

*Continuous interior penalty finite element method for  
the transient convection–diffusion–reaction equation*

Erik Burman — Miguel A. Fernández

**N° 6543**

May 2008

Thème BIO



*Rapport  
de recherche*





# Continuous interior penalty finite element method for the transient convection–diffusion–reaction equation

Erik Burman\*, Miguel A. Fernández†

Thème BIO — Systèmes biologiques  
Projet REO

Rapport de recherche n° 6543 — May 2008 — 25 pages

**Abstract:** We consider time-stepping methods for continuous interior penalty (CIP) stabilized finite element approximations of singularly perturbed parabolic problems or hyperbolic problems. We focus on methods for which the linear system obtained after discretization has the same matrix pattern as a standard Galerkin method. We prove that an iterative method using only the standard Galerkin matrix stencil is convergent. We also prove that the combination of the CIP stabilized finite element method with some known  $\mathcal{A}$ -stable time discretizations leads to unconditionally stable and optimally convergent schemes. In particular, we show that the contribution from the gradient jumps leading to the extended stencil may be extrapolated from previous time steps, and hence handled explicitly without loss of stability and accuracy. With these variants, unconditional stability and optimal accuracy is obtained for first order schemes, whereas for the second order backward differencing scheme a CFL-like condition has to be respected. The CFL condition is related to the size of the stabilization parameter of the stabilized method but is independent of the diffusion coefficient.

**Key-words:** Time discretization, stabilized finite element method, continuous interior penalty method, standard Galerkin stencil, transient convection-reaction-diffusion equation

Preprint submitted to *Comput. Methods Appl. Mech. Engrg.*

\* University of Sussex, Dep. of Mathematics, UK; e-mail: [E.N.Burman@sussex.ac.uk](mailto:E.N.Burman@sussex.ac.uk)

† INRIA, REO project-team; e-mail: [miguel.fernandez@inria.fr](mailto:miguel.fernandez@inria.fr)

Unité de recherche INRIA Rocquencourt  
Domaine de Voluceau, Rocquencourt, BP 105, 78153 Le Chesnay Cedex (France)  
Téléphone : +33 1 39 63 55 11 — Télécopie : +33 1 39 63 53 30

## Méthode d'éléments finis avec pénalisation intérieure pour l'équation de convection–diffusion–réaction transitoire

**Résumé :** On considère des méthodes de marche en temps pour des problèmes hyperboliques ou de perturbation singulière paraboliques, approchés en espace par la méthode d'éléments finis stabilisée avec pénalisation intérieure continue (CIP). En particulier, on s'intéresse à des méthodes pour lesquelles le système linéaire, obtenu après discrétisation, a la même structure creuse que la méthode de Galerkin standard. On propose d'abord un méthode itérative, comportant seulement la résolution des systèmes avec structure creuse standard, pour laquelle on montre convergence. D'autre part, on montre que la combinaison de la méthode d'éléments finis stabilisée CIP avec quelques méthodes connues de discrétisation en temps  $\mathcal{A}$ -stables, donne lieu à des schémas inconditionnellement stables et convergents. En outre, la contribution des sauts du gradient (donnant lieu à la structure creuse étendue de la méthode CIP) peut être extrapolée à partir des pas de temps précédents, et donc traitée explicitement, sans perte de stabilité et de précision. Dans ce cadre, on montre stabilité inconditionnelle pour des schémas d'ordre un, tandis que pour un schémas d'ordre 2 (BDF2) la stabilité dépend d'une condition CFL. Cette condition fait intervenir le paramètre de stabilisation de la méthode CIP, mais elle est indépendante du coefficient de diffusion.

**Mots-clés :** Discrétisation en temps, méthode d'éléments finis stabilisée, méthode de pénalisation intérieure, structure creuse de Galerkin standard, équation de convection-diffusion-réaction transitoire

## 1 Introduction

The interior penalty method for elliptic and parabolic problems was introduced in 1976 by Douglas and Dupont in the seminal work [8]. In 2004 Burman and Hansbo [5] proved that the method was robust at high Peclet numbers and enjoyed the same stability properties as the Streamline-Diffusion (SD) method. A number of extensions to various problems in fluid mechanics were then proposed by Burman and co-workers. An extension to non-conforming approximation spaces was proposed in [3]. The pressure stabilization for Stokes' problem was considered in [6] and stability and convergence of the Oseen's problem at high Reynolds numbers in [4]. The method has several advantages compared to the SD-method, mainly thanks to the fact that the stabilizing term does not couple to the low order residual and is therefore independent of both time derivatives and source terms. Hence, space and time discretization commute and the method can be combined with any type of time discretisation and nodal quadrature leads to a diagonal matrix contribution from stiff source terms. Another important feature is that the stabilization parameter is independent of the diffusion parameter and more generally has less dependence of the problem parameters than the SD method, since consistency of the SD-method depends on the residual of the differential equation, whereas the CIP method is weakly consistent, depending only on the regularity of the exact solution.

However these advantages come with a price: the size of the system matrix doubles in two space dimensions and triples in three. This is a consequence of the added interelement couplings introduced by the gradient jump stabilization operator. In this paper we consider the problem of the extended stencil and show that the equations may in fact be solved using a matrix containing only the part with the same pattern as a standard Galerkin method. More precisely, we propose three different approaches to time-stepping where only the standard Galerkin pattern is considered:

- (A). Time-stepping with relaxed fixed point iterations.
- (B). First order unconditionally stable timestepping to a steady state solution.
- (C). Second order conditionally stable timestepping for transient flow.

Other relevant results concerning the analysis of stabilized finite element formulations for the transient convection-diffusion-equation can be found in the literature. In particular, the analysis of the GaLS (Galerkin Least-Square) stabilization and the  $\theta$ -scheme is reported in [12], the subgrid viscosity method in a semi-discrete formulation is investigated in [9], and in [7] the orthogonal transient sub-scales stabilization is combined with the backward Euler scheme.

The remainder of this paper is organized as follows. The next section introduces the two problems under consideration and some useful notation. The CIP stabilized finite element discretization is introduced in section §3. Section §4 is devoted to the stationary problem. We show that an iterative method, where only a linear system with the standard Galerkin pattern is solved at each iteration, converges. In section §5, we address the transient case by considering two  $\mathcal{A}$ -stable time-stepping schemes, the  $\theta$ -scheme and the second order backward difference formula (BDF2). We prove that the contribution from the gradient jumps leading to the extended stencil may be extrapolated from previous

time steps, and hence handled explicitly without loss of stability and accuracy. Numerical results illustrating the theory are reported in section §6 and some conclusions in section §7.

It is our hope that the results in the present paper will give some indication on how to make efficient implementations of the continuous interior penalty method.

## 2 Problem setting

Let  $\Omega$  be a domain in  $\mathbb{R}^d$  ( $d = 1, 2$  or  $3$ ), with a polyhedral boundary  $\partial\Omega$ , and  $T > 0$ . We consider the following two problems:

- Solving for  $u : \Omega \longrightarrow \mathbb{R}$ :

$$\begin{cases} \boldsymbol{\beta} \cdot \nabla u + \sigma u - \mu \Delta u = f, & \text{in } \Omega, \\ u = 0, & \text{on } \partial\Omega. \end{cases} \quad (1)$$

- Solving for  $u : \Omega \times [0, T] \longrightarrow \mathbb{R}$ :

$$\begin{cases} \partial_t u + \boldsymbol{\beta} \cdot \nabla u + \sigma u - \mu \Delta u = f, & \text{in } \Omega \times (0, T), \\ u = 0, & \text{on } \partial\Omega \times (0, T), \\ u(\cdot, 0) = u_0, & \text{in } \Omega. \end{cases} \quad (2)$$

where  $\boldsymbol{\beta}$  is a given, Lipschitz continuous, velocity field satisfying  $\nabla \cdot \boldsymbol{\beta} = 0$  (and that may depend on  $t$ ),  $f$  is a source function,  $u_0$  the initial data and  $\sigma, \mu \geq 0$  are given bounded functions. The case  $\mu = 0$  (hyperbolic regime) requires a suitable modification of the boundary conditions of (1) and (2).

Let us introduce some standard notation. For a given domain  $\omega \subset \mathbb{R}^d$ , the space of functions whose distributional derivatives of order up to  $m \geq 0$  belong to  $L^2(\omega)$  is denoted by  $H^m(\omega)$ . The subspace of  $H^1(\omega)$  consisting of functions vanishing on the boundary is denoted as  $H_0^1(\omega)$ . The norm of  $H^m(\omega)$  is denoted by  $\|\cdot\|_{m,\omega}$ . The  $L^2$  norm is denoted by  $\|\cdot\|_{0,\Omega}$  and its inner product by  $(\cdot, \cdot)_\omega$ . The latter being simplified in the case  $\omega = \Omega$  as  $(\cdot, \cdot) \stackrel{\text{def}}{=} (\cdot, \cdot)_\Omega$ .

In order to introduce a variational setting for (1) and (2) we consider the space  $V \stackrel{\text{def}}{=} H_0^1(\Omega)$  and we define the continuous bi-linear form

$$a(u, v) = (\boldsymbol{\beta} \cdot \nabla u + \sigma u, v) + (\mu \nabla u, \nabla v), \quad \forall u, v \in V,$$

where  $(\cdot, \cdot)$  indicates the inner product of  $L^2(\Omega)$ . The above problems can then be cast in the weak form, respectively, as follows:

$$\begin{cases} \text{Find } u \in V \text{ such that} \\ a(u, v) = (f, v), \quad \forall v \in V, \end{cases} \quad (3)$$

and

$$\begin{cases} \text{For all } t \in (0, T), \text{ find } u(t) \in V \text{ such that} \\ (\partial_t u, v) + a(u, v) = (f, v), \quad \forall v \in V, \\ u(0) = u_0. \end{cases} \quad (4)$$

### 3 Space discretization

Let  $\{\mathcal{T}_h\}_{0 < h \leq 1}$  denote a family of shape regular triangulations of the domain  $\Omega$ . In order to simplify the presentation, and without loss of generality, we assume that the triangulations  $\mathcal{T}_h$  are quasi-uniform. The main results of the next sections can be proved assuming only that the family of triangulations is shape regular, under a constraint on the local variation of the mesh size (see *e.g.* [2, 1]). For each triangulation  $\mathcal{T}_h$ , the subscript  $h \in (0, 1]$  refers to the level of refinement of the triangulation, which is defined by

$$h \stackrel{\text{def}}{=} \max_{K \in \mathcal{T}_h} h_K, \quad h_K \stackrel{\text{def}}{=} \max_{f \subset \partial K} h_f,$$

with  $h_f$  the diameter of the face  $f$ .

From now on, we let  $V_h$  denote the standard finite element space of continuous piecewise polynomial functions of degree  $k \geq 1$ . In order to simplify the analysis below, Dirichlet boundary conditions will be enforced in a weak sense, using Nitsche's method [13] (see also [14, Page 24]). Then, the finite element CIP stabilized approximations of (3) and (4) are defined, respectively, as follows:

$$\begin{cases} \text{Find } u_h \in V_h \text{ such that} \\ a_h(u_h, v_h) + j(u_h, v_h) = (f, v_h), \quad \forall v_h \in V_h, \end{cases} \quad (5)$$

and

$$\begin{cases} \text{For all } t \in (0, T), \text{ find } u_h(t) \in V_h \text{ such that} \\ (\partial_t u_h, v_h) + a_h(u_h, v_h) + j(u_h, v_h) = (f, v_h), \quad \forall v_h \in V_h, \\ (u_h(0), v_h) = (u_0, v_h), \quad \forall v_h \in V_h, \end{cases} \quad (6)$$

where

$$\begin{aligned} a_h(u_h, v_h) &\stackrel{\text{def}}{=} a(u_h, v_h) - (\mu \nabla u_h \cdot \mathbf{n}, v_h)_{\partial\Omega} - (u_h, \mu \nabla v_h \cdot \mathbf{n})_{\partial\Omega} \\ &\quad + \frac{\gamma_{bc}\mu}{h} (u_h, v_h)_{\partial\Omega} - (\boldsymbol{\beta} \cdot \mathbf{n} u_h, v_h)_{\Gamma_{in}}, \\ j(u_h, v_h) &\stackrel{\text{def}}{=} \gamma \sum_{K \in \mathcal{T}_h} \sum_{f \in \partial K} \int_f h_f^2 |\boldsymbol{\beta} \cdot \mathbf{n}_f| [[\mathbf{n}_f \cdot \nabla u_h]] [[\mathbf{n}_f \cdot \nabla v_h]] ds, \end{aligned} \quad (7)$$

with  $\gamma_{bc} > 0$  the Nitsche's penalty parameter,  $\Gamma_{in} \stackrel{\text{def}}{=} \{\mathbf{x} \in \partial\Omega : (\boldsymbol{\beta} \cdot \mathbf{n})(\mathbf{x}) < 0\}$  the inlet boundary,  $\gamma > 0$  the stabilization parameter and  $[[\mathbf{n}_f \cdot \nabla u_h]] \stackrel{\text{def}}{=} \mathbf{n}_f \cdot (\nabla u_h^- - \nabla u_h^+)$  denoting the jump of the normal gradient over an interior face  $f$  ( $f \not\subset \partial\Omega$ ) of the simplex  $K \in \mathcal{T}_h$ . Here,  $\nabla u^+|_{f \in \partial K}$  stands for the value  $\lim_{\epsilon \rightarrow 0^+} \nabla u(x + \epsilon \mathbf{n}_K)$  and similarly for  $\nabla u^-|_{f \in \partial K}$ . For faces  $f \in \partial\Omega$ ,  $[[\mathbf{n}_f \cdot \nabla u_h]] = 0$ .

As pointed out in [5] (see also [4]), where a priori error estimates independent of the viscosity coefficient are provided for (5), the gradient jump (7) serves to stabilize the convective term. The price to pay is an extended matrix stencil of the stiffness matrix, due to the fact that the jump term couples neighboring elements. In practice, this leads to a higher computational cost of the linear system solution, for instance, if an incomplete LU factorization is used as preconditioner.

For the convergence analysis below, we introduce the following discrete *triple-norm*

$$\|v_h\|^2 \stackrel{\text{def}}{=} \|\sigma^{\frac{1}{2}} v_h\|_{0,\Omega}^2 + \|\mu^{\frac{1}{2}} \nabla v_h\|_{0,\Omega}^2 + j(v_h, v_h),$$

for all  $v_h \in V_h$ . Note that for  $\sigma = \mu = 0$  this is only a semi-norm. From the analysis in the steady case, see [5, 4], we know that the bilinear form  $a_h(\cdot, \cdot)$  satisfies the following coercivity and continuity estimate, where  $\pi_h$  denotes the standard  $L^2$ -projection onto  $V_h$ :

$$\|v_h\| \lesssim a_h(v_h, v_h) + j(v_h, v_h), \quad (8)$$

$$a_h(v - \pi_h v, v_h) + j(\pi_h v, v_h) \lesssim h^k (|\beta|_{\infty}^{\frac{1}{2}} h^{\frac{1}{2}} + |\sigma|_{\infty}^{\frac{1}{2}} h + |\mu|_{\infty}^{\frac{1}{2}}) \|v\|_{k+1,\Omega} \|v_h\| \quad (9)$$

$$+ h^{k+1} |\beta|_{1,\infty,\Omega} \|v\|_{k+1,\Omega} \|v_h\|_{0,\Omega}, \quad (10)$$

for all  $v_h \in V_h$ ,  $v \in H^{k+1}(\Omega)$  and with the notations  $|\beta|_{\infty} \stackrel{\text{def}}{=} \|\beta\|_{0,\infty,\Omega}$ ,  $|\sigma|_{\infty} \stackrel{\text{def}}{=} \|\sigma\|_{0,\infty,\Omega}$  and  $|\mu|_{\infty} \stackrel{\text{def}}{=} \|\mu\|_{0,\infty,\Omega}$ . Here, and in what follows, the symbol  $\lesssim$  indicates an inequality up to a multiplicative constant independent of the discretization and physical parameters.

### 3.1 Preliminary results

In this paper we propose a couple of strategies that allow to avoid the extended stencil of the CIP stabilization when solving (5) or (6). To this aim, let us first rewrite the interior penalty operator (7) into two parts, one that gives a contribution only to the standard Galerkin stencil,  $j_{sG}(\cdot, \cdot)$ , and another part,  $j_X(\cdot, \cdot)$ , that contributes both to the standard Galerkin stencil and to the extended stencil. This is stated by the following Lemma.

**Lemma 3.1 (Stencil decomposition)** *For each  $u_h, v_h \in V_h$ , we have*

$$j(u_h, v_h) = j_{sG}(u_h, v_h) - j_X(u_h, v_h),$$

where

$$j_{sG}(u_h, v_h) \stackrel{\text{def}}{=} \gamma \sum_{K \in \mathcal{T}_h} \sum_{f \in \partial K} \int_f h_f^2 |\beta \cdot \mathbf{n}_f| \mathbf{n}_f \cdot \nabla u_h^- \mathbf{n}_f \cdot \nabla v_h^- \, ds,$$

$$j_X(u_h, v_h) \stackrel{\text{def}}{=} \gamma \sum_{K \in \mathcal{T}_h} \sum_{f \in \partial K} \int_f h_f^2 |\beta \cdot \mathbf{n}_f| \mathbf{n}_f \cdot \nabla u_h^- \mathbf{n}_f \cdot \nabla v_h^+ \, ds.$$

*Proof.* The proof is elementary noting that by summation

$$\begin{aligned} j_{sG}(u_h, v_h) - j_X(u_h, v_h) &= \gamma \sum_{K \in \mathcal{T}_h} \sum_{f \in \partial K} \int_f h_f^2 |\beta \cdot \mathbf{n}_f| \mathbf{n}_f \cdot \nabla u_h^- \llbracket \mathbf{n}_f \cdot \nabla v_h \rrbracket \, ds \\ &= \sum_{K \in \mathcal{T}_h} \sum_{f \in \partial K} \int_f h_f^2 |\beta \cdot \mathbf{n}_f| \llbracket \mathbf{n}_f \cdot \nabla u_h \rrbracket \llbracket \mathbf{n}_f \cdot \nabla v_h \rrbracket \, ds, \quad (11) \end{aligned}$$

which concludes the proof.  $\square$

The following Cauchy-Schwarz inequalities for the stabilization operators will be useful

$$j_{sG}(u_h, v_h) \leq j_{sG}(u_h, u_h)^{\frac{1}{2}} j_{sG}(v_h, v_h)^{\frac{1}{2}},$$

$$j_X(u_h, v_h) \leq j_{sG}(u_h, u_h)^{\frac{1}{2}} j_{sG}(v_h, v_h)^{\frac{1}{2}},$$

for all  $u_h, v_h \in V_h$ . We will also make use of the following estimate.

**Lemma 3.2** Let  $v_h, w_h, z_h \in V_h$  be given. Then, there holds

$$j_X(v_h - w_h, z_h) \leq j_{sG}(v_h - w_h, z_h) + \frac{1}{2}(j_{sG}(v_h - w_h, v_h - w_h) + j(z_h, z_h)).$$

*Proof.* The proof follows by adding and subtracting suitable terms as follows

$$\begin{aligned} j_X(v_h - w_h, z_h) &= \sum_{K \in \mathcal{T}_h} \sum_{f \in \partial K} \int_f h_f^2 |\boldsymbol{\beta} \cdot \mathbf{n}_K| \mathbf{n}_K \cdot \nabla(v_h - w_h)^- \mathbf{n}_K \cdot \nabla z_h^+ \, ds \\ &= \sum_{K \in \mathcal{T}_h} \sum_{f \in \partial K} \int_f h_f^2 |\boldsymbol{\beta} \cdot \mathbf{n}_K| \mathbf{n}_K \cdot \nabla(v_h - w_h)^- \mathbf{n}_K \cdot (\nabla z_h^+ - \nabla z_h^-) \, ds + j_{sG}(v_h - w_h, z_h). \end{aligned}$$

We conclude by applying Cauchy-Schwarz inequality and the arithmetic-geometric inequality.  $\square$

Finally, the Galerkin part of the jump can be bounded as follows.

**Lemma 3.3** There holds

$$\begin{aligned} j_X(u_h, v_h) &\lesssim C_T \gamma |\boldsymbol{\beta}|_\infty \|\nabla u_h\|_{0,\Omega} \|v_h\|_{0,\Omega}, \\ j_{sG}(u_h, v_h) &\lesssim C_T \gamma |\boldsymbol{\beta}|_\infty \|\nabla u_h\|_{0,\Omega} \|v_h\|_{0,\Omega}, \\ j_{sG}(u_h, u_h) &\lesssim C_T \gamma |\boldsymbol{\beta}|_\infty h^{-1} \|u_h\|_{0,\Omega}^2, \end{aligned}$$

for all  $u_h, v_h \in V_h$ .

*Proof.* An immediate consequence of the Cauchy-Schwarz inequality and the trace inequality followed by an inverse inequality.  $\square$

## 4 An iterative scheme

Based on the jump stencil decomposition provided by Lemma 3.1, we consider the following iterative method for solving problem (5):

$$\begin{cases} \text{Given } u_h^0 \in V_h, \text{ for } j \geq 0, \text{ find } u_h^{j+1} \in V_h \text{ such that} \\ a_h(u_h^{j+1}, v_h) + \alpha j_{sG}(u_h^{j+1}, v_h) = (f, v_h) \\ \quad + (\alpha - 1)j_{sG}(u_h^j, v_h) + j_X(u_h^j, v_h), \quad \forall v_h \in V_h, \end{cases} \quad (12)$$

where  $\alpha > 1$  is a given relaxation parameter.

Note that, contrary to (5), in the iterative procedure (12) the stiffness matrix of the corresponding linear system has a standard Galerkin stencil.

**Theorem 4.1** For  $\alpha \geq 3$  we have

$$\lim_{N \rightarrow \infty} \|u_h - u_h^N\| = 0.$$

*Proof.* Subtracting formulation (1) from formulation (12) yields

$$a_h(u_h^{j+1} - u_h, v_h) + j(u_h^{j+1} - u_h, v_h) + (\alpha - 1)j_{sG}(u_h^{j+1} - u_h^j, v_h) + j_X(u_h^{j+1} - u_h^j, v_h) = 0,$$

for all  $v_h \in V_h$ . Denoting now, for simplicity,  $e_h^{j+1} \stackrel{\text{def}}{=} u_h^{j+1} - u_h$  and taking  $v_h = e_h^{j+1}$  in the previous expression, we have

$$a_h(e_h^{j+1}, e_h^{j+1}) + j(e_h^{j+1}, e_h^{j+1}) + (\alpha - 1)j_{sG}(e_h^{j+1} - e_h^j, e_h^{j+1}) + j_X(e_h^{j+1} - e_h^j, e_h^{j+1}) = 0.$$

By applying Lemma 3.2 we get

$$a_h(e_h^{j+1}, e_h^{j+1}) + \frac{1}{2}j(e_h^{j+1}, e_h^{j+1}) + (\alpha - 2)j_{sG}(e_h^{j+1} - e_h^j, e_h^{j+1}) - \frac{1}{2}j_{sG}(e_h^{j+1} - e_h^j, e_h^{j+1} - e_h^j) \leq 0. \quad (13)$$

On the other hand, using the symmetry of  $j_{sG}(\cdot, \cdot)$ , we have

$$j_{sG}(e_h^{j+1} - e_h^j, e_h^{j+1}) = \frac{1}{2} \left[ j_{sG}(e_h^{j+1}, e_h^{j+1}) - j_{sG}(e_h^j, e_h^j) + j_{sG}(e_h^{j+1} - e_h^j, e_h^{j+1} - e_h^j) \right].$$

Thus, from (13), we obtain

$$a_h(e_h^{j+1}, e_h^{j+1}) + \frac{1}{2}j(e_h^{j+1}, e_h^{j+1}) + \frac{\alpha - 2}{2} \left( j_{sG}(e_h^{j+1}, e_h^{j+1}) - j_{sG}(e_h^j, e_h^j) \right) + \frac{\alpha - 3}{2} j_{sG}(e_h^{j+1} - e_h^j, e_h^{j+1} - e_h^j) \leq 0.$$

In particular, using the coercivity of the bi-linear form (8) and the positivity of  $j_{sG}(\cdot, \cdot)$ , for  $\alpha \geq 3$  we get

$$\|e_h^{j+1}\|^2 + j_{sG}(e_h^{j+1}, e_h^{j+1}) - j_{sG}(e_h^j, e_h^j) \leq 0.$$

Finally, summing over  $j = 1, \dots, N - 1$  yields

$$j_{sG}(e_h^N, e_h^N) + \sum_{j=0}^{N-1} \|e_h^{j+1}\|^2 \leq j_{sG}(e_h^0, e_h^0),$$

and we may conclude by letting  $N \rightarrow \infty$ .  $\square$

## 5 Time discretization

In this section we address the stability and convergence of a class of fully discrete formulations obtained from (4) and some known  $\mathcal{A}$ -stable schemes for ODE's. Particular focus is made on the formulations which only involve the solution of linear systems with a standard Galerkin stencil.

The next paragraph is devoted to the  $\theta$ -scheme variants. Paragraph §5.2 addresses the second order backward difference formula (BDF2). Let  $N > 0$  be a given positive integer. In what follows, we consider a uniform partition of the time interval of interest  $[0, T]$  with time step size  $\tau \stackrel{\text{def}}{=} T/N$ . In addition, the discrete value  $u_h^n \in V_h$  stands for an approximation of  $u(t^n)$  in  $V_h$ , with  $t^n \stackrel{\text{def}}{=} n\tau$  and  $0 \leq n \leq N - 1$ .

### 5.1 Timestepping with the $\theta$ -scheme

For  $\theta \in [1/2, 1]$ , the classical  $\theta$ -scheme time discretization of (4) reads as follows:

$$\begin{cases} \text{For } 0 \leq n \leq N - 1, \text{ find } u_h^{n+1} \in V_h \text{ such that} \\ (\partial_\tau u_h^{n+1}, v_h) + \tau a_h(u_h^{n+\theta}, v_h) + \tau j(u_h^{n+1}, v_h) = \tau (f^{n+\theta}, v_h), \quad \forall v_h \in V_h, \\ u_h^0 = \pi_h u_0. \end{cases} \quad (14)$$

with the notations  $\partial_\tau u_h^{n+1} \stackrel{\text{def}}{=} (u_h^{n+1} - u_h^n)/\tau$ ,  $u_h^{n+\theta} \stackrel{\text{def}}{=} \theta u_h^{n+1} + (1 - \theta)u_h^n$ ,  $f^{n+\theta} \stackrel{\text{def}}{=} f(t^{n+\theta})$ , and  $t^{n+\theta} \stackrel{\text{def}}{=} \theta t^{n+1} + (1 - \theta)t^n$ .

At each time-level  $0 \leq n \leq N - 1$ , problem (14) is a particular case of (5) with  $\sigma \neq 0$ , which involves the solution of a linear system with a non-standard Galerkin stencil. Therefore, the iterative scheme introduced in Section §4 may be applied at each time-level of (14). This provides a solution procedure which only involves linear systems with a standard Galerkin matrix pattern. In this section, however, we shall see that, for  $\theta \in (1/2, 1]$ , no iteration is necessary to assure stability and optimal convergence rate. To this aim, let us consider the following time-stepping formulation:

$$\begin{cases} \text{For } 0 \leq n \leq N - 1, \text{ find } u_h^{n+1} \in V_h \text{ such that:} \\ (\partial_\tau u_h^{n+1}, v_h) + a_h(u_h^{n+\theta}, v_h) + J_{\alpha,\lambda}(u_h^n, u_h^{n+1}; v_h) = (f^{n+\theta}, v_h), \quad \forall v_h \in V_h, \\ u^0 = \pi_h u_0. \end{cases} \quad (15)$$

where  $J_{\alpha,\lambda}$  stands for the relaxed CIP stabilization operator defined by

$$\begin{aligned} J_{\alpha,\lambda}(u_h^n, u_h^{n+1}; v_h) &\stackrel{\text{def}}{=} \alpha j_{sG}(u_h^{n+1}, v_h) + (1 - \alpha)j_{sG}(u_h^n, v_h) \\ &\quad - \lambda j_X(u_h^{n+1}, v_h) - (1 - \lambda)j_X(u_h^n, v_h), \end{aligned} \quad (16)$$

with  $\alpha \geq 1$  and  $\lambda \in \{0, 1\}$  two given relaxation parameters.

For  $\theta = \alpha = \lambda = 1$  (CIP stencil), the discrete formulation (15) reduces to the standard Backward Euler scheme, and for  $\theta = \frac{1}{2}$ ,  $\alpha = \lambda = 1$  (CIP stencil) to the Crank-Nicholson variant. On the contrary, for  $\lambda = 0$  only the standard Galerkin contribution of the stabilization term is treated in an implicit fashion (Galerkin stencil).

### 5.1.1 Stability

The following result states the stability properties of the fully discrete scheme (15).

**Theorem 5.1 (Stability  $\theta$ -scheme)** *Assume that one of the following three conditions holds:*

- CIP stencil:

$$\frac{1}{2} \leq \theta \leq 1, \quad \lambda = 1, \quad \alpha \geq 1. \quad (17)$$

- standard Galerkin stencil (high over-relaxation):

$$\frac{1}{2} < \theta \leq 1, \quad \lambda = 0, \quad \alpha \geq \frac{4\theta - 1}{2\theta - 1}. \quad (18)$$

- standard Galerkin stencil (low over-relaxation):

$$\begin{aligned} \frac{1}{2} < \theta \leq 1, \quad \lambda = 0, \quad 1 \leq \alpha < \frac{4\theta - 1}{2\theta - 1}, \\ C_T \gamma \tau |\beta|_\infty \leq h \min \left\{ \frac{1}{4\theta}, \theta - \frac{1}{2} \right\}. \end{aligned} \quad (19)$$

Then, for all  $\{e_h^n\}_{n=0}^N \in [V_h]^{N+1}$  there holds

$$\begin{aligned} \sum_{m=0}^{n-1} \tau \left[ (\partial_\tau e_h^{m+1}, e_h^{m+\theta}) + a_h(e_h^{m+\theta}, e_h^{m+\theta}) + J_{\theta,\lambda}(e_h^m, e_h^{m+1}; e_h^{m+\theta}) \right] \\ \geq \frac{1}{4} \|e_h^n\|_{0,\Omega}^2 + \frac{1}{2} \sum_{m=0}^{n-1} \tau \|e_h^{m+\theta}\|^2, \end{aligned} \quad (20)$$

for  $1 \leq n \leq N-1$ .

*Proof.* For  $1 \leq n \leq N-1$ , we first define

$$\begin{aligned} I_1 &= \sum_{m=0}^{n-1} [(e_h^{m+1} - e_h^m, e_h^{m+\theta}) + \tau a_h(e_h^{m+1}, e_h^{m+\theta})], \\ I_2 &= \tau \sum_{m=0}^{n-1} J_{\alpha,\lambda}(e_h^m, e_h^{m+1}; e_h^{m+\theta}). \end{aligned}$$

Note that

$$(e_h^{m+1} - e_h^m, e_h^{m+\theta}) = \frac{1}{2} \|e_h^{m+1}\|_{0,\Omega}^2 - \frac{1}{2} \|e_h^m\|_{0,\Omega}^2 + \frac{2\theta-1}{2} \|e_h^{m+1} - e_h^m\|_{0,\Omega}^2, \quad (21)$$

and, therefore,

$$I_1 \geq \frac{1}{2} \|e_h^n\|_{0,\Omega}^2 - \frac{1}{2} \|e_h^0\|_{0,\Omega}^2 + \sum_{m=0}^{n-1} \left[ \frac{2\theta-1}{2} \|e_h^{m+1} - e_h^m\|_{0,\Omega}^2 + \tau a_h(e_h^{m+\theta}, e_h^{m+\theta}) \right]. \quad (22)$$

On the other hand, from (3.1) and Lemma 3.2 (note that  $\lambda \in \{0, 1\}$ ), we have

$$\begin{aligned} J_{\alpha,\lambda}(e_h^m, e_h^{m+1}; e_h^{m+\theta}) &= j(e_h^{m+1}, e_h^{m+\theta}) + (\alpha-1)j_{sG}(e_h^{m+1} - e_h^m, e_h^{m+\theta}) \\ &\quad + (1-\lambda)j_X(e_h^{m+1} - e_h^m, e_h^{m+\theta}) \\ &\geq j(e_h^{m+1}, e_h^{m+\theta}) + \frac{\lambda-1}{2} j(e_h^{m+\theta}, e_h^{m+\theta}) \\ &\quad + (\alpha+\lambda-2)j_{sG}(e_h^{m+1} - e_h^m, e_h^{m+\theta}) \\ &\quad + \frac{\lambda-1}{2} j_{sG}(e_h^{m+1} - e_h^m, e_h^{m+1} - e_h^m). \end{aligned} \quad (23)$$

Since  $e_h^{m+1} = e_h^{m+\theta} + (1-\theta)(e_h^{m+1} - e_h^m)$ , the symmetry of  $j(\cdot, \cdot)$  and an identity similar to (21) yield

$$\begin{aligned} j(e_h^{m+1}, e_h^{m+\theta}) &= j(e_h^{m+\theta}, e_h^{m+\theta}) + \frac{(1-\theta)}{2} \left[ j(e_h^{m+1}, e_h^{m+1}) - j(e_h^m, e_h^m) \right. \\ &\quad \left. + (2\theta-1)j(e_h^{m+1} - e_h^m, e_h^{m+1} - e_h^m) \right]. \end{aligned}$$

Similarly, thanks to the symmetry of  $j_{sG}(\cdot, \cdot)$ , we have

$$\begin{aligned} j_{sG}(e_h^{m+1} - e_h^m, e_h^{m+\theta}) &= \frac{1}{2} j_{sG}(e_h^{m+1}, e_h^{m+1}) - \frac{1}{2} j_{sG}(e_h^m, e_h^m) \\ &\quad + \frac{2\theta-1}{2} j_{sG}(e_h^{m+1} - e_h^m, e_h^{m+1} - e_h^m). \end{aligned}$$

Therefore, from (23) and summing over  $m = 0, \dots, n - 1$ , we have

$$\begin{aligned}
 I_2 &\geq \frac{\tau}{2}(\lambda + 1) \sum_{m=0}^{n-1} j(e_h^{m+\theta}, e_h^{m+\theta}) + \frac{\tau}{2}(1 - \theta) (j(e_h^n, e_h^n) - j(e_h^0, e_h^0)) \\
 &\quad + \frac{\tau}{2}(1 - \theta)(2\theta - 1) \sum_{m=0}^{n-1} j(e_h^{m+1} - e_h^m, e_h^{m+1} - e_h^m) \\
 &\quad + \frac{\tau}{2}(\alpha + \lambda - 2) (j_{sG}(e_h^n, e_h^n) - j_{sG}(e_h^0, e_h^0)) \\
 &\quad + \frac{\tau}{2}[(2\theta - 1)\alpha + 2\theta\lambda - 4\theta + 1] \sum_{m=0}^{n-1} j_{sG}(e_h^{m+1} - e_h^m, e_h^{m+1} - e_h^m).
 \end{aligned} \tag{24}$$

Under condition (17), or condition (18), and since  $e_h^0 = 0$ , there follows that

$$I_2 \geq \frac{\tau}{2} \sum_{m=0}^{n-1} j(e_h^{m+\theta}, e_h^{m+\theta}).$$

which, in combination with (22), yields (20).

Finally, under condition (19), we use Lemma 3.3 and (24) to obtain

$$\begin{aligned}
 I_2 &\geq \frac{\tau}{2} \sum_{m=0}^{n-1} j(e_h^{m+\theta}, e_h^{m+\theta}) - \frac{\tau}{2}|\alpha - 2|C_T\gamma|\beta|_\infty h^{-1} \|e_h^n\|_{0,\Omega}^2 \\
 &\quad - \frac{\tau}{2} |(2\theta - 1)\alpha - 4\theta + 1|C_T\gamma|\beta|_\infty h^{-1} \sum_{m=0}^{n-1} \|e_h^{m+1} - e_h^m\|_{0,\Omega}^2 \\
 &\geq \frac{\tau}{2} \sum_{m=0}^{n-1} j(e_h^{m+\theta}, e_h^{m+\theta}) - \frac{1}{4} \|e_h^n\|_{0,\Omega}^2 - \frac{1}{4} \sum_{m=0}^{n-1} \|e_h^{m+1} - e_h^m\|_{0,\Omega}^2
 \end{aligned}$$

By combining this estimate with (22) we obtain (20), which completes the proof.  $\square$

We conclude this paragraph with two remarks.

**Remark 5.2** Under condition (17) (CIP stencil), Theorem 5.1 provides the expected unconditional stability of the  $\theta$ -scheme (14). As a result, the Crank-Nicholson variant ( $\theta = 1/2$ ) combined with the CIP stabilization is unconditionally stable. It is worth noting that this is not the case for finite element stabilizations involving the residual of the PDE, like GaLS (see e.g. [12]), in which the time derivative included in the residual perturbs the stability of the traditional Crank-Nicholson scheme.

**Remark 5.3** Theorem 5.1, under condition (18), states the unconditional stability of the Galerkin stencil time-stepping. Note that the Crank-Nicholson variant is excluded since, for  $\theta = 1/2$ , the condition on the relaxation parameter  $\alpha$  blows up. Finally, Theorem 5.1 shows that one can (partially) bypass the blow up of  $\alpha$  with the payoff of the CFL condition (19) (independent of the viscosity coefficient  $\mu$ ). The Crank-Nicholson variant is also excluded in this case, since for  $\theta = 1/2$  the CFL condition requires the time-step  $\tau$  to be 0.

### 5.1.2 Convergence

In this paragraph we prove optimal *a priori* error estimates for the fully discrete formulation (15).

**Theorem 5.4 (Convergence  $\theta$ -scheme)** *Let  $u$  be the solution of (2) and  $\{u_h^n\}_{n=0}^N$  be the solution of (15). Assume that the hypothesis of Theorem 5.1 are satisfied and let  $e_h^n \stackrel{\text{def}}{=} \pi_h u(t^n) - u_h^n$ . Then, there holds*

$$\begin{aligned} \|e^n\|_{0,\Omega}^2 + \sum_{m=0}^{n-1} \tau \|e_h^{m+1}\|^2 &\lesssim \mathcal{H}^2(T, \mu, \beta, \sigma) \|u\|_{H^1(0,T;H^{k+1}(\Omega))}^2 + T\tau^4 \|\partial_{ttt}u\|_{L^2(0,T;L^2(\Omega))}^2 \\ &+ T\tau^2 \left[ T|(1-\theta)^2 - \theta^2| \|\partial_{tt}u\|_{H^1(0,T;L^2(\Omega))}^2 + |\beta|_\infty^2 ((\alpha-1)^2 + (1-\lambda)^2) \|\partial_t u(t)\|_{L^2(0,T;H^1(\Omega))}^2 \right], \end{aligned}$$

for  $1 \leq n \leq N$  and with  $\mathcal{H}(T, \mu, \beta, \sigma) \stackrel{\text{def}}{=} h^k (|\beta|_\infty^{\frac{1}{2}} h^{\frac{1}{2}} + T^{\frac{1}{2}} |\beta|_{1,\infty,\Omega} h + |\sigma|_\infty^{\frac{1}{2}} h + |\mu|_\infty^{\frac{1}{2}})$ .

*Proof.* From the stability result of Theorem 5.1 we have

$$\begin{aligned} \frac{1}{4} \|e_h^n\|_{0,\Omega}^2 + \frac{1}{2} \sum_{m=0}^{n-1} \tau \|e_h^{m+\theta}\|^2 &\leq \tau \sum_{m=0}^{n-1} \{(\partial_\tau e_h^{m+1}, e_h^{m+\theta}) + a_h(e_h^{m+\theta}, e_h^{m+\theta})\} \\ &+ \tau \sum_{m=0}^{n-1} J_{\alpha,\lambda}(e_h^m, e_h^{m+1}; e_h^{m+\theta}). \end{aligned} \quad (25)$$

On the other hand, subtracting (15) from (4), at  $t = t^{m+\theta}$ , yields

$$\begin{aligned} \tau (\partial_\tau u(t^{m+1}) - \partial_\tau u_h^{m+1}, v_h) + \tau a_h(u(t^{m+\theta}) - u_h^{m+\theta}, v_h) \\ = \tau J_{\alpha,\lambda}(u_h^m, u_h^{m+1}; v_h) + \tau (\partial_\tau u(t^{m+1}) - \partial_t u(t^{m+\theta}), v_h), \end{aligned} \quad (26)$$

for  $v_h \in V_h$ . Thus, from (25) and using the  $L^2$ -orthogonality of  $\pi_h$ , we have

$$\begin{aligned} \frac{1}{4} \|e_h^n\|_{0,\Omega}^2 + \sum_{m=0}^{n-1} \tau \|e_h^{m+\theta}\|^2 &\leq \sum_{m=0}^{n-1} \tau (\partial_\tau u(t^{m+1}) - \partial_t u(t^{m+\theta}), e_h^{m+\theta}) \\ &+ \sum_{m=0}^{n-1} \tau [a((\pi_h u - u)(t^{m+1}), e_h^{m+\theta}) + J_{\alpha,\lambda}(\pi_h u(t^m), \pi_h u(t^{m+1}); e_h^{m+\theta})]. \end{aligned}$$

Finally, using (23), we obtain

$$\begin{aligned}
 \frac{1}{4} \|e_h^n\|_{0,\Omega}^2 + \sum_{m=0}^{n-1} \tau \|e_h^{m+\theta}\|^2 &\leq \underbrace{\sum_{m=0}^{n-1} \tau (\partial_\tau u(t^{m+1}) - \partial_t u(t^{m+\theta}), e_h^{m+\theta})}_{I_1} \\
 &+ \underbrace{\sum_{m=0}^{n-1} \tau [a((\pi_h u - u)(t^{m+\theta}), e_h^{m+\theta}) + j(\pi_h u(t^{m+1}), e_h^{m+\theta})]}_{I_2} \\
 &+ \underbrace{\sum_{m=0}^{n-1} \tau (\alpha - 1) j_{sG}(\pi_h(u(t^{m+1}) - u(t^m)), e_h^{m+\theta})}_{I_3} \\
 &+ \underbrace{\sum_{m=0}^{n-1} \tau (1 - \lambda) j_X(\pi_h(u(t^{m+1}) - u(t^m)), e_h^{m+\theta})}_{I_4}. \quad (27)
 \end{aligned}$$

We now treat the the four contributions in the right hand side term by term.

The first term is standard, using a Taylor expansion we have

$$\begin{aligned}
 I_1 &= \sum_{m=0}^{n-1} (u(t^{m+1}) - u(t^m) - \tau \partial_t u(t^{m+\theta}), e_h^{m+\theta}) \\
 &\leq \sum_{m=0}^{n-1} \|u(t^{m+1}) - u(t^m) - \tau \partial_t u(t^{m+\theta})\|_{0,\Omega} \|e_h^{m+\theta}\|_{0,\Omega} \\
 &\leq \sum_{m=0}^{n-1} \tau^2 \frac{|(1-\theta)^2 - \theta^2|}{2} \|\partial_{tt} u(t^{m+\theta})\|_{0,\Omega} \|e_h^{m+\theta}\|_{0,\Omega} \\
 &\quad + \frac{1}{2} \sum_{m=0}^{n-1} \left\| \int_{t^m}^{t^{m+\theta}} (t-t^m)^2 \partial_{ttt} u(t) dt + \int_{t^{m+\theta}}^{t^{m+1}} (t-t^{m+1})^2 \partial_{ttt} u(t) dt \right\|_{0,\Omega} \|e_h^{m+\theta}\|_{0,\Omega} \\
 &\lesssim T \left( \tau^2 |(1-\theta)^2 - \theta^2| \sum_{m=0}^{n-1} \tau \|\partial_{tt} u(t^{m+\theta})\|_{0,\Omega}^2 + \tau^4 \|\partial_{ttt} u\|_{L^2(0,T;\Omega)}^2 \right) \\
 &\quad + \frac{1}{T} (1 + |(1-\theta)^2 - \theta^2|) \sum_{m=0}^{n-1} \tau \|e_h^{m+\theta}\|_{0,\Omega}^2.
 \end{aligned}$$

For  $I_2$  we use the continuity estimate (10) to obtain

$$\begin{aligned}
I_2 &\lesssim h^k (|\boldsymbol{\beta}|_{\infty}^{\frac{1}{2}} h^{\frac{1}{2}} + |\sigma|_{\infty}^{\frac{1}{2}} h + |\mu|_{\infty}^{\frac{1}{2}}) \left( \tau \sum_{m=0}^{n-1} \|u(t^{m+\theta})\|_{k+1,\Omega}^2 \right)^{\frac{1}{2}} \left( \sum_{m=0}^{n-1} \tau \|e_h^{m+\theta}\|^2 \right)^{\frac{1}{2}} \\
&\quad + h^{k+1} |\boldsymbol{\beta}|_{1,\infty,\Omega} \left( \tau \sum_{m=0}^{n-1} \|u(t^{m+\theta})\|_{k+1,\Omega}^2 \right)^{\frac{1}{2}} \left( \sum_{m=0}^{n-1} \tau \|e_h^{m+\theta}\|_{0,\Omega}^2 \right)^{\frac{1}{2}} \\
&\lesssim h^{2k} (|\boldsymbol{\beta}|_{\infty}^{\frac{1}{2}} h^{\frac{1}{2}} + T^{\frac{1}{2}} |\boldsymbol{\beta}|_{1,\infty,\Omega} h + |\sigma|_{\infty}^{\frac{1}{2}} h + |\mu|_{\infty}^{\frac{1}{2}})^2 \sum_{m=1}^{m-1} \tau \|u(t^{m+\theta})\|_{k+1,\Omega}^2 \\
&\quad + \frac{1}{4} \sum_{m=0}^{m-1} \tau \|e_h^{m+\theta}\|^2 + \frac{1}{T} \sum_{m=0}^{n-1} \tau \|e_h^{m+\theta}\|_{0,\Omega}^2.
\end{aligned}$$

For the third term we use Lemma 3.3, the  $H^1$ -stability of  $\pi_h$  (see *e.g.* [2, 1]), and a Taylor expansion. This yields

$$\begin{aligned}
I_3 &= \sum_{m=0}^{n-1} \tau (\alpha - 1) j_{sG} \left( \int_{t^m}^{t^{m+1}} \partial_t \pi_h u(t) \, dt, e_h^{m+\theta} \right) \\
&= \sum_{m=0}^{n-1} \tau (\alpha - 1) \int_{t^m}^{t^{m+1}} j_{sG}(\pi_h \partial_t u(t), e_h^{m+\theta}) \, dt \\
&\leq C_{T\gamma} \sum_{m=0}^{n-1} \tau |\alpha - 1| |\boldsymbol{\beta}|_{\infty} \left( \int_{t^m}^{t^{m+1}} \|\nabla \pi_h \partial_t u(t)\|_{0,\Omega} \, dt \right) \|e_h^{m+\theta}\|_{0,\Omega} \quad (28) \\
&\leq C_{T\gamma} \sum_{m=0}^{n-1} \tau |\alpha - 1| |\boldsymbol{\beta}|_{\infty} \left( \tau \int_{t^m}^{t^{m+1}} \|\nabla \pi_h \partial_t u(t)\|_{0,\Omega}^2 \, dt \right)^{\frac{1}{2}} \|e_h^{m+\theta}\|_{0,\Omega} \\
&\lesssim \left( |\boldsymbol{\beta}|_{\infty}^2 |\alpha - 1|^2 \tau^2 T \|\partial_t u\|_{L^2(0,T;H^1(\Omega))}^2 + \frac{1}{T} \sum_{m=0}^{n-1} \tau \|e_h^{m+\theta}\|_{0,\Omega}^2 \right).
\end{aligned}$$

and similarly for the last term

$$I_4 \lesssim \left( |1 - \lambda|^2 \tau^2 T \|\partial_t u(t)\|_{L^2(0,T;H^1(\Omega))}^2 + \frac{1}{T} \sum_{m=0}^{n-1} \tau \|e_h^{m+\theta}\|_{0,\Omega}^2 \right). \quad (29)$$

We conclude by absorbing the triple norm contribution  $\frac{1}{4} \sum_{m=0}^{n-1} \tau \|e_h^{m+\theta}\|^2$  in the left hand side and applying Gronwall's lemma (see *e.g.* [11, Lemma 5.1]). This proves the error estimate in the norm  $l^\infty(\|\cdot\|_{0,\Omega})$ . The bound for the  $l^2(\|\cdot\|)$ -norm contribution is obtained applying the  $l^\infty$ -estimate directly in equation (27).  $\square$

**Remark 5.5** For  $\alpha = \lambda = 1$ , Theorem 5.4 provides the expected convergence rate of the  $\theta$ -scheme: first order for  $\theta \in (1/2, 1]$  and second order for the Crank-Nicholson variant,  $\theta = 1/2$ . For completeness, we also observe that in the hypothetical case  $\alpha > 1$ ,  $\lambda = 1$  the convergence rate is always  $O(\tau)$ , due to the first order extrapolation involved in the relaxed CIP jump term (16). Second order extrapolations are exploited in Paragraph §5.2.

**Remark 5.6** *The previous theorem ensures that the stable Galerkin stencil variants of (15), i.e. under conditions (18) or (19), provide first order accuracy, which is optimal since  $\theta \in (1/2, 1]$ .*

## 5.2 Timestepping with BDF2

Let  $\tilde{u}_h^n \stackrel{\text{def}}{=} 2u_h^n - u_h^{n-1}$  and

$$\bar{\partial}_\tau u_h^{n+1} \stackrel{\text{def}}{=} \frac{1}{\tau} \left( \frac{3}{2}u_h^{n+1} - 2u_h^n + \frac{1}{2}u_h^{n-1} \right),$$

be the standard second order difference formula (BDF2). Consider the timestepping formulation:

$$\begin{cases} \text{For } 1 \leq n \leq N-1, \text{ find } u_h^{n+1} \in V_h \text{ such that:} \\ (\bar{\partial}_\tau u_h^{n+1}, v_h) + a_h(u_h^{n+1}, v_h) + J_{\alpha, \lambda}(\tilde{u}_h^n, u_h^{n+1}; v_h) = (f, v_h), \quad \forall v_h \in V_h, \end{cases} \quad (30)$$

with  $u_h^0 = \pi_h u_0$  and  $u_h^1$  an approximation of  $u(\tau)$  that will be discussed in the next section.

For  $\alpha = \lambda = 1$  (CIP stencil), (30) reduces to the standard BDF2 time-discretization of (6). For  $\lambda = 0$ , one readily verifies that the left hand side of (30) results in an algebraic system with the same structure as the standard Galerkin method.

The next three paragraphs address the initialization, stability and convergence analysis of the fully discrete formulation (30). We will see that, if  $\alpha = \lambda = 1$  the formulation is unconditionally stable and optimally convergent, whereas for  $\lambda = 0$  a CFL-like condition (independent of the viscosity coefficient  $\mu$ ) is required due to the explicit treatment of the non-Galerkin part of the stabilization. If one wishes to take large time steps keeping a reduced stencil, (30) can be combined with the iterative method proposed in section §4.

### 5.2.1 Initialization

Since BDF2 is a multistep method an approximation  $u_h^1 \approx u(\tau)$  is needed to start the time marching. This can be obtained either by a step of the Crank-Nicholson scheme in which case we have the bound from Theorem 5.4, with  $\theta = \frac{1}{2}$ ,  $T = \tau$  and  $\lambda = \alpha = 1$ ,

$$\|u_h^1 - u(\tau)\|_{0, \Omega}^2 \lesssim \mathcal{H}^2(\tau, \mu, \beta, \sigma) \|u\|_{H^1(0, \tau; H^{k+1}(\Omega))}^2 + \tau^5 \|\partial_{ttt} u\|_{L^2(0, T; L^2(\Omega))}^2,$$

or by a step of the first order backward differentiation formula (BDF1) in which case the estimate of Theorem 5.4 with  $\theta = 1$  and  $T = \tau$  yields

$$\begin{aligned} \|u_h^1 - u(\tau)\|_{0, \Omega}^2 &\lesssim \mathcal{H}^2(\tau, \mu, \beta, \sigma) \|u\|_{H^1(0, \tau; H^{k+1}(\Omega))}^2 + \tau^3 \left[ \tau \|\partial_{tt} u\|_{H^1(0, \tau; L^2(\Omega))}^2 \right. \\ &\quad \left. + |\beta|_\infty^2 ((\alpha - 1)^2 + (1 - \lambda)^2) \|\partial_t u\|_{L^2(0, \tau; H^1(\Omega))}^2 \right]. \end{aligned}$$

Noting that  $\|\partial_t u\|_{L^2(0, \tau; H^1(\Omega))}^2 \lesssim \tau \|\partial_t u\|_{L^\infty(0, \tau; H^1(\Omega))}^2$  we conclude that for sufficiently regular solutions

$$\|u_h^1 - u(\tau)\|_{0, \Omega}^2 \lesssim \mathcal{H}^2(\tau, \mu, \beta, \sigma) + \tau^4.$$

This is exactly the convergence order needed to use any one of the various  $\theta$ -scheme methods analysed above for the initialization of the BDF2 scheme, without losing order.

### 5.2.2 Stability

The next theorem states the stability properties of the fully discrete formulation (30).

**Theorem 5.7 (Stability BDF2)** *Assume that one of the following two conditions holds:*

- *CIP stencil:*

$$\lambda = 1, \quad \alpha = 1. \quad (31)$$

- *standard Galerkin stencil:*

$$\lambda = 0, \quad \alpha = 2, \quad 4C_T\gamma\tau|\beta|_\infty \leq h. \quad (32)$$

Then for  $\{e_h^n\}_{n=0}^N \in [V_h]^{N+1}$ , there holds

$$\begin{aligned} & \sum_{m=1}^{n-1} \tau \left[ (\bar{\partial}_\tau e_h^{m+1}, e_h^{m+1}) + a_h(e_h^{m+1}, e_h^{m+1}) + J_{\alpha,\lambda}(e_h^m, e_h^{m+1}; e_h^{m+1}) \right] \\ & \geq \frac{1}{4} (\|e_h^n\|_{0,\Omega}^2 + \|\tilde{e}^n\|_{0,\Omega}^2) + \sum_{m=1}^{N-1} \tau \|e_h^{m+1}\|^2 - \frac{1}{4} (\|e_h^1\|_{0,\Omega}^2 + \|\tilde{e}_h^1\|_{0,\Omega}^2), \end{aligned} \quad (33)$$

for all  $2 \leq n \leq N$ .

*Proof.* The proof is similar to the proof of Theorem 5.1. For  $2 \leq n \leq N-1$ , we define the following quantities

$$\begin{aligned} I_1 & \stackrel{\text{def}}{=} \sum_{m=1}^{n-1} [\tau(\bar{\partial}_\tau e_h^{m+1}, e_h^{m+1}) + \tau a_h(e_h^{m+1}, e_h^{m+1})], \\ I_2 & \stackrel{\text{def}}{=} \sum_{m=1}^{n-1} \tau J_{\alpha,\lambda}(\tilde{e}_h^m, e_h^{m+1}; e_h^{m+1}). \end{aligned}$$

First we note that

$$\tau(\bar{\partial}_\tau e_h^{m+1}, e_h^{m+1}) = \frac{1}{4} \left[ \|e_h^{m+1}\|_{0,\Omega}^2 + \|\tilde{e}_h^{m+1}\|_{0,\Omega}^2 - (\|e_h^m\|_{0,\Omega}^2 + \|\tilde{e}_h^m\|_{0,\Omega}^2) + \|e_h^{m+1} - \tilde{e}_h^m\|_{0,\Omega}^2 \right],$$

and, therefore,

$$\begin{aligned} I_1 & \geq \frac{1}{4} (\|e_h^n\|_{0,\Omega}^2 + \|\tilde{e}_h^n\|_{0,\Omega}^2) - \frac{1}{4} (\|e_h^1\|_{0,\Omega}^2 + \|\tilde{e}_h^1\|_{0,\Omega}^2) \\ & \quad + \sum_{m=1}^{n-1} \left[ \frac{1}{4} \|e_h^{m+1} - \tilde{e}_h^m\|^2 + \tau a_h(e_h^{m+1}, e_h^{m+1}) \right]. \end{aligned} \quad (34)$$

As in (23), we have

$$\begin{aligned} J_{\alpha,\lambda}(\tilde{e}_h^m, e_h^{m+1}; e_h^{m+1}) & = j(e_h^{m+1}, e_h^{m+1}) + (\alpha - 1)j_{sG}(e_h^{m+1} - \tilde{e}_h^m, e_h^{m+1}) \\ & \quad + (1 - \lambda)j_X(e_h^{m+1} - \tilde{e}_h^m, e_h^{m+1}). \end{aligned}$$

Therefore, applying Lemma 3.2 (noting that  $\lambda \in \{0, 1\}$ ), we have

$$\begin{aligned} I_2 &\geq \frac{\tau}{2}(1 + \lambda) \sum_{m=1}^{n-1} j(e_h^{m+1}, e_h^{m+1}) \\ &\quad + \tau(\alpha + \lambda - 2) \sum_{m=1}^{n-1} j_{sG}(e_h^{m+1} - \tilde{e}_h^{m+1}, e_h^{m+1}) \\ &\quad - \frac{\tau}{2}(1 - \lambda) \sum_{m=1}^{n-1} j_{sG}(e_h^{m+1} - \tilde{e}_h^m, e_h^{m+1} - \tilde{e}_h^m). \end{aligned} \quad (35)$$

Under condition (31) or (32) it follows that

$$I_2 \geq \frac{\tau}{2} \sum_{m=1}^{n-1} j(e_h^{m+1}, e_h^{m+1}),$$

which, in combination with (34), yields (33).

Finally, under condition (32), we use Lemma 3.3 and (24) to obtain

$$\begin{aligned} I_2 &\geq \frac{\tau}{2} \sum_{m=1}^{n-1} j(e_h^{n+1}, e_h^{n+1}) - \frac{\tau}{2} C_T \gamma |\beta|_\infty h^{-1} \sum_{m=1}^{n-1} \|e_h^{m+1} - \tilde{e}_h^m\|_{0,\Omega}^2 \\ &\geq \frac{\tau}{2} \sum_{m=0}^{n-1} j(e_h^{m+1}, e_h^{m+1}) - \frac{1}{8} \sum_{m=1}^{n-1} \|e_h^{m+1} - \tilde{e}_h^m\|_{0,\Omega}^2. \end{aligned}$$

By combining this estimate with (34) we obtain (33), which completes the proof.  $\square$

**Remark 5.8** Under condition (31) (extended stencil), the previous theorem provides the expected unconditional stability of the BDF2 scheme. For  $\lambda = 0$  a CFL-like condition, independent of the viscosity coefficient  $\mu$ , is demanded to ensure stability. Note also that, for  $\tilde{u}_h^n = 2u_h^n - u_h^{n-1}$ , the stability condition (32) requires a fixed value of  $\alpha = 2$ , since the term  $\sum_{m=1}^{n-1} j_{sG}(e_h^{m+1} - \tilde{e}_h^{m+1}, e_h^{m+1})$  in (35) becomes a telescopic sum only for the case  $\tilde{u}_h^n = u_h^n$ , i.e. first order extrapolation (as for the  $\theta$ -scheme (15)).

### 5.2.3 Convergence

The next theorem states an optimal *a priori* error estimate for the fully discrete formulation (30).

**Theorem 5.9 (Convergence BDF2)** *Let the hypothesis of Lemma 5.7 be satisfied and let  $e_h^n = \pi_h u(t^n) - u_h^n$ . Then, the solution of the numerical method (30) satisfies the following error estimate*

$$\begin{aligned} \|e^n\|_{0,\Omega}^2 + \sum_{m=1}^{n-1} \tau \|e_h^{m+1}\|^2 &\lesssim T \mathcal{H}^2(T, \mu, \beta, \sigma) \|\partial_t u\|_{H^1(0,T;H^{k+1}(\Omega))}^2 \\ &\quad + T \tau^4 \left( \|\partial_{ttt} u\|_{L^2(0,T;L^2(\Omega))}^2 + |\beta|_\infty^2 \|\partial_{tt} u\|_{L^2(0,T;H^1(\Omega))}^2 \right) + \|e^1\|_{0,\Omega}^2, \end{aligned} \quad (36)$$

for all  $2 \leq n \leq N$ . We recall that  $\mathcal{H}(T, \mu, \beta, \sigma) \stackrel{\text{def}}{=} h^k (|\beta|_{\infty}^{\frac{1}{2}} h^{\frac{1}{2}} + T^{\frac{1}{2}} |\beta|_{1, \infty, \Omega} h + |\sigma|_{\infty}^{\frac{1}{2}} h + |\mu|_{\infty}^{\frac{1}{2}})$  and  $\|e^1\|_{0, \Omega}^2$  is the error induced by the initialization step (see discussion in section §5.2.1).

*Proof.* The proof is similar to that of Theorem 5.4. Hence, applying Lemma 5.7 and using the notation  $\tilde{u}^n \stackrel{\text{def}}{=} 2u(t^n) - u(t^{n-1})$ , we arrive at the following error representation

$$\begin{aligned}
& \frac{1}{4} \|e_h^n\|_{0, \Omega}^2 - \frac{1}{4} (\|e_h^1\|_{0, \Omega}^2 + \|\tilde{e}_h^1\|_{0, \Omega}^2) + \sum_{m=1}^{n-1} \tau \|e_h^{m+1}\|^2 \\
& \lesssim \sum_{m=1}^{n-1} \tau (\bar{\partial}_\tau u(t^{m+1}) - \partial_t u(t^{m+1}), e_h^{m+1}) + \sum_{m=1}^{n-1} \tau a_h((\pi_h u - u)(t^{m+1}), e_h^{m+1}) \\
& \quad + \sum_{m=1}^{n-1} \tau J_{\alpha, \lambda}(\pi_h \tilde{u}^m, \pi_h u(t^{m+1}); e_h^{m+\theta}) \\
& = \sum_{m=1}^{n-1} \left[ \underbrace{\tau (\bar{\partial}_\tau u(t^{m+1}) - \partial_t u(t^{m+1}), e_h^{m+1})}_{I_1} \right. \\
& \quad + \underbrace{\tau a_h((\pi_h u - u)(t^{m+1}), e_h^{m+1}) + j(\pi_h u(t^{m+1}), e_h^{m+1})}_{I_2} \\
& \quad + \underbrace{\tau (\alpha - 1) j_{sG}(\pi_h u(t^{m+1}) - \pi_h \tilde{u}^{m+1}, e_h^{m+1})}_{I_3} \\
& \quad \left. + \underbrace{\tau j_X(\pi_h u(t^{m+1}) - \pi_h \tilde{u}^{m+1}, e_h^{m+1})}_{I_4} \right].
\end{aligned} \tag{37}$$

The term  $I_2$  is treated as in the proof of Theorem 5.4. Thus, we only need to consider the terms  $I_1$ ,  $I_3$  and  $I_4$ .

Following [14, Page 17], using a Taylor expansion we have

$$\tau \|\bar{\partial}_\tau u(t^{m+1}) - \partial_t u(t^{m+1})\|_{0, \Omega} \lesssim \tau^2 \int_{t^{m-1}}^{t^{m+1}} \|\partial_{ttt} u(s)\|_{0, \Omega} ds.$$

As a result, it follows that

$$\begin{aligned}
I_1 & \leq \tau \|\bar{\partial}_\tau u(t^{m+1}) - \partial_t u(t^{m+1})\|_{0, \Omega} \|e_h^{m+1}\|_{0, \Omega} \\
& \lesssim \tau^2 \int_{t^{m-1}}^{t^{m+1}} \|\partial_{ttt} u(s)\|_{0, \Omega} ds \|e_h^{m+1}\|_{0, \Omega} \\
& \lesssim \tau \left( \int_{t^{m-1}}^{t^{m+1}} \tau^3 \|\partial_{ttt} u(s)\|_{0, \Omega}^2 ds \right)^{\frac{1}{2}} \|e_h^{m+1}\|_{0, \Omega} \\
& \lesssim T \tau^4 \int_{t^{m-1}}^{t^{m+1}} \|\partial_{ttt} u\|_{0, \Omega}^2 dt + \frac{\tau}{T} \|e_h^{m+1}\|_{0, \Omega}^2
\end{aligned}$$

In a similar fashion we may estimate the extrapolation error in  $I_3$  and  $I_4$  for brevity we here only consider the former term. By a Taylor development we

have

$$u(t^{m+1}) - \tilde{u}^{m+1} = u(t^{m+1}) - 2u(t^m) + u(t^{m-1}) = \int_{t^{m-1}}^{t^{m+1}} (t^{m+1} - s) \partial_{tt} u(s) \, ds,$$

and in a similar fashion as in (28) we have

$$\begin{aligned} I_3 &= \tau j_{sG}(\pi_h u(t^{m+1}) - \pi_h \tilde{u}^{m+1}, e_h^{m+1}) \\ &= \tau \int_{t^{m-1}}^{t^{m+1}} (t^{m+1} - s) j_{sG}(\partial_{tt} u(s) \, ds, e_h^{n+1}) \\ &= \tau \int_{t^{m-1}}^{t^{m+1}} (t^{m+1} - s) j_{sG}(\partial_{tt} \pi_h u(s), e_h^{n+1}) \, ds \\ &\leq C_T \gamma \tau \left( \int_{t^{m-1}}^{t^{m+1}} |\beta|_\infty \tau \|\nabla \pi_h \partial_{tt} u(s)\|_{0,\Omega} \, ds \right) \|e_h^{m+1}\|_{0,\Omega} \\ &\lesssim \tau \left( \int_{t^{m-1}}^{t^{m+1}} |\beta|_\infty^2 \tau^3 \|\nabla \pi_h \partial_{tt} u(s)\|_{0,\Omega}^2 \, ds \right)^{\frac{1}{2}} \|e_h^{m+1}\|_{0,\Omega} \\ &\lesssim \tau^4 T |\beta|_\infty^2 \int_{t^{m-1}}^{t^{m+1}} \|\partial_{tt} u(s)\|_{1,\Omega}^2 \, ds + \frac{\tau}{T} \|e_h^{m+1}\|_{0,\Omega}^2. \end{aligned}$$

We conclude in the same way as for Theorem 5.4.  $\square$

## 6 Numerical experiments

In this section we illustrate with numerical computations some of the theoretical results obtained in the previous sections. All computations have been performed using FreeFem++ [10].

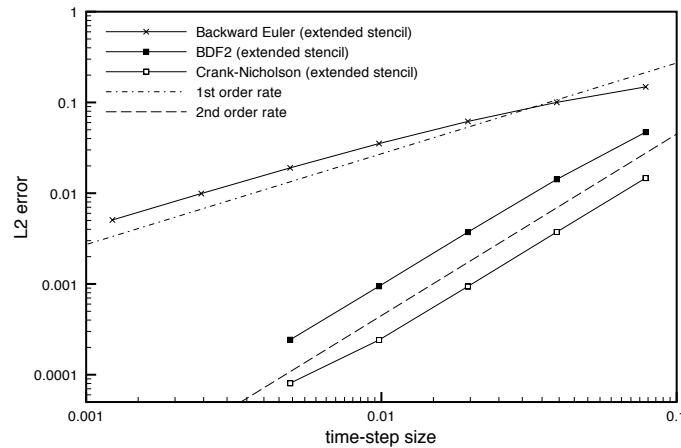


Figure 1: Convergence history (in time) of the Backward-Euler ( $\theta = 1$ ), Crank-Nicholson ( $\theta = 1/2$ ) and BDF2 schemes with the CIP matrix stencil ( $\alpha = \lambda = 1$ ).

We consider as example a pure transport problem in two dimensions: a rotating Gaussian benchmark. Hence, in (2), we take

$$\beta = (-y, x)^T, \quad \sigma = \mu = f = 0, \quad \Omega = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\},$$

with the Gaussian initial condition centered at  $(0.3, 0.3)$  given by

$$u_0(x, y) = e^{-10[(x-0.3)^2 + (y-0.3)^2]}.$$

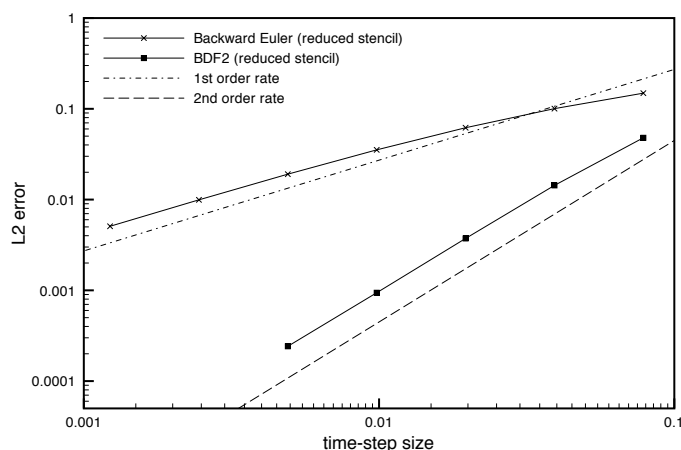


Figure 2: Convergence history (in time) of the Backward-Euler ( $\theta = 1, \alpha = 3$ ) and BDF2 ( $\alpha = 2$ ) schemes with the standard Galerkin matrix stencil ( $\lambda = 0$ ).

In order to illustrate the convergence rate in time of the discrete solutions, we have used quadratic approximations in space on a fixed mesh with 3510 triangles. The stabilization parameter was chosen as  $\gamma = 0.01$ . In figure 1 we report the convergence of the  $L^2$ -errors at time  $T = 2\pi$  (one rotation) of the discrete solutions obtained with the Backward-Euler, Crank-Nicholson and BDF2 schemes for  $\alpha = \lambda = 1$  (extended stencil). In all cases the three numerical solutions converge at the optimal rate ( $O(\tau)$  for Backward-Euler and  $O(\tau^2)$  for Crank-Nicholson and BDF2), which is agreement with the results of Theorems 5.4 and 5.9. The BDF2 scheme was initialized using one step of Backward-Euler.

Finally, in figure 2 we report the convergence history obtained with the variants of the Backward-Euler and BDF2 schemes involving a standard Galerkin stencil, *i.e.* by taking  $\lambda = 0$ . For the Backward-Euler scheme the relaxation parameter was  $\alpha = 3$  and for BDF2 we have taken  $\alpha = 2$ . Optimal convergence with respect to the  $L^2$ -norm was again obtained, which is also in agreement with the results of Theorems 5.4 and 5.9.

To compare the properties of the methods for a problem with non-smooth data we now propose to use the initial data obtained by taking

$$u_0(x, y) = \frac{1}{2} \left[ \tanh \left( \frac{e^{-10[(x-0.3)^2 + (y-0.3)^2]} - 0.5}{0.001} \right) + 1 \right].$$

This function is smooth but has a sharp layer that has thickness of order 0.001. The mesh is uniform with 400 elements along the circumference of  $\Omega$  and hence

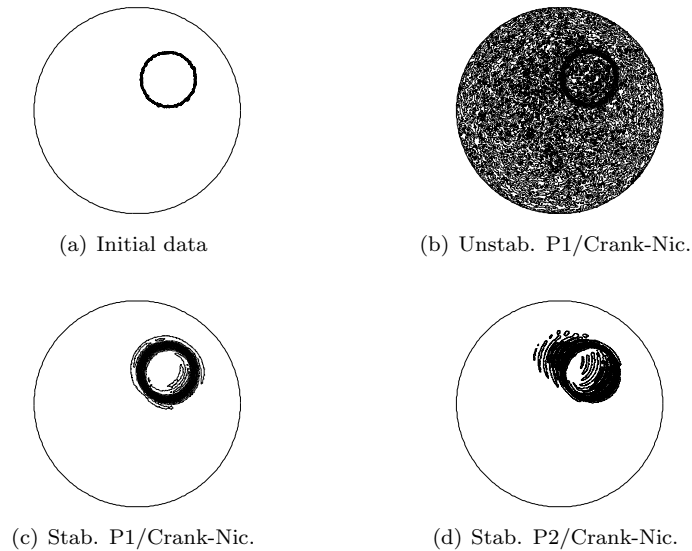


Figure 3: Contour lines for initial data and final solution using piecewise affine or quadratic finite elements and Crank-Nicholson time discretization

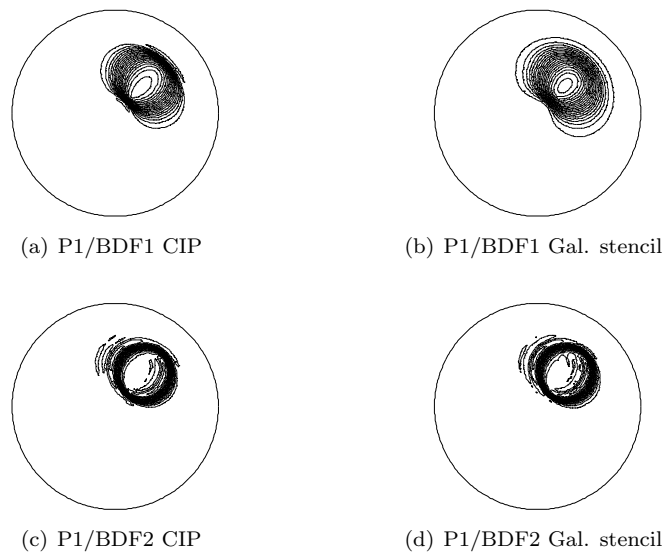


Figure 4: Contour lines for final solutions using piecewise affine finite elements and different backward differentiation formulas, with ("Gal.") or without (CIP) extrapolation time discretization

the layer is underresolved. The contourlines of the interpolant of the initial data are given in Figure 3, left plot. The initial data take the form of a cylinder of height 1 centered at  $(0.3, 0.3)$ . This cylinder has been transported one full turn using the above analyzed methods. Due to the sharp variation of the solution

that is not resolved by the computational mesh we are not in the asymptotic regime where our analysis is valid. However it is known that stabilized methods have the property to limit the propagation of perturbations induced by sharp layers (see [5, 4] for an analysis of the CIP-method) and we wish to study how the methods proposed in this paper behave with respect to this important property. To give a qualitative comparison we report the final solutions obtained using the split and unsplit backward differentiation type formulas and the unsplit Crank-Nicholson formula on discretizations using piecewise affine or piecewise quadratic elements. The time interval was decomposed in 500 timesteps for both discretizations. In

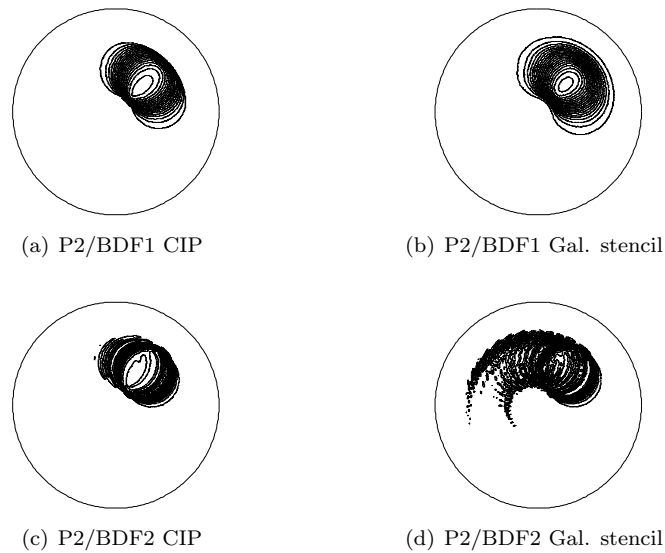


Figure 5: Contour lines for final solutions using piecewise quadratic finite elements and different backward differentiation formulas, with ("Gal.") or without (CIP) extrapolation time discretization

Figure 3 we show the initial data followed by the final solution when using the standard Galerkin method with piecewise affine approximation in space and a Crank-Nicholson discretization in time, the third and fourth plots show the final solution using Crank-Nicholson and the CIP method and piecewise affine approximation and piecewise quadratic approximation respectively. Note the global spurious oscillations that pollute the unstabilized solution. In Figures 4 and 5 we give the final solutions for the fully coupled or semi explicit versions of the first and second order backward differentiation schemes. In Figures 4 we consider piecewise affine approximation and in Figures 5 piecewise quadratic approximation. It is clear from the graphics that the higher polynomial order does not pay in this case. The fully coupled and semi-explicit method give similar results in all cases except the semi-explicit method using quadratic finite elements. In this case however the oscillations are due to a too large timestep. If we instead take 2000 timesteps the oscillations vanish as shown in Figure 6, illustrating the dependence of the constant in CFL condition (32) on the polynomial order. Another technique that is popular for the reduction of oscillations

for parabolic problems is the use of nodal quadrature for the mass matrix (see [14]). This is not covered by the above analysis. Indeed the inconsistency of the integration leads to a term with the gradient of the error that can not be controlled. The right plot of Figure 6 shows the deterioration of the solution due to the mass-lumping. In case no stabilization is used the result is even poorer than that of Figure 3b.

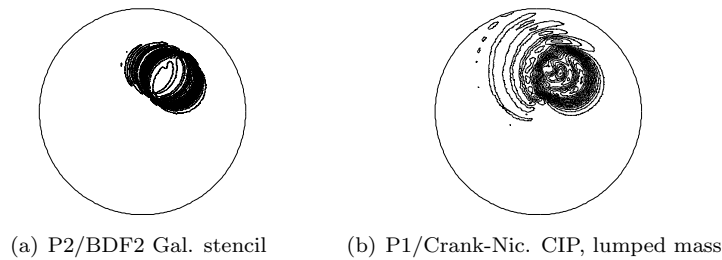


Figure 6: Left: Contour lines for using piecewise quadratic finite elements and the second order backward differencing formula using the semi-explicit treatment (with extrapolation) of the penalty term. The time interval is divided into 2000 timesteps. Stability is fully recovered. Right: Computation using Crank-Nicholson, CIP-stabilization and nodal quadrature for the discrete time-derivative. Stability is lost.

The best results both for the smooth and the non-smooth solution were obtained by using a CIP stabilized (piecewise affine in the non-smooth case) discretization in space and Crank-Nicholson in time. It seems that the instabilities that sometimes haunt the Crank-Nicholson scheme when initial data are non-smooth are efficiently controlled by the interior penalty stabilization. On the other hand the results obtained using the semi-implicit BDF2 scheme with CIP stabilization and piecewise affine approximation in space are not that much poorer. The fact that using this method only a matrix with the standard Galerkin stencil has to be inverted is expected to make this approach more efficient than the Crank-Nicholson approach with the full CIP-matrix.

## 7 Conclusion

We have analysed the  $\theta$ -timestepping method and the second order backward differentiation formula for the convection dominated convection-diffusion-reaction equation. The analysis is robust with respect to the reaction and diffusion coefficients and therefore extends to the case of a pure transport equation.

The continuous interior penalty method has an extended stencil compared to the standard Galerkin method. In this paper we prove that the linear system arising from CIP discretization can be solved using a relaxed iterative procedure. Moreover for time stepping methods that adds some dissipation to the numerical scheme we prove that optimal order can be retained while solving a linear system with the same matrix as the standard Galerkin by extrapolating the CIP-extended part from previous time-steps. To give a concise overview of our results we collect them in Table 1.

|                  | $\theta$ -scheme   | BDF2                          |
|------------------|--|-------------------------------|
| full CIP stencil | $\theta \in (\frac{1}{2}, 1], O(\tau)$<br>$\theta = \frac{1}{2}, O(\tau^2)$ uncond stable          | $O(\tau^2)$ , uncond stable   |
| Galerkin stencil | $\theta \in (\frac{1}{2}, 1], O(\tau)$ , stability: (18)-(19)<br>$\theta = \frac{1}{2}$ , unstable | $O(\tau^2)$ , stability: (32) |

Table 1: A recollection of the main stability results and convergence orders

## References

- [1] M. Boman. Estimates for the  $L_2$ -projection onto continuous finite element spaces in a weighted  $L_p$ -norm. *BIT*, 46(2):249–260, 2006.
- [2] J.H. Bramble, J.E. Pasciak, and O. Steinbach. On the stability of the  $L^2$  projection in  $H^1(\Omega)$ . *Math. Comp.*, 71(237):147–156 (electronic), 2002.
- [3] E. Burman. A unified analysis for conforming and nonconforming stabilized finite element methods using interior penalty. *SIAM J. Numer. Anal.*, 43(5):2012–2033 (electronic), 2005.
- [4] E. Burman, M. A. Fernández, and P. Hansbo. Continuous interior penalty finite element method for Oseen’s equations. *SIAM J. Numer. Anal.*, 44(3):1248–1274 (electronic), 2006.
- [5] E. Burman and P. Hansbo. Edge stabilization for Galerkin approximations of convection-diffusion-reaction problems. *Comput. Methods Appl. Mech. Engrg.*, 193(15-16):1437–1453, 2004.
- [6] E. Burman and P. Hansbo. Edge stabilization for the generalized Stokes problem: a continuous interior penalty method. *Comput. Methods Appl. Mech. Engrg.*, 195(19-22):2393–2410, 2006.
- [7] R. Codina and J. Blasco. Analysis of a stabilized finite element approximation of the transient convection-diffusion-reaction equation using orthogonal subscales. *Comput. Vis. Sci.*, 4(3):167–174, 2002.
- [8] J. Douglas Jr. and T. Dupont. *Interior Penalty Procedures for Elliptic and Parabolic Galerkin Methods*, volume 58 of *Lecture Notes in Physics*. Springer-Verlag, Berlin, 1976.
- [9] J.-L. Guermond. Subgrid stabilization of Galerkin approximations of linear contraction semi-groups of class  $C^0$  in Hilbert spaces. *Numer. Methods Partial Differential Equations*, 17(1):1–25, 2001.
- [10] F. Hecht, O. Pironneau, A. Le Hyaric, and K. Ohtsuka. *FreeFem++ v. 2.11. User’s Manual*. LJLL, University of Paris 6.
- [11] J.G. Heywood and R. Rannacher. Finite element approximation of the nonstationary Navier-Stokes problem. IV. Error analysis for second-order time discretization. *SIAM J. Numer. Anal.*, 27(2):353–384, 1990.
- [12] G. Lube and D. Weiss. Stabilized finite element methods for singularly perturbed parabolic problems. *Appl. Numer. Math.*, 17(4):431–459, 1995.

- [13] J. Nitsche. über ein Variationsprinzip zur Lösung von Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unterworfen sind. *Abh. Math. Sem. Univ. Hamburg*, 36:9–15, 1971.
- [14] V. Thomée. *Galerkin finite element methods for parabolic problems*, volume 25 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 1997.

## Contents

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Introduction</b>                              | <b>3</b>  |
| <b>2</b> | <b>Problem setting</b>                           | <b>4</b>  |
| <b>3</b> | <b>Space discretization</b>                      | <b>5</b>  |
| 3.1      | Preliminary results . . . . .                    | 6         |
| <b>4</b> | <b>An iterative scheme</b>                       | <b>7</b>  |
| <b>5</b> | <b>Time discretization</b>                       | <b>8</b>  |
| 5.1      | Timestepping with the $\theta$ -scheme . . . . . | 8         |
| 5.1.1    | Stability . . . . .                              | 9         |
| 5.1.2    | Convergence . . . . .                            | 12        |
| 5.2      | Timestepping with BDF2 . . . . .                 | 15        |
| 5.2.1    | Initialization . . . . .                         | 15        |
| 5.2.2    | Stability . . . . .                              | 16        |
| 5.2.3    | Convergence . . . . .                            | 17        |
| <b>6</b> | <b>Numerical experiments</b>                     | <b>19</b> |
| <b>7</b> | <b>Conclusion</b>                                | <b>23</b> |



---

Unité de recherche INRIA Rocquencourt  
Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes  
4, rue Jacques Monod - 91893 ORSAY Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique  
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)

Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)

Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)

Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

---

Éditeur  
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)  
<http://www.inria.fr>  
ISSN 0249-6399