde{C}_1 \sum_{n=1}^{N} \tau_n \|u(t_n) - P_n u(t_n)\|^2_V + \widetilde{C}_2 \left\| u - \widehat{\Pi} u \right\|^2_{L^2(V)} \\
&\qquad + \widetilde{C}_3 \left(\|P_1(I - P_0)u(t_0)\|^2_W + \sum_{n=1}^{N-1} \left\| P_n^+ (I - P_n) u(t_n) \right\|^2_W \right),
\end{aligned}$$

where $\widetilde{C}_1 := 6\kappa_\sigma^2$, $\widetilde{C}_2 := 3(k_\sigma/\sqrt{5} + 1)^2$, $\widetilde{C}_3 := 6\nu_a^{-1}\kappa_\sigma \max\{1, C_a^2\sigma^2\}$.

Proof. We insert $v = \widehat{\mathcal{I}} u$ in the infimum on the right-hand side of (6.16). As in (6.10) we use triangle inequality and split

$$\left\| u - \widehat{\mathcal{I}} u \right\|_{1,\mathcal{P}} \le \|u - \Xi u\|_{1,\mathcal{P}} + \|Z\|_{1,\mathcal{P}} .$$

Concerning $\|Z\|_{1,\mathcal{P}}$, we notice that it can be bounded as in Proposition 6.5 with $u(t_n)$ in place of $\Pi^n u$. Concerning $\|u - \Xi u\|_{1,\mathcal{P}}$ we observe that

$$(\Xi u|_{I_n})' = P_n \frac{u(t_n) - u(t_{n-1})}{\tau_n} = P_n \frac{1}{\tau_n} \int_{I_n} u',$$

and that $(\Xi u)(t_n) = P_n u(t_n)$, $n = 1, \dots, N$. □

6.2 Natural formulation

Assume that $V \subset W \subset V^*$, a and A are as in Section 2.1, while, as in Section 2.3, $X_1 = L^2(V)$, $X_2 = \{\varphi \in H^1(V, V^*), \varphi(T) = 0\}$ with norms

$$\|v\|_1^2 = \int_0^T \|v\|_V^2 , \qquad \|\varphi\|_2^2 = \int_0^T \|\varphi'\|_{V^*}^2 + \|\varphi\|_V^2 ,$$

and the bilinear form $b : X_1 \times X_2 \to \mathbb{R}$ is given by

$$b(v, \varphi) = \int_0^T - \langle \varphi', v \rangle + \langle Av, \varphi \rangle .$$

Moreover, as in Sections 4.2.1 and 5.2, $\ell \in X_2^*$ is of the following form

$$\ell(\varphi) = \langle g_0, \varphi(0)\rangle + \sum_{j=1}^{\jmath-1} \left\langle g_j, \varphi(\tilde{t}_j)\right\rangle + \int_0^T \langle f, \varphi\rangle \,,$$

with $f \in L^2(V^*)$, $(g_j)_{j=0}^{\jmath-1} \subset W$ and $0 = \tilde{t}_0 < \tilde{t}_1 < \ldots < \tilde{t}_\jmath = T$. Given $N \in \mathbb{N}$ the partition $\mathcal{P}$

$$0 = t_0 < t_1 < \ldots < t_N = T$$

of $I = (0, T)$ into N subintervals $I_n = [t_{n-1}, t_n)$, satisfies

$$\mu_{\mathcal{P}} = \sup_n \frac{\tau_{n-1}}{\tau_n} < \infty,$$

and it is subordinate to $(\tilde{t}_j)_{j=1}^{\jmath}$. In addition, as in Section 5.2, we consider a sequence of finite-dimensional subspaces $\{\mathbb{V}_n\}_{n=1}^N \subset V$. For $n = 1, \ldots, N$, the approximation in the n-th interval I_n occurs in $\mathbb{V}_n$. The W-orthogonal projection on $\mathbb{V}_n$ is denoted by P_n, and $\{P_n\}_{n=1}^N$ is assumed to be uniformly stable in V, with

$$\sigma = \sup_{n=1,\ldots,N} \sup_{v \in V} \frac{\|P_n v\|_V}{\|v\|_V}.$$

Finally, as in Section 6.1, the operator $\mathcal{A}_n : \mathbb{V}_n \to \mathbb{V}_n^*$, is defined by

$$\mathcal{A}_n v = \frac{1}{\tau_n} \int_{I_n} \mathbb{A}_n(t) v \, \mathrm{d}t.$$

We consider the spaces

$$M_1 := \{v \in L^2(V), v|_{I_n} \in \mathbb{V}_n, n = 1, \ldots, N\}, \tag{6.18a}$$

$$M_2 := \{\varphi \in L^2(V), \varphi|_{I_n} \in \mathbb{P}^1(I_n, \mathbb{V}_n), \varphi(T) = 0, \quad \varphi(t_n^-) = P_n \varphi(t_n), n = 1, \ldots, N-1\}. \tag{6.18b}$$

We notice that $M_1 \subset X_1$, while in general $M_2 \nsubseteq X_2$. They are finite-dimensional spaces with $\dim(M_1) = \dim(M_2)$. We remark that the constraints $\varphi(t_n^-) = P_n \varphi(t_n)$ can be seen as a discrete replacement of the embedding $X_2 \subset C^0(W)$. The space M_1 inherits the $\|\cdot\|_1$-norm, while M_2 is endowed with

$$\|\varphi\|_{2,\mathcal{P}}^2 := \sum_{n=1}^N \int_{I_n} \|\varphi'\|_{V^*}^2 + \left\|\breve{\Pi}\varphi\right\|_V^2,$$

where $\breve{\Pi}$ is defined for $n = 1, \ldots, N$ as in Section 4.2:

$$\breve{\Pi} v|_{I_n} = \Pi_n v = \int_{I_n} v \psi_n, \quad \text{with} \quad \psi_n(t) = \frac{-6(t - t_{n-1})}{\tau_n^2} + \frac{4}{\tau_n}.$$

We replace b with

$$b_M(v,\varphi) := \sum_{n=1}^{N} \int_{I_n} - \langle v_n, \varphi' \rangle + \left\langle A v_n, \breve{\Pi}\varphi \right\rangle, \tag{6.19}$$

so that b_M is well-defined also on $M_1 \times M_2$, and ℓ with

$$\ell_M(\varphi) := \langle g_0, \varphi(0) \rangle + \sum_{j=1}^{\jmath-1} \left\langle g_j, \varphi(\tilde{t}_j) \right\rangle + \int_0^T \left\langle f, \breve{\Pi}\varphi \right\rangle. \tag{6.20}$$

We prove that the discrete problem is well-posed.

Proposition 6.9. *The bilinear form* (6.19) *is continuous and satisfies the inf-sup condition* (1.9) *on* $M_1 \times M_2$ *with*

$$C_M \le \sqrt{2} \max\{1, C_a\}, \qquad c_M \ge \frac{\nu_a}{\sqrt{2}} \min\{1, C_a^{-1}\sigma^{-1}\}.$$

Moreover, it is also continuous on $(X_1, \|\cdot\|_{1,\mathcal{P}}) \times (M_2, \|\cdot\|_2)$ *and* $C_{X_1 \times M_2}$ *satisfies the same bound as* C_M*.*

Proof. The proof mimics those of Propositions 4.7 and 5.8. We derive the lower bound for c_M and the non-degeneracy condition (1.9b) follows since $\dim(M_1) = \dim(M_2) < \infty$. We choose, for every $v \in M_1$,

$$v_n = \Pi_n \varphi - \mathcal{A}_n^{-1}(\varphi'|_{I_n}), \quad n = 1, \dots, N.$$

In addition, we exploit $\|\varphi'\|_{\mathbb{V}_n^*} \ge \sigma^{-1} \|\varphi'\|_{V^*}$ and

$$\begin{aligned}
\sum_{n=1}^{N} \int_{I_n} -2 \langle \varphi', \varphi(t_{n-1}) \rangle &= \sum_{n=1}^{N} 2 \|\varphi(t_{n-1})\|_W^2 - 2 \left\langle \varphi(t_n^-), \varphi(t_{n-1}) \right\rangle \\
&\ge \|\varphi(0)\|_W^2 + \sum_{n=1}^{N} \|\varphi(t_n)\|_W^2 - 2 \langle P_n \varphi(t_n), \varphi(t_{n-1}) \rangle + \|\varphi(t_{n-1})\|_W^2 \\
&= \|\varphi(0)\|_W^2 + \sum_{n=1}^{N} \|\varphi(t_n) - \varphi(t_{n-1})\|_W^2 \ge 0.
\end{aligned} \tag{6.21}$$

□

In order to derive an abstract error estimate, we need to bound the consistency error. We write, for every $\varphi \in M_2$

$$\begin{aligned} b_M(u,\varphi) - \ell_M(\varphi) \\ &= \sum_{n=1}^{N} \int_{I_n} - \langle \varphi', u\rangle + \Big\langle Au, \breve{\Pi}\varphi \Big\rangle - \sum_{j=0}^{\jmath-1} \big\langle g_j, \varphi(\tilde{t}_j)\big\rangle - \int_0^T \Big\langle f, \breve{\Pi}\varphi \Big\rangle \\ &= \sum_{n=1}^{N} - \big\langle \varphi(t_n^-), u(t_n^-)\big\rangle + \langle \varphi(t_{n-1}), u(t_{n-1})\rangle \\ &\qquad + \int_{I_n} \langle u', \varphi\rangle + \Big\langle Au - f, \breve{\Pi}\varphi \Big\rangle - \sum_{j=0}^{\jmath-1} \big\langle g_j, \varphi(\tilde{t}_j)\big\rangle \\ &= \int_0^T \Big\langle Au - f, \breve{\Pi}\varphi - \varphi \Big\rangle + \sum_{n=1}^{N-1} \big\langle \varphi(t_n), u(t_n^-)\big\rangle - \big\langle P_n\varphi(t_n), u(t_n^-)\big\rangle . \end{aligned} \tag{6.22}$$

We split the right-hand side into two contributions and we mimic the bounds in Sections 4.2.1 and 5.2.1. We obtain the following

Theorem 6.10. *The Galerkin solution U_M of method* (6.18) *with b_M as in* (6.19) *and ℓ_M as in* (6.20) *satisfies the following estimate*

$$\begin{aligned} \|u - U_M\|_1 \le 2\nu_a^{-1} \max\{1, C_a\sigma, C_a^2\sigma\} \inf_{v\in M_1} \|u - v\|_1 \\ &+ \sqrt{2}\nu_a^{-1} \max\{1, C_a\sigma\} \sup_{\varphi\in M_2} \frac{\int_0^T \Big\langle Au - f, \breve{\Pi}\varphi - \varphi \Big\rangle}{\|\varphi\|_{2,\mathcal{P}}} \\ &+ 2\nu_a^{-1} \max\{1, C_a\sigma\} \sqrt{\mu_{\mathcal{P}} + 2} \left(\sum_{n=1}^{N-1} \big\| (I - P_n) P_n^+ u(t_n^-) \big\|_W^2 \right)^{1/2} . \end{aligned}$$

Proof. Combine (6.22) with (5.22) to obtain, for every $\varphi \in M_2$,

$$\begin{aligned} b_M(u,\varphi) - \ell_M(\varphi) \\ &\le \int_0^T \Big\langle Au - f, \varphi - \breve{\Pi}\varphi \Big\rangle \\ &\quad + \left(\sum_{n=1}^{N-1} \|\varphi(t_n) - P_n\varphi(t_n)\|_W^2 \right)^{1/2} \left(\sum_{n=1}^{N-1} \big\| (I - P_n) P_n^+ u(t_n^-) \big\|_W^2 \right)^{1/2} . \end{aligned}$$

Moreover, we recall from (5.20) that

$$\sum_{n=1}^{N-1} \|\varphi(t_n) - P_n\varphi(t_n)\|_W^2 \le \int_{I_n} \|\varphi'\|_{V^*}^2 + \|\varphi\|_V^2 \le 2(\mu_{\mathcal{P}} + 2) \|\varphi\|_{2,\mathcal{P}}^2 .$$

Therefore

$$\sup_{\varphi\in M_2}\frac{b_M(u,\varphi)-\ell_M(\varphi)}{\|\varphi\|_{2,\mathcal{P}}} \leq \sup_{\varphi\in M_2}\frac{\int_0^T\left\langle Au-f,\varphi-\breve{\Pi}\varphi\right\rangle}{\|\varphi\|_{2,\mathcal{P}}} + \sqrt{2}\sqrt{\mu_{\mathcal{P}}+2}\left(\sum_{n=1}^{N-1}\left\|(I-P_n)P_n^+u(t_n^-)\right\|_W^2\right)^{1/2}.$$

The thesis follows by the results in Section 1.3. □

As alternative to (6.22) we can also write, for every $\varphi\in M_2$

$$\begin{aligned}
&b_M(u,\varphi)-\ell_M(\varphi)\\
&\quad=\sum_{n=1}^{N}\int_{I_n}-\langle\varphi',u\rangle+\left\langle Au,\breve{\Pi}\varphi\right\rangle-\sum_{j=0}^{\jmath-1}\left\langle g_j,\varphi(\tilde{t}_j)\right\rangle-\int_0^T\left\langle f,\breve{\Pi}\varphi\right\rangle\\
&\quad=\sum_{n=1}^{N}\int_{I_n}-\langle\varphi',u\rangle-\left\langle u',\breve{\Pi}\varphi\right\rangle-\sum_{j=0}^{\jmath-1}\left\langle g_j,\varphi(\tilde{t}_j)\right\rangle\\
&\quad=\sum_{n=1}^{N}\left\langle u(t_{n-1}^+),\varphi(t_{n-1})\right\rangle-\left\langle u(t_n^-),\varphi(t_{n-1})\right\rangle+\int_{I_n}-\langle\varphi',u\rangle\\
&\qquad\qquad-\langle u(0),\varphi(0)\rangle-\sum_{j=1}^{\jmath-1}\left\langle u(\tilde{t}_j^+)-u(\tilde{t}_j^-),\varphi(\tilde{t}_j)\right\rangle\\
&\quad=\sum_{n=1}^{N}\left\langle u(t_n^-),\varphi(t_n)-\varphi(t_{n-1})\right\rangle+\int_{I_n}-\langle\varphi',u\rangle\,. \qquad (6.23)
\end{aligned}$$

Furthermore, for every $\varphi\in M_2$ and for every $W_n\in\mathbb{V}_n$, we have

$$\begin{aligned}
\int_{I_n}-\langle\varphi',W_n\rangle&=\langle\varphi(t_{n-1}),W_n\rangle-\left\langle\varphi(t_n^-),W_n\right\rangle\\
&=\langle\varphi(t_{n-1}),W_n\rangle-\langle\varphi(t_n),W_n\rangle\,.
\end{aligned}$$

Choosing $W_n = P_n u(t_n^-)$, and adding to (6.23) for $n = 1, \dots, N$, we get

$$\begin{aligned} & b_M(u,\varphi) - \ell_M(\varphi) \\ & \quad = \sum_{n=1}^{N} \left\langle u(t_n^-) - P_n u(t_n^-), \varphi(t_n) - \varphi(t_{n-1}) \right\rangle + \int_{I_n} - \left\langle \varphi', u - P_n u(t_n^-) \right\rangle \\ & \quad \le \sum_{n=1}^{N-1} \left\| P_n^+ (I - P_n) u(t_n^-) \right\|_W \left\| \varphi(t_n) - \varphi(t_{n-1}) \right\|_W \\ & \qquad\qquad + \sum_{n=1}^{N} \left\| \varphi' \right\|_{L^2(I_n, V^*)} \left\| u - P_n u(t_n^-) \right\|_{L^2(I_n, V)} . \end{aligned}$$

We notice that, from (6.21), we have, for every $\varphi \in M_2$

$$\sum_{n=1}^{N} \left\| \varphi(t_n) - \varphi(t_{n-1}) \right\|_W^2 \le \sum_{n=1}^{N} \int_{I_n} -2 \left\langle \varphi', \varphi(t_{n-1}) \right\rangle \le \left\| \varphi \right\|_{2,\mathcal{P}}^2 .$$

Therefore we can bound the consistency error independently of $\mu_{\mathcal{P}}$:

$$\begin{aligned} & \sup_{\varphi \in M_2} \frac{b_M(u,\varphi) - \ell_M(\varphi)}{\left\| \varphi \right\|_{2,\mathcal{P}}} \\ & \quad \le \left(\sum_{n=1}^{N-1} \left\| P_n^+ (I - P_n) u(t_n^-) \right\|_W^2 \right)^{1/2} + \left(\sum_{n=1}^{N} \int_{I_n} \left\| u - P_n u(t_n^-) \right\|_V^2 \right)^{1/2} . \end{aligned}$$

Applying the results in Section 1.3 we obtain the following theorem.

Theorem 6.11. *The Galerkin solution U_M of method* (6.18) *with b_M as in* (6.19) *and ℓ_M as in* (6.20) *satisfies the following estimate*

$$\begin{aligned} \left\| u - U_M \right\|_1 \le{}& 2\nu_a^{-1} \max\{1, C_a\sigma, C_a^2\sigma\} \inf_{v \in M_1} \left\| u - v \right\|_1 \\ & + \sqrt{2}\nu_a^{-1} \max\{1, C_a\sigma\} \left(\sum_{n=1}^{N} \int_{I_n} \left\| u - P_n u(t_n^-) \right\|_V^2 \right)^{1/2} \\ & + \sqrt{2}\nu_a^{-1} \max\{1, C_a\sigma\} \left(\sum_{n=1}^{N-1} \left\| (I - P_n) P_n^+ u(t_n^-) \right\|_W^2 \right)^{1/2} . \end{aligned}$$

Chapter 7

A Priori Error Estimates for FEM

In this chapter we derive error estimates in the case the spatial discretization occurs by means of finite elements, exploiting the results in Chapters 3–6.

The chapter is divided into two parts. In the first part, we consider exact solutions with integer regularity, that is, they belong to Sobolev spaces of integer order. Notice that, in the standard formulation, the approximation of the time derivative involves the H^{-1}-norm in space. We therefore use a suitable interpolation operator, well-defined in L^2, which allows for duality arguments. In the second part, we consider more general exact solutions, that can have only fractional regularity. In this setting, the interpolation operator mentioned above cannot be used for functions with regularity less than L^2. We therefore define an interpolation operator that acts on H^{-1} and has values in the space of continuous and piecewise polynomial functions of degree one.

The structure of the two parts is similar. We start with fixing the notation and giving some auxiliary results. We then describe the interpolation operator and its approximation properties in particular in H^{-1}. Finally we derive the error estimates for both the standard and the natural formulation.

7.1 Notation and auxiliary results

Functional spaces

Let $\Omega \subset \mathbb{R}^d$ be a bounded polyhedral domain, with Lipschitz boundary $\partial\Omega$. We denote with $C^0(\Omega)$ the space of continuous functions over Ω and with $C^k(\Omega)$ the space of functions such that, for every multi-index α with $|\alpha| = k$,

$D^\alpha f \in C^0(\Omega)$. Moreover we set $C^\infty(\Omega) := \bigcap_{k\in\mathbb{N}} C^k(\Omega)$, and C_0^∞ indicates the subspace of $C^\infty(\Omega)$ of functions with compact support in Ω.

Given $p \in [1,\infty)$, we indicate with $L^p(\Omega)$ the space of functions whose absolute value, raised to the p-th power, has finite integral over Ω, and with $L^\infty(\Omega)$ the space of essentially bounded functions. With $L_0^p(\Omega)$ we indicate the subspace of $L^p(\Omega)$ of functions with mean value zero. Given a subdomain $\omega \subset \Omega$, we denote with $\|\cdot\|_{0,p;\omega}$ the L^p-norm, and with $\|\cdot\|_{0,\infty;\omega}$ the L^∞-norm.

Moreover, the Sobolev space $W^{m,p}(\Omega)$, $m \in \mathbb{N}$, $p \in [1,\infty]$, consists of all functions f in $L^p(\Omega)$, such that, for every multi-index α with order $|\alpha| = m$, $D^\alpha f$ exists in the weak sense and belongs to $L^p(\Omega)$. For $p = 2$ we write $H^m(\Omega) := W^{m,2}(\Omega)$. For $m = 0$ we set $H^0(\Omega) := L^2(\Omega)$. We denote by $|\cdot|_{m,p;\omega}$ and $\|\cdot\|_{m,p;\omega}$ the $W^{m,p}$-seminorm and the $W^{m,p}$-norm on $\omega \subset \Omega$, respectively. Furthermore, we denote with $H_0^1(\Omega)$ the subspace of $H^1(\Omega)$ of those functions that vanish on the boundary $\partial\Omega$, and $H^{-1}(\Omega)$ is the dual of $H_0^1(\Omega)$. For every $f \in H^{-1}(\Omega)$ the dual norm is defined as

$$\|f\|_{-1;\Omega} := \sup_{\varphi\in H_0^1(\Omega)} \frac{\langle f,\varphi\rangle}{|\varphi|_{1,2;\Omega}}.$$

Finite element spaces

Let $\mathcal{T}$ be a conforming simplicial mesh of Ω. We denote by

$$\sigma_{\mathcal{T}} := \max_{K\in\mathcal{T}} \frac{\operatorname{diam}(K)}{\rho_K}$$

the shape parameter of $\mathcal{T}$, where $\operatorname{diam}(K)$ and ρ_K indicate, respectively, the diameter of the element K and the maximum diameter of a ball inscribed in K.

We indicate with $\mathcal{V}$ the set of vertices of $\mathcal{T}$. A subscript K, Ω, etc. to $\mathcal{V}$ indicates that only the vertices contained in the index-set are considered. Similarly, $\mathcal{T}_\Omega$ indicates the set of elements contained in Ω, while $\mathcal{T}_{\partial\Omega}$ denotes the set of elements with at least a vertex on $\partial\Omega$. For every vertex $z \in \mathcal{V}$ and for every element $K \in \mathcal{T}$ we set

$$\omega_z := \bigcup_{\substack{K\in\mathcal{T}\\ K\ni z}} K, \quad \omega_K := \bigcup_{z\in\mathcal{V}_K} \omega_z, \quad \widetilde{\omega}_z := \bigcup_{K\subset\omega_z} \omega_K. \tag{7.1}$$

When we write $K \subset \omega_z$, we intend that $K \in \mathcal{T}$, even if not explicitly specified.

We denote by $\#\omega_K := \#\{\widetilde{K} \in \mathcal{T}, \widetilde{K} \subset \omega_K\}$ the number of simplices in the patch ω_K, and with

$$\nu_{\mathcal{T}} := \max_{K\in\mathcal{T}} \#\omega_K$$

the maximum number of simplices in a patch.

The space

$$S^{\ell,0}(\mathcal{T}) := \{v \in C^0(\Omega),\, v \in \mathbb{P}^\ell(K),\, \forall\, K \in \mathcal{T}\}$$

is the space of continuous piecewise polynomial functions on $\mathcal{T}$, while

$$S_0^{\ell,0}(\mathcal{T}) := S^{\ell,0}(\mathcal{T}) \cap H_0^1(\Omega)$$

is the subspace of $S^{\ell,0}(\mathcal{T})$ of those functions with zero boundary values. Furthermore, we denote by $\mathcal{N}$ the set of nodes of $S^{\ell,0}(\mathcal{T})$. Again, a subscript K, Ω, etc. to $\mathcal{N}$ indicates that only the nodes contained in the index-set are considered. We denote by $\{\phi_z\}_{z\in\mathcal{N}}$ the nodal basis, that is, for every $z \in \mathcal{N}$,

$$\phi_z \in S^{\ell,0}(\mathcal{T}) \text{ and for every } y \in \mathcal{N}, \quad \phi_z(y) = \delta_{yz}.$$

We recall that $\{\phi_z\}_{z\in\mathcal{N}}$ forms a partition of unity, that is,

$$\sum_{z\in\mathcal{N}} \phi_z = 1.$$

We denote by $\omega_z := \operatorname{supp}(\phi_z)$, the support of ϕ_z, $z \in \mathcal{N}$. Note that, if z is a vertex node, ω_z coincides with the star around z defined in (7.1). The local $L^2(K)$-dual basis functions $\{\psi_z^K\}_{z\in\mathcal{N}_K}$, $K \in \mathcal{T}$, have the crucial property that, for every y, $z \in \mathcal{N}_K$,

$$\int_K \psi_z^K \phi_y = \delta_{zy}.$$

Norms under affine transformations

Let the reference d-simplex be defined as $\hat{K} := \operatorname{convhull}\{\mathbf{0}, e_1, \ldots, e_d\}$, where $e_1, \ldots, e_d$ denotes the canonical basis in $\mathbb{R}^d$. Moreover let $\hat{h} := \operatorname{diam}(\hat{K})$ be the diameter of $\hat{K}$, and $\hat{\rho}$ be the maximum diameter of a ball inscribed in $\hat{K}$. For every $K \in \mathcal{T}$, there exists an affine transformation $F : \mathbb{R}^d \to \mathbb{R}^d$ such that $F(\hat{K}) = K$. We denote with B the non-singular matrix associated to F. We recall (see [15] p. 117-120) that, for every $v \in W^{m,p}(K)$, there hold

$$|v|_{m,p;K} \le C(m,p,d)\, \left\|B^{-1}\right\|^m |\det(B)|^{1/p}\, |v\circ F|_{m,p;\hat{K}}\,, \tag{7.2a}$$

$$|v\circ F|_{m,p;\hat{K}} \le C(m,p,d)\, \|B\|^m |\det(B)|^{-1/p}\, |v|_{m,p;K}\,, \tag{7.2b}$$

and that

$$\|B\| \le \frac{\operatorname{diam}(K)}{\hat{\rho}}, \qquad \left\|B^{-1}\right\| \le \frac{\operatorname{diam}(\hat{K})}{\rho_K}, \qquad \det(B) = \frac{|K|}{|\hat{K}|}. \tag{7.2c}$$

In particular, we consider the change of norms for the basis and dual basis functions. We denote by $\{\hat{\phi}_{\hat{z}}\}$ and $\{\hat{\psi}_{\hat{z}}\}$ respectively the basis and dual basis functions on $\hat{K}$. For every $K \in \mathcal{T}$, and for every $z \in \mathcal{N}_K$, there exists an affine transformation $F : \mathbb{R}^d \to \mathbb{R}^d$ with $F(\hat{K}) = K$, and $F(\hat{z}) = z$. There may be different choices for $\hat{z}$, which nevertheless lead to the same value of $\|\hat{\phi}_{\hat{z}}\|_{\hat{K}}$. However, $\|\nabla\hat{\phi}_{\hat{z}}\|_{\hat{K}}$ depends on the chosen node. For this reason, we take a $\hat{z}$ with minimal sum of the coordinates, so that $\|\nabla\hat{\phi}_{\hat{z}}\|_{\hat{K}}$ is unique. Since $\hat{\psi}_{\hat{z}} = (\det B)\psi_z^K \circ F$, we have

$$\|\phi_z\|_{0,2;K} = \frac{|K|^{1/2}}{|\hat{K}|^{1/2}}\|\hat{\phi}_{\hat{z}}\|_{0,2;\hat{K}}, \qquad \|\phi_z\|_{0,2;\omega_z} = \frac{|\omega_z|^{1/2}}{|\hat{K}|^{1/2}}\|\hat{\phi}_{\hat{z}}\|_{0,2;\hat{K}}, \tag{7.3}$$

$$\|\nabla\phi_z\|_{0,2;K} \leq \frac{\hat{h}|K|^{1/2}}{\rho_K|\hat{K}|^{1/2}}\|\nabla\hat{\phi}_{\hat{z}}\|_{0,2;\hat{K}}, \quad \|\nabla\phi_z\|_{0,\infty;K} \leq \frac{\hat{h}}{\rho_K}\|\nabla\hat{\phi}_{\hat{z}}\|_{0,\infty;\hat{K}}, \tag{7.4}$$

$$\left\|\psi_z^K\right\|_{0,2;K} = \frac{|\hat{K}|^{1/2}}{|K|^{1/2}}\|\hat{\psi}_{\hat{z}}\|_{0,2;\hat{K}}. \tag{7.5}$$

Polynomial approximation in Sobolev spaces

We recall some basic results. Assume ω is star-shaped with respect to a ball. Set

$$\rho_{\max} := \sup\{\rho,\ \omega \text{ is star-shaped with respect to a ball of radius } \rho\}$$

and

$$\gamma := \frac{\operatorname{diam}(\omega)}{\rho_{\max}}.$$

Assume k, $m \in \mathbb{N}$ with $k \leq m$. Then it holds, see [11, Lemma (4.3.8)],

$$\inf_{P\in\mathbb{P}^m} |f - P|_{k,2;\omega} \leq C(m, d, \gamma)\operatorname{diam}(\omega)^{m-k}\,|f|_{m,2;\omega}\,. \tag{7.6}$$

For convex domains, the constant can be taken independently of γ, see [46] where it is also explicitly expressed.

Moreover, we recall from [43] that the squared global best error in $S_0^{\ell,0}(\mathcal{T})$ with respect to the H^1-seminorm is equivalent, up to a constant, to the sum of the squared local best errors in $\mathbb{P}^\ell(K)$, that is, for every $u \in H_0^1(\Omega)$,

$$\inf_{v\in S_0^{\ell,0}(\mathcal{T})} |u - v|_{1,2;\Omega}^2 \approx \sum_{K\in\mathcal{T}} \inf_{P\in\mathbb{P}^\ell(K)} |u - P|_{1,2;K}^2\,. \tag{7.7}$$

Th constant in the $\gtrsim$-direction is given by one, while the one in the $\lesssim$-direction depends on d, ℓ, and $\sigma_{\mathcal{T}}$.

Bounds for the Poincaré and Friedrichs constants

The Poincaré constant of a patch of elements can be bounded explicitly, applying Proposition 2.10 of [45] with a decomposition of ω_K into elements. We recall that the Poincaré constant of a simplex is given by π^{-1}, see [5, 35]. Recalling that $\#\omega_K$ stands for the numbers of elements in ω_K, we get

$$C_{P,\omega_K} \leq \frac{4}{\pi}(\#\omega_K - 1)^{1/2}\left(\frac{1}{2}+\pi\right)^{1/2} \max_{1\leq i\leq \#\omega_K} \frac{\operatorname{diam}(K_i)|\omega_K|^{1/2}}{\operatorname{diam}(\omega_K)|K_i|^{1/2}}.$$

As in [44] we can also bound the Friedrichs constant of a patch in terms of the corresponding Poincaré constant. In fact, if $\partial\omega_K \cap \partial\Omega$ is a set of non-zero $(d-1)$-dimensional measure, and $f \in H_0^1(\Omega)$, we have

$$\begin{aligned}\|f\|_{0,2;\omega_K} &\leq \|f - c_K\|_{0,2;\omega_K} + \|c_K\|_{0,2;\omega_K} \\ &\leq \|f - c_K\|_{0,2;\omega_K} + |c_K||\omega_K|^{1/2} = \frac{|\omega_K|^{1/2}}{|\partial\omega_K \cap \partial\Omega|}\left|\int_{\partial\omega_K\cap\partial\Omega} f - c_K\right|,\end{aligned} \tag{7.8}$$

with $c_K := \frac{1}{|\omega_K|}\int_{\omega_K} f$. For every face $E \subset \partial\omega_K \cap \partial\Omega$, let K_E the element such that $E \subset \partial K_E$ and $K_E \subset \omega_K$. Exploiting the Trace Theorem [44, Corollary 4.5], we get

$$\int_E |f - c_K| \leq \frac{|E|}{|K_E|^{1/2}}\left(\|f - c_K\|_{0,2;K_E} + d^{-1}\operatorname{diam}(K_E)\,\|\nabla f\|_{0,2;K_E}\right). \tag{7.9}$$

Summing over $E \subset \partial\omega_K \cap \partial\Omega$ gives

$$\begin{aligned}&\int_{\partial\omega_K\cap\partial\Omega} |f - c_K| \\ &\quad\leq (d+1)|\partial\omega_K \cap \partial\Omega|^{1/2} \max_{E\subset\partial\omega_K\cap\partial\Omega}\left(\frac{|E|}{|K_E|}\right)^{1/2}\|f - c_K\|_{0,2;\omega_K} \\ &\qquad+ \frac{d+1}{d}|\partial\omega_K \cap \partial\Omega|^{1/2} \max_{E\subset\partial\omega_K\cap\partial\Omega}\left(\frac{|E|\operatorname{diam}(K_E)^2}{|K_E|}\right)^{1/2}\|\nabla f\|_{0,2;\omega_K}.\end{aligned} \tag{7.10}$$

Combining (7.8)–(7.10), we obtain

$$\begin{aligned}C_{F,\omega_K} \leq &\left(1 + (d+1)\max_{E\subset\partial\omega_K\cap\partial\Omega}\left(\frac{|E|}{|K_E|}\right)^{1/2}\frac{|\omega_K|^{1/2}}{|\partial\omega_K\cap\partial\Omega|^{1/2}}\right)C_{P,\omega_K} \\ &+ \frac{d+1}{d}\max_{E\subset\partial\omega_K\cap\partial\Omega}\left(\frac{|E|\operatorname{diam}(K_E)^2}{|K_E|\operatorname{diam}(\omega_K)^2}\right)^{1/2}\frac{|\omega_K|^{1/2}}{|\partial\omega_K\cap\partial\Omega|^{1/2}}.\end{aligned}$$

We set

$$C_P := \max_{K\in\mathcal{T}_\Omega} C_{P,\omega_K}, \qquad C_F := \max_{K\in\mathcal{T}_{\partial\Omega}} C_{F,\omega_K},$$

and note that they are bounded in terms of $\sigma_{\mathcal{T}}$.

7.2 Interpolation and dual norms I

When bounding the error in the $H^1(H^1, H^{-1})$-norm, we have to deal with an approximation problem in H^{-1}. To this end we need an interpolation operator allowing for duality arguments. Since we are working with functions of integer regularity, we can consider an interpolation operator that is well-defined in L^2.

We define $\Pi_0 : L^2(\Omega) \to S_0^{\ell,0}(\mathcal{T})$ as

$$\Pi_0 f := \sum_{z\in\mathcal{N}_\Omega} \left(\int_{\omega_z} f\phi_z^*\right)\phi_z. \tag{7.11}$$

For every $z \in \mathcal{N}_\Omega$, the function $\phi_z^* \in L^2(\omega_z)$ is given by

$$\phi_z^* := \sum_{K:\ \mathcal{N}_K\ni z} \frac{|K|}{|\omega_z|}\psi_z^K, \tag{7.12}$$

where $\{\psi_z^K\}_{z\in\mathcal{N}_K}$ are the local $L^2(K)$-dual basis functions.

Moreover we define $\Pi_0^* : L^2(\Omega) \to \operatorname{span}\{\phi_z^*\}_{z\in\mathcal{N}_\Omega}$ as

$$\Pi_0^* g := \sum_{z\in\mathcal{N}_\Omega} \left(\int_{\omega_z} g\phi_z\right)\phi_z^*. \tag{7.13}$$

For every f,$g \in L^2(\Omega)$, we have

$$\begin{aligned}(\Pi_0 f, g)_{L^2(\Omega)} &= \int_\Omega \left(\sum_{z\in\mathcal{N}_\Omega}\left(\int_{\omega_z} f\phi_z^*\right)\phi_z\right) g = \sum_{z\in\mathcal{N}_\Omega}\left(\int_{\omega_z} f\phi_z^*\right)\left(\int_\Omega \phi_z g\right)\\ &= \int_\Omega f\left(\sum_{z\in\mathcal{N}_\Omega}\left(\int_{\omega_z}\phi_z g\right)\phi_z^*\right) = (f, \Pi_0^* g)_{L^2(\Omega)}.\end{aligned} \tag{7.14}$$

The following properties of $\{\phi_z^*\}_{z\in\mathcal{N}_\Omega}$ are useful.

Remark 7.1. The functions $\{\phi_z^*\}_{z\in\mathcal{N}_\Omega}$ defined in (7.12) satisfy

(i) $\displaystyle\int_\Omega \phi_y\phi_z^* = \delta_{yz}$, for every $y \in \mathcal{N}_\Omega$.

(ii) $\|\phi_z^*\|_{0,2;K} = \dfrac{|\hat{K}|^{1/2}|K|^{1/2}}{|\omega_z|}\|\hat{\psi}_{\hat{z}}\|_{0,2;\hat{K}}$ and $\|\phi_z^*\|_{0,2;\omega_z} = \dfrac{|\hat{K}|^{1/2}}{|\omega_z|^{1/2}}\|\hat{\psi}_{\hat{z}}\|_{0,2;\hat{K}}$.

Proof. The definition of ϕ_z^* in terms of the dual basis functions implies (i):

$$\int_K \phi_y \phi_z^* = \int_K \phi_y \frac{|K|}{|\omega_z|}\psi_z^K = \frac{|K|}{|\omega_z|}\delta_{yz},$$

for every $K \subset \omega_z$. Property (ii) is a consequence of the definition of ϕ_z^* and (7.5). □

As a consequence, we get the following proposition.

Proposition 7.2 (Properties of Π_0 and Π_0^*)**.** *The interpolation operator* Π_0 *defined in* (7.11) *satisfies the following properties*

(i) Invariance over $S_0^{\ell,0}(\mathcal{T})$*. For every* $f \in S_0^{\ell,0}(\mathcal{T})$*,*

$$\Pi_0 f = f.$$

(ii) Stability in L^2*. For every* $f \in L^2(\Omega)$*, for every* $K \in \mathcal{T}$*,*

$$\|\Pi_0 f\|_{0,2;K} \leq C(d,\ell)\,\|f\|_{0,2;\omega_K}.$$

(iii) Stability in H^1*. For every* $f \in H_0^1(\Omega)$*,*

$$\forall K \in \mathcal{T}_\Omega, \qquad |\Pi_0 f|_{1,2;K} \leq C(d,\ell,C_P)\frac{\operatorname{diam}(\omega_K)}{\rho_K}\,|f|_{1,2;\omega_K},$$
$$\forall K \in \mathcal{T}_{\partial\Omega}, \qquad |\Pi_0 f|_{1,2;K} \leq C(d,\ell,C_F)\frac{\operatorname{diam}(\omega_K)}{\rho_K}\,|f|_{1,2;\omega_K}.$$

The interpolation operator Π_0^* *defined in* (7.13) *satisfies the following properties:*

(iv) Local invariance over constants. For every $c \in \mathbb{R}$*, for every* $K \in \mathcal{T}_\Omega$*,*

$$(\Pi_0^* c\chi_{\omega_K})|_K = c\chi_K.$$

(v) Stability in L^2*. For every* $\varphi \in L^2(\Omega)$*, for every* $K \in \mathcal{T}$*,*

$$\|\Pi_0^*\varphi\|_{0,2;K} \leq C(d,\ell)\,\|\varphi\|_{0,2;\omega_K}.$$

Proof. Invariance over $S_0^{\ell,0}(\mathcal{T})$ is a consequence of Property (i) of Remark 7.1.

Concerning stability of Π_0 in L^2, we use the Cauchy-Schwarz inequality and Property (ii) of Remark 7.1 to obtain

$$\begin{aligned}\|\Pi_0 f\|_{0,2;K} &\leq \sum_{z\in\mathcal{N}_K} \left|\int_{\omega_z} f\phi_z^*\right| \|\phi_z\|_{0,2;K} \\ &\leq \sum_{z\in\mathcal{N}_K} \|f\|_{0,2;\omega_z} \|\phi_z^*\|_{0,2;\omega_z} \|\phi_z\|_{0,2;K} \\ &\leq \sum_{z\in\mathcal{N}_K} \|f\|_{0,2;\omega_z} \frac{|K|^{1/2}}{|\omega_z|^{1/2}} \|\hat{\psi}_{\hat{z}}\|_{0,2;\hat{K}} \|\hat{\phi}_{\hat{z}}\|_{0,2;\hat{K}} \\ &\leq C(d,\ell)\, \|f\|_{0,2;\omega_K}\,. \end{aligned} \tag{7.15}$$

Concerning stability of Π_0 in H^1, we set, for every $f \in H_0^1(\Omega)$,

$$c_K(f) := \begin{cases} \dfrac{1}{|\omega_K|}\displaystyle\int_{\omega_K} f & \text{if } K \in \mathcal{T}_\Omega \\ 0 & \text{if } K \in \mathcal{T}_{\partial\Omega} \end{cases}. \tag{7.16}$$

We use the invariance of Π over $S_0^{1,0}(\mathcal{T})$, the Cauchy-Schwarz inequality, (7.4), Property (ii) of Remark 7.1 and we get

$$\begin{aligned}|\Pi_0 f|_{1,2;K} &= |\Pi_0 f - c_K(f)|_{1,2;K} = |\Pi_0 f - \Pi_0 c_K(f)|_{1,2;K} \\ &\leq \sum_{z\in\mathcal{N}_K} \|f - c_K(f)\|_{0,2;\omega_z} \|\phi_z^*\|_{0,2;\omega_z} \|\nabla\phi_z\|_{0,2;K} \\ &\leq C(d,\ell) \sum_{z\in\mathcal{N}_K} \|f - c_K(f)\|_{0,2;\omega_z} \frac{|K|^{1/2}}{\rho_K |\omega_z|^{1/2}} \|\hat{\psi}_{\hat{z}}\|_{0,2;\hat{K}} \|\nabla\hat{\phi}_{\hat{z}}\|_{0,2;\hat{K}} \\ &\leq C(d,\ell)\rho_K^{-1}\, \|f - c_K(f)\|_{0,2;\omega_K}\,. \end{aligned} \tag{7.17}$$

To obtain (iii), for every $K \in \mathcal{T}_\Omega$, we exploit the Poincaré inequality, while for every $K \in \mathcal{T}_{\partial\Omega}$, we observe that ω_K has one or more faces lying on the boundary $\partial\Omega$, and therefore we exploit the Friedrichs inequality.

Concerning (iv), by linearity of Π_0^*, it is sufficient to show the assertion for $c = 1$. We compute first

$$\int_{\omega_z} \phi_z = \sum_{K\subset\omega_z} \int_K \phi_z = \sum_{K\subset\omega_z} |K| d! \int_{\hat{K}} \hat{\phi}_{\hat{z}} = |\omega_z| d! \int_{\hat{K}} \hat{\phi}_{\hat{z}}.$$

Therefore we have

$$(\Pi_0^* \chi_{\omega_K})|_K = \sum_{z\in\mathcal{N}_K} \left(\int_{\omega_z} \phi_z\right) \phi_z^*|_K = \sum_{z\in\mathcal{N}_K} d!|K| \left(\int_{\hat{K}} \hat{\phi}_{\hat{z}}\right) \psi_z^K.$$

We denote by $\zeta := \sum_{z\in\mathcal{N}_K} \int_{\hat{K}} \hat{\phi}_{\hat{z}} \psi_z^K \in \mathbb{P}^\ell(K)$. For every $y \in \mathcal{N}_K$, we have

$$(\zeta, \phi_y)_{L^2(K)} = \sum_{z\in\mathcal{N}_K} \left(\int_{\hat{K}} \hat{\phi}_{\hat{z}} \right) \int_K \psi_z^K \phi_y = \int_{\hat{K}} \hat{\phi}_{\hat{y}} = \frac{1}{|K|d!} (\chi_K, \phi_y)_{L^2(K)}.$$

Consequently $\zeta = \frac{1}{|K|d!}\chi_K$ and $(\Pi_0^* \chi_{\omega_K})|_K = \chi_K$.

Concerning stability of Π_0^* in L^2, we proceed as in (7.15) and get

$$\begin{aligned}
\|\Pi_0^* \varphi\|_{0,2;K} &\le \sum_{z\in\mathcal{N}_K} \left| \int_{\omega_z} \varphi \phi_z \right| \|\phi_z^*\|_{0,2;K} \\
&\le \sum_{z\in\mathcal{N}_K} \|\varphi\|_{0,2;\omega_z} \|\phi_z\|_{0,2;\omega_z} \|\phi_z^*\|_{0,2;K} \\
&\le \sum_{z\in\mathcal{N}_K} \|\varphi\|_{0,2;\omega_z} \frac{|K|^{1/2}}{|\omega_z|^{1/2}} \|\hat{\psi}_{\hat{z}}\|_{0,2;\hat{K}} \|\hat{\phi}_{\hat{z}}\|_{0,2;\hat{K}} \\
&\le C(d,\ell) \|\varphi\|_{0,2;\omega_K}.
\end{aligned}$$

□

Approximation properties

With the following propositions we analyse the approximation properties of Π_0 in the H^1-seminorm, in the L^2-norm and in the H^{-1}-norm.

Proposition 7.3 (Approximation in H^1)**.** *The interpolation operator* Π_0 *defined in* (7.11) *satisfies, for every* $f \in H_0^1(\Omega)$*,*

$$|f - \Pi_0 f|_{1,2;\Omega} \le C(d, C_P, C_F, \sigma_\mathcal{T}) \left(\sum_{K\in\mathcal{T}_\Omega} \inf_{P\in\mathbb{P}^\ell(K)} |f - P|_{1,2;K}^2 \right)^{1/2}.$$

Proof. We denote by $Q_K \in S_0^{\ell,0}(\mathcal{T})|_{\omega_K}$ a best approximation to f in the H^1-seminorm. By Properties (i) and (iii) of Proposition 7.2 we get

$$\begin{aligned}
|f - \Pi_0 f|_{1,2;\Omega}^2 &= \sum_{K\in\mathcal{T}} |f - \Pi_0 f|_{1,2;K}^2 = \sum_{K\in\mathcal{T}} |f - Q_K|_{1,2;K}^2 + |Q_K - \Pi_0 f|_{1,2;K}^2 \\
&\le C(d, C_P, C_F) \sum_{K\in\mathcal{T}} \frac{\operatorname{diam}(\omega_K)^2}{\rho_K^2} |f - Q_K|_{1,2;\omega_K}^2.
\end{aligned}$$

The assertion follows thanks to (7.7) applied to the triangulation induced by $\mathcal{T}$ on ω_K. □

Next we investigate approximation in L^2, with the help of the variant of Π_0 that has value in $S^{\ell,0}(\mathcal{T})$. For every $f \in L^2(\Omega)$ we set

$$\widetilde{\Pi}_0 f := \sum_{z \in \mathcal{N}} \left(\int_{\omega_z} f \phi_z^* \right) \phi_z, \tag{7.18}$$

where $\{\phi_z^*\}_{z \in \mathcal{N}}$ are defined in (7.12). The difference with Π_0 is that the sum is on $\mathcal{N}$ and not only on $\mathcal{N}_\Omega$. For this reason $\widetilde{\Pi}_0$ is invariant on $S^{\ell,0}(\mathcal{T})$, and enjoys the same properties of Π_0.

Proposition 7.4 (Approximation in L^2)**.** *The interpolation operator* Π_0 *defined in* (7.11) *satisfies, for every* $f \in H_0^1(\Omega)$*,*

$$\|f - \Pi_0 f\|_{0,2;\Omega} \leq C(d, \ell, \sigma_{\mathcal{T}}) \Big(\sum_{K \in \mathcal{T}} \inf_{P \in S^{\ell,0}(\mathcal{T})|_{\omega_K}} \|f - P\|_{0,2;\omega_K}^2 + \sum_{K \in \mathcal{T}_{\partial\Omega}} \operatorname{diam}(K)^2 \inf_{P \in S^{\ell,0}(\mathcal{T})|_{\omega_K}} |f - P|_{1,2;\omega_K}^2 \Big)^{1/2}.$$

Proof. As a consequence of Properties (i) and (ii) of Proposition 7.2, we have, for every $K \in \mathcal{T}_\Omega$,

$$\|f - \Pi_0 f\|_{0,2;K} \leq C(d, \ell) \inf_{P \in S^{\ell,0}(\mathcal{T})|_{\omega_K}} \|f - P\|_{0,2;\omega_K}. \tag{7.19}$$

For the elements in $\mathcal{T}_{\partial\Omega}$, we use the strategy of [16], and insert $\widetilde{\Pi}_0 f$:

$$\|f - \Pi_0 f\|_{0,2;K} \leq \left\| f - \widetilde{\Pi}_0 f \right\|_{0,2;K} + \left\| \widetilde{\Pi}_0 f - \Pi_0 f \right\|_{0,2;K}. \tag{7.20}$$

The first term can be bounded as in (7.19). The second term is the norm of a polynomial on K that can be expressed by means of the basis functions. Since $(\widetilde{\Pi}_0 f)(z) = (\Pi_0 f)(z)$, for every $z \in \mathcal{N}_\Omega$, we have

$$\left\| \widetilde{\Pi}_0 f - \Pi_0 f \right\|_{0,2;K} \leq \sum_{z \in \mathcal{N}_{K \cap \partial\Omega}} |(\widetilde{\Pi}_0 f)(z)| \, \|\phi_z\|_{0,2;K}.$$

We choose an element K_z with a face E_z on the boundary $\partial\Omega$ and such that $z \in \mathcal{N}_{E_z}$. First we map to the reference $(d-1)$-simplex, use the equivalence of norms on a finite dimensional space, and map back to E_z. Then, inserting $f|_{\partial\Omega} = 0$, and exploiting the Trace Theorem [44, Corollary 4.5] and the

Young inequality, we get

$$
\begin{aligned}
|(\widetilde{\Pi}_0 f)(z)| &\le \left\|\widetilde{\Pi}_0 f\right\|_{0,\infty;E_z} \le C(d,\ell)|E_z|^{-1/2}\left\|\widetilde{\Pi}_0 f\right\|_{0,2;E_z} \\
&= C(d,\ell)|E_z|^{-1/2}\left\|\widetilde{\Pi}_0 f - f\right\|_{0,2;E_z} \\
&\le C(d,\ell)\frac{1}{|K_z|^{1/2}}\left(\left\|\widetilde{\Pi}_0 f - f\right\|_{0,2;K_z}\right. \\
&\qquad \left. +\mathrm{diam}(K_z)^{1/2}\left\|\widetilde{\Pi}_0 f - f\right\|^{1/2}_{0,2;K_z}\left|\widetilde{\Pi}_0 f - f\right|^{1/2}_{1,2;K_z}\right) \\
&\le C(d,\ell)\frac{1}{|K_z|^{1/2}}\left(\left\|\widetilde{\Pi}_0 f - f\right\|_{0,2;K_z} + \mathrm{diam}(K_z)\left|\widetilde{\Pi}_0 f - f\right|_{1,2;K_z}\right).
\end{aligned}
$$

Recalling (7.3), and invoking invariance of $\widetilde{\Pi}_0$ over $S^{\ell,0}(\mathcal{T})$ and its stability in L^2 and H^1, we obtain

$$
\begin{aligned}
&\sum_{z\in\mathcal{N}_{K\cap\partial\Omega}} |(\widetilde{\Pi}_0 f)(z)|\,\|\phi_z\|_{0,2;K} \\
&\quad\le C(d,\ell)\sum_{z\in\mathcal{N}_{K\cap\partial\Omega}}\frac{|K|^{1/2}}{|K_z|^{1/2}}\left(\left\|\widetilde{\Pi}_0 f - f\right\|_{0,2;K_z} + \mathrm{diam}(K_z)\left|\widetilde{\Pi}_0 f - f\right|_{1,2;K_z}\right) \\
&\quad\le C(d,\ell,\sigma_{\mathcal{T}})\sum_{z\in\mathcal{N}_{K\cap\partial\Omega}}\left(\inf_{P\in S^{\ell,0}(\mathcal{T})|_{\omega_{K_z}}}\|f-P\|_{0,2;\omega_{K_z}}\right. \\
&\qquad\qquad \left. +\mathrm{diam}(K_z)\inf_{P\in S^{\ell,0}(\mathcal{T})|_{\omega_{K_z}}}|f-P|_{1,2;\omega_{K_z}}\right).
\end{aligned}
$$

For every $K\in\mathcal{T}_{\partial\Omega}$ with a face on the boundary $\partial\Omega$ the number of elements $\widetilde{K}$ such that $\mathcal{N}_{\widetilde{K}\cap\partial\Omega}\cap\mathcal{N}_{K\cap\partial\Omega}\neq\emptyset$ is bounded in terms of $\sigma_{\mathcal{T}}$. Therefore, we

have

$$\begin{aligned}
&\sum_{K\in\mathcal{T}_{\partial\Omega}} \left\|\widetilde{\Pi}_0 f - \Pi_0 f\right\|^2_{0,2;K} \\
&\quad\le C(d,\ell,\sigma_{\mathcal{T}}) \sum_{K\in\mathcal{T}_{\partial\Omega}} \sum_{z\in\mathcal{N}_{K\cap\partial\Omega}} \Bigg(\inf_{P\in S^{\ell,0}(\mathcal{T})|_{\omega_{K_z}}} \|f-P\|^2_{0,2;\omega_{K_z}} \\
&\qquad\qquad + \operatorname{diam}(K_z)^2 \inf_{P\in S^{\ell,0}(\mathcal{T})|_{\omega_{K_z}}} |f-P|^2_{1,2;\omega_{K_z}} \Bigg) \\
&\quad\le C(d,\ell,\sigma_{\mathcal{T}}) \sum_{K\in\mathcal{T}_{\partial\Omega}} \Bigg(\inf_{P\in S^{\ell,0}(\mathcal{T})|_{\omega_K}} \|f-P\|^2_{0,2;\omega_K} \\
&\qquad\qquad + \operatorname{diam}(K) \inf_{P\in S^{\ell,0}(\mathcal{T})|_{\omega_K}} |f-P|^2_{1,2;\omega_K} \Bigg). \qquad (7.21)
\end{aligned}$$

□

The next proposition concerns the H^{-1}-norm.

Proposition 7.5 (Approximation in H^{-1})**.** *The interpolation operator* Π_0 *defined in* (7.11) *satisfies, for every* $f \in L^2(\Omega)$*,*

$$\|f-\Pi_0 f\|_{-1;\Omega} \le C(d,\ell,C_P,\nu_{\mathcal{T}}) \Bigg(\sum_{K\in\mathcal{T}} \operatorname{diam}(\omega_K)^2 \|f\|^2_{0,2;K} \Bigg)^{1/2}. \qquad (7.22)$$

Moreover, if $f \in H^1_0(\Omega)$*, it holds*

$$\begin{aligned}
&\|f-\Pi_0 f\|_{-1;\Omega} \\
&\quad\le C(d,\ell,C_P,\sigma_{\mathcal{T}}) \Bigg(\sum_{K\in\mathcal{T}} \operatorname{diam}(\omega_K)^2 \inf_{P\in S^{\ell,0}(\mathcal{T})|_{\omega_K}} \|f-P\|^2_{0,2;\omega_K} \\
&\qquad\qquad + \sum_{K\in\mathcal{T}_{\partial\Omega}} \operatorname{diam}(\omega_K)^4 \inf_{P\in S^{\ell,0}(\mathcal{T})|_{\omega_K}} |f-P|^2_{0,2;\omega_K} \Bigg)^{1/2}. \qquad (7.23)
\end{aligned}$$

Proof. We start with (7.22). We exploit first (7.14) and that $f \in L^2(\Omega)$ to

get

$$\begin{aligned}\|f-\Pi_0 f\|_{-1;\Omega} &= \sup_{\varphi\in H_0^1(\Omega)} \frac{\langle f-\Pi_0 f,\varphi\rangle}{|\varphi|_{1,2;\Omega}} = \sup_{\varphi\in H_0^1(\Omega)} \frac{\langle f,\varphi-\Pi_0^*\varphi\rangle}{|\varphi|_{1,2;\Omega}} \\ &\leq \sup_{\varphi\in H_0^1(\Omega)} \frac{\sum_{K\in\mathcal{T}}\int_K f(\varphi-\Pi_0^*\varphi)}{|\varphi|_{1,2;\Omega}} \\ &\leq \sup_{\varphi\in H_0^1(\Omega)} \frac{\sum_{K\in\mathcal{T}}\|f\|_{0,2;K}\,\|\varphi-\Pi_0^*\varphi\|_{0,2;K}}{|\varphi|_{1,2;\Omega}}. \end{aligned} \tag{7.24}$$

Recalling the definition of $c_K(\cdot)$ in (7.16), by means of Properties (iv) and (v) and the Poincarè or Friedrichs inequality, we get

$$\begin{aligned}\|\varphi-\Pi_0^*\varphi\|_{0,2;K} &\leq C(d,\ell)\,\|\varphi-c_K(\varphi)\|_{0,2;\omega_K} \\ &\leq C(d,\ell,C_P,C_F)\mathrm{diam}(\omega_K)\,|\varphi|_{1,2;\omega_K}\,. \end{aligned} \tag{7.25}$$

Finally, using the Cauchy-Schwarz inequality for sums in (7.24) and (7.25), we arrive at

$$\begin{aligned}&\|f-\Pi_0 f\|_{-1;\Omega} \\ &\quad\leq C(d,C_P,C_F)\sup_{\varphi\in H_0^1(\Omega)}\frac{\sum_{K\in\mathcal{T}}\|f\|_{0,2;K}\,\mathrm{diam}(\omega_K)\,|\varphi|_{1,2;\omega_K}}{|\varphi|_{1,2;\Omega}} \\ &\quad\leq C(d,C_P,C_F,\nu_{\mathcal{T}})\left(\sum_{K\in\mathcal{T}}\mathrm{diam}(\omega_K)^2\,\|f\|_{0,2;K}^2\right)^{1/2},\end{aligned}$$

where the constant also depends on the number of elements in a patch.

Concerning (7.23) we exploit also Property (i) of Proposition 7.2 and, similarly as in (7.24), we get

$$\begin{aligned}\|f-\Pi_0 f\|_{-1;\Omega} &= \sup_{\varphi\in H_0^1(\Omega)}\frac{\langle f-\Pi_0 f,\varphi-\Pi_0^*\varphi\rangle}{|\varphi|_{1,2;\Omega}} \\ &\leq C(d,C_P,C_F)\sup_{\varphi\in H_0^1(\Omega)}\frac{\sum_{K\in\mathcal{T}}\|f-\Pi_0 f\|_{0,2;K}\,\|\varphi-\Pi_0^*\varphi\|_{0,2;K}}{|\varphi|_{1,2;\Omega}}.\end{aligned}$$

By means of (7.25), (7.19), (7.20) and (7.21), we obtain

$$\begin{aligned}
&\|f-\Pi_0 f\|_{-1;\Omega}\\
&\quad\le C(d,C_P,C_F)\sup_{\varphi\in H_0^1(\Omega)}\frac{\sum_{K\in\mathcal{T}}\|f-\Pi_0 f\|_{0,2;K}\operatorname{diam}(\omega_K)\,|\varphi|_{1,2;\omega_K}}{|\varphi|_{1,2;\Omega}}\\
&\quad\le C(d,C_P,C_F,\nu_{\mathcal{T}})\left(\sum_{K\in\mathcal{T}}\operatorname{diam}(\omega_K)^2\,\|f-\Pi_0 f\|_{0,2;K}^2\right)^{1/2}\\
&\quad\le C(d,C_P,C_F,\nu_{\mathcal{T}})\left(\sum_{K\in\mathcal{T}}\operatorname{diam}(\omega_K)^2\inf_{P\in S^{\ell,0}(\mathcal{T})}\|f-P\|_{0,2;\omega_K}^2\right.\\
&\qquad\qquad\left.+\operatorname{diam}(\omega_K)^4\inf_{P\in S^{\ell,0}(\mathcal{T})}|f-P|_{0,2;\omega_K}^2\right)^{1/2}. \qquad \square
\end{aligned}$$

As in [39], we apply (7.6) together with Theorem 7.1 of [23], where the subdomains are the interior of pairs of elements that share a common face. In this way the constant depends on $\sigma_{\mathcal{T}}$. Combining with Propositions 7.5–7.3 we get the following corollary.

Corollary 7.6. *Assume $1\le m\le \ell+1$ and assume $f\in H^m(\Omega)\cap H_0^1(\Omega)$. Then, the interpolation operator Π_0 defined in* (7.11) *satisfies*

$$\begin{aligned}
\|f-\Pi_0 f\|_{-1;\Omega}&\le C(d,\ell,C_P,C_F,\sigma_{\mathcal{T}})\left(\sum_{K\in\mathcal{T}}\operatorname{diam}(\omega_K)^{2m+2}\,|f|_{m,2;\omega_K}^2\right)^{1/2},\\
\|f-\Pi_0 f\|_{0,2;\Omega}&\le C(d,\ell,C_P,C_F,\sigma_{\mathcal{T}})\left(\sum_{K\in\mathcal{T}}\operatorname{diam}(\omega_K)^{2m}\,|f|_{m,2;\omega_K}^2\right)^{1/2},\\
|f-\Pi_0 f|_{1,2;\Omega}&\le C(d,\ell,C_P,C_F,\sigma_{\mathcal{T}})\left(\sum_{K\in\mathcal{T}}\operatorname{diam}(K)^{2m-2}|f|_{m,2;K}^2\right)^{1/2}.
\end{aligned}$$

7.3 Standard formulation and integer regularity

In this section we derive error estimates for the approximation of the solution u of the parabolic problem in the standard formulation. We assume for u integer regularity.

We recall that $u \in H^1(H^1_0, H^{-1})$ satisfies, for every $(\varphi_0, \varphi_1) \in L^2 \times L^2(H^1_0)$,

$$\langle u(0), \varphi_0\rangle + \int_0^T \langle u', \varphi_1\rangle + \langle Au, \varphi_1\rangle = \langle u_0, \varphi_0\rangle + \int_0^T \langle f, \varphi_1\rangle\,.$$

7.3.1 Spatial semidiscretization

In order to apply the results in Chapter 3, we require that $\mathcal{T}$ belongs to a family of triangulations for which the L^2-projection onto $S^{\ell,0}_0$ is H^1-stable. Conditions that guarantee this assumption can be found in [10, 17]. Moreover in [32] it is proven that the L^2-projection is H^1-stable on $S^{1,0}_0$ where the meshes are adaptively generated by newest vertex bisection in $2d$.

We recall that the semidiscrete solution $U \in H^1(S^{\ell,0}_0(\mathcal{T}))$ satisfies, for every $(\varphi_0, \varphi_1) \in S^{\ell,0}_0(\mathcal{T}) \times L^2(S^{\ell,0}_0(\mathcal{T}))$,

$$\langle U(0), \varphi_0\rangle + \int_0^T \langle U', \varphi_1\rangle + \langle AU, \varphi_1\rangle = \langle u_0, \varphi_0\rangle + \int_0^T \langle f, \varphi_1\rangle\,.$$

We combine the results in Section 3.2 and in 7.2 to obtain the following theorem.

Theorem 7.7. *Assume $2 \le m \le \ell + 1$. Assume $u \in L^2(H^m)$ and, if $m = 2$, $u' \in L^2(H^{m-2})$ otherwise $u' \in L^2(H^{m-2} \cap H^1_0)$. Then we have*

$$\begin{aligned}
&\|u(0) - U(0)\|^2_{0,2;\Omega} + \int_0^T \|u'(t) - U'(t)\|^2_{-1;\Omega} + |u(t) - U(t)|^2_{1,2;\Omega}\,\mathrm{d}t \\
&\quad\lesssim \sum_{K\in\mathcal{T}} \operatorname{diam}(\omega_K)^{2m-2}\,|u(0)|^2_{m-1,2;\omega_K} \\
&\qquad + \int_0^T \sum_{K\in\mathcal{T}} \operatorname{diam}(\omega_K)^{2m-2}\,|u'(t)|^2_{m-2,2;\omega_K} + \operatorname{diam}(K)^{2m-2}\,|u(t)|^2_{m,2;K}\,\mathrm{d}t.
\end{aligned}$$

The hidden constant depends on the H^1-norm of the L^2-projection on $S^{\ell,0}_0(\mathcal{T})$, the coercivity and continuity constants of the parabolic problem, the dimension d, the polynomial degree ℓ and the shape-parameter $\sigma_{\mathcal{T}}$.

Proof. We recall that Theorem 3.7 in particular states that

$$\begin{aligned}
&\|u(0) - U(0)\|^2_{0,2;\Omega} + \int_0^T \|u'(t) - U'(t)\|^2_{-1;\Omega} + |u(t) - U(t)|^2_{1,2;\Omega}\,\mathrm{d}t \\
&\quad\lesssim \inf_{V\in H^1(S^{\ell,0}_0(\mathcal{T}))} \Big(\|u(0) - V(0)\|^2_{0,2;\Omega} \\
&\qquad\qquad + \int_0^T \|u'(t) - V'(t)\|^2_{-1;\Omega} + |u(t) - V(t)|^2_{1,2;\Omega}\,\mathrm{d}t\Big).
\end{aligned}$$

We insert a particular choice in the infimum on the right-hand side:

$$\forall t \in [0,T], \qquad V(t) := \Pi_0 u(t).$$

Since $u \in H^1(H^m, H^{m-2})$ entails $u \in C^0(H^{m-1})$, the assertion follows by applying Corollary 7.6. □

7.3.2 Semidiscretization in time

We recall that $\mathcal{P}$ is a partition $0 = t_0 < t_1 < \ldots < t_N = T$ of the time interval $(0,T)$ into subintervals $I_n = (t_{n-1}, t_n]$, and that $\mu_{\mathcal{P}} := \sup_n \tau_n/\tau_{n+1}$. Moreover we recall that

$$\begin{aligned}\mathcal{S}^{1,0}(\mathcal{P}, H_0^1) &= \{v \in C^0(H_0^1), v|_{I_n} \in \mathbb{P}^1(I_n, H_0^1), n = 1, \ldots, N\},\\ \mathcal{S}^{0,-1}(\mathcal{P}, H_0^1) &= \{\phi \in L^2(H_0^1), \phi|_{I_n} \in H_0^1, n = 1, \ldots, N\}.\end{aligned}$$

The semidiscrete solution $U \in \mathcal{S}^{1,0}(\mathcal{P}, H_0^1)$ satisfies, for every $\varphi \in L^2 \times \mathcal{S}^{0,-1}(\mathcal{P}, H_0^1)$,

$$\langle U(0), \varphi_0\rangle + \sum_{n=1}^N \int_{I_n} \langle U', \varphi_n\rangle + \langle AU(t_n), \varphi_n\rangle = \langle u_0, \varphi_0\rangle + \sum_{n=1}^N \int_{I_n} \langle f, \varphi_n\rangle\,.$$

We exploit the results in Section 4.1 and obtain the following theorem.

Theorem 7.8. *Assume $u' \in H^1(H_0^1, H^{-1})$. Then*

$$\begin{aligned}\|u(0) - U(0)\|_{0,2;\Omega}^2 + \sum_{n=1}^N \int_{I_n} \|u' - U'\|_{-1;\Omega}^2 + |u - U(t_n)|_{1,2;\Omega}^2 \\ \lesssim \sum_{n=1}^N \tau_n^2 \int_{I_n} \|u''\|_{-1;\Omega}^2 + \tau_n^2 \int_{I_n} |u'|_{1,2;\Omega}^2\,,\end{aligned}$$

where the hidden constant depends on the coercivity and continuity constants of the parabolic problem.

Proof. Since $u \in H^1(H^1)$ implies $u \in C^0(H^1)$, we can apply Theorem 4.5, which states that

$$\begin{aligned}\|u(0) - U(0)\|_{0,2;\Omega}^2 + \sum_{n=1}^N \int_{I_n} \|u'(t) - U'(t)\|_{-1;\Omega}^2 + |u(t) - U(t_n)|_{1,2;\Omega}^2\,\mathrm{d}t \\ \lesssim \sum_{n=1}^N \int_{I_n} |u(t) - u(t_n)|_{1,2;\Omega}^2\,\mathrm{d}t + \inf_{v \in \mathcal{S}^{1,0}(\mathcal{P}, H_0^1)} \Bigg(\|u(0) - v(0)\|_{0,2;\Omega}^2 \\ + \sum_{n=1}^N \int_{I_n} \|u'(t) - v'(t)\|_{-1;\Omega}^2 + |u(t_n) - v(t_n)|_{1,2;\Omega}^2\,\mathrm{d}t \Bigg)\,.\end{aligned}$$

Choosing $v \in \mathcal{S}^{1,0}(\mathcal{P}, H^1_0)$ such that

$$v(t_n) = u(t_n), \qquad n = 0, \dots, N,$$

we have that

$$v'(t)|_{I_n} = \frac{u(t_n) - u(t_{n-1})}{\tau_n} = \frac{1}{\tau_n} \int_{I_n} u'.$$

Exploiting the Poincaré or Friedrichs inequality on the subintervals I_n, we get the assertion. □

7.3.3 Varying the spatial discretization

We consider a sequence $\{\mathcal{T}_n\}_{n=0}^N$ of triangulations that belongs to a family for which the L^2-projection is uniformly H^1-stable. The sequence of finite-dimensional spaces $\{\mathbb{V}_n\}_{n=0}^N \subset H^1_0(\Omega)$ is given by $\{S^{\ell,0}_0(\mathcal{T}_n)\}_{n=1}^N$. The L^2-projection onto $S^{\ell,0}_0(\mathcal{T}_n)$ is denoted by P_n and P_n^+ indicates the L^2-projection onto $S^{\ell,0}_0(\mathcal{T}_n) \oplus S^{\ell,0}_0(\mathcal{T}_{n+1})$. We recall that

$$\mathcal{S}^{L^2}(\mathcal{P}, \mathbb{V}) = \{v \in L^2(H^1_0),\ v|_{I_n} \in L^2(S^{\ell,0}_0(\mathcal{T}_n)),\ n = 1, \dots, N\},$$
$$\mathcal{S}^{H^1}(\mathcal{P}, \mathbb{V}) = \{v \in L^2(H^1_0),\ v|_{I_n} \in H^1(S^{\ell,0}_0(\mathcal{T}_n)),\ n = 1, \dots, N\}.$$

The semidiscrete solution U belongs to the space

$$\{v \in \mathcal{S}^{H^1}(\mathcal{P}, \mathbb{V}),\ v(0) \in S^{\ell,0}_0(\mathcal{T}_0),\ v(t_{n-1}^+) = P_n v(t_{n-1}),\ n = 1, \dots, N\}$$

and satisfies, for every $(\varphi_0, \varphi) \in S^{\ell,0}_0(\mathcal{T}_0) \times \mathcal{S}^{L^2}(\mathcal{P}, \mathbb{V})$,

$$\langle U(0), \varphi_0 \rangle + \sum_{n=1}^N \int_{I_n} \langle U', \varphi \rangle + \langle AU, \varphi \rangle = \langle u_0, \varphi_0 \rangle + \int_0^T \langle f, \varphi \rangle .$$

We resort to the results in Section 5.1 and derive the following theorem.

Theorem 7.9. *Assume $2 \le m \le \ell + 1$, and assume $u \in C^0(H^m)$. Moreover, if $m = 2$ assume $u' \in L^2(L^2)$, otherwise, $u' \in L^2(H^{m-2} \cap H^1_0)$. Then we have*

$$\begin{aligned}
&\|u(0) - U(0)\|^2_{0,2;\Omega} + \sum_{n=1}^N \int_{I_n} \|u' - U'\|^2_{-1;\Omega} + |u - U|^2_{1,2;\Omega} \\
&\quad \lesssim \sum_{K \in \mathcal{T}_0} \operatorname{diam}(\omega_K)^{2m} \, |u(0)|^2_{m,2;\omega_K} \\
&\qquad + \sum_{n=1}^N \int_{I_n} \sum_{K \in \mathcal{T}_n} \operatorname{diam}(\omega_K)^{2m-2} |u'(t)|^2_{m-2,2;\omega_K} + \operatorname{diam}(K)^{2m-2} |u(t)|^2_{m,2;K} \, \mathrm{d}t \\
&\qquad + \sum_{n=0}^{N-1} \sum_{K \in \mathcal{T}_n} \operatorname{diam}(\omega_K)^{2m} |u(t_n)|^2_{m,2;\omega_K}.
\end{aligned}$$

The hidden constant depends on the maximum of the H^1-norms of the L^2-projection on $S_0^{\ell,0}(\mathcal{T}_n)$, the coercivity and continuity constants of the parabolic problem, the dimension d, the polynomial degree ℓ and the shape parameters $\sigma_{\mathcal{T}_n}$.

Proof. We exploit Theorem 5.5, which states that

$$\begin{aligned}
&\|u(0)-U(0)\|_{0,2;\Omega}^2+\sum_{n=1}^N\int_{I_n}\|u'-U'\|_{-1;\Omega}^2+|u-U|_{1,2;\Omega}^2\\
&\quad\lesssim\inf_{\substack{v\in\{v\in\mathcal{S}^{H^1}(\mathcal{P},\mathbb{V}),\\ v(0)\in S_0^{\ell,0}(\mathcal{T}_0)\}}}\|u(0)-v(0)\|_{0,2;\Omega}^2+\sum_{n=1}^N\int_{I_n}\|u'-v'\|_{-1;\Omega}^2+|u-v|_{1,2;\Omega}^2\\
&\qquad+\|P_1(I-P_0)u(t_0)\|_{0,2;\Omega}^2+\sum_{n=1}^{N-1}\left\|P_n^+(I-P_n)u(t_n)\right\|_{0,2;\Omega}^2. \quad (7.26)
\end{aligned}$$

The infimum on the right-hand side of (7.26) can be bounded as in Theorem 7.7. Given $n=0,\dots,N$, we indicate with Π_0^n the interpolation operator Π_0 that acts onto $S_0^{\ell,0}(\mathcal{T}_n)$ and choose $v\in\mathcal{S}^{H^1}(\mathcal{P},\mathbb{V})$ such that,

$$v(0)=\Pi_0^0u(0),$$

and, for every $n=1,\dots,N$,

$$\forall t\in I_n\qquad v(t)=\Pi_0^nu(t).$$

Concerning the terms $\|P_n^+(I-P_n)u(t_n^-)\|_{0,2;\Omega}$, we have, for every $n=1,\dots,N-1$,

$$\begin{aligned}
\left\|P_n^+(I-P_n)u(t_n)\right\|_{0,2;\Omega}^2&\le\|(I-P_n)u(t_n)\|_{0,2;\Omega}^2\\
&\le\|(I-\Pi_0^n)u(t_n)\|_{0,2;\Omega}^2\\
&\le C(d,\ell)\sum_{K\in\mathcal{T}_n}\operatorname{diam}(\omega_K)^{2m}\,|u(t_n)|_{m,2;\omega_K}^2, \quad (7.27)
\end{aligned}$$

and similarly for $\left\|P_1^+(I-P_0)u(t_0)\right\|_{0,2;\Omega}$. □

Remark 7.10 (Dupont's example)**.** We consider the example presented by Dupont in [22, Sect. 4]. There, the time partition is uniform, and the spatial partitions in 1d are also uniform, with possible exceptions next to the endpoints of the spatial domain. Every time step, the endpoints of the new spatial partition coincide with the midpoints of the previous intervals. The

spatial discretization occurs with continuous piecewise affine functions, and denoted with τ the uniform time-step and with h the mesh-size, there is no convergence as $h, \tau \to 0$, if h^4/τ goes to infinity. Since the exact solution is smooth, we can apply Theorem 7.9 with $\ell = 1$ and $m = 2$. We notice that

$$\sum_{n=0}^{N-1} \sum_{K \in \mathcal{T}_n} \operatorname{diam}(\omega_K)^{2m} |u(t_n)|^2_{m,2;\omega_K} \lesssim \frac{h^4}{\tau} \sup_{n=1,\dots,N} \|u(t_n)\|^2_{2,2;\Omega} .$$

Therefore if $h, \tau \to 0$ in such a way that $h^4/\tau \to 0$, the semidiscrete solution converges to the exact one.

7.3.4 Full discretization with the backward-Euler Galerkin method

Assume $\mathcal{P}$ to be as in Section 7.3.2 and $\{\mathcal{T}_n\}_{n=0}^N$ to be as in Section 7.3.3. We recall that

$$\mathcal{S}^{0,-1}(\mathcal{P}, \mathbb{V}) = \{v \in L^2(H^1_0), v|_{I_n} \in S^{\ell,0}_0(\mathcal{T}_n), n = 1, \dots, N\}.$$

The discrete solution U belongs to the space

$$\{v \in L^2(H^1_0),\ v(0) \in S^{\ell,0}_0(\mathcal{T}_0),\ v|_{I_n} \in \mathbb{P}^1(I_n, S^{\ell,0}_0(\mathcal{T}_n)),\\ v(t^+_{n-1}) = P_n v(t_{n-1}),\ n = 1, \dots, N\}$$

and satisfies, for every $\varphi \in S^{\ell,0}_0(\mathcal{T}_0) \times \mathcal{S}^0(\mathcal{P}, \mathbb{V})$,

$$\langle U(0), \varphi_0 \rangle + \sum_{n=1}^N \int_{I_n} \langle U', \varphi_n \rangle + \langle AU(t_n), \varphi_n \rangle = \langle u_0, \varphi_0 \rangle + \sum_{n=1}^N \int_{I_n} \langle f, \varphi_n \rangle .$$

In view of the results in Section 6.1, we get the following theorem.

Theorem 7.11. *Assume* $2 \le m \le \ell + 1$. *Assume* $u \in L^2(H^m)$, $u' \in$

$H^1(H^1_0, H^{-1}) \cap L^2(H^{m-2})$. *Then*

$$\begin{aligned}
&\|u(0) - U(0)\|^2_{0,2;\Omega} + \sum_{n=1}^{N} \int_{I_n} \|u' - U'\|^2_{-1;\Omega} + |u - U(t_n)|^2_{1,2;\Omega} \\
&\quad \lesssim \sum_{K \in \mathcal{T}_0} \operatorname{diam}(\omega_K)^{2m-2} \, |u(0)|^2_{m-1,2;\omega_K} \\
&\qquad + \sum_{n=1}^{N} \tau_n^2 \int_{I_n} \|u''\|^2_{-1;\Omega} + \sum_{K \in \mathcal{T}_n} \operatorname{diam}(\omega_K)^{2m-2} \int_{I_n} |u'|^2_{m-2,2;\omega_K} \\
&\qquad\quad + \tau_n^2 \int_{I_n} |u'|^2_{1,2;\Omega} + \sum_{K \in \mathcal{T}_n} \operatorname{diam}(\omega_K)^{2m-2} \int_{I_n} |u|^2_{m,2;\omega_K} \\
&\qquad + \sum_{n=1}^{N-1} \tau_n^{-1} \sum_{K \in \mathcal{T}_n} \operatorname{diam}(\omega_K)^{2m} \int_{I_n} |u|^2_{m,2;\omega_K}
\end{aligned}$$

The hidden constant depends on the maximum of the H^1-norms of the L^2-projection on $S^{\ell,0}_0(\mathcal{T}_n)$, the coercivity and continuity constants of the parabolic problem, the dimension d, the polynomial degree ℓ, the shape parameters $\sigma_{\mathcal{T}_n}$, and the parameter $\mu_{\mathcal{P}}$.

Proof. We exploit Theorem 6.6, which states that

$$\begin{aligned}
&\|u(0) - U(0)\|^2_{0,2;\Omega} + \sum_{n=1}^{N} \int_{I_n} \|u' - U'\|^2_{-1;\Omega} + |u - U(t_n)|^2_{1,2;\Omega} \\
&\quad \lesssim \inf_{v_0 \in S^{\ell,0}_0(\mathcal{T}_0)} \|u(0) - v_0\|^2_{0,2;\Omega} + \inf_{v \in \mathcal{S}^{0,-1}(\mathcal{P},\mathbb{V})} \|u' - v\|^2_{L^2(H^{-1})} \\
&\qquad + \inf_{w \in \mathcal{S}^{L^2}(\mathcal{P},\mathbb{V})} \|u - w\|^2_{L^2(H^1)} + \inf_{z \in \mathcal{S}^{0,-1}(\mathcal{P},H^1)} \|u - z\|^2_{L^2(H^1)} \\
&\qquad + \|P_1(I - P_0)u(0)\|^2_{0,2;\Omega} + \sum_{n=1}^{N-1} \big\|P_n^+(I - P_n)\Pi^n u\big\|^2_{0,2;\Omega} .
\end{aligned}$$

Concerning the infima on the right-hand side, we insert $v_0 = \Pi^0_0 u(0) \in S^{\ell,0}_0(\mathcal{T}_0)$, $v \in \mathcal{S}^{0,-1}(\mathcal{P},\mathbb{V})$, $w \in \mathcal{S}^{L^2}(\mathcal{P},\mathbb{V})$ and $z \in \mathcal{S}^{0,-1}(\mathcal{P},H^1)$ such that, for $n = 1, \dots, N$,

$$v|_{I_n} = P_n \frac{1}{\tau_n} \int_{I_n} u', \qquad w|_{I_n} = \Pi^n_0 u|_{I_n}, \qquad z|_{I_n} = \frac{1}{\tau_n} \int_{I_n} u.$$

The terms $\|u(0) - \Pi^0_0 u(0)\|_{0,2;\Omega}$ and $\|u - \Pi^n_0 u\|_{L^2(I_n,H^1_0)}$ can be bounded as in Theorem 7.9, while $\left\|u - \frac{1}{\tau_n}\int_{I_n} u\right\|_{L^2(I_n,H^1_0)}$ can be bounded as in Theorem 7.8.

Concerning $\left\|u' - P_n \frac{1}{\tau_n}\int_{I_n} u'\right\|_{L^2(I_n,H^{-1})}$, because of the H^{-1}-stability of P_n, we have

$$\begin{aligned}
&\left\|u' - P_n\frac{1}{\tau_n}\int_{I_n} u'\right\|^2_{L^2(I_n,H^{-1})} \\
&\quad\le 2\left\|u' - P_n u'\right\|^2_{L^2(I_n,H^{-1})} + 2\left\|P_n\left(u' - \frac{1}{\tau_n}\int_{I_n} u'\right)\right\|^2_{L^2(I_n,H^{-1})} \\
&\quad\lesssim \left\|u' - \Pi_0^n u'\right\|^2_{L^2(I_n,H^{-1})} + \left\|u' - \frac{1}{\tau_n}\int_{I_n} u'\right\|^2_{L^2(I_n,H^{-1})} \\
&\quad\lesssim \int_{I_n}\sum_{K\in\mathcal{T}_n} \operatorname{diam}(\omega_K)^{2m-2}\,|u'|^2_{m-2,2;\omega_K} + \tau_n^2\int_{I_n}\|u''\|^2_{-1;\Omega}\,.
\end{aligned}$$

Regarding the terms $\|P_n^+(I-P_n)\Pi^n u\|_{0,2;\Omega}$, we notice that $P_n^+(I-P_n)\Pi^n u = \Pi^n(P_n^+(I-P_n)u)$, and therefore

$$\begin{aligned}
\left\|P_n^+(I-P_n)\Pi^n u\right\|_{0,2;\Omega} &\le \int_{I_n}\|(I-P_n)u\|_{0,2;\Omega}\,|\psi_n| \\
&\le 2\tau_n^{-1/2}\left\|(I-\Pi_0^n)u\right\|_{L^2(I_n,L^2)}\,.
\end{aligned}$$

Summing over n and exploiting Corollary 7.6, we get

$$\sum_{n=1}^{N-1}\left\|P_n^+(I-P_n)\Pi^n u\right\|^2_{0,2;\Omega} \lesssim \sum_{n=1}^{N-1}\tau_n^{-1}\int_{I_n}\sum_{K\in\mathcal{T}_n}\operatorname{diam}(\omega_K)^{2m}\,|u|^2_{m,2;\omega_K}\,. \quad (7.28)$$

For $\|P_1(I-P_0)u(0)\|_{0,2;\Omega}$ we proceed as in the proof of Theorem 7.9 and get

$$\|P_1(I-P_0)u(0)\|_{0,2;\Omega} \le C(\ell,d)\sum_{K\in\mathcal{T}_n}\operatorname{diam}(\omega_K)^{2m-2}|u(0)|^2_{m-1,2;\omega_K}. \qquad \square$$

Assume that $\mathcal{T}_0 = \mathcal{T}_1 = \ldots = \mathcal{T}_N =: \mathcal{T}$ are uniform meshes with mesh-size h and that the time partition is also uniform, with time-step τ. If $\ell = 1$ and $m = 2$, the upper barrier is of the form

$$C(u)\left(h^2 + \tau^2 + \frac{h^4}{\tau}\right). \quad (7.29)$$

If we assume some relation between h and τ, then (7.29) becomes

$$\begin{aligned}
&C(u)\left(h^2+h^3\right) && \text{if } h\sim\tau,\\
&C(u)\left(h^2+h^4\right) && \text{if } h^2\sim\tau.
\end{aligned}$$

In case $\ell = 2$ and $m = 3$, the upper barrier is of the form

$$C(u)\left(h^4 + \tau^2 + \frac{h^6}{\tau}\right)$$

and if $h^2 \sim \tau$ then the three terms converge with the same rate $C(u)h^4$.

7.4 Natural formulation and integer regularity

In this section we derive error estimate for the approximation of solution u of the parabolic problem in the natural formulation. As in Section 7.3 we assume for u integer regularity.

We recall that $0 = \tilde{t}_0 < \tilde{t}_1 < \ldots < \tilde{t}_\jmath = T$ is a partition of the time interval $(0, T)$, and that $u \in L^2(H_0^1)$ satisfies, for every $\varphi \in \{\varphi \in H^1(H_0^1, H^{-1}), \varphi(T) = 0\}$,

$$\int_0^T -\langle \varphi', u\rangle + \langle Au, \varphi\rangle = \sum_{j=0}^{\jmath-1} \left\langle g_j, \varphi(\tilde{t}_j)\right\rangle + \int_0^T \langle f, \varphi\rangle \,.$$

7.4.1 Spatial semidiscretization

As in 7.3.1 we require that $\mathcal{T}$ belongs to a family for which the L^2-projection onto $S_0^{\ell,0}$ is uniformly stable. We recall that the semidiscrete solution $U \in L^2(S_0^{\ell,0}(\mathcal{T}))$ satisfies, for every $\varphi \in \{H^1(S_0^{\ell,0}(\mathcal{T})), \varphi(T) = 0\}$,

$$\int_0^T -\langle \varphi', U\rangle + \langle AU, \varphi\rangle = \sum_{j=0}^{\jmath-1} \left\langle g_j, \varphi(\tilde{t}_j)\right\rangle + \int_0^T \langle f, \varphi\rangle \,.$$

We combine the results in Section 3.3 and in 7.2 to obtain the following theorem.

Theorem 7.12. *Assume $u \in L^2(H^m)$, with $1 \le m \le \ell + 1$. Then we have*

$$\int_0^T |u(t) - U(t)|_{1,2;\Omega}^2 \lesssim \int_0^T \sum_{K\in\mathcal{T}} \operatorname{diam}(K)^{2m-2} \, |u(t)|_{m,2;K}^2 \,.$$

The hidden constant depends on the H^1-norm of the L^2-projection on $S_0^{\ell,0}(\mathcal{T})$, the coercivity and continuity constants of the parabolic problem, the dimension d, the polynomial degree ℓ and the shape parameter $\sigma_\mathcal{T}$.

Proof. We recall that Theorem 3.10 in particular states that

$$\int_0^T |u(t) - U(t)|^2_{1,2;\Omega} \lesssim \inf_{V \in L^2(S_0^{\ell,0}(\mathcal{T}))} \int_0^T |u(t) - V(t)|^2_{1,2;\Omega} \,.$$

The assertion follows from Corollary 7.6, by inserting

$$\forall t \in [0,T], \qquad V(t) := \Pi_0 u(t)$$

in the infimum on the right-hand side. □

7.4.2 Semidiscretization in time

We recall that $\mathcal{P}$ is a partition $0 = t_0 < t_1 < \ldots < t_N = T$ of $(0,T)$ into subintervals $I_n = [t_{n-1}, t_n)$, subordinate to $0 = \tilde{t_0} < \tilde{t_1} < \ldots < \tilde{t_J} = T$ and that

$$\begin{aligned} \mathcal{S}^{1,0}(\mathcal{P}, H^1_0) &= \{v \in C^0(H^1_0), v|_{I_n} \in \mathbb{P}^1(I_n, H^1_0), n = 1, \ldots, N\}, \\ \mathcal{S}^{0,-1}(\mathcal{P}, H^1_0) &= \{\phi \in L^2(H^1_0), \phi|_{I_n} \in H^1_0, n = 1, \ldots, N\}. \end{aligned}$$

The semidiscrete solution $U \in \mathcal{S}^{0,-1}(\mathcal{P}, H^1_0)$ satisfies, for every $\varphi \in \{\phi \in \mathcal{S}^{1,0}(\mathcal{P}, H^1_0), \phi(T) = 0\}$,

$$\sum_{n=1}^N \int_{I_n} -\langle \varphi', U\rangle + \langle AU, \varphi(t_{n-1})\rangle = \sum_{j=0}^{J-1} \left\langle g_j, \varphi(\tilde{t_j})\right\rangle + \sum_{n=1}^N \int_{I_n} \langle f, \varphi(t_{n-1})\rangle \,.$$

Moreover we recall that the discrete test space is endowed with

$$\|\varphi\|^2_{2,\mathcal{P}} = \sum_{n=1}^N \int_{I_n} \|\varphi'\|^2_{-1;\Omega} + |\varphi(t_{n-1})|^2_{1,2;\Omega} \,.$$

We observe that, given a Hilbert space Y,

$$\begin{aligned} \|\varphi - \varphi(t_{n-1})\|^2_{L^2(I_n,Y)} &= \int_{I_n} \frac{(t - t_{n-1})^2}{\tau_n^2} \|\varphi(t_n) - \varphi(t_{n-1})\|^2_Y \\ &= \frac{\tau_n}{3} \|\varphi(t_n) - \varphi(t_{n-1})\|^2_Y \,. \end{aligned} \tag{7.30}$$

In case $Y = H^{-1}(\Omega)$, we have

$$\begin{aligned} \sum_{n=1}^N \tau_n^{-2} \|\varphi - \varphi(t_{n-1})\|^2_{L^2(I_n,H^{-1})} &\le \frac{1}{3} \sum_{n=1}^N \int_{I_n} \left\| \frac{\varphi(t_n) - \varphi(t_{n-1})}{\tau_n} \right\|^2_{-1;\Omega} \\ &\le \frac{1}{3} \|\varphi\|^2_{2,\mathcal{P}} \,. \end{aligned} \tag{7.31}$$

Exploiting the results in Section 4.2.1 we get the following theorem.

Theorem 7.13. *Assume that, for every $n = 1, \dots, N$, $u|_{I_n} \in H^1(I_n, H^1_0)$. Then,*

$$\|u - U\|^2_{L^2(H^1_0)} \lesssim \sum_{n=1}^{N} \tau_n^2 \int_{I_n} |u'|^2_{1,2;\Omega}\,.$$

The hidden constant depends on the coercivity and continuity constants of the parabolic problem.

Proof. We apply Proposition 4.9, which states that

$$\|u - U\|_{L^2(H^1_0)} \lesssim \inf_{v \in \mathcal{S}^{0,-1}(\mathcal{P}, H^1_0)} \|u - v\|_{L^2(H^1_0)} + \sup_{\substack{\varphi \in \mathcal{S}^{1,0}(\mathcal{P}, H^1_0) \\ \varphi(T) = 0}} \frac{\displaystyle\sum_{n=1}^{N} \int_{I_n} \langle Au - f, \varphi - \varphi(t_{n-1})\rangle}{\|\varphi\|_{2,\mathcal{P}}}. \tag{7.32}$$

We insert v in the infimum on the right-hand side of (7.32) such that, for every $n = 1, \dots, N$,

$$v|_{I_n} = \frac{1}{\tau_n} \int_{I_n} u,$$

and we exploit the Poincaré inequality on every I_n. Concerning the supremum on the right-hand side of (7.32), we exploit Cauchy-Schwarz inequality for integrals and for sums to get

$$\begin{aligned}\sum_{n=1}^{N} \int_{I_n} \langle Au - f, \varphi - \varphi(t_{n-1})\rangle &\le \sum_{n=1}^{N} \int_{I_n} |Au - f|_{1,2;\Omega} \, \|\varphi - \varphi(t_{n-1})\|_{-1;\Omega} \\ &\le \left(\sum_{n=1}^{N} \tau_n^2 \, \|Au - f\|^2_{L^2(I_n, H^1)} \right)^{1/2} \left(\sum_{n=1}^{N} \tau_n^{-2} \, \|\varphi - \varphi(t_{n-1})\|^2_{L^2(I_n, H^{-1})} \right)^{1/2}.\end{aligned}$$

We recall that $Au - f = -u'$ on I_n. The thesis follows by (7.31) and taking the supremum over $\varphi \in \{\phi \in \mathcal{S}^{1,0}(\mathcal{P}, H^1_0),\ \phi(T) = 0\}$. □

7.4.3 Varying the spatial discretization

We consider a sequence $\{\mathcal{T}_n\}_{n=0}^N$ of triangulations that belongs to a family for which the L^2-projection is uniformly H^1-stable. The sequence of finite-dimensional spaces $\{\mathbb{V}_n\}_{n=1}^N \subset H^1_0(\Omega)$ is given by $\{S^{\ell,0}_0(\mathcal{T}_n)\}_{n=1}^N$. The L^2-projection onto $S^{\ell,0}_0(\mathcal{T}_n)$ is denoted by P_n and P_n^+ indicates the L^2-projection

onto $S_0^{\ell,0}(\mathcal{T}_n) \oplus S_0^{\ell,0}(\mathcal{T}_{n+1})$. We recall that

$$\mathcal{S}^{L^2}(\mathcal{P},\mathbb{V}) = \{v \in L^2(H_0^1),\ v|_{I_n} \in L^2(S_0^{\ell,0}(\mathcal{T}_n)),\ n = 1,\dots,N\}$$
$$\mathcal{S}^{H^1}(\mathcal{P},\mathbb{V}) = \{v \in L^2(H_0^1),\ v|_{I_n} \in H^1(S_0^{\ell,0}(\mathcal{T}_n)),\ n = 1,\dots,N\}.$$

The semidiscrete solution $U \in \mathcal{S}^{L^2}(\mathcal{P},\mathbb{V})$ satisfies, for every $\varphi \in \{\phi \in \mathcal{S}^{H^1}(\mathcal{P},\mathbb{V}),\ \phi(t_n^-) = P_n\phi(t_n), n = 1,\dots,N-1,\ \phi(T) = 0\}$,

$$\sum_{n=1}^{N} \int_{I_n} -\langle \varphi', U\rangle + \langle AU, \varphi\rangle = \sum_{j=1}^{\jmath-1} \left\langle g_j, \varphi(\tilde{t}_j)\right\rangle + \int_0^T \langle f, \varphi\rangle\,.$$

We resort to the results in Section 5.2 and derive the following theorem.

Theorem 7.14. *Assume $1 \le m \le \ell+1$, and assume that, for every $n = 1,\dots,N$, $u|_{I_n} \in C^0(I_n, H^m)$. Then we have*

$$\|u-U\|^2_{L^2(H_0^1)} \lesssim \sum_{n=1}^{N} \int_{I_n} \operatorname{diam}(K)^{2m-2} |u(t)|^2_{m,2;K}\,\mathrm{d}t + \sum_{n=1}^{N-1} \sum_{K\in\mathcal{T}_n} \operatorname{diam}(\omega_K)^{2m} |u(t_n^-)|^2_{m,2;\omega_K}.$$

The hidden constant depends on the maximum of the H^1-norms of the L^2-projection on $S_0^{\ell,0}(\mathcal{T}_n)$, the coercivity and continuity constants of the parabolic problem, the dimension d, the polynomial degree ℓ and the shape parameters $\sigma_{\mathcal{T}_n}$.

Proof. The proof mimics the one of Theorem 7.9. We exploit Theorem 5.10, which states that

$$\|u-U\|^2_{L^2(H_0^1)} \lesssim \inf_{v\in\mathcal{S}^{L^2}(\mathcal{P},\mathbb{V})} \|u-v\|^2_{L^2(H_0^1)} + \sum_{n=1}^{N-1} \left\|P_n^+(I-P_n)u(t_n^-)\right\|^2_{0,2;\Omega}. \tag{7.33}$$

In the infimum on the right-hand side of (7.33), we insert $v \in \mathcal{S}^{L^2}(\mathcal{P},\mathbb{V})$ such that, for every $n = 1,\dots,N$,

$$\forall t \in I_n \qquad v(t) = \Pi_0^n u(t).$$

The terms $\|P_n^+(I-P_n)u(t_n^-)\|_{0,2;\Omega}$ can be bounded as in (7.27). □

7.4.4 Full discretization with the backward-Euler Galerkin method

Assume $\mathcal{P}$ to be as in Section 7.4.2 and $\{\mathcal{T}_n\}_{n=0}^N$ to be as in Section 7.4.3. We recall that

$$\mathcal{S}^{0,-1}(\mathcal{P},\mathbb{V}) = \{v \in L^2(H_0^1), v|_{I_n} \in S_0^{\ell,0}(\mathcal{T}_n), n = 1,\dots,N\},$$

and we set

$$\mathcal{S}^{1,0^-}(\mathcal{P},\mathbb{V}) = \{\varphi \in L^2(H_0^1), \varphi|_{I_n} \in \mathbb{P}^1(I_n, S_0^{\ell,0}(\mathcal{T}_n)), \\ \varphi(t_n^-) = P_n\varphi(t_n),\ n = 1,\dots,N\}.$$

The discrete solution $U \in \mathcal{S}^{0,-1}(\mathcal{P},\mathbb{V})$ satisfies

$$\sum_{n=1}^N \int_{I_n} -\langle \varphi', U\rangle + \langle AU, \varphi(t_{n-1})\rangle = \sum_{j=0}^{\jmath-1} \langle g_j, \varphi(\tilde{t}_j)\rangle + \sum_{n=1}^N \int_{I_n} \langle f, \varphi(t_{n-1})\rangle$$

for every $\varphi \in \{\phi \in \mathcal{S}^{1,0^-}(\mathcal{P},\mathbb{V}),\ \phi(T) = 0\}$. In view of the results in Section 6.2, we get the following theorem.

Theorem 7.15. *Assume $1 \le m \le \ell + 1$. Assume that, for every $n = 1,\dots,N$, $u|_{I_n} \in C^0(I_n, H^m) \cap H^1(I_n, H_0^1)$. Then*

$$\|u - U\|^2_{L^2(H_0^1)} \\ \lesssim \sum_{n=1}^N \tau_n^2 \int_{I_n} |u'|^2_{1,2;\Omega} + \sum_{K\in\mathcal{T}_n} \operatorname{diam}(\omega_K)^{2m-2} \int_{I_n} |u|^2_{m,2;\omega_K} \\ + \sum_{n=1}^{N-1} \sum_{K\in\mathcal{T}_n} \operatorname{diam}(\omega_K)^{2m} \left|u(t_n^-)\right|^2_{m,2;\omega_K}.$$

The hidden constant depends on the maximum of the H^1-norms of the L^2-projection on $S_0^{\ell,0}(\mathcal{T}_n)$, the coercivity and continuity constants of the parabolic problem, the dimension d, the polynomial degree ℓ, the shape parameters $\sigma_{\mathcal{T}_n}$.

Proof. We exploit Theorem 6.11, which states that

$$\|u - U\|^2_{L^2(H_0^1)} \\ \lesssim \inf_{v\in\mathcal{S}^{0,-1}(\mathcal{P},\mathbb{V})} \|u - v\|^2_{L^2(H_0^1)} + \sum_{n=1}^N \int_{I_n} \left|u - P_n u(t_n^-)\right|^2_{1,2;\Omega} \\ + \sum_{n=1}^{N-1} \left\|P_n^+(I - P_n)u(t_n^-)\right\|^2_{0,2;\Omega}.$$

Concerning the infimum on the right-hand side, we insert $v \in \mathcal{S}^{0,-1}(\mathcal{P}, \mathbb{V})$, such that, for $n = 1, \dots, N$,

$$v|_{I_n} = \Pi_0^n \frac{1}{\tau_n} \int_{I_n} u.$$

Because of the stability of Π_0 in H^1, we have

$$\begin{aligned}
&\left\| u - \Pi_0^n \frac{1}{\tau_n} \int_{I_n} u \right\|^2_{L^2(I_n, H_0^1)} \\
&\quad \le 2 \left\| u - \Pi_0^n u \right\|^2_{L^2(I_n, H_0^1)} + 2 \left\| \Pi_0^n \left(u - \frac{1}{\tau_n} \int_{I_n} u \right) \right\|^2_{L^2(I_n, H_0^1)} \\
&\quad \lesssim \left\| u - \Pi_0^n u \right\|^2_{L^2(I_n, H_0^1)} + \left\| u - \frac{1}{\tau_n} \int_{I_n} u \right\|^2_{L^2(I_n, H_0^1)} \\
&\quad \lesssim \int_{I_n} \sum_{K \in \mathcal{T}_n} \operatorname{diam}(\omega_K)^{2m-2} \left| u \right|^2_{m,2;\omega_K} + \tau_n^2 \int_{I_n} \left| u' \right|^2_{1,2;\Omega} . \qquad (7.34)
\end{aligned}$$

Regarding $\int_{I_n} |u - P_n u(t_n^-)|^2_{1,2;\Omega}$ we insert $P_n u$ and by stability of P_n in H^1 we get, for $n = 1, \dots, N$,

$$\begin{aligned}
\int_{I_n} \left| u - P_n u(t_n^-) \right|^2_{1,2;\Omega} &\le 2 \int_{I_n} |u - P_n u|^2_{1,2;\Omega} + \left| P_n u - P_n u(t_n^-) \right|^2_{1,2;\Omega} \\
&\lesssim \int_{I_n} |u - \Pi_0^n u|^2_{1,2;\Omega} + \int_{I_n} \left| u - u(t_n^-) \right|^2_{1,2;\Omega} \\
&\lesssim \int_{I_n} \sum_{K \in \mathcal{T}_n} \operatorname{diam}(\omega_K)^{2m-2} \left| u \right|^2_{m,2;\omega_K} + \tau_n^2 \int_{I_n} \left| u' \right|^2_{1,2;\Omega} .
\end{aligned}$$

Regarding the terms $\| P_n^+ (I - P_n) u(t_n^-) \|_{0,2;\Omega}$, they can be bounded as in (7.27). □

7.5 Additional notation

We introduce some more notation, in order to derive error bounds in terms of fractional regularity.

For $0 < s < 1$, we indicate with $H^s(\Omega)$ the space of functions for which

$$|f|^2_{s,2;\Omega} := (1 - s) \int_{\Omega \times \Omega} \frac{|f(x) - f(y)|^2}{|x - y|^{d+2s}} \, \mathrm{d}x \, \mathrm{d}y$$

is finite, and we endow it with $\|\cdot\|^2_{s,2;\Omega} := \|\cdot\|^2_{0,2;\Omega} + |\cdot|^2_{s,2;\Omega}$. The coefficient $\sqrt{1-s}$ is motivated by the results in [9], where it is proven that, for a smooth domain Ω,

$$\lim_{s\to 1}(1-s)\int_{\Omega\times\Omega}\frac{|f(x)-f(y)|^2}{|x-y|^{d+2s}}\,\mathrm{d}x\,\mathrm{d}y \approx |f|^2_{1,2;\Omega}\,.$$

The space $H^s_0(\Omega)$ denotes the completion of $C^\infty_0(\Omega)$ with respect to $\|\cdot\|_{s,2;\Omega}$. We recall that $H^s_0(\Omega) = H^s(\Omega)$ for $s < 1/2$, see [29, Cor. 1.4.4.5], while $H^s_0(\Omega) = \{f \in H^s(\Omega),\ f|_{\partial\Omega} = 0\}$, for $s > 1/2$, see [29, Cor. 1.5.1.6]. The dual space $H^{-s}(\Omega)$ of $H^s_0(\Omega)$ is endowed with

$$\|f\|_{-s;\Omega} := \sup_{\varphi\in H^s(\Omega)}\frac{\langle f,\varphi\rangle}{\|\varphi\|_{s,2;\Omega}}.$$

If $\theta \in \mathbb{R}^+$, and $\theta = m + s$ with $m \in \mathbb{N}$, $s \in (0,1)$, the space $H^\theta(\Omega)$ is defined by

$$H^\theta(\Omega) := \{f \in H^m(\Omega),\ |D^\alpha f|_{s,2;\Omega} < \infty, \forall|\alpha| = m\}.$$

We set

$$|f|^2_{\theta,2;\Omega} := \sum_{|\alpha|=m}|D^\alpha f|^2_{s,2;\Omega}$$

and we endow $H^\theta(\Omega)$ with $\|f\|^2_{\theta,2;\Omega} := \|f\|^2_{m,2;\Omega} + |f|^2_{\theta,2;\Omega}$.

We remark that, for every $m \in \mathbb{N}$, and for every ω_1, $\omega_2 \subset \Omega$, such that $|\omega_1 \cap \omega_2| = 0$, it holds

$$\|\cdot\|^2_{m,2;\omega_1} + \|\cdot\|^2_{m,2;\omega_2} = \|\cdot\|^2_{m,2;\omega_1\cup\omega_2}\,.$$

However, for noninteger θ, in general we can only affirm that

$$\|\cdot\|^2_{\theta,2;\omega_1} + \|\cdot\|^2_{\theta,2;\omega_2} \le \|\cdot\|^2_{\theta,2;\omega_1\cup\omega_2}\,. \tag{7.35}$$

Given a Hilbert space Y and a proper time interval J, we define for $s \in (0,1)$ the space $H^s(J,Y)$ as the subspace of $L^2(J,Y)$ of those functions for which

$$|f|^2_{H^s(J,Y)} := \int_{J\times J}\frac{\|f(t)-f(\tau)\|^2_Y}{|t-\tau|^{2s+1}}\,\mathrm{d}t\,\mathrm{d}\tau$$

is finite.

The real method of interpolation

Let (B_0, B_1) be a couple of Banach spaces with $B_1 \subset B_0$. For every $t > 0$ and every $f \in B_0$, the K-functional of the couple (B_0, B_1) is given by

$$K(f, t, B_0, B_1) := K(f, t) := \inf_{g \in B_1} \|f - g\|_{B_0} + t \|g\|_{B_1} .$$

We recall that, as a function of t, $K(f, t)$ is increasing and subadditive, see [18, Ch. 6, Prop. 1.1].

Given two parameters $s \in (0, 1)$ and $q \in [1, \infty)$, the intermediate spaces $(B_0, B_1)_{s,q}$ are defined by

$$(B_0, B_1)_{s,q} := \{f \in B_0, \|f\|_{(B_0,B_1)_{s,q}} < \infty\},$$

where

$$\|f\|^q_{(B_0,B_1)_{s,q}} := s(1-s)q \int_0^\infty \left(t^{-s} K(f, t)\right)^q \frac{\mathrm{d}t}{t}.$$

As in [34], we put the non-standard coefficient $s(1-s)q$ in the definition of $\|\cdot\|^q_{(B_0,B_1)_{s,q}}$. This guarantees that

$$\|f\|_{B_0} = \|f\|_{(B_0,B_0)_{s,q}} , \qquad \forall s \in (0, 1), q \in [1, \infty).$$

Moreover, the following result, see [6, Sect. 3.5], holds with $C = 1$:

$$\forall f \in B_1, \qquad \|f\|_{(B_0,B_1)_{s,q}} \le C \|f\|^{1-s}_{B_0} \|f\|^s_{B_1} . \tag{7.36}$$

In fact, exploiting [1, Thm. 7.16 (a), Lemma 7.19 (b)] it holds, for every $f \in B_1$ and $t > 0$,

$$\|f\|_{(B_0,B_1)_{s,q}} \le t^{-s} \max\{\|f\|_{B_0} , t \|f\|_{B_1}\}.$$

Choosing $t = \|f\|_{B_0} / \|f\|_{B_1}$ as in [6], we get (7.36) with $C = 1$.

We also recall the following result about interpolation of operators. See, for example, [6, Thm. 3.1.2].

Lemma 7.16. *Assume that (A_0, A_1) and (B_0, B_1) are two pairs of Banach spaces as above, and that T is a linear operator that maps A_i to B_i, $i = 0$, 1. Then T maps $A_{s,q}$ to $B_{s,q}$ and there holds*

$$\|T\|_{\mathcal{L}(A_{s,q},B_{s,q})} \le \|T\|^{1-s}_{\mathcal{L}(A_0,B_0)} \|T\|^s_{\mathcal{L}(A_1,B_1)} .$$

Moduli of continuity and K-functional

We recall that the modulus of continuity of a function $f \in L^2(\Omega)$ is defined by

$$\omega(f,t)_2 := \sup_{|h|<t} \|\Delta_h(f)\|_{0,2;\Omega}\,, \qquad t \geq 0,$$

where Δ_h is the difference operator:

$$\Delta_h(f)(x) := \begin{cases} f(x+h) - f(x) & \text{if } x, x+h \in \Omega, \\ 0 & \text{otherwise.} \end{cases}$$

We recall that, as a function of t, $\omega(f,t)_2$ is non-decreasing and subadditive, see [18, Ch. 2, Sect. 6]. We also recall the averaged modulus of continuity, that is given by

$$w(f,t)_2 := \left(\frac{1}{t^d} \int_{|h|<t} \|\Delta_h(f)\|_{0,2;\Omega}^2 \, \mathrm{d}h \right)^{1/2}, \qquad t > 0.$$

A straightforward inequality yields, for every $t > 0$,

$$w(f,t)_2^2 \leq \min\left\{1, \frac{\operatorname{diam}(\Omega)^d}{t^d}\right\} C(d)\omega(f,t)_2^2. \tag{7.37}$$

Concerning the converse inequality, if $\Omega \subset \mathbb{R}$ is an interval, we have, see [18, Ch. 6, Lemma 5.1],

$$\omega(f,t)_2 \leq C w(f,t)_2, \qquad \forall t \leq |\Omega|. \tag{7.38}$$

The modulus of continuity is related to the variant of the K-functional of the couple (L^2, H^1), where the H^1-norm is replaced by the H^1-seminorm. More precisely, if $f \in L^2(\Omega)$ and $t > 0$, we set

$$K_{|\cdot|}(f,t,L^2,H^1) := \inf_{g \in H^1} \|f-g\|_{0,2;\Omega} + t\,|g|_{1,2;\Omega}\,.$$

If $\Omega \subset \mathbb{R}$ is an interval or if $\Omega \subset \mathbb{R}^d$ is a Lipschitz domain, there hold, see [18, Ch. 6, Thm. 2.4] and [31, Thm. 1],

$$\forall t \in (0,1) \qquad \omega(f,t)_2 \gtrsim K_{|\cdot|}(f,t,L^2(\Omega),H^1(\Omega)), \tag{7.39a}$$

$$\forall t \in (0,\infty) \qquad \omega(f,t)_2 \lesssim K_{|\cdot|}(f,t,L^2(\Omega),H^1(\Omega)). \tag{7.39b}$$

The hidden constant in (7.39a) depends on the geometry of Ω.

We recall that $L^2_0(\Omega)$ denotes the subspace of $L^2(\Omega)$ of functions with mean value zero. We provide a proof of

$$(L^2_0(\Omega), L^2_0(\Omega) \cap H^1(\Omega))_{s,2} = L^2_0(\Omega) \cap H^s(\Omega),$$

see [11, Thm. 14.2.3], in the case $\Omega \subset \mathbb{R}$ is an interval, in order to highlight how the constants depend on s.

Lemma 7.17. *Let $\Omega \subset \mathbb{R}$ be an interval with* $\operatorname{diam}(\Omega) \leq 1$. *Then,*

$$(L^2_0(\Omega), L^2_0(\Omega) \cap H^1(\Omega))_{s,2} = L^2_0(\Omega) \cap H^s(\Omega),$$

where $L^2_0(\Omega) \cap H^1(\Omega)$ is endowed with $|\cdot|_{1,2;\Omega}$. Furthermore, for every $f \in L^2_0(\Omega) \cap H^s(\Omega)$,

$$\|f\|_{(L^2_0(\Omega), H^1(\Omega)\cap L^2_0(\Omega))_{s,2}} \leq C\,|f|_{s,2;\Omega}\,, \tag{7.40}$$

where the constant is independent of s, and

$$|f|_{s,2;\Omega} \leq C(s)\,\|f\|_{(L^2_0(\Omega), H^1(\Omega)\cap L^2_0(\Omega))_{s,2}}\,, \tag{7.41}$$

where $C(s) \lesssim s^{-1/2}$ for $s \to 0$.

Proof. We start with (7.40). We first notice that, for every $t \geq \operatorname{diam}(\Omega)$ and for every $g \in H^1(\Omega) \cap L^2_0(\Omega)$, the Poincaré inequality yields

$$K(f,t) \leq \|f\|_{0,2;\Omega} \leq \|f-g\|_{0,2;\Omega} + \|g\|_{0,2;\Omega} \leq \|f-g\|_{0,2;\Omega} + \operatorname{diam}(\Omega)\,|g|_{1,2;\Omega}\,.$$

Thus $K(f,t) \leq K(f, \operatorname{diam}(\Omega))$ and thanks also the subadditivity and monotonicity of $K(f,\cdot)$, we have

$$\begin{aligned} s\int_{\operatorname{diam}(\Omega)}^{\infty} t^{-2s-1}K(f,t)^2\,\mathrm{d}t &\leq \frac{\operatorname{diam}(\Omega)^{-2s}}{2}K(f,\operatorname{diam}(\Omega))^2 \\ &\leq 2\operatorname{diam}(\Omega)^{-2s}K\left(f, \frac{\operatorname{diam}(\Omega)}{2}\right)^2 \\ &\leq \frac{4s}{2^{2s}-1}\int_{\frac{\operatorname{diam}(\Omega)}{2}}^{\operatorname{diam}(\Omega)} t^{-2s-1}K(f,t)^2\,\mathrm{d}t \\ &\leq \frac{2}{\ln 2}\int_0^{\operatorname{diam}(\Omega)} \left(t^{-s}K(f,t)\right)^2 \frac{\mathrm{d}t}{t}. \end{aligned} \tag{7.42}$$

Therefore,

$$\|f\|^2_{(L^2_0(\Omega), H^1(\Omega)\cap L^2_0(\Omega))_{s,2}} \leq C(1-s)\int_0^{\operatorname{diam}(\Omega)} \left(t^{-s}K(f,t)\right)^2 \frac{\mathrm{d}t}{t}.$$

We observe that, for $f \in L^2_0(\Omega)$, we have $K(f,t,L^2_0(\Omega), L^2_0(\Omega)\cap H^1) = K_{|\cdot|}(f,t,L^2,H^1)$. We can exploit thus (7.39a) and (7.38), to get

$$\|f\|^2_{(L^2_0(\Omega), H^1(\Omega)\cap L^2_0(\Omega))_{s,2}} \leq C(1-s)\int_0^{\infty} \left(t^{-s}w(f,t)_2\right)^2 \frac{\mathrm{d}t}{t}. \tag{7.43}$$

Finally, by Fubini's Theorem, see [19] for details, we have

$$|f|_{s,2;\Omega}^2 = \frac{1-s}{2s+d}\int_0^\infty [t^{-s}w(f,t)_2]^2\frac{\mathrm{d}t}{t}. \tag{7.44}$$

Combining (7.43)–(7.44) gives (7.40).

On the other hand, by (7.44), (7.37) and (7.39b), we have

$$|f|_{s,2;\Omega} \le C(1-s)\int_0^\infty t^{-2s}K(f,t)^2\frac{\mathrm{d}t}{t},$$

from which (7.41) follows readily. □

We notice that (7.44), (7.37) and (7.39b) also hold for Lipschitz domains in $\mathbb{R}^d$, and therefore also does (7.41).

Fractional Poincaré inequality in 1d

We apply Lemma 7.16 with $T = I$, $q = 2$, $(A_0, A_1) = (L_0^2(\Omega), H^1(\Omega)\cap L_0^2(\Omega))$, and $(B_0, B_1) = (L_0^2(\Omega), L_0^2(\Omega))$ where $\Omega \subset \mathbb{R}$ is an interval. We have

$$\|T\|_{\mathcal{L}(A_0,B_0)} = 1, \qquad \|T\|_{\mathcal{L}(A_1,B_1)} \le \frac{1}{\pi}\mathrm{diam}(\Omega).$$

Hence, for $s \in (0,1)$,

$$\|T\|_{\mathcal{L}(A_{s,2},B_{s,2})} \le \frac{1}{\pi^s}\mathrm{diam}(\Omega)^s,$$

and, for every $f \in (L_0^2(\Omega), H^1(\Omega)\cap L_0^2(\Omega))_{s,2}$,

$$\|f\|_{0,2;\Omega} = \|f\|_{(L_0^2(\Omega),L_0^2(\Omega))_{s,2}} \le \frac{1}{\pi^s}\mathrm{diam}(\Omega)^s\,\|f\|_{(L_0^2(\Omega),H^1(\Omega)\cap L_0^2(\Omega))_{s,2}}.$$

Taking into account (7.40) we get, for every $f \in H^s(\Omega)$ with $\int_\Omega f = 0$,

$$\|f\|_{0,2;\Omega} \le C\mathrm{diam}(\Omega)^s\,|f|_{s,2;\Omega}. \tag{7.45}$$

Useful inequalities

The following lemma helps bounding the H^s-seminorm of a product on a finite element star ω.

Lemma 7.18. *Assume ω is a star of $\mathcal{T}$. Assume $f \in W^{1,\infty}(\omega)$ and $g \in H^s(\omega)$ with $s \in (0,1)$. Then $fg \in H^s(\omega)$ with*

$$|fg|_{s,2;\omega}^2 \le 2\,\|f\|_{0,\infty;\omega}^2\,|g|_{s,2;\omega}^2 + C(d,\sigma_{\mathcal{T}})(\mathrm{diam}(\omega))^{2-2s}\,\|\nabla f\|_{0,\infty;\omega}^2\,\|g\|_{0,2;\omega}^2. \tag{7.46}$$

Proof. Adding and subtracting $f(x)g(y)$ in the definition of $|fg|^2_{s,2;\omega}$ we get

$$|fg|^2_{s,2;\omega} \leq 2(1-s)\int_{\omega\times\omega} \frac{|f(x)|^2|g(x)-g(y)|^2}{|x-y|^{d+2s}}\,\mathrm{d}(x,y) + 2(1-s)\int_{\omega\times\omega} \frac{|g(y)|^2|f(x)-f(y)|^2}{|x-y|^{d+2s}}\,\mathrm{d}(x,y).$$

We bound the two terms on the right-hand side separately. First we have

$$2(1-s)\int_{\omega\times\omega} \frac{|f(x)|^2|g(x)-g(y)|^2}{|x-y|^{d+2s}}\,\mathrm{d}(x,y) \leq 2\,\|f\|^2_{0,\infty;\omega}\,|g|^2_{s,2;\omega}\,.$$

Concerning the second term we observe that

$$\sup_{x,y\in\omega} \frac{|f(x)-f(y)|}{|x-y|} \leq C(\sigma_{\mathcal{T}})\,\|\nabla f\|_{0,\infty;\omega}\,.$$

In fact, if the segment $[x,y]\subset\omega$ we can apply the Mean Value Theorem and get

$$|f(x)-f(y)| \leq \|\nabla f\|_{0,\infty;\omega}\,|x-y|.$$

Otherwise, let K_x and K_y be two elements such that $x\in K_x$, and $y\in K_y$. Following the ideas in [37, Lemma 3.4, Case 2], we take $m\in K_x\cap K_y$ such that the convex angle α between $x-m$ and $y-m$ is maximum. This angle is bounded away from zero in terms of $\sigma_{\mathcal{T}}$. By the Cosine Theorem we get

$$(1-\max\{0,\cos\alpha\})(|x-m|^2+|y-m|^2) \leq |x-y|^2,$$

so that

$$\begin{aligned}|f(x)-f(y)| &\leq |f(x)-f(m)|+|f(m)-f(y)|\\ &\leq \|\nabla f\|_{0,2;\omega}\,(|x-m|+|y-m|)\\ &\leq C(\sigma_{\mathcal{T}})\,\|\nabla f\|_{0,2;\omega}\,|x-y|.\end{aligned}$$

Therefore, we have

$$\begin{aligned}&2(1-s)\int_{\omega\times\omega} \frac{|g(y)|^2|f(x)-f(y)|^2}{|x-y|^{d+2s}}\,\mathrm{d}(x,y)\\ &\quad\leq 2(1-s)\sup_{x,y\in\omega}\frac{|f(x)-f(y)|^2}{|x-y|^2}\int_\omega |g(y)|^2\int_\omega |x-y|^{2-d-2s}\,\mathrm{d}x\,\mathrm{d}y\\ &\quad\leq C(d,\sigma_{\mathcal{T}})(1-s)\,\|\nabla f\|^2_{0,\infty,\omega}\int_\omega |g(y)|^2\int_0^{\mathrm{diam}(\omega)}\rho^{2-2s-1}\,\mathrm{d}\rho\,\mathrm{d}y\\ &\quad\leq C(d,\sigma_{\mathcal{T}})(\mathrm{diam}(\omega))^{2-2s}\,\|\nabla f\|^2_{0,\infty;\omega}\,\|g\|^2_{0,2;\omega}\,.\end{aligned}$$

□

For $s \in (0,1)$, the space $H^1(\Omega)$ is embedded in $H^s(\Omega)$, see, for example, [20]. We provide a proof, in order to underline that the constant is independent of s.

Lemma 7.19. *Let $\Omega \subset \mathbb{R}^d$ be a Lipschitz domain. Then, there exists a constant that depends on Ω but independent of s, such that, for every $f \in H^1(\Omega)$ and $s \in (0,1)$,*

$$|f|_{s,2;\Omega} \leq C \, \|f\|_{1,2;\Omega} \, .$$

Proof. We recall (7.44), which also holds for Lipschitz domain, and we exploit (7.37), to get

$$|f|^2_{s,2;\Omega} \leq C(d)(1-s)\Big(\omega(f, \mathrm{diam}(\Omega))^2 \mathrm{diam}(\Omega)^{-2s} + \int_0^{\mathrm{diam}(\Omega)} t^{-2s-1} \omega(f,t)_2^2 \, \mathrm{d}t\Big).$$

By the monotonicity and subadditivity of $\omega(f,\cdot)$, reasoning similarly as in (7.42), we get

$$\begin{aligned} \omega(f, \mathrm{diam}(\Omega))^2 \mathrm{diam}(\Omega)^{-2s} &\leq C \int_0^{\mathrm{diam}(\Omega)} t^{-2s-1} \omega(f,t)^2 \, \mathrm{d}t \\ &\leq C \max\{1, \mathrm{diam}(\Omega)^{2-2s}\} \int_0^1 t^{-2s-1} \omega(f,t)^2 \, \mathrm{d}t. \end{aligned}$$

The thesis follows by (7.39b) and

$$K(f,t) \leq t \, \|f\|_{1,2;\Omega} \, , \qquad \forall t \in (0,1), \ f \in H^1(\Omega). \qquad \square$$

We conclude with an upper bound for $\|f\|_{s,2;\Omega}$, in the spirit of (7.36).

Lemma 7.20. *Assume $s \in (0,1)$. For every $f \in H^1(\Omega)$ we have*

$$\|f\|_{s,2;\Omega} \leq C(s) \left(\|f\|_{0,2;\Omega} + \|f\|^{1-s}_{0,2;\Omega} \, |f|^s_{1,2;\Omega} \right), \tag{7.47}$$

where the constant grows with $s^{-1/2}$.

Proof. As in the proof of (7.41) we get

$$|f|_{s,2;\Omega} \lesssim \frac{1}{\sqrt{s}} \, \|f\|_{(L^2,H^1)_{s,2}} \, .$$

Thanks to (7.36) we obtain

$$\begin{aligned} |f|_{s,2;\Omega} &\lesssim \frac{1}{\sqrt{s}} \, \|f\|^{1-s}_{0,2;\Omega} \, \|f\|^s_{1,2;\Omega} \\ &\lesssim \frac{1}{\sqrt{s}} \left(\|f\|_{0,2;\Omega} + \|f\|^{1-s}_{0,2;\Omega} \, |f|^s_{1,2;\Omega} \right). \end{aligned}$$

$\square$

Simplices and barycentric coordinates

We recall that a d-simplex $K \subset \mathbb{R}^d$ is the convex hull of the $d+1$ points $a_j = (a_{ij})_{i=1}^d$, $j = 0, \dots, d$:

$$K := \left\{ x = \sum_{j=0}^{d} \lambda_j a_j, 0 \le \lambda_j \le 1, j = 0, \dots, d, \sum_{j=0}^{d} \lambda_j = 1 \right\}.$$

The points a_j are called the vertices of the simplex and are such that $a_0 - a_1, \dots, a_0 - a_d$ are linearly independent.

The barycentric coordinates $\lambda_j = \lambda_j(x)$, $0 \le j \le d$, of any point $x = (x_i)_{i=1}^d \in K$ are the unique solutions of the linear system

$$\begin{cases} \displaystyle\sum_{j=0}^{d} \lambda_j a_j = x_i, \quad i = 1, \dots, d \\ \displaystyle\sum_{j=0}^{d} \lambda_j = 1 \end{cases}.$$

We denote by $\Lambda_K : K \to \mathbb{R}^{d+1}$ the operator that maps any point in K to its barycentric coordinates:

$$x \mapsto \Lambda_K(x) := (\lambda_0(x), \dots, \lambda_d(x)).$$

Given a mutiindex $\alpha = (\alpha_0, \dots, \alpha_d)$, the integral of the barycentric monomial $\boldsymbol{\lambda}^\alpha = \Pi_{i=0}^d \lambda_i^{\alpha_i}$ over a simplex K is given by

$$\int_K \boldsymbol{\lambda}^\alpha = \frac{\alpha!}{(|\alpha| + d)!} d! |K|, \tag{7.48}$$

where $\alpha! := \Pi_{i=0}^d \alpha_i!$, and $|\alpha| := \sum_{i=0}^d \alpha_i$.

Reference stars

Because of (7.35), when bounding in terms of the H^θ-(semi)norm, it is in general not possible to split the norm on contributions from simplices. In (7.64) below we need the analogous of (7.2a) for stars ω_z, $z \in \mathcal{V}$. To this end, as in [37], we divide them in equivalence classes, and fix for each a reference star. First of all, there are interior stars, for which z is an interior point of ω_z, and boundary stars for which z is a boundary point of ω_z. By construction, z is a common vertex, that is, a vertex that is shared by all d-dimensional simplices of the star ω_z. For boundary stars, it may be that z is not the

only common vertex. Two stars ω_1 and ω_2 are topologically equivalent if and only if there exists a bijection $F : \omega_1 \to \omega_2$, such that F and its inverse F^{-1} are continuous and affine on each simplex. An equivalence class can be characterized by the number of d-dimensional simplices in the star and by the lists of vertices shared by every pair of the simplices. Since the number of simplices in a star of $\mathcal{T}$ is bounded uniformly in terms of the shape parameter $\sigma_{\mathcal{T}}$, the number of equivalence classes appearing in $\mathcal{T}$ is finite. For each such equivalence class, we fix a reference star with the common vertex, or one common vertex, in the origin. Moreover, if the equivalence class consists of interior stars, all other vertices are on the unit sphere S^{d-1}, otherwise they are on the semisphere $S^{d-1} \cap \mathbb{R}^d_+$ with $\mathbb{R}^d_+ := \{x = (x_1, \dots, x_d),\ x_i \geq 0 \text{ for } i = 1, \dots, d\}$.

Polynomial approximation in fractional Sobolev spaces

We conclude this section with an analogous of (7.6) for estimates in fractional order Sobolev spaces. Assume $\theta \in \mathbb{R}$, with $\theta = m + s$, $m \in \mathbb{N}$ and $s \in (0, 1)$. Then it holds, see [23, Prop. 6.1]

$$\inf_{P \in \mathbb{P}^{m+1}} \|f - P\|_{0,2;\omega} \leq C(m, d, \gamma, s) \mathrm{diam}(\omega)^{\theta} \, |f|_{\theta,2;\omega} \,. \tag{7.49}$$

The constant in (7.49) grows with $s^{-1/2}$ and $(1 - s)^{-1/2}$ if $s \to 0$ or $s \to 1$.

7.6 Interpolation and dual norms II

The interpolation operator of Section 7.2 is well-defined for functions in L^2. In order to approximate less regular functions, we introduce an interpolator operator with the same structure of Π_0, but that acts on H^{-1} and has values in $S_0^{1,0}(\mathcal{T})$.

For every $z \in \mathcal{V}_\Omega$ we define $\phi_z^* \in H_0^1(\omega_z)$ as

$$\forall K \subset \omega_z, \forall x \in K, \qquad \phi_z^*(x) := \frac{1}{|\omega_z|} p_\ell(\Lambda_K(x)), \tag{7.50a}$$

where p_ℓ is given by

$$p_\ell((\lambda_0, \dots, \lambda_d)) := \lambda_\ell \bigg(a_1 \lambda_\ell^2 + a_2 \sum_{\substack{j=0 \\ j \neq \ell}}^{d} \lambda_j^2 + a_3 \lambda_\ell \sum_{\substack{j=0 \\ j \neq \ell}}^{d} \lambda_j + a_4 \sum_{\substack{j=0 \\ j \neq \ell}}^{d-1} \sum_{\substack{i=j+1 \\ i \neq \ell}}^{d} \lambda_i \lambda_j \bigg). \tag{7.50b}$$

The index ℓ refers to the barycentric coordinate associate to the vertex z in K and the coefficients are given by

$$a_1 := d+1, \qquad a_2 := -\frac{1}{2}(d+1)(d+2)(d+5),$$
$$a_3 := \frac{1}{2}(d+1)(d^2+7d+16), \qquad a_4 := 2(d+1).$$

Remark 7.21. The definition (7.50b) of the polynomial p_ℓ ensures that

(i) $\displaystyle\sum_{\ell=0}^{d} p_\ell = d+1;$

(ii) $\displaystyle\int_K p_\ell \lambda_k = |K|\delta_{k\ell}$, for every $k = 0, \dots, d;$

(iii) $\phi_z^* \in H_0^1(\omega_z);$

(iv) $\|\phi_z^*\|_{0,2;K} = C(d)\dfrac{|K|^{1/2}}{|\omega_z|}$ and $\|\phi_z^*\|_{0,2;\omega_z} = \dfrac{C(d)}{|\omega_z|^{1/2}};$

(v) $|\phi_z^*|_{1,2;K} \le C(d)\dfrac{|K|^{1/2}}{|\omega_z|\rho_K}.$

Proof. Since $\sum_{\ell=0}^{d} \lambda_\ell = 1$ we have

$$\begin{aligned}
d+1 &= (d+1)\left(\sum_{\ell=0}^{d}\lambda_\ell\right)^3 \\
&= (d+1)\sum_{\ell=0}^{d}\lambda_\ell^3 + 3(d+1)\sum_{\substack{j,\ell=0\\ \ell\neq j}}^{d}\lambda_j^2\lambda_\ell + 6(d+1)\sum_{\ell=0}^{d-2}\sum_{j=\ell+1}^{d-1}\sum_{i=j+1}^{d}\lambda_i\lambda_j\lambda_\ell \\
&= a_1\sum_{\ell=0}^{d}\lambda_\ell^3 + (a_2+a_3)\sum_{\substack{j,\ell=0\\ j\neq\ell}}^{d}\lambda_j^2\lambda_\ell + 3a_4\sum_{\ell=0}^{d-2}\sum_{j=\ell+1}^{d-1}\sum_{i=j+1}^{d}\lambda_i\lambda_j\lambda_\ell \\
&= a_1\sum_{\ell=0}^{d}\lambda_\ell^3 + a_2\sum_{\ell=0}^{d}\sum_{\substack{j=0\\ j\neq\ell}}^{d}\lambda_j^2\lambda_\ell + a_3\sum_{\ell=0}^{d}\sum_{\substack{j=0\\ j\neq\ell}}^{d}\lambda_j\lambda_\ell^2 + a_4\sum_{\ell=0}^{d}\sum_{\substack{j=0\\ j\neq\ell}}^{d-1}\sum_{\substack{i=j+1\\ i\neq\ell}}^{d}\lambda_i\lambda_j\lambda_\ell \\
&= \sum_{\ell=0}^{d} p_\ell.
\end{aligned}$$

Concerning Property (ii), if $k \neq \ell$, we exploit (7.48) and we compute

$$(4+d)! \int_K a_1\lambda_\ell^3\lambda_k + a_2\lambda_k^3\lambda_\ell + a_2 \sum_{\substack{j=0\\ j\neq k,\ell}}^{d} \lambda_j^2\lambda_k\lambda_\ell + a_3\lambda_\ell^2\lambda_k^2 + a_3 \sum_{\substack{j=0\\ j\neq k,\ell}}^{d} \lambda_\ell^2\lambda_k\lambda_j$$
$$+ a_4\lambda_\ell\lambda_k^2 \sum_{\substack{j=0\\ j\neq k,\ell}}^{d} \lambda_j + a_4\lambda_\ell\lambda_k \sum_{\substack{j=0\\ j\neq k,\ell}}^{d-1} \sum_{\substack{i=j+1\\ i\neq k,\ell}}^{d} \lambda_j\lambda_i$$
$$= d!|K| \Bigg(6a_1 + 6a_2 + 2(d-1)a_2 + 4a_3 + 2(d-1)a_3 + 2(d-1)a_4$$
$$+ \frac{(d-1)(d-2)}{2} a_4 \Bigg) = 0. \tag{7.51}$$

Instead, if $j = \ell$, we test (i) with λ_ℓ and integrate over K. Exploiting (7.51) we get

$$\int_K p_\ell\lambda_\ell = (d+1) \int_K \lambda_\ell = |K|.$$

Concerning (iii), we remark that $\phi_z^* \in H^1(K)$ for every $K \subset \omega_z$, and that p_ℓ is a multiple of λ_ℓ, so that $\phi_z^*|_{\partial\omega_z} = 0$. Moreover p_ℓ is symmetric with respect to permutations where ℓ is fixed. This guarantees that $\phi_z^* \in H_0^1(\omega_z)$. In fact, let K_1 and K_2 be two simplices such that $K_i \subset \omega_z$, $i = 1, 2$ and $K_1 \cap K_2$ is an m-simplex, $0 \leq m \leq d-1$. Let ℓ_1 and ℓ_2 be the indices of the barycentric coordinate associated to z in K_1 and K_2 respectively. Let σ be the permutation that exchanges only ℓ_1 with ℓ_2, and let $\Sigma \in \mathbb{R}^{(d+1)\times(d+1)}$ be the corresponding matrix. More precisely, $\sigma(\ell_1) = \ell_2$, $\sigma(\ell_2) = \ell_1$ and $\sigma(j) = j$ for every $j = 0, \ldots, d$, $j \neq \ell_1, \ell_2$. Moreover there exists a permutation σ' with matrix Σ' such that $\sigma(\ell_2) = \ell_2$ and for every $x \in K_1 \cap K_2$

$$\Lambda_{K_2}(x) = \Sigma'\Sigma\Lambda_{K_1}(x).$$

Therefore, for every $x \in K_1 \cap K_2$

$$p_{\ell_1}(\Lambda_{K_1}(x)) = p_{\sigma(\ell_1)}(\Sigma\Lambda_{K_1}(x)) = p_{\ell_2}(\Sigma\Lambda_{K_1}(x)) = p_{\ell_2}(\Sigma'\Sigma\Lambda_{K_1}(x))$$
$$= p_{\ell_2}(\Lambda_{K_2}(x)).$$

Properties (iv) and (v) follow from the definition of ϕ_z^*, (7.48) and (7.2). □

We define $\Pi : H^{-1}(\Omega) \to S_0^{1,0}(\mathcal{T})$ as

$$\Pi f := \sum_{z\in\mathcal{V}_\Omega} \langle f, \phi_z^* \rangle \, \phi_z. \tag{7.52}$$

Moreover we define $\Pi^* : H^{-1}(\Omega) \to \operatorname{span}\{\phi_z^*\}_{z\in\mathcal{V}_\Omega}$ as

$$\Pi^* g := \sum_{z\in\mathcal{V}_\Omega} \langle g, \phi_z\rangle \, \phi_z^*. \tag{7.53}$$

For every $f \in H^{-1}(\Omega)$, $g \in H_0^1(\Omega)$, we have

$$\begin{aligned}\langle \Pi f, g\rangle &= \left\langle \sum_{z\in\mathcal{V}_\Omega} \langle f, \phi_z^*\rangle \, \phi_z, g \right\rangle = \sum_{z\in\mathcal{V}_\Omega} \langle f, \phi_z^*\rangle \, \langle \phi_z, g\rangle \\ &= \left\langle f, \sum_{z\in\mathcal{V}_\Omega} \langle g, \phi_z\rangle \, \phi_z^* \right\rangle = \langle f, \Pi^* g\rangle \,. \end{aligned} \tag{7.54}$$

With the following proposition we analyse the properties of Π and Π^*.

Proposition 7.22 (Properties of Π and Π^*)**.** *The interpolation operator* Π *defined in* (7.52) *satisfies the following properties:*

(i) Invariance over $S_0^{1,0}(\mathcal{T})$*. For every* $f \in S_0^{1,0}(\mathcal{T})$*,*

$$\Pi f = f.$$

(ii) Stability in L^2*. For every* $f \in L^2(\Omega)$*, for every* $K \in \mathcal{T}$*,*

$$\|\Pi f\|_{0,2;K} \le C(d) \, \|f\|_{0,2;\omega_K} \,.$$

(iii) Stability in H^1*. For every* $f \in H_0^1(\Omega)$*,*

$$\begin{aligned} \forall K \in \mathcal{T}_\Omega \qquad & |\Pi f|_{1,2;K} \le C(d, C_P) \frac{\operatorname{diam}(\omega_K)}{\rho_K} \, |f|_{1,2;\omega_K} \,, \\ \forall K \in \mathcal{T}_{\partial\Omega} \qquad & |\Pi f|_{1,2;K} \le C(d, C_F) \frac{\operatorname{diam}(\omega_K)}{\rho_K} \, |f|_{1,2;\omega_K} \,. \end{aligned}$$

The interpolation operator Π^* *defined in* (7.53) *satisfies the following properties:*

(iv) Local invariance over constants. For every $c \in \mathbb{R}$*, for every* $K \in \mathcal{T}_\Omega$*,*

$$(\Pi^* c\chi_{\omega_K})|_K = c\chi_K.$$

(v) Stability in L^2*. For every* $\varphi \in L^2(\Omega)$*, for every* $K \in \mathcal{T}$

$$\|\Pi^* \varphi\|_{0,2;K} \le C(d) \, \|\varphi\|_{0,2;\omega_K} \,.$$

(vi) Stability in H^1. For every $\varphi \in H^1_0(\Omega)$,

$$\forall K \in \mathcal{T}_\Omega, \qquad |\Pi^* \varphi|_{1,2;K} \leq C(d, C_P) \frac{\operatorname{diam}(\omega_K)}{\rho_K} |\varphi|_{1,2;\omega_K},$$
$$\forall K \in \mathcal{T}_{\partial\Omega}, \qquad |\Pi^* \varphi|_{1,2;K} \leq C(d, C_F) \frac{\operatorname{diam}(\omega_K)}{\rho_K} |\varphi|_{1,2;\omega_K}.$$

Proof. We start with (i), which is equivalent to $\Pi\phi_z = \phi_z$ for every $z \in \mathcal{N}_\Omega$. This in turn is equivalent to

$$\forall y, z \in \mathcal{N}_\Omega \qquad \int_\Omega \phi_z \phi_y^* = \delta_{yz}. \tag{7.55}$$

This is a direct consequence of Property (ii) of Remark 7.21.

Concerning (iv), by linearity of Π^* it is sufficient to show the assertion for $c = 1$. We exploit Property (i) of Remark 7.21 and get

$$\begin{aligned}(\Pi^* \chi_{\omega_K})|_K &= \sum_{z \in \mathcal{V}_K} \left(\int_{\omega_z} \phi_z \right) \phi_z^*|_K = \frac{|\omega_z|}{d+1} \sum_{z \in \mathcal{V}_K} \phi_z^*|_K = \frac{1}{d+1} \sum_{\ell=0}^{d} p_\ell \chi_K \\ &= \chi_K.\end{aligned}$$

Concerning Property (ii) and (v), we can proceed as in the proof of Proposition 7.2, using Property (iv) of Remark 7.21.

Property (iii), is also proved in the same way as Property (iii) of Proposition 7.2, exploiting the invariance over $S_0^{1,0}(\mathcal{T})$ and the Friedrichs or Poincaré inequality. The proof of Property (vi) also goes along the same lines, exploiting Property (iv), and recalling the definition of $c_K(\cdot)$ of (7.16), we get

$$\begin{aligned}|\Pi^* \varphi|_{1,2;K} &= |\Pi^* \varphi - c_K(\varphi)|_{1,2;K} = |\Pi^* \varphi - \Pi^* c_K(\varphi)|_{1,2;K} \\ &\leq \sum_{z \in \mathcal{V}_K} \|\varphi - c_K(\varphi)\|_{0,2;\omega_z} \|\phi_z\|_{0,2;\omega_z} |\phi_z^*|_{1,2;K} \\ &\leq C(d) \sum_{z \in \mathcal{V}_K} \|\varphi - c_K(\varphi)\|_{0,2;\omega_z} \frac{|K|^{1/2}}{\rho_K |\omega_z|^{1/2}} \\ &\leq C(d) \rho_K^{-1} \|\varphi - c_K(\varphi)\|_{0,2;\omega_K}.\end{aligned}$$

Property (vi) follows by means of the Poincaré inequality, for every $K \in \mathcal{T}_\Omega$, or the Friedrichs inequality, for every $K \in \mathcal{T}_{\partial\Omega}$. □

Approximation properties

The following propositions investigate the approximation properties of Π in the H^1-seminorm, in the L^2-norm and in the H^1-norm.

We start with the H^1-seminorm. The following proposition is the counterpart of Proposition 7.3. The proof goes along the same lines.

Proposition 7.23 (Approximation in H^1)**.** *The interpolation operator* Π *defined in* (7.52) *satisfies, for every* $f \in H^1_0(\Omega)$*,*

$$|f - \Pi f|_{1,2;\Omega} \leq C(d, C_P, C_F, \sigma_{\mathcal{T}}) \left(\sum_{K \in \mathcal{T}_\Omega} \inf_{P \in \mathbb{P}^1(K)} |f - P|^2_{1,2;K} \right)^{1/2}.$$

Combining with (7.49) we get, for every $f \in H^{1+s}(\Omega) \cap H^1_0(\Omega)$ with $s \in (0,1)$,

$$|f - \Pi f|_{1,2;\Omega} \leq C(d, s, C_P, C_F, \sigma_{\mathcal{T}}) \left(\sum_{K \in \mathcal{T}} \operatorname{diam}(K)^{2s} |f|^2_{1+s,2;K} \right)^{1/2}. \quad (7.56)$$

Next we investigate approximation in L^2. Mimicking (7.18), we define the variant of Π that has value in $S^{1,0}(\mathcal{T})$ as

$$\widetilde{\Pi} f := \sum_{z \in \mathcal{V}} \langle f, \phi^*_z \rangle \, \phi_z,$$

where $\{\phi^*_z\}_{z \in \mathcal{V}}$ are as in (7.50). The difference with Π is that the sum is on $\mathcal{V}$ and not only on $\mathcal{V}_\Omega$. For this reason $\widetilde{\Pi}$ is invariant on $S^{1,0}(\mathcal{T})$, and enjoys the same properties of Π.

Proposition 7.24 (Approximation in L^2)**.** *The interpolation operator* Π *defined in* (7.52) *satisfies, for every* $f \in H^s_0(\Omega)$*,* $s \in (0,1)$*,* $s \neq 1/2$*,*

$$\|f - \Pi f\|_{0,2;\Omega} \leq C(d, \sigma_{\mathcal{T}}) \Big(\sum_{K \in \mathcal{T}} \operatorname{diam}(\omega_K)^{2s} \, |f|^2_{s,2;\omega_K} \Big)^{1/2}.$$

Proof. Properties (i) and (ii) of Proposition 7.22 imply, for every $K \in \mathcal{T}_\Omega$,

$$\|f - \Pi f\|_{0,2;K} \leq C(d) \inf_{P \in S^{1,0}(\mathcal{T})|_{\omega_K}} \|f - P\|_{0,2;\omega_K}. \quad (7.57)$$

For the elements in $\mathcal{T}_{\partial\Omega}$, we insert $\widetilde{\Pi} f$, as in (7.20), so that we are left with $\left\| \widetilde{\Pi} f - \Pi f \right\|_{0,2;K}$. Since $f \in L^2(\Omega)$, we get

$$\left\| \widetilde{\Pi} f - \Pi f \right\|_{0,2;K} \leq \sum_{z \in \mathcal{V}_{K \cap \partial\Omega}} \left| \int_{\omega_z} f \phi^*_z \right| \|\phi_z\|_{0,2;K}.$$

We denote by ρ_z the distance from $\partial\omega_z \cap \partial\Omega$. The function $\rho_z^{-s} f$ belongs to $L^2(\omega_z)$, see [29, Thm. 1.4.4.3], and we have

$$\begin{aligned}\int_{\omega_z} f\phi_z^* &= \int_{\omega_z} \rho_z^{-s} f \phi_z^* \rho_z^s \le \operatorname{diam}(\omega_z)^s \left\|\phi_z^*\right\|_{0,2;\omega_z} \left\|\rho_z^{-s} f\right\|_{0,2;\omega_z} \\ &\le C \frac{\operatorname{diam}(\omega_z)^s}{|\omega_z|^{1/2}} \left|f\right|_{s,2;\omega_z}.\end{aligned}$$

Recalling also (7.3), we get

$$\begin{aligned}\sum_{K\in\mathcal{T}_{\partial\Omega}} \left(\sum_{z\in\mathcal{V}_{K\cap\partial\Omega}} \frac{|K|^{1/2}}{|\omega|^{1/2}} \operatorname{diam}(\omega_z)^s \left|f\right|_{s,2;\omega_z} \right)^2 \\ \le C(d) \sum_{K\in\mathcal{T}_{\partial\Omega}} \operatorname{diam}(\omega_K)^{2s} \left|f\right|^2_{s,2;\omega_K}.\end{aligned}$$

The thesis follows combining (7.57) with [23, Thm. 7.1] and (7.49). □

For $s = 1/2$ we have

$$\begin{aligned}\left\|f - \Pi f\right\|_{0,2;\Omega} \le C(d,\sigma_{\mathcal{T}}) \Big(\sum_{K\in\mathcal{T}} \operatorname{diam}(\omega_K) \left|f\right|^2_{1/2,2;\omega_K} \\ + \sum_{K\in\mathcal{T}_{\partial\Omega}} \operatorname{diam}(\omega_K) \left\|\rho^{-1/2} f\right\|^2_{0,2;\omega_K} \Big)^{1/2}. \quad (7.58)\end{aligned}$$

The term $\left\|\rho^{-1/2} f\right\|_{0,2;\omega_K}$ cannot in general be bounded by $\left|f\right|_{1/2,2;\omega_K}$, see [33, Thm. 11.7].

Remark 7.25. Proceeding as in the proof of Proposition 7.24 we also get a bound for $\left\|f - \Pi_0 f\right\|_{0,2;\Omega}$ in case $f \in H^s_0(\Omega)$, $s \in (0,1)$, $s \neq 1/2$

$$\left\|f - \Pi_0 f\right\|_{0,2;\Omega} \le C(d,\ell,\sigma_{\mathcal{T}}) \Big(\sum_{K\in\mathcal{T}} \operatorname{diam}(\omega_K)^{2s} \left|f\right|^2_{s,2;\omega_K} \Big)^{1/2}.$$

Proposition 7.26 (Approximation in H^{-1})**.** *The interpolation operator* Π *defined in* (7.52) *satisfies:*

(i) For every $f \in H^{-s}(\Omega)$ *with* $s \in (0,1)$*,*

$$\left\|f - \Pi f\right\|_{-1;\Omega} \le C(d, C_P, \sigma_{\mathcal{T}}) \left(\sum_{z\in\mathcal{V}} \operatorname{diam}(\omega_z)^{2-2s} \left\|f\right\|^2_{-s;\omega_z} \right)^{1/2}. \quad (7.59)$$

(ii) For every $f \in L^2(\Omega)$

$$\|f - \Pi f\|_{-1;\Omega} \leq C(d, C_P, \nu_{\mathcal{T}}) \left(\sum_{K \in \mathcal{T}} \operatorname{diam}(\omega_K)^2 \, \|f\|^2_{0,2;K} \right)^{1/2}.$$

Proof. We start with (i) exploiting (7.54) and the fact that $\{\phi_z\}_{z \in \mathcal{V}}$ form a partition of unity:

$$\begin{aligned}
\|f - \Pi f\|_{-1;\Omega} &= \sup_{\varphi \in H^1_0(\Omega)} \frac{\langle f - \Pi f, \varphi \rangle}{|\varphi|_{1,2;\Omega}} = \sup_{\varphi \in H^1_0(\Omega)} \frac{\langle f, \varphi - \Pi^* \varphi \rangle}{|\varphi|_{1,2;\Omega}} \\
&= \sup_{\varphi \in H^1_0(\Omega)} \sum_{z \in \mathcal{V}} \frac{\langle f, (\varphi - \Pi^* \varphi) \phi_z \rangle}{|\varphi|_{1,2;\Omega}} \\
&\leq \sup_{\varphi \in H^1_0(\Omega)} \sum_{z \in \mathcal{V}} \sup_{\psi \in H^s_0(\omega_z)} \frac{\langle f, \psi \rangle}{\|\psi\|_{s,2;\omega_z} \, |\varphi|_{1,2;\Omega}} \, \|(\varphi - \Pi^* \varphi) \phi_z\|_{s,2;\omega_z} \\
&\leq \left(\sum_{z \in \mathcal{V}} \operatorname{diam}(\omega_z)^{2-2s} \left(\sup_{\psi \in H^s_0(\omega_z)} \frac{\langle f, \psi \rangle}{\|\psi\|_{s,2;\omega_z}} \right)^2 \right)^{1/2} \\
&\quad \cdot \sup_{\varphi \in H^1_0(\Omega)} \left(\sum_{z \in \mathcal{V}} \operatorname{diam}(\omega_z)^{2s-2} \frac{\|(\varphi - \Pi^* \varphi) \phi_z\|^2_{s,2;\omega_z}}{|\varphi|^2_{1,2;\Omega}} \right)^{1/2}. \qquad (7.60)
\end{aligned}$$

The thesis follows by bounding the supremum over $\varphi \in H^1_0(\Omega)$ on the right-hand side. To this end, we set $c_K(\cdot)$ as in (7.16). By means of Properties (iii) and (iv) of Proposition 7.22 and the Poincaré or Friedrichs inequality, we get, for every $z \in \mathcal{V}$,

$$\begin{aligned}
\|(\varphi - \Pi^* \varphi) \phi_z\|^2_{0,2;\omega_z} &\leq \|\varphi - \Pi^* \varphi\|^2_{0,2;\omega_z} \\
&= \sum_{K \subset \omega_z} \|\varphi - c_K(\varphi) - \Pi^*(\varphi - c_K(\varphi))\|^2_{0,2;K} \\
&\leq C(d) \sum_{K \subset \omega_z} \|\varphi - c_K(\varphi)\|^2_{0,2;\omega_K} \\
&\leq C(d, C_P, C_F) \sum_{K \in \omega_z} \operatorname{diam}(\omega_K)^2 \, |\varphi|^2_{1,2;\omega_K}. \qquad (7.61)
\end{aligned}$$

Taking into account the multiple counting in the sum over the elements of a star, we obtain

$$\begin{aligned}
\|(\varphi - \Pi^* \varphi) \phi_z\|^2_{0,2;\omega_z} &\leq \|\varphi - \Pi^* \varphi\|^2_{0,2;\omega_z} \\
&\leq C(d, C_P, C_F, \nu_{\mathcal{T}}) \operatorname{diam}(\widetilde{\omega}_z)^2 \, |\varphi|^2_{1,2;\widetilde{\omega}_z}. \qquad (7.62)
\end{aligned}$$

Moreover we apply (7.46) and get

$$|(\varphi - \Pi^*\varphi)\phi_z|^2_{s,2;\omega_z} \leq 2\,\|\phi_z\|^2_{0,\infty;\omega_z}\,|\varphi - \Pi^*\varphi|^2_{s,2;\omega_z} + C(d,\sigma_{\mathcal{T}})\mathrm{diam}(\omega_z)^{2-2s}\,\|\nabla\phi_z\|^2_{0,\infty;\omega_z}\,\|\varphi - \Pi^*\varphi\|^2_{0,2;\omega_z}\,. \tag{7.63}$$

Concerning the first term on the right-hand side, we map to the reference stars and we exploit the embedding $H^1(\hat{\omega}) \subset H^s(\hat{\omega})$, see Lemma 7.19:

$$\begin{aligned}|\varphi - \Pi^*\varphi|^2_{s,2;\omega_z} &\leq C(d)\max_{K\subset\omega_z}\rho_K^{-2s}\frac{|\omega_z|}{|\hat{\omega}|}\,|(\varphi - \Pi^*\varphi)\circ F|^2_{s,2;\hat{\omega}}\\ &\leq C(d)\max_{K\subset\omega_z}\rho_K^{-2s}\frac{|\omega_z|}{|\hat{\omega}|}\,\|(\varphi - \Pi^*\varphi)\circ F\|^2_{1,2;\hat{\omega}}\,.\end{aligned} \tag{7.64}$$

We bound the two terms of the $\|\cdot\|_{1,2;\hat{\omega}}$-norm separately. Concerning the L^2-norm we map back to every simplex in ω_z and proceed as in (7.61)

$$\begin{aligned}\|(\varphi - \Pi^*\varphi)\circ F\|^2_{0,2;\hat{\omega}} &= \sum_{\hat{T}\in\hat{\omega}}\|(\varphi - \Pi^*\varphi)\circ F\|^2_{0,2;\hat{T}}\\ &\leq \sum_{K\subset\omega_z}\frac{|\hat{T}|}{|K|}\,\|\varphi - c_K - \Pi^*(\varphi - c_K)\|^2_{0,2;K}\\ &\leq C(d,C_P,C_F)\sum_{K\subset\omega_z}\frac{|\hat{T}|}{|K|}\mathrm{diam}(\omega_K)^2\,|\varphi|^2_{1,2;\omega_K}\,.\end{aligned} \tag{7.65}$$

Concerning the H^1-seminorm we exploit property (vi) of Proposition 7.22 and get

$$\begin{aligned}|(\varphi - \Pi^*\varphi)\circ F|^2_{1,2;\hat{\omega}} &= \sum_{\hat{T}\in\hat{\omega}}|(\varphi - \Pi^*\varphi)\circ F|^2_{1,2;\hat{T}}\\ &\leq \sum_{K\subset\omega_z}\frac{|\hat{T}|}{|K|}\mathrm{diam}(K)^2\,|\varphi - \Pi^*\varphi|^2_{1,2;K}\\ &\leq C(d,C_P,C_F)\sum_{K\subset\omega_z}\frac{\mathrm{diam}(\omega_K)^2}{\rho_K^2|K|}\mathrm{diam}(K)^2\,|\varphi|^2_{1,2;\omega_K}\,.\end{aligned} \tag{7.66}$$

Combining (7.64)–(7.66), and taking into account the multiple counting in the sum over the elements of a star, we obtain

$$\begin{aligned}&|\varphi - \Pi^*\varphi|^2_{s,2;\omega_z}\\ &\quad\lesssim \mathrm{diam}(\widetilde{\omega}_z)^{2-2s}\max_{K\subset\omega_z}\frac{\mathrm{diam}(\widetilde{\omega}_z)^{2s}}{\rho_K^{2s}}\max_{K\subset\omega_z}\frac{|\omega_z|\mathrm{diam}(K)^2}{|K|\rho_K^2}\,|\varphi|^2_{1,2;\widetilde{\omega}_z}\,,\end{aligned} \tag{7.67}$$

where the hidden constant depends on d, C_P, C_F, and $\nu_{\mathcal{T}}$. Concerning the second term on the right-hand side of (7.63), we combine (7.62) with

$$\|\nabla\phi_z\|^2_{0,\infty;\omega_z} \leq C(d) \max_{K\subset\omega_z} \rho_K^{-2}.$$

Combining this with (7.63) and (7.67), we get

$$|(\varphi - \Pi^*\varphi)\phi_z|^2_{s,2;\omega_z} \leq C(d, C_P, C_F, \nu_{\mathcal{T}}, \sigma_{\mathcal{T}})\mathrm{diam}(\omega_z)^{2-2s}|\varphi|^2_{1,2;\widetilde{\omega}_z}.$$

Inserting this and (7.61) into (7.60), and taking into account the multiple counting in the sum over $z \in \mathcal{V}$, gives the first assertion.

Regarding (ii), since $f \in L^2(\Omega)$, we can write

$$\begin{aligned}
\|f - \Pi f\|_{-1;\Omega} &= \sup_{\varphi\in H^1_0(\Omega)} \frac{\langle f - \Pi f, \varphi\rangle}{|\varphi|_{1,2;\Omega}} = \sup_{\varphi\in H^1_0(\Omega)} \frac{\langle f, \varphi - \Pi^*\varphi\rangle}{|\varphi|_{1,2;\Omega}} \\
&\leq \sup_{\varphi\in H^1_0(\Omega)} \frac{\sum_{K\in\mathcal{T}} \int_K f(\varphi - \Pi^*\varphi)}{|\varphi|_{1,2;\Omega}} \\
&\leq \sup_{\varphi\in H^1_0(\Omega)} \frac{\sum_{K\in\mathcal{T}} \|f\|_{0,2;K} \|\varphi - \Pi^*\varphi\|_{0,2;K}}{|\varphi|_{1,2;\Omega}}.
\end{aligned}$$

We proceed as in (7.65) and use the Cauchy-Schwarz inequality for sums to obtain

$$\begin{aligned}
\|f - \Pi f\|_{-1;\Omega} &\leq C(d, C_P, C_F) \sup_{\varphi\in H^1_0(\Omega)} \frac{\sum_{K\in\mathcal{T}} \|f\|_{0,2;K} \,\mathrm{diam}(\omega_K)\, |\varphi|_{1,2;\omega_K}}{|\varphi|_{1,2;\Omega}} \\
&\leq C(d, C_P, C_F, \nu_{\mathcal{T}}) \left(\sum_{K\in\mathcal{T}} \mathrm{diam}(\omega_K)^2 \|f\|^2_{0,2;K} \right)^{1/2}. \qquad \square
\end{aligned}$$

7.7 Standard formulation and fractional regularity

In this section we derive error estimates for the approximation of the solution u of the parabolic problem in the standard formulation. We extend the results of Section 7.3 to exact solutions belonging to fractional order Sobolev spaces.

7.7.1 Spatial semidiscretization

As in Section 7.3.1 we require that $\mathcal{T}$ belongs to a family of triangulations for which the L^2-projection onto $S_0^{\ell,0}$ is H^1-stable. We assume $U \in S_0^{\ell,0}(\mathcal{T})$ to be as in Section 7.3.1.

With the tools in Sections 7.5–7.6 we obtain the following extension of Theorem 7.7.

Theorem 7.27. *Assume $1 < \theta \le \ell + 1$ with $\theta = m + s$, $m \in \mathbb{N}$, $s \in (0,1)$, $\theta \neq 3/2, 5/2$. Moreover assume $u \in L^2(H^\theta)$, and, if $\theta < 5/2$, $u' \in L^2(H^{\theta-2})$ otherwise $u' \in L^2(H_0^{\theta-2})$. Then we have*

$$\begin{aligned}\|u(0) - U(0)\|_{0,2;\Omega}^2 + \int_0^T \|u'(t) - U'(t)\|_{-1;\Omega}^2 + |u(t) - U(t)|_{1,2;\Omega}^2 \\ \lesssim \sum_{K\in\mathcal{T}} \operatorname{diam}(K)^{2\theta-2} \, |u(0)|_{\theta-1,2;\omega_K}^2 \\ + \int_0^T \sum_{K\in\mathcal{T}} \operatorname{diam}(\omega_K)^{2\theta-2} \, |u'(t)|_{\theta-2,2;\omega_K}^2 + \operatorname{diam}(K)^{2\theta-2} \, |u(t)|_{\theta,2;K}^2 \, .\end{aligned}$$

The hidden constant depends on the H^1-norm of the L^2-projection on $S_0^{\ell,0}(\mathcal{T})$, the coercivity and continuity constants of the parabolic problem, the dimension d, the parameter s, the polynomial degree ℓ and the shape-parameter $\sigma_{\mathcal{T}}$.

Proof. The proof mimics the one of Theorem 7.7. In case $\theta \ge 2$, we still interpolate with Π_0 of Section 7.2 and invoke Propositions 7.3–7.5 combined with (7.49) together with [23, Thm. 7.1]; recall also Remark 7.25. In case $\theta \in (1,2)$, instead, u' does not belong to $L^2(L^2)$ and we take

$$\forall t \in [0,T] \qquad V(t) := \Pi u(t) \in S_0^{1,0}(\mathcal{T}) \subset S_0^{\ell,0}(\mathcal{T}).$$

Because of the low regularity of the function, we cannot exploit the full approximation power of $S_0^{\ell,0}(\mathcal{T})$, and we do not waste in taking $V \in H^1(S_0^{1,0}(\mathcal{T}))$. The assertion follows from Propositions 7.24–7.26 and (7.56). □

Remark 7.28. In case $\theta = 3/2$ or $\theta = 5/2$ we apply (7.58) and get similar bounds. The difference consists in an additional term on the right-hand side, which is

$$\sum_{K\in\mathcal{T}_{\partial\Omega}} \operatorname{diam}(\omega_K) \left\|\rho^{-1/2} u(0)\right\|_{0,2;\omega_K}^2$$

in case $\theta = 3/2$, or

$$\int_0^T \sum_{K\in\mathcal{T}_{\partial\Omega}} \operatorname{diam}(\omega_K)^3 \left\|\rho^{-1/2} u'(t)\right\|_{0,2;\omega_K}^2$$

in case $\theta = 5/2$. Similar considerations are valid also for the theorems below.

7.7.2 Semidiscretization in time

Let $U \in \mathcal{S}^{1,0}(\mathcal{P}, H^1_0)$ be as in Section 7.3.2, and assume $\sup_n \tau_n < 1$. By means of the fractional Poincaré inequality, we obtain the following extension of Theorem 7.8.

Theorem 7.29. *Assume $s \in (0,1)$, $u \in H^s(H^1_0)$ and $u' \in H^s(H^{-1})$. Then*

$$\begin{aligned}\|u(0) - U(0)\|^2_{0,2;\Omega} + \int_0^T \|u' - U'\|^2_{-1;\Omega} + |u - U(t_n)|^2_{1,2;\Omega} \\ \lesssim \sum_{n=1}^N \tau_n^{2s} \|u'\|^2_{H^s(I_n,H^{-1})} + \tau_n^{2s} \|u\|^2_{H^s(I_n,H^1)},\end{aligned}$$

where the hidden constant depends on the coercivity and continuity constant of the parabolic problem and on the parameter $\mu_{\mathcal{P}}$, but it is independent of s.

Proof. We apply Theorem 4.4, which states that

$$\begin{aligned}\|u(0) - U(0)\|^2_{0,2;\Omega} + \sum_{n=1}^N \int_{I_n} \|u' - U'\|^2_{-1;\Omega} + |u - U(t_n)|^2_{1,2;\Omega} \\ \lesssim \inf_{v \in \mathcal{S}^{0,-1}(\mathcal{P},H^{-1})} \|u' - v\|^2_{L^2(H^{-1})} + \inf_{z \in \mathcal{S}^{0,-1}(\mathcal{P},H^1_0)} \|u - z\|^2_{L^2(H^1_0)}.\end{aligned}$$

We insert $v \in \mathcal{S}^{0,-1}(\mathcal{P}, H^{-1})$ and $z \in \mathcal{S}^{0,-1}(\mathcal{P}, H^1_0)$ such that, for $n = 1, \dots, N$,

$$v|_{I_n} = \frac{1}{\tau_n} \int_{I_n} u', \qquad \text{and} \qquad z|_{I_n} = \frac{1}{\tau_n} \int_{I_n} u$$

in the infima on the right-hand side. The thesis follows applying (7.45). □

7.7.3 Full discretization with the backward Euler-Galerkin method

We assume that the sequence $\{\mathcal{T}_n\}_{n=0}^N$ of triangulations belongs to a family for which the L^2-projection is uniformly H^1-stable.

Let U be as in Section 7.3.4. Combining the results in Sections 7.7.1–7.7.2 we obtain the following extension of Theorem 7.11.

Theorem 7.30. *Assume $s \in (0,1)$ and $\theta \in (1, \ell+1)$, with $\theta = m + \bar{s}$, $m \in \mathbb{N}$, $\bar{s} \in (0,1)$ and $\theta \neq 3/2, 5/2$. Assume $u \in L^2(H^\theta) \cap H^s(H^1_0)$, and, if $\theta < 5/2$,*

$u' \in L^2(H^{\theta-2}) \cap H^s(H^{-1})$, *otherwise* $u' \in L^2(H_0^{\theta-2}) \cap H^s(H^{-1})$. *Then*

$$\begin{aligned}
&\|u(0) - U(0)\|_{0,2;\Omega}^2 + \sum_{n=1}^{N} \int_{I_n} \|u' - U'\|_{-1;\Omega}^2 + |u - U(t_n)|_{1,2;\Omega}^2 \\
&\quad \lesssim \sum_{K \in \mathcal{T}_0} \operatorname{diam}(\omega_K)^{2\theta-2} \, |u(0)|_{\theta-1,2;\omega_K}^2 \\
&\qquad + \sum_{n=1}^{N} \tau_n^{2s} \, \|u'\|_{H^s(I_n,H^{-1})}^2 + \int_{I_n} \sum_{K \in \mathcal{T}_n} \operatorname{diam}(\omega_K)^{2\theta-2} \, |u'|_{\theta-2,2;\omega_K}^2 \\
&\qquad\quad + \tau_n^{2s} \, \|u\|_{H^s(I_n,H_0^1)}^2 + \int_{I_n} \sum_{K \in \mathcal{T}_n} \operatorname{diam}(\omega_K)^{2\theta-2} \, |u|_{\theta,2;\omega_K}^2 \\
&\qquad + \sum_{n=1}^{N-1} \tau_n^{-1} \int_{I_n} \sum_{K \in \mathcal{T}_n} \operatorname{diam}(\omega_K)^{2\theta} \, |u|_{\theta,2;\omega_K}^2 \, .
\end{aligned}$$

The hidden constant depends on the maximum of the H^1-norms of the L^2-projection on $S_0^{\ell,0}(\mathcal{T}_n)$, the coercivity and continuity constants of the parabolic problem, the dimension d, the parameter $\bar{s}$, the polynomial degree ℓ, the shape parameters $\sigma_{\mathcal{T}_n}$ and the parameter $\mu_{\mathcal{P}}$, but it is independent of s.

Proof. The proof mimics the one of Theorem 7.11. As in Theorem 7.29 we use (7.45) to bound the terms

$$\left\| u - \frac{1}{\tau_n} \int_{I_n} u \right\|_{L^2(I_n,H^1)} \qquad \text{and} \qquad \left\| u' - \frac{1}{\tau_n} \int_{I_n} u' \right\|_{L^2(I_n,H^{-1})} .$$

As in Theorem 7.27, if $\theta \geq 2$ we bound the contributions involving the interpolation operator Π_0 with the help of (7.49), [23, Thm. 7.1] and Remark 7.25. If $\theta \in (1,2)$ we use Π in place of Π_0 in the treatment of the time derivative. □

7.8 Natural formulation and fractional regularity

In this section we derive error estimate for the approximation of the solution u of the parabolic problem in the natural formulation. We extend the results of Section 7.4 to exact solutions belonging to fractional order Sobolev spaces.

7.8.1 Spatial semidiscretization

Since Theorem 7.12 does not involve the approximation of the time derivative in the $L^2(H^{-1})$-norm, we can easier extend this result. With the help of (7.49) with [23, Thm. 7.1], we get

Theorem 7.31. *Assume $1 \le \theta \le \ell+1$, with $\theta = m+s$, $m \in \mathbb{N}$ and $s \in (0,1)$. Moreover assume $u \in L^2(H^\theta)$. Then we have*

$$\int_0^T |u(t) - U(t)|_{1,2;\Omega}^2 \lesssim \int_0^T \sum_{K\in\mathcal{T}} \operatorname{diam}(K)^{2\theta-2} \, |u(t)|_{\theta,2;K}^2 \, .$$

The hidden constant depends on the H^1-norm of the L^2-projection on $S_0^{\ell,0}(\mathcal{T})$, the coercivity and continuity constants of the parabolic problem, the dimension d, the parameter s, the polynomial degree ℓ and the shape parameter $\sigma_{\mathcal{T}}$.

7.8.2 Semidiscretization in time

Assume $\sup_n \tau_n < 1$ and let $U \in \mathcal{S}^{0,-1}(\mathcal{P}, H_0^1)$ be as in Section 7.4.2. We recall (4.33) and (4.30) and use (7.30) with $Y = L^2(\Omega)$ and $Y = H_0^1(\Omega)$, to obtain

$$\begin{aligned} \sum_{n=1}^N \tau_n^{-1} \, \|\varphi - \varphi(t_{n-1})\|_{L^2(I_n,L^2)}^2 &\le \frac{1}{3} \sum_{n=1}^N \int_{I_n} \|\varphi(t_n) - \varphi(t_{n-1})\|_{0,2;\Omega}^2 \\ &\le \frac{1}{3} \, \|\varphi\|_{2,\mathcal{P}}^2 \, , \end{aligned} \tag{7.68a}$$

and

$$\begin{aligned} \sum_{n=1}^N \|\varphi - \varphi(t_{n-1})\|_{L^2(I_n,H^1)}^2 &\le \frac{1}{3} \sum_{n=1}^N \int_{I_n} |\varphi(t_n) - \varphi(t_{n-1})|_{1,2;\Omega}^2 \\ &\le \frac{2}{3}(\mu_{\mathcal{P}} + 1) \, \|\varphi\|_{2,\mathcal{P}}^2 \, . \end{aligned} \tag{7.68b}$$

Theorem 7.32. *Assume $s \in (0,1)$ and that, for every $n = 1, \dots, N$, $u|_{I_n} \in H^s(I_n, H_0^1)$ and, if $s \in (0, 1/2]$, $u'|_{I_n} \in L^2(I_n, H^{2s-1})$, otherwise $u'|_{I_n} \in H^s(I_n, H^{-1})$. Then,*

$$\|u - U\|_{L^2(H_0^1)}^2 \lesssim \sum_{n=1}^N \tau_n^{2s} \, \|u\|_{H^s(H_0^1)}^2 + \begin{cases} \tau_n^{2s} \, \|u'\|_{L^2(I_n,H^{2s-1})}^2 & \text{if } s \in (0,1/2] \\ \tau_n^{2s} \, \|u'\|_{H^s(I_n,H^{-1})}^2 & \text{if } s \in (1/2,1) \end{cases} .$$

The hidden constant depends on the coercivity and continuity constants of the parabolic problem, on s and on the parameter $\mu_{\mathcal{P}}$ if $s \neq 1/2$.

Proof. We apply Proposition 4.9, which states that

$$\|u-U\|_{L^2(H^1_0)} \lesssim \inf_{v\in\mathcal{S}^{0,-1}(\mathcal{P},H^1_0)} \|u-v\|_{L^2(H^1_0)} + \sup_{\substack{\varphi\in\mathcal{S}^{1,0}(\mathcal{P},H^1_0)\\ \varphi(T)=0}} \frac{\displaystyle\sum_{n=1}^N \int_{I_n} \langle Au-f, \varphi-\varphi(t_{n-1})\rangle}{\|\varphi\|_{2,\mathcal{P}}}. \tag{7.69}$$

We insert $v\in\mathcal{S}^{0,-1}(\mathcal{P},H^1_0)$ such that, for every $n=1,\dots,N$,

$$v|_{I_n} = \frac{1}{\tau_n}\int_{I_n} u,$$

in the infimum on the right-hand side of (7.32), and we exploit (7.45) on every I_n. Concerning the supremum on the right-hand side of (7.69), we first consider the case $s=1/2$. Exploiting (7.68a), we get, for every $\varphi\in\mathcal{S}^{1,0}(\mathcal{P},H^1_0)$ with $\varphi(T)=0$,

$$\begin{aligned}
&\sum_{n=1}^N \int_{I_n} \langle Au-f, \varphi-\varphi(t_{n-1})\rangle \\
&\quad\le \sum_{n=1}^N \|Au-f\|_{L^2(I_n,L^2)} \|\varphi-\varphi(t_{n-1})\|_{L^2(I_n,L^2)} \\
&\quad\le \left(\sum_{n=1}^N \tau_n \|Au-f\|^2_{L^2(I_n,L^2)}\right)^{1/2} \left(\sum_{n=1}^N \tau_n^{-1} \|\varphi-\varphi(t_{n-1})\|^2_{L^2(I_n,L^2)}\right)^{1/2} \\
&\quad\le \frac{1}{3}\left(\sum_{n=1}^N \tau_n \|Au-f\|^2_{L^2(I_n,L^2)}\right)^{1/2} \|\varphi\|_{2,\mathcal{P}}\,.
\end{aligned} \tag{7.70}$$

Dividing by $\|\varphi\|_{2,\mathcal{P}}$ and taking the supremum over φ, we get the assertion for $s=1/2$.

For the case $s\in(0,1/2)$, we exploit (7.47) and we obtain, for every

$\varphi \in \mathcal{S}^{1,0}(\mathcal{P}, H_0^1)$ with $\varphi(T) = 0$,

$$\begin{aligned}
&\sum_{n=1}^{N} \int_{I_n} \langle Au - f, \varphi - \varphi(t_{n-1}) \rangle \\
&\qquad \leq \sum_{n=1}^{N} \int_{I_n} \|Au - f\|_{2s-1;\Omega} \, \|\varphi - \varphi(t_{n-1})\|_{1-2s,2;\Omega} \\
&\qquad \leq C(s) \left(\sum_{n=1}^{N} \int_{I_n} \|Au - f\|_{2s-1;\Omega} \, |\varphi - \varphi(t_{n-1})|_{1,2;\Omega}^{1-2s} \, \|\varphi - \varphi(t_{n-1})\|_{0,2;\Omega}^{2s} \right. \\
&\qquad\qquad \left. + \sum_{n=1}^{N} \int_{I_n} \|Au - f\|_{2s-1;\Omega} \, \|\varphi - \varphi(t_{n-1})\|_{0,2;\Omega} \right). \qquad (7.71)
\end{aligned}$$

In order to bound the first sum on the right-hand side, we apply twice the Hölder inequality for three functions with $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1$, being $p = 2$, $q = \frac{2}{1-2s}$ and $r = \frac{1}{s}$:

$$\begin{aligned}
&\sum_{n=1}^{N} \int_{I_n} \|Au - f\|_{2s-1;\Omega} \, |\varphi - \varphi(t_{n-1})|_{1,2;\Omega}^{1-2s} \, \|\varphi - \varphi(t_{n-1})\|_{0,2;\Omega}^{2s} \\
&\qquad \leq \sum_{n=1}^{N} \tau_n^s \, \|u'\|_{L^2(I_n, H^{2s-1})} \, \|\varphi - \varphi(t_{n-1})\|_{L^2(I_n, H_0^1)}^{1-2s} \cdot \\
&\qquad\qquad \cdot \tau_n^{-s} \, \|\varphi - \varphi(t_{n-1})\|_{L^2(I_n, L^2)}^{2s} \\
&\qquad \leq \left(\sum_{n=1}^{N} \tau_n^{2s} \, \|u'\|_{L^2(I_n, H^{2s-1})}^2 \right)^{\frac{1}{2}} \left(\sum_{n=1}^{N} \|\varphi - \varphi(t_{n-1})\|_{L^2(I_n, H_0^1)}^2 \right)^{\frac{1-2s}{2}} \cdot \\
&\qquad\qquad \cdot \left(\sum_{n=1}^{N} \tau_n^{-1} \, \|\varphi - \varphi(t_{n-1})\|_{L^2(I_n, L^2)}^2 \right)^{s} .
\end{aligned}$$

We exploit (7.68) to get

$$\begin{aligned}
&\sum_{n=1}^{N} \int_{I_n} \|Au - f\|_{2s-1;\Omega} \, |\varphi - \varphi(t_{n-1})|_{1,2;\Omega}^{1-2s} \, \|\varphi - \varphi(t_{n-1})\|_{0,2;\Omega}^{2s} \\
&\qquad \lesssim \left(\sum_{n=1}^{N} \tau_n^{2s} \, \|u'\|_{L^2(I_n, H^{2s-1})}^2 \right)^{\frac{1}{2}} \|\varphi\|_{2,\mathcal{P}}^2 . \qquad (7.72)
\end{aligned}$$

Concerning the second sum on the right-hand side of (7.71), as in (7.70), we

get

$$\sum_{n=1}^{N} \int_{I_n} \|Au - f\|_{2s-1;\Omega} \, \|\varphi - \varphi(t_{n-1})\|_{0,2;\Omega} \\ \lesssim \left(\sum_{n=1}^{N} \tau_n \, \|u'\|^2_{L^2(I_n,H^{2s-1})} \right)^{\frac{1}{2}} \|\varphi\|^2_{2,\mathcal{P}} \,. \tag{7.73}$$

Since $\sup_n \tau_n < 1$, we have $\tau_n < \tau_n^{2s}$ for every $n = 1, \dots, N$, and this term can be bounded by (7.72). The assertion then follows for $s \in (0, 1/2)$.

For $s \in (1/2, 1)$, we write, for every $\varphi \in \{\phi \in \mathcal{S}^{1,0}(\mathcal{P}, H^1_0),\ \phi(T) = 0\}$, and every $n = 1, \dots, N$,

$$\begin{aligned} \int_{I_n} \langle f - Au, \varphi - \varphi(t_{n-1}) \rangle &= \int_{I_n} \langle u', \varphi - \varphi(t_{n-1}) \rangle \\ &= \big\langle u(t_n^-), \varphi(t_n) - \varphi(t_{n-1}) \big\rangle + \int_{I_n} - \langle \varphi', u \rangle \\ &= \big\langle u(t_n^-) - \Pi^n u, \varphi(t_n) - \varphi(t_{n-1}) \big\rangle + \int_{I_n} - \langle \varphi', u - \Pi^n u \rangle \\ &\le \big\| u(t_n^-) - \Pi^n u \big\|_{-1;\Omega} \, |\varphi(t_n) - \varphi(t_{n-1})|_{1,2;\Omega} \\ &\qquad + \|\varphi'\|_{L^2(I_n,H^{-1})} \, \|u - \Pi^n u\|_{L^2(I_n,H^1_0)} \,. \end{aligned} \tag{7.74}$$

Summing over $n = 1, \dots, N$, we get

$$\begin{aligned} &\sum_{n=1}^{N} \int_{I_n} \langle f - Au, \varphi - \varphi(t_{n-1}) \rangle \\ &\quad \le \left(\sum_{n=1}^{N} \tau_n^{-1} \big\| u(t_n^-) - \Pi^n u \big\|^2_{-1;\Omega} \right)^{1/2} \left(\sum_{n=1}^{N} \tau_n \, |\varphi(t_n) - \varphi(t_{n-1})|^2_{1,2;\Omega} \right)^{1/2} \\ &\qquad + \left(\sum_{n=1}^{N} \|u - \Pi^n u\|^2_{L^2(I_n,H^1_0)} \right)^{1/2} \|\varphi'\|_{L^2(H^{-1})} \,. \end{aligned}$$

We bound the second contribution by means of (4.11) and (7.45). Concerning the first contribution, we exploit Property (iii) of Remark 4.1 and (7.45) to obtain

$$\tau_n^{-1} \big\| u(t_n^-) - \Pi^n u \big\|^2_{-1;\Omega} \lesssim \inf_{c \in H^{-1}} \|u' - c\|^2_{L^2(I_n,H^{-1})} \lesssim \tau_n^{2s} \, \|u'\|_{H^s(I_n,H^{-1})} \,.$$

Combining this with (7.68b) completes the proof for $s \in (1/2, 1)$. □

7.8.3 Full discretization with the backward-Euler Galerkin method

Assume $\sup_n \tau_n < 1$ and let $U \in \mathcal{S}^{0,-1}(\mathcal{P}, \mathbb{V})$ be as in Section 7.4.4. Thanks to the H^1-stability of the L^2-projection, proceeding as in (4.30) gives, for every $\varphi \in \mathcal{S}^{1,0^-}(\mathcal{P}, \mathbb{V})$,

$$\sum_{n=1}^{N} \tau_n \left|\varphi(t_n^-) - \varphi(t_{n-1})\right|_{1,2;\Omega}^2 = \sum_{n=1}^{N} \tau_n \left|P_n\varphi(t_n) - \varphi(t_{n-1})\right|_{1,2;\Omega}^2 \lesssim \|\varphi\|_{2,\mathcal{P}}^2 . \tag{7.75}$$

Theorem 7.33. *Assume $s \in (0,1)$ and $\theta \in (1, \ell+1)$, with $\theta = m + \bar{s}$, $m \in \mathbb{N}$, $\bar{s} \in (0,1)$. Assume that, for every $n = 1, \dots, N$, $u|_{I_n} \in C^0(I_n, H^\theta) \cap H^s(I_n, H_0^1)$, and, if $s \in (0, 1/2]$, $u'|_{I_n} \in L^2(I_n, H^{2s-1})$, otherwise $u'|_{I_n} \in H^s(I_n, H^{-1})$. Then*

$$\begin{aligned}
&\|u - U\|_{L^2(H_0^1)}^2 \\
&\quad\lesssim \sum_{n=1}^{N} \tau_n^{2s} \|u\|_{H^s(I_n,H_0^1)}^2 + \sum_{K\in\mathcal{T}_n} \operatorname{diam}(\omega_K)^{2\theta-2} \int_{I_n} |u|_{\theta,2;\omega_K}^2 \\
&\qquad + \sum_{n=1}^{N-1} \sum_{K\in\mathcal{T}_n} \operatorname{diam}(\omega_K)^{2\theta} \left|u(t_n^-)\right|_{\theta,2;\omega_K}^2 \\
&\qquad + \sum_{n=1}^{N} \begin{cases} \tau_n^{2s} \displaystyle\int_{I_n} \|u'\|_{2s-1;\Omega}^2 & \text{if } s \in (0,1/2] \\ \tau_n^{2s} \|u'\|_{H^s(I_n,H^{-1})}^2 & \text{if } s \in (1/2,1) \end{cases} .
\end{aligned}$$

The hidden constants depend on the maximum of the H^1-norms of the L^2-projection on $S_0^{\ell,0}(\mathcal{T}_n)$, the coercivity and continuity constants of the parabolic problem, the parameters s and $\bar{s}$, the dimension d, the polynomial degree ℓ, the shape parameters $\sigma_{\mathcal{T}_n}$.

Proof. We exploit Theorem 6.10, which states that

$$\begin{aligned}
\|u - U\|_{L^2(H_0^1)} \lesssim{}& \inf_{v\in\mathcal{S}^{0,-1}(\mathcal{P},\mathbb{V})} \|u - v\|_{L^2(H_0^1)} \\
&+ \sup_{\substack{\varphi\in\mathcal{S}^{1,0^-}(\mathcal{P},\mathbb{V}) \\ \varphi(T)=0}} \frac{\displaystyle\sum_{n=1}^{N} \int_{I_n} \langle Au - f, \varphi - \varphi(t_{n-1})\rangle}{\|\varphi\|_{2,\mathcal{P}}} \\
&+ \left(\sum_{n=1}^{N-1} \left\|(I - P_n)P_n^+ u(t_n^-)\right\|_{0,2;\Omega}^2\right)^{1/2} .
\end{aligned} \tag{7.76}$$

We insert $v \in \mathcal{S}^{0,-1}(\mathcal{P}, \mathbb{V})$ such that, for $n = 1, \dots, N$,

$$v|_{I_n} = \Pi_0^n \frac{1}{\tau_n} \int_{I_n} u$$

in the infimum on the right-hand side. Proceeding as in (7.34), we obtain with the help of (7.45) and (7.49):

$$\left\| u - \Pi_0^n \frac{1}{\tau_n} \int_{I_n} u \right\|^2_{L^2(I_n, H_0^1)} \lesssim \int_{I_n} \sum_{K \in \mathcal{T}_n} \operatorname{diam}(\omega_K)^{2\theta-2} \, |u|^2_{\theta,2;\omega_K} + \tau_n^{2s} \, \|u\|^2_{H^s(I_n, H_0^1)} \, .$$

The terms $\|(I - P_n) P_n^+ u(t_n^-)\|^2_{0,2;\Omega}$ can be bounded as in (7.27). Concerning the supremum on the right-hand side of (7.76), we proceed as in the proof of Theorem 7.32. If $s \in (0, 1/2)$, we recall that (7.68a)–(7.68b) are still valid, so that we obtain

$$\sup_{\substack{\varphi \in \mathcal{S}^{1,0^-}(\mathcal{P}, \mathbb{V}) \\ \varphi(T) = 0}} \frac{\displaystyle\sum_{n=1}^N \int_{I_n} \langle Au - f, \varphi - \varphi(t_{n-1}) \rangle}{\|\varphi\|_{2,\mathcal{P}}} \lesssim \left(\sum_{n=1}^N \tau_n^{2s} \, \|u'\|^2_{L^2(I_n, H^{2s-1})} \right)^{1/2} .$$

In case $s \in (1/2, 1)$, as in (7.74), we get, for every $n = 1, \dots, N$,

$$\begin{aligned} \int_{I_n} \langle f - Au, \varphi - \varphi(t_{n-1}) \rangle \le \left\| u(t_n^-) - \Pi^n u \right\|_{-1;\Omega} \left| \varphi(t_n^-) - \varphi(t_{n-1}) \right|_{1,2;\Omega} \\ + \|\varphi'\|_{L^2(I_n, H^{-1})} \, \|u - \Pi^n u\|_{L^2(H_0^1)} \, , \end{aligned}$$

By means of (7.75) we can conclude

$$\begin{aligned} \sup_{\substack{\varphi \in \mathcal{S}^{1,0^-}(\mathcal{P}, \mathbb{V}) \\ \varphi(T) = 0}} & \frac{\displaystyle\sum_{n=1}^N \int_{I_n} \langle Au - f, \varphi - \varphi(t_{n-1}) \rangle}{\|\varphi\|_{2,\mathcal{P}}} \\ & \lesssim \left(\sum_{n=1}^N \tau_n^{2s} \, \|u'\|^2_{H^s(I_n, H^{-1})} + \tau_n^{2s} \, \|u\|^2_{H^s(I_n, H_0^1)} \right)^{1/2} . \end{aligned}$$

□

Bibliography

[1] Robert A. Adams and John J. F. Fournier. *Sobolev spaces*, volume 140 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, second edition, 2003.

[2] Ivo Babuška. Error-bounds for finite element method. *Numer. Math.*, 16:322–333, 1970/1971.

[3] C. Baiocchi and F. Brezzi. Optimal error estimates for linear parabolic problems under minimal regularity assumptions. *Calcolo*, 20(2):143–176 (1984), 1983.

[4] Randolph E. Bank and Rafael F. Santos. Analysis of some moving space-time finite element methods. *SIAM J. Numer. Anal.*, 30(1):1–18, 1993.

[5] M. Bebendorf. A note on the Poincaré inequality for convex domains. *Z. Anal. Anwendungen*, 22(4):751–756, 2003.

[6] Jöran Bergh and Jörgen Löfström. *Interpolation spaces. An introduction*. Springer-Verlag, Berlin, 1976. Grundlehren der Mathematischen Wissenschaften, No. 223.

[7] G. Birkhoff, M. H. Schultz, and R. S. Varga. Piecewise Hermite interpolation in one and two variables with applications to partial differential equations. *Numer. Math.*, 11:232–256, 1968.

[8] Wilhelm Blaschke. *Kreis und Kugel*. Walter de Gruyter & Co., Berlin, 1956. 2te Aufl.

[9] J. Bourgain, H. Brezis, and P. Mironescu. Another look at Sobolev spaces. In *Optimal Control and Partial Differential Equation*, pages 439–455. IOS Press, 2001.

[10] James H. Bramble, Joseph E. Pasciak, and Olaf Steinbach. On the stability of the L^2 projection in $H^1(\Omega)$. *Math. Comp.*, 71(237):147–156 (electronic), 2002.

[11] Susanne C. Brenner and L. Ridgway Scott. *The mathematical theory of finite element methods*, volume 15 of *Texts in Applied Mathematics*. Springer, New York, third edition, 2008.

[12] Jean Céa. Approximation variationnelle des problèmes aux limites. *Ann. Inst. Fourier (Grenoble)*, 14(fasc. 2):345–444, 1964.

[13] K. Chrysafinos and L. S. Hou. Error estimates for semidiscrete finite element approximations of linear and semilinear parabolic equations under minimal regularity assumptions. *SIAM J. Numer. Anal.*, 40(1):282–306, 2002.

[14] K. Chrysafinos and Noel J. Walkington. Error estimates for the discontinuous Galerkin methods for parabolic equations. *SIAM J. Numer. Anal.*, 44(1):349–366 (electronic), 2006.

[15] Philippe G. Ciarlet. *The finite element method for elliptic problems*, volume 4 of *Studies in Mathematics and its Applications*. North–Holland, Amsterdam, 1978.

[16] Ph. Clément. Approximation by finite element functions using local regularization. *Rev. Française Automat. Informat. Recherche Opérationnelle Sér. RAIRO Analyse Numérique*, 9(R-2):77–84, 1975.

[17] M. Crouzeix and V. Thomée. The stability in L_p and W_p^1 of the L_2-projection onto finite element function spaces. *Math. Comp.*, 48(178):521–532, 1987.

[18] Ronald A. DeVore and George G. Lorentz. *Constructive approximation*, volume 303 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1993.

[19] Ronald A. DeVore and Robert C. Sharpley. Besov spaces on domains in $\mathbf{R}^d$. *Trans. Amer. Math. Soc.*, 335(2):843–864, 1993.

[20] Eleonora Di Nezza, Giampiero Palatucci, and Enrico Valdinoci. Hitchhiker's guide to the fractional Sobolev spaces. *Bull. Sci. Math.*, 136(5):521–573, 2012.

[21] Jim Douglas, Jr. and Todd Dupont. Galerkin methods for parabolic equations. *SIAM J. Numer. Anal.*, 7:575–626, 1970.

[22] Todd Dupont. Mesh modification for evolution equations. *Math. Comp.*, 39(159):85–107, 1982.

[23] Todd Dupont and Ridgway Scott. Polynomial approximation of functions in Sobolev spaces. *Math. Comp.*, 34(150):441–463, 1980.

[24] Todd F. Dupont and Yingjie Liu. Symmetric error estimates for moving mesh Galerkin methods for advection-diffusion equations. *SIAM J. Numer. Anal.*, 40(3):914–927 (electronic), 2002.

[25] Todd F. Dupont and Itir Mogultay. A symmetric error estimate for Galerkin approximations of time-dependent Navier-Stokes equations in two dimensions. *Math. Comp.*, 78(268):1919–1927, 2009.

[26] Todd F. Dupont and Itir Mogultay. A new symmetric error estimate for a discrete-time moving mesh method. Technical report, University of Chicago, 09 2011.

[27] Alexandre Ern and Jean-Luc Guermond. *Theory and practice of finite elements*, volume 159 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2004.

[28] Lawrence C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1998.

[29] P. Grisvard. *Elliptic problems in nonsmooth domains*, volume 24 of *Monographs and Studies in Mathematics*. Pitman (Advanced Publishing Program), Boston, MA, 1985.

[30] W. Hackbusch. Optimal $H^{p,p/2}$ error estimates for a parabolic Galerkin method. *SIAM J. Numer. Anal.*, 18(4):681–692, 1981.

[31] H. Johnen and K. Scherer. On the equivalence of the K-functional and moduli of continuity and some applications. In *Constructive theory of functions of several variables (Proc. Conf., Math. Res. Inst., Oberwolfach, 1976)*, pages 119–140. Lecture Notes in Math., Vol. 571. Springer, Berlin, 1977.

[32] M. Karkulik, D. Pavlicek, and D. Praetorius. On 2d newest vertex bisection: Optimality of mesh-closure and H^1-stability of L_2-projection. *Constructive Approximation*, 2013.

[33] Jacques-Louis Lions and E. Magenes. *Non–Homogeneous Boundary Value Problems and Applications I.* Number 181 in Die Grundlehren der mathematischen Wissenschaften. Springer, Berlin, 1972.

[34] Mario Milman. Notes on limits of Sobolev spaces and the continuity of interpolation scales. *Trans. Amer. Math. Soc.*, 357(9):3425–3442 (electronic), 2005.

[35] L. E. Payne and H. F. Weinberger. An optimal Poincaré inequality for convex domains. *Arch. Rational Mech. Anal.*, 5:286–292 (1960), 1960.

[36] Michael Renardy and Robert C. Rogers. *An introduction to partial differential equations*, volume 13 of *Texts in Applied Mathematics*. Springer-Verlag, New York, second edition, 2004.

[37] Roberta Sacchi and Andreas Veeser. Locally efficient and reliable a posteriori error estimators for Dirichlet problems. *Math. Models Methods Appl. Sci.*, 16(3):319–346, 2006.

[38] Christoph Schwab and Rob Stevenson. Space-time adaptive wavelet methods for parabolic evolution problems. *Math. Comp.*, 78(267):1293–1318, 2009.

[39] L. Ridgway Scott and Shangyou Zhang. Finite element interpolation of nonsmooth functions satisfying boundary conditions. *Math. Comp.*, 54(190):483–493, 1990.

[40] Roger Temam. *Infinite-dimensional dynamical systems in mechanics and physics*, volume 68 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1988.

[41] Vidar Thomée. *Galerkin finite element methods for parabolic problems*, volume 25 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second edition, 2006.

[42] Franco Tomarelli. Regularity theorems and optimal error estimates for linear parabolic Cauchy problems. *Numer. Math.*, 45(1):23–50, 1984.

[43] Andreas Veeser. Approximating gradients with continuous piecewise polynomial functions. Preprint, Dipartimento di Matematica, submitted.

[44] Andreas Veeser and Rüdiger Verfürth. Explicit upper bounds for dual norms of residuals. *SIAM J. Numer. Anal.*, 47(3):2387–2405, 2009.

[45] Andreas Veeser and Rüdiger Verfürth. Poincaré constants for finite element stars. *IMA J. Numer. Anal.*, 32(1):30–47, 2012.

[46] Rüdiger Verfürth. A note on polynomial approximation in Sobolev spaces. *M2AN Math. Model. Numer. Anal.*, 33(4):715–719, 1999.

[47] J. Wloka. *Partial differential equations*. Cambridge University Press, Cambridge, 1987. Translated from the German by C. B. Thomas and M. J. Thomas.

[48] Jinchao Xu and Ludmil Zikatanov. Some observations on Babuška and Brezzi theories. *Numer. Math.*, 94(1):195–202, 2003.