Rapport de recherche du CERMICS Janvier 2006

A continuous finite element method with face penalty to approximate Friedrichs' systems

Erik Burman & Alexandre Ern

Department of Mathematics, Ecole Polytechnique Fédérale de Lausanne, Switzerland

CERMICS, Ecole nationale des ponts et chaussées, 6-8 avenue Blaise Pascal, Champs sur Marne, 77455 Marne la Vallée Cedex 2, France

CERMICS — ENPC
6 et 8 avenue Blaise Pascal
Cité Descartes - Champs sur Marne
77455 Marne la Vallée Cedex 2

http://cermics.enpc.fr

Mathematical Modelling and Numerical Analysis Will be set by the publisher
Modélisation Mathématique et Analyse Numérique

A CONTINUOUS FINITE ELEMENT METHOD WITH FACE PENALTY TO APPROXIMATE FRIEDRICHS' SYSTEMS

Erik Burman[1] and Alexandre Ern[2]

Abstract. A continuous finite element method to approximate Friedrichs' systems is proposed and analyzed. Stabilization is achieved by penalizing the jumps across mesh interfaces of the gradient of some components of the discrete solution. The convergence analysis leads to optimal convergence rates in the graph norm and suboptimal of order $\frac{1}{2}$ convergence rates in the L^2-norm. A variant of the method specialized to Friedrichs' systems associated with elliptic PDE's in mixed form is also proposed and analyzed. Finally, numerical results are presented to illustrate the theoretical analysis.

Résumé. On analyse une méthode d'éléments finis continus pour approcher les systèmes de Friedrichs. La stabilisation est obtenue en pénalisant le saut à travers les interfaces du maillage du gradient de certaines composantes de la solution discrète. L'analyse de convergence conduit à des estimations optimales en norme du graphe et à des estimations sous-optimales d'ordre $\frac{1}{2}$ dans la norme L^2. On propose également une variante de la méthode adaptée aux systèmes de Friedrichs associés aux EDP elliptiques sous forme mixte. Enfin, des résultats numériques sont présentés afin d'illustrer l'analyse théorique.

1991 Mathematics Subject Classification. 65N30, 65N12, 35F15.

Draft version: 6th January 2006.

1. Introduction

Friedrichs' systems are systems of first-order PDE's endowed with a symmetry and a positivity property. The mathematical analysis of such systems, which was initiated by Friedrichs in 1958 [13], has made considerable progress in the last decades; see, e.g., the review by Jensen [16]. Recently, the theory was revisited by Ern and Guermond [10] where the well–posedness of the Friedrichs' system was established whenever a suitable boundary operator can be defined on the graph of the differential operator. Friedrichs' systems are encountered in many applications, including advection–reaction equations, advection–diffusion–reaction equations, the linear elasticity equations, the wave equation, the linearized Euler equations, and the Maxwell equations in the so-called elliptic regime, to cite a few examples.

The finite element approximation of Friedrichs' systems was initiated by Lesaint and Raviart in 1974 [19,20] where the Discontinuous Galerkin method (DGM) was analyzed. The convergence estimate was subsequently improved by Johnson et al. [17], and more recently a thorough systematic analysis generalizing [17,19,20] was

Keywords and phrases: finite elements – interior penalty – stabilization methods – Friedrichs' systems – first-order PDE's

[1] Department of Mathematics, Ecole Polytechnique Fédérale de Lausanne, Switzerland; e-mail: `erik.burman@epfl.ch`

[2] CERMICS, Ecole nationale des ponts et chaussées, Champs sur Marne, 77455 Marne la Vallée Cedex 2, France; e-mail: `ern@cermics.enpc.fr`

proposed by Ern and Guermond [10,11]. From a practical viewpoint, the DGM offers various advantages, including the flexibility in using non-matching grids, handling heterogeneous media, and performing hp-refinement. However, a drawback is that keeping the mesh fixed, the method involves a much larger number of degrees of freedom than the continuous finite element method (CFEM). There is therefore a clear motivation to design and analyze suitable approximation schemes for Friedrichs' systems based on continuous finite elements.

To approximate satisfactorily the simplest example of Friedrichs' systems, namely an advection–reaction equation, using continuous finite elements, it is well-known that a stabilization technique must be used. Drawing on earlier ideas by Babuška [1], Babuška and Zlámal [2], and Douglas and Dupont [7], the analysis of face penalty finite element methods has been recently extended to advection–diffusion equations [3,4]. The principle of the method consists of stabilizing the continuous finite element approximation by penalizing the jumps of the advective derivative of the discrete solution across mesh interfaces. The degrees of freedom in the resulting stabilized continuous finite element method (SCFEM) are those of the CFEM on the same mesh, which represents a substantial saving with respect to a DGM. However, the penalty term acting on the gradient jumps extends the discretization stencil, since a mesh node ν is coupled to the nodes located in the set $\mathcal{T}_\nu$ of the elements to which ν belongs, but also to the nodes located in the neighboring elements sharing a face with the elements in $\mathcal{T}_\nu$. In two space dimensions, when working with first-order finite elements, the number of nonzero entries in the stiffness matrix scales as 7, 13, and 72 times the number of mesh vertices for CFEM, SCFEM, and DGM, respectively, and when working with second-order finite elements, this number scales as 63, 100, and 288 times the number of mesh vertices for CFEM, SCFEM, and DGM, respectively.

The goal of this work is to generalize the above face penalty technique to approximate satisfactorily Friedrichs' systems using continuous finite elements. In §2 the main results on Friedrichs' systems derived in [10, 11] are briefly restated and four examples of Friedrichs' systems are discussed. In §3 the SCFEM with face penalty is designed and analyzed. In §4 the setting is specialized to a certain class of Friedrichs' systems associated with elliptic-like PDE's written in mixed form. Approximating the mixed form of the PDE presents some advantages: it provides a more accurate reconstruction of the fluxes (the gradient of the primal variable for diffusion-like problems and the stress tensor for linear elasticity problems), it reduces the condition number of the stiffness matrix from a multiple of h^{-2} to a multiple of h^{-1} (see, e.g., [12]), and it is often the only viable formulation whenever complex constitutive laws such as those of viscoelastic fluids are considered. Finally, in §5 numerical results are presented to illustrate the convergence estimates and the fact that oscillations produced by CFEM without stabilization can effectively be controlled by the present face penalty technique.

2. Friedrichs' systems

2.1. The setting

Let Ω be a bounded, open, and connected Lipschitz domain in $\mathbb{R}^d$ and let m be a positive integer. Let $\mathcal{K}$ and $\{\mathcal{A}^k\}_{1\le k\le d}$ be $(d+1)$ functions on Ω with values in $\mathbb{R}^{m,m}$. Assume that these fields satisfy

$$\mathcal{K} \in [L^\infty(\Omega)]^{m,m}, \tag{A1}$$

$$\mathcal{A}^k \in [L^\infty(\Omega)]^{m,m} \quad \text{and} \quad \sum_{k=1}^{d} \partial_k \mathcal{A}^k \in [L^\infty(\Omega)]^{m,m}, \tag{A2}$$

$$\mathcal{A}^k = (\mathcal{A}^k)^t \ \text{ a.e. in } \Omega, \tag{A3}$$

$$\exists \mu_0 > 0, \quad \mathcal{K} + \mathcal{K}^t - \sum_{k=1}^{d} \partial_k \mathcal{A}^k \ge 2\mu_0 \mathcal{I}_m \ \text{ a.e. on } \Omega, \tag{A4}$$

where $\mathcal{I}_m$ is the identity matrix in $\mathbb{R}^{m,m}$. Set $L = [L^2(\Omega)]^m$ and let $\mathfrak{D}(\Omega)$ the space of $\mathfrak{C}^\infty$ functions that are compactly supported in Ω. Let $w \in L$. If the linear form $[\mathfrak{D}(\Omega)]^m \ni \varphi \longmapsto -\int_\Omega \sum_{k=1}^{d} w^t \partial_k(\mathcal{A}^k\varphi) \in \mathbb{R}$, is bounded on L, the function w is said to have an A-weak derivative in L, and the function in L that can be associated with the above linear form by means of the Riesz representation theorem is denoted by Aw. Clearly, if $w \in [\mathfrak{C}^1(\Omega)]^m$, $Aw = \sum_{k=1}^d \mathcal{A}^k \partial_k w$. Define the graph space $W = \{w \in L;\ Aw \in L\}$. Equipped with the graph norm $\|w\|_W^2 = \|Aw\|_L^2 + \|w\|_L^2$, W is a Hilbert space. Define the operators $T \in \mathcal{L}(W;L)$ and $\tilde{T} \in \mathcal{L}(W;L)$ as

$$Tw = \mathcal{K}w + \sum_{k=1}^{d} \mathcal{A}^k \partial_k w, \qquad \tilde{T}w = \mathcal{K}^t w - \sum_{k=1}^{d} \partial_k(\mathcal{A}^k w). \tag{1}$$

Let $D \in \mathcal{L}(W;W')$ be the operator such that for all $(v,w) \in W \times W$,

$$\langle Dv, w\rangle_{W',W} = (Tv,w)_L - (v,\tilde{T}w)_L. \tag{2}$$

One readily verifies that the operator D is self-adjoint; D is also a boundary operator since $[\mathfrak{D}(\Omega)]^m \subset \mathrm{Ker}(D)$.

Consider the following problem: For f in L, seek $z \in W$ such that $Tz = f$. In general, boundary conditions must be enforced for this problem to be well-posed. In other words, one must find a closed subspace V of W such that the restricted operator $T: V \rightarrow L$ is an isomorphism. To specify the space V, the key assumption consists of assuming that there exists an operator $M \in \mathcal{L}(W;W')$ such that

$$\langle Mw, w\rangle_{W',W} \geq 0 \text{ for all } w \text{ in } W, \tag{M1}$$

$$W = \mathrm{Ker}(D-M) + \mathrm{Ker}(D+M). \tag{M2}$$

Assumptions (M1)–(M2) have been introduced in [10,11] drawing on earlier ideas by Friedrichs [13]. In particular, they imply that $\mathrm{Ker}(D) = \mathrm{Ker}(M)$ so that M is also a boundary operator. Define the following bilinear form: For all $(v,w) \in W \times W$,

$$a(v,w) = (Tv,w)_L + \tfrac{1}{2}\langle (M-D)v, w\rangle_{W',W}. \tag{3}$$

In this framework, the main result proven in [10] is the following

Theorem 2.1. *Assume* (A1)–(A4) *and* (M1)–(M2). *Then, for all $f \in L$, the following problem is well-posed:*

$$\text{Seek } z \in W \text{ such that } a(z,y) = (f,y)_L,\ \forall y \in W, \tag{4}$$

and the unique solution to (4) *is such that $z \in V := \mathrm{Ker}(D-M)$ and $Tz = f$ in L.*

On $\partial\Omega$, define the $\mathbb{R}^{m,m}$-valued field $\mathcal{D} = \sum_{k=1}^d n_k \mathcal{A}^k$ where $n = (n_1, \dots, n_d)^t$ is the unit outward normal vector to $\partial\Omega$. Then, it is clear that for v, w smooth enough,

$$\langle Dv, w\rangle_{W',W} = \int_{\partial\Omega} w^t \mathcal{D} v. \tag{5}$$

Henceforth, we assume that the boundary operator M can be associated with a matrix-valued field $\mathcal{M}: \partial\Omega \longrightarrow \mathbb{R}^{m,m}$ such that for v, w smooth enough,

$$\langle Mv, w\rangle_{W',W} = \int_{\partial\Omega} w^t \mathcal{M} v. \tag{6}$$

This assumption holds true for the various examples presented in the following section.

Remark 2.1. In some situations, assumption (A4) can be relaxed. For instance, this is the case for Friedrichs' systems endowed with a 2×2 block structure such that a Poincaré-like inequality holds for some components of the dependent variable; see Remarks 2.2 and 2.3 for more details.

2.2. **Examples**

This section briefly presents four examples of Friedrichs' systems: the advection–reaction equation, the advection–diffusion–reaction equation, the linear elasticity equations in the mixed stress–pressure–displacement form, and the Maxwell equations in the so-called elliptic regime.

2.2.1. *Advection–reaction*

Let $\mu \in L^\infty(\Omega)$, let $\beta \in [L^\infty(\Omega)]^d$ with $\nabla\cdot\beta \in L^\infty(\Omega)$, and assume that $\mu(x) - \frac{1}{2}\nabla\cdot\beta(x) \geq \mu_0 > 0$ a.e. in Ω. Let $f \in L^2(\Omega)$. The PDE

$$\mu u + \beta\cdot\nabla u = f \tag{7}$$

falls into the category of Friedrichs' systems by setting $m = 1$, $\mathcal{K} = \mu$ and $\mathcal{A}^k = \beta^k$ for $k \in \{1, \dots, d\}$. The graph space is $W = \{w \in L^2(\Omega);\ \beta\cdot\nabla w \in L^2(\Omega)\}$.

To enforce boundary conditions, define $\partial\Omega^\pm = \{x \in \partial\Omega;\ \pm\beta(x)\cdot n(x) > 0\}$. Assume that $\mathfrak{C}^1(\overline{\Omega})$ is dense in W and that $\partial\Omega^-$ and $\partial\Omega^+$ are well-separated, i.e., $\mathrm{dist}(\partial\Omega^-, \partial\Omega^+) > 0$. Then, the boundary operator D admits the following representation [10]: For all $v, w \in W$,

$$\langle Dv, w\rangle_{W',W} = \int_{\partial\Omega} vw(\beta\cdot n). \tag{8}$$

Let $M \in \mathcal{L}(W; W')$ be defined such that for $v, w \in W$,

$$\langle Mv, w\rangle_{W',W} = \int_{\partial\Omega} vw|\beta\cdot n|. \tag{9}$$

Then, (M1)–(M2) hold and $V = \{v \in W;\ v|_{\partial\Omega^-} = 0\}$, i.e., homogeneous Dirichlet boundary conditions are enforced at the inflow boundary. Observe that (5)–(6) hold with

$$\mathcal{D} = \beta\cdot n, \qquad \mathcal{M} = |\beta\cdot n|. \tag{10}$$

2.2.2. *Advection–diffusion–reaction*

Let μ, β, and f be as above. The PDE $-\Delta u + \beta\cdot\nabla u + \mu u = f$ written in the following mixed form

$$\begin{cases} \sigma + \nabla u = 0, \\ \mu u + \nabla\cdot\sigma + \beta\cdot\nabla u = f, \end{cases} \tag{11}$$

falls into the category of Friedrichs' systems by setting $m = d + 1$ and

$$\mathcal{K} = \left[\begin{array}{c|c} \mathcal{I}_d & 0 \\ \hline 0 & \mu \end{array}\right], \qquad \mathcal{A}^k = \left[\begin{array}{c|c} 0 & e^k \\ \hline (e^k)^t & \beta^k \end{array}\right], \tag{12}$$

where $\mathcal{I}_d$ is the identity matrix in $\mathbb{R}^{d,d}$ and e^k is the k-th vector in the canonical basis of $\mathbb{R}^d$. The graph space is $W = H(\mathrm{div};\Omega)\times H^1(\Omega)$.

The boundary operator D is such that for all $(\sigma, u), (\tau, v) \in W$,

$$\langle D(\sigma, u), (\tau, v)\rangle_{W',W} = \langle \sigma\cdot n, v\rangle_{-\frac{1}{2},\frac{1}{2}} + \langle \tau\cdot n, u\rangle_{-\frac{1}{2},\frac{1}{2}} + \int_{\partial\Omega} (\beta\cdot n)uv, \tag{13}$$

where $\langle , \rangle_{-\frac{1}{2},\frac{1}{2}}$ denotes the duality pairing between $H^{-\frac{1}{2}}(\partial\Omega)$ and $H^{\frac{1}{2}}(\partial\Omega)$. Let $M \in \mathcal{L}(W; W')$ be defined such that for all $(\sigma, u), (\tau, v) \in W$,

$$\langle M(\sigma, u), (\tau, v)\rangle_{W',W} = \langle \sigma\cdot n, v\rangle_{-\frac{1}{2},\frac{1}{2}} - \langle \tau\cdot n, u\rangle_{-\frac{1}{2},\frac{1}{2}}. \tag{14}$$

Then, (M1)–(M2) hold and $V = H(\mathrm{div};\Omega)\times H^1_0(\Omega)$, i.e., homogeneous Dirichlet boundary conditions are enforced. Neumann and Robin boundary conditions can be treated as well; see [10]. Observe that (5)–(6) hold with

$$\mathcal{D} = \left[\begin{array}{c:c} 0 & n \\ \hdashline n^t & \beta{\cdot}n \end{array}\right], \qquad \mathcal{M} = \left[\begin{array}{c:c} 0 & -n \\ \hdashline n^t & 0 \end{array}\right]. \tag{15}$$

Remark 2.2. Using a Poincaré inequality, one can show that the well–posedness of (4) still holds if $\mu(x) - \frac{1}{2}\nabla{\cdot}\beta(x) \geq 0$ a.e. in Ω.

2.2.3. *Linear elasticity*

Let γ_1 and γ_2 be two positive functions in $L^\infty(\Omega)$ uniformly bounded away from zero. Let $f \in [L^2(\Omega)]^d$. Let u be the $\mathbb{R}^d$-valued displacement field and let σ be the $\mathbb{R}^{d,d}$-valued stress tensor. The PDE's $\sigma = \frac{1}{2}(\nabla u + (\nabla u)^t) + \frac{1}{\gamma_1}(\nabla{\cdot}u)\mathcal{I}_d$ and $-\nabla{\cdot}\sigma + \gamma_2 u = f$ can be written in the following mixed stress–pressure–displacement form

$$\begin{cases} \sigma + p\mathcal{I}_d - \frac{1}{2}(\nabla u + (\nabla u)^t) = 0, \\ \mathrm{tr}(\sigma) + (d+\gamma_1)p = 0, \\ -\frac{1}{2}\nabla{\cdot}(\sigma + \sigma^t) + \gamma_2 u = f. \end{cases} \tag{16}$$

The tensor σ in $\mathbb{R}^{d,d}$ can be identified with the vector $\overline{\sigma} \in \mathbb{R}^{d^2}$ by setting $\overline{\sigma}_{[ij]} = \sigma_{ij}$ with $1 \leq i,j \leq d$ and $[ij] = d(j-1)+i$. Then, (16) falls into the category of Friedrichs' systems by setting $m = d^2 + 1 + d$ and

$$\mathcal{K} = \left[\begin{array}{c:c:c} \mathcal{I}_{d^2} & \mathcal{Z} & 0 \\ \hdashline (\mathcal{Z})^t & (d{+}\gamma_1) & 0 \\ \hdashline 0 & 0 & \gamma_2\mathcal{I}_d \end{array}\right], \qquad \mathcal{A}^k = \left[\begin{array}{c:c:c} 0 & 0 & \mathcal{E}^k \\ \hdashline 0 & 0 & 0 \\ \hdashline (\mathcal{E}^k)^t & 0 & 0 \end{array}\right], \tag{17}$$

where $\mathcal{Z} \in \mathbb{R}^{d^2}$ has components given by $\mathcal{Z}_{[ij]} = \delta_{ij}$, and for all $k \in \{1,\dots,d\}$, $\mathcal{E}^k \in \mathbb{R}^{d^2,d}$ has components given by $\mathcal{E}^k_{[ij],l} = -\frac{1}{2}(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})$; here, $i,j,l \in \{1,\dots,d\}$ and the δ's denote Kronecker symbols. The graph space is $W = H_{\overline{\sigma}}\times L^2(\Omega)\times[H^1(\Omega)]^d$ with $H_{\overline{\sigma}} = \{\overline{\sigma} \in [L^2(\Omega)]^{d^2};\ \nabla{\cdot}(\sigma+\sigma^t) \in [L^2(\Omega)]^d\}$.

The boundary operator D is such that for all $(\overline{\sigma},p,u), (\overline{\tau},q,v) \in W$,

$$\langle D(\overline{\sigma},p,u),(\overline{\tau},q,v)\rangle_{W',W} = -\langle \tfrac{1}{2}(\tau+\tau^t){\cdot}n, u\rangle_{-\frac{1}{2},\frac{1}{2}} - \langle \tfrac{1}{2}(\sigma+\sigma^t){\cdot}n, v\rangle_{-\frac{1}{2},\frac{1}{2}}, \tag{18}$$

where $\langle,\rangle_{-\frac{1}{2},\frac{1}{2}}$ denotes the duality pairing between $[H^{-\frac{1}{2}}(\partial\Omega)]^d$ and $[H^{\frac{1}{2}}(\partial\Omega)]^d$. Let $M \in \mathcal{L}(W;W')$ be defined such that for all $(\overline{\sigma},p,u), (\overline{\tau},q,v) \in W$,

$$\langle M(\overline{\sigma},p,u),(\overline{\tau},q,v)\rangle_{W',W} = \langle \tfrac{1}{2}(\tau+\tau^t){\cdot}n, u\rangle_{-\frac{1}{2},\frac{1}{2}} - \langle \tfrac{1}{2}(\sigma+\sigma^t){\cdot}n, v\rangle_{-\frac{1}{2},\frac{1}{2}}. \tag{19}$$

Then, (M1)–(M2) hold and $V = H_{\overline{\sigma}}\times L^2(\Omega)\times[H^1_0(\Omega)]^d$, i.e., homogeneous Dirichlet boundary conditions are enforced on the displacement. Observe that (5)–(6) hold with

$$\mathcal{D} = \left[\begin{array}{c:c:c} 0 & 0 & \mathcal{H} \\ \hdashline 0 & 0 & 0 \\ \hdashline \mathcal{H}^t & 0 & 0 \end{array}\right], \qquad \mathcal{M} = \left[\begin{array}{c:c:c} 0 & 0 & -\mathcal{H} \\ \hdashline 0 & 0 & 0 \\ \hdashline \mathcal{H}^t & 0 & 0 \end{array}\right], \tag{20}$$

where $\mathcal{H} = \sum_{k=1}^d n_k\mathcal{E}^k \in \mathbb{R}^{d^2,d}$ is such that $\mathcal{H}\xi = -\frac{1}{2}(\xi\otimes n + n\otimes\xi)$ for all $\xi \in \mathbb{R}^d$.

Remark 2.3. Using a Korn inequality, one can show that the well–posedness of (4) still holds if $\gamma_2 \geq 0$ a.e. in Ω.

2.2.4. *Maxwell's equations in the elliptic regime*

Let σ and μ be two positive functions in $L^\infty(\Omega)$ uniformly bounded away from zero. A simplified form of Maxwell's equations in $\mathbb{R}^3$ in the elliptic regime, i.e., when displacement currents are negligible, consists of the PDE's

$$\begin{cases} \mu H + \nabla\times E = f, \\ \sigma E - \nabla\times H = g, \end{cases} \tag{21}$$

with data f and g in $[L^2(\Omega)]^3$. The above PDE's fall into the category of Friedrichs' systems by setting $m = 6$ and

$$\mathcal{K} = \left[\begin{array}{c|c} \mu\mathcal{I}_3 & 0 \\ \hline 0 & \sigma\mathcal{I}_3 \end{array}\right], \qquad \mathcal{A}^k = \left[\begin{array}{c|c} 0 & \mathcal{R}^k \\ \hline (\mathcal{R}^k)^t & 0 \end{array}\right], \tag{22}$$

with $\mathcal{R}^k_{ij} = \epsilon_{ikj}$ for $i, j, k \in \{1, 2, 3\}$, ϵ_{ikj} being the Levi–Civita permutation tensor. The graph space is $W = H(\text{curl}; \Omega)\times H(\text{curl}; \Omega)$.

The boundary operator D is such that for all $(H, E), (h, e) \in W$,

$$\langle D(H, E), (h, e)\rangle_{W',W} = (\nabla\times E, h)_{[L^2(\Omega)]^3} - (E, \nabla\times h)_{[L^2(\Omega)]^3} + (H, \nabla\times e)_{[L^2(\Omega)]^3} - (\nabla\times H, e)_{[L^2(\Omega)]^3}. \tag{23}$$

Let $M \in \mathcal{L}(W; W')$ be defined such that for all $(H, E), (h, e) \in W$,

$$\langle M(H, E), (h, e)\rangle_{W',W} = -(\nabla\times E, h)_{[L^2(\Omega)]^3} + (E, \nabla\times h)_{[L^2(\Omega)]^3} + (H, \nabla\times e)_{[L^2(\Omega)]^3} - (\nabla\times H, e)_{[L^2(\Omega)]^3}. \tag{24}$$

Then, (M1)–(M2) hold and $V = \{(H, E) \in W;\ (E\times n)|_{\partial\Omega} = 0\}$, i.e., homogeneous Dirichlet boundary conditions are enforced on the tangential component of the electric field. Observe that (5)–(6) hold with

$$\mathcal{D} = \left[\begin{array}{c|c} 0 & \mathcal{N} \\ \hline \mathcal{N}^t & 0 \end{array}\right], \qquad \mathcal{M} = \left[\begin{array}{c|c} 0 & -\mathcal{N} \\ \hline \mathcal{N}^t & 0 \end{array}\right], \tag{25}$$

where $\mathcal{N} = \sum_{k=1}^{3} n_k\mathcal{R}^k \in \mathbb{R}^{3,3}$ is such that $\mathcal{N}\xi = n\times\xi$ for all $\xi \in \mathbb{R}^3$.

3. The continuous finite element method with face penalty

The purpose of this section is to design and analyze a continuous finite element method to approximate Friedrichs' systems. The two main features of the method are that boundary conditions are enforced weakly and that some components of the gradient jump across mesh interfaces are penalized.

3.1. The discrete setting

Let $\{\mathcal{T}_h\}_{h>0}$ be a shape-regular family of affine meshes of Ω. We assume that the meshes do not possess hanging nodes and that Ω is a polyhedron so that the meshes cover Ω exactly. The notation $A \lesssim B$ represents the inequality $A \le cB$ with c positive and independent of h.

Let $\mathcal{F}_h^{\mathrm{i}}$ be the set of interior faces in the mesh, let $\mathcal{F}_h^{\partial}$ the set of the faces that separate the mesh from the exterior of Ω, and set $\mathcal{F}_h = \mathcal{F}_h^{\mathrm{i}} \cup \mathcal{F}_h^{\partial}$. For all $F \in \mathcal{F}_h^{\mathrm{i}}$, let $T_1(F)$ and $T_2(F) \in \mathcal{T}_h$ be such that $F = T_1(F) \cap T_2(F)$ and set $\mathcal{T}(F) = T_1(F) \cup T_2(F)$. Let n_F be the unit normal vector to F pointing from $T_1(F)$ to $T_2(F)$ (nothing that is said hereafter depends on the orientation of n_F) and set $\mathcal{D}_F = \sum_{k=1}^{d} \mathcal{A}^k n_{F,k}$; then, $|\mathcal{D}_F|$ is well–defined. Furthermore, for an $\mathbb{R}^m$-valued function v such that ∇v admits a (possibly two-valued) trace on F, define the $\mathbb{R}^m$-valued normal jump of its gradient as

$$[\![\nabla v]\!]_F = (\nabla v|_{T_1(F)} - \nabla v|_{T_2(F)})\cdot n_F. \tag{26}$$

The subscript F in jumps is omitted if there is no ambiguity.

For $T \in \mathcal{T}_h$ (resp., $F \in \mathcal{F}_h$), h_T (resp., h_F) denotes the diameter of T (resp., of F). Let $\mathfrak{h}$ be the continuous, piecewise affine function equal on each vertex ν of $\mathcal{T}_h$ to the mean-value of the elements of the set $\{h_T;\ T \ni \nu\}$. Owing to the shape–regularity of the mesh family, for all $T \in \mathcal{T}_h$ and for all $T' \in \mathcal{T}_h$ such that $T' \cap T \neq \emptyset$, $h_{T'} \lesssim \mathfrak{h}|_T \lesssim h_{T'}$.

Let p be a positive integer and set

$$V_h = \{v_h \in \mathfrak{C}^0(\Omega);\ \forall T \in \mathcal{T}_h,\ v_h|_T \in \mathbb{P}_p\}, \tag{27}$$

where $\mathbb{P}_p$ denotes the vector space of polynomials of total degree less than or equal to p. Set $W_h = [V_h]^m$ and $W(h) = W_h + [H^1(\Omega)]^m$.

For any measurable subset of Ω, say E, $(\cdot,\cdot)_E$ denotes the usual L^2-scalar product on E, and $\|\cdot\|_E$ the associated norm. The same notation is used for vector-valued functions. Since the mesh family is shape-regular, for all $v_h \in V_h$ and for all $T \in \mathcal{T}_h$,

$$\|\nabla v_h\|_T \lesssim h_T^{-1}\|v_h\|_T, \tag{28}$$

$$\|v_h\|_F \lesssim h_T^{-\frac{1}{2}}\|v_h\|_T, \quad \forall F \subset \partial T. \tag{29}$$

To enforce boundary conditions weakly, we introduce for all $F \in \mathcal{F}_h^\partial$ an $\mathbb{R}^{m,m}$-valued field $\mathcal{M}_F$ such that for all $v, w \in [L^2(F)]^m$,

$$0 \leq \mathcal{M}_F \leq \mathcal{I}_m, \tag{30}$$

$$(\mathcal{M}v = \mathcal{D}v) \implies (\mathcal{M}_F v = \mathcal{D}v), \tag{31}$$

$$|((\mathcal{M}_F - \mathcal{D})v, w)_{L,F}| \lesssim |v|_{M,F}\|w\|_F, \tag{32}$$

$$|((\mathcal{M}_F + \mathcal{D})v, w)_{L,F}| \lesssim \|v\|_F|w|_{M,F}, \tag{33}$$

where we have introduced for all $v \in W(h)$ the semi-norms $|v|_{M,F} = (\mathcal{M}_F v, v)_F^{\frac{1}{2}}$. Furthermore, to penalize gradient jumps across interfaces, we introduce for all $F \in \mathcal{F}_h^{\mathrm{i}}$ an $\mathbb{R}^{m,m}$-valued field $\mathcal{S}_F$ such that

$$\mathcal{S}_F \text{ is symmetric}, \tag{34}$$

$$h_F^2|\mathcal{D}_F| \lesssim \mathcal{S}_F \lesssim h_F^2\mathcal{I}_m, \tag{35}$$

and we introduce for all $v \in W(h)$ the semi-norms $|v|_{S,F} = (\mathcal{S}_F v, v)_F^{\frac{1}{2}}$.

On $W(h)\times W(h)$ define the bilinear form

$$a_h(v,w) = (Tv, w)_\Omega + \sum_{F\in\mathcal{F}_h^\partial} \tfrac{1}{2}((\mathcal{M}_F - \mathcal{D})v, w)_F + \sum_{F\in\mathcal{F}_h^{\mathrm{i}}} (\mathcal{S}_F[\![\nabla v]\!], [\![\nabla w]\!])_F. \tag{36}$$

In the last term, $(\mathcal{S}_F[\![\nabla v]\!], [\![\nabla w]\!])_F = \sum_{k=1}^d (\mathcal{S}_F[\![\partial_k v]\!], [\![\partial_k w]\!])_F = (\mathcal{S}_F[\![\partial_n v]\!], [\![\partial_n w]\!])_F$ where ∂_n denotes the normal derivative across F since functions in $W(h)$ are continuous. Then, to approximate the solution z of (4), the following problem is considered:

$$\text{Seek } z_h \in W_h \text{ such that } a_h(z_h, y_h) = (f, y_h)_\Omega,\ \forall y_h \in W_h. \tag{37}$$

Remark 3.1. The design conditions on the boundary field $\mathcal{M}_F$ are similar to those introduced for the DGM by Ern and Guermond [10]. The design of the interface field $\mathcal{S}_F$ is, however, different, since in the DGM, this operator penalizes the jumps of the discrete solution and scales independently of h.

3.2. Convergence analysis

To perform the error analysis we introduce the following norm on $W(h)$,

$$|||v|||^2 = \|v\|_\Omega^2 + \sum_{F\in\mathcal{F}_h^\partial} |v|_{M,F}^2 + \sum_{F\in\mathcal{F}_h^{\mathrm{i}}} |[\![\nabla v]\!]|_{S,F}^2 + \|\mathfrak{h}^{\frac{1}{2}} Av\|_\Omega^2. \tag{38}$$

Using integration by parts yields for all $v, w \in W(h)$,

$$a_h(v,w) = (v, \tilde{T}w)_\Omega + \sum_{F\in\mathcal{F}_h^\partial} \tfrac{1}{2}((\mathcal{M}_F + \mathcal{D})v, w)_F + \sum_{F\in\mathcal{F}_h^{\mathrm{i}}} (\mathcal{S}_F[\![\nabla v]\!], [\![\nabla w]\!])_F. \tag{39}$$

Hence, owing to (A4), for all $v_h \in W_h$,

$$a_h(v_h, v_h) \gtrsim \|v_h\|_\Omega^2 + \sum_{F\in\mathcal{F}_h^\partial} |v_h|_{M,F}^2 + \sum_{F\in\mathcal{F}_h^{\mathrm{i}}} |[\![\nabla v_h]\!]|_{S,F}^2, \tag{40}$$

which shows that the bilinear form a_h is at least L-coercive on W_h. To control the graph norm, a sharper stability result is needed. This is the purpose of the following

Lemma 3.1 (Stability)**.** *Assume that for all $k \in \{1,\dots,d\}$, $\mathcal{A}^k \in [\mathcal{C}^{0,\frac{1}{2}}(\Omega)]^{m,m}$. Then, the following holds:*

$$\forall v_h \in W_h, \qquad \sup_{w_h\in W_h\setminus\{0\}} \frac{a_h(v_h,w_h)}{|||w_h|||} \gtrsim |||v_h|||. \tag{41}$$

Proof. Let $v_h \in W_h$.
(i) For all $T \in \mathcal{T}_h$, denote by $\overline{\mathcal{A}_T^k}$ the mean-value of $\mathcal{A}^k$ on T. Then, by assumption

$$\|\overline{\mathcal{A}_T^k} - \mathcal{A}^k\|_{[L^\infty(T)]^{m,m}} \lesssim h_T^{\frac{1}{2}}.$$

Define $\overline{A}v_h|_T = \sum_{k=1}^d \overline{\mathcal{A}_T^k}\partial_k v_h$. Set $\zeta_h' = \mathfrak{h}\overline{A}v_h$ and observe that for all $T \in \mathcal{T}_h$, $\zeta_h'|_T \in [\mathbb{P}_p]^m$ and that

$$\|\zeta_h'\|_T \lesssim \min(\|v_h\|_T, h_T^{\frac{1}{2}}\|\mathfrak{h}^{\frac{1}{2}}Av_h\|_T + h_T^{\frac{1}{2}}\|v_h\|_T). \tag{42}$$

Let $\zeta_h = \pi_h\zeta_h'$ where π_h is the Oswald interpolation operator defined as follows: For all $w_h \in [L^2(\Omega)]^m$ such that $w_h|_T \in [\mathbb{P}_p]^m$ for all $T \in \mathcal{T}_h$, $\pi_h w_h \in W_h$ is defined by its values at the usual Lagrange interpolation nodes by setting

$$\pi_h w_h(\nu) = \frac{1}{\mathrm{card}(\mathcal{T}_\nu)} \sum_{T\in\mathcal{T}_\nu} w_h|_T(\nu),$$

where ν is a Lagrange interpolation node and $\mathcal{T}_\nu$ is the the set of elements to which ν belongs. Recall the following local stability and interpolation results [3, 8, 9, 15, 18]:

$$\|\pi_h w_h\|_T \lesssim \|w_h\|_{\Delta_1(T)}, \tag{43}$$

$$\|w_h - \pi_h w_h\|_T \lesssim \sum_{F\in\Delta_2(T)} h_F^{\frac{1}{2}}\|[\![w_h]\!]\|_F, \tag{44}$$

where $\Delta_1(T) = \{T' \in \mathcal{T}_h;\ T'\cap T \neq \emptyset\}$, $\Delta_2(T) = \{F \in \mathcal{F}_h^{\mathrm{i}};\ F\cap T \neq \emptyset\}$, and $[\![w_h]\!] = w_h|_{T_1(F)} - w_h|_{T_2(F)}$. The shape-regularity of the mesh family implies that $\mathrm{card}(\Delta_1(T)) \lesssim 1$ and $\mathrm{card}(\Delta_2(T)) \lesssim 1$. Furthermore, using (28), (29), (30) (upper bound), (35) (upper bound), (42) (both bounds), and (43), it is inferred that

$$|||\zeta_h||| \lesssim |||v_h|||.$$

(ii) Observe that

$$\|\mathfrak{h}^{\frac{1}{2}}Av_h\|_\Omega^2 = a_h(v_h,\zeta_h) - (\mathcal{K}v_h,\zeta_h)_\Omega - \sum_{F\in\mathcal{F}_h^\partial}\tfrac{1}{2}((\mathcal{M}_F-\mathcal{D})v_h,\zeta_h)_F - \sum_{F\in\mathcal{F}_h^{\rm i}}(\mathcal{S}_F[\![\nabla v_h]\!],[\![\nabla\zeta_h]\!])_F$$
$$+ (Av_h,\mathfrak{h}Av_h-\zeta_h)_\Omega := a_h(v_h,\zeta_h) + R_1+R_2+R_3+R_4.$$

We now bound the remainder terms R_1 to R_4. Using (42) (first bound) and (43) yields

$$|R_1| \lesssim \sum_{T\in\mathcal{T}_h}\|v_h\|_T\|\zeta_h\|_T \lesssim \|v_h\|_\Omega^2.$$

Using (32), (29), (42) (second bound), (43), and Young's inequality leads to

$$|R_2| \lesssim \|v_h\|_\Omega^2 + \sum_{F\in\mathcal{F}_h^\partial}|v_h|_{M,F}^2 + \gamma\|\mathfrak{h}^{\frac{1}{2}}Av_h\|_\Omega^2,$$

where γ can be chosen as small as needed. Similarly, using (35) (upper bound), (28), (29), (43), (42) (second bound), and Young's inequality yields

$$|R_3| \lesssim \|v_h\|_\Omega^2 + \sum_{F\in\mathcal{F}_h^{\rm i}}|[\![\nabla v_h]\!]|_{S,F}^2 + \gamma\|\mathfrak{h}^{\frac{1}{2}}Av_h\|_\Omega^2.$$

Finally, observe that

$$R_4 = (Av_h,\mathfrak{h}Av_h-\zeta_h')_\Omega + (Av_h,\zeta_h'-\zeta_h)_\Omega := R_{4,1}+R_{4,2}.$$

Using (28) yields

$$|R_{4,1}| \lesssim \|v_h\|_\Omega^2 + \gamma\|\mathfrak{h}^{\frac{1}{2}}Av_h\|_\Omega^2.$$

Using (44) yields

$$|R_{4,2}| \lesssim \sum_{T\in\mathcal{T}_h}\|\mathfrak{h}^{\frac{1}{2}}Av_h\|_T\left(\sum_{F\in\Delta_2(T)}\|[\![\zeta_h']\!]\|_F\right).$$

For all $F\in\mathcal{F}_h^{\rm i}$, using the continuity of $\mathfrak{h}$ leads to

$$\|[\![\zeta_h']\!]\|_F \le \|[\![\mathfrak{h}(\overline{A}-A)v_h]\!]\|_F + \|\mathfrak{h}[\![Av_h]\!]\|_F \lesssim \|v\|_{\mathcal{T}(F)} + |[\![\nabla v_h]\!]|_{S,F},$$

owing to (35) (lower bound). This yields

$$|R_{4,2}| \lesssim \|v_h\|_\Omega^2 + \sum_{F\in\mathcal{F}_h^{\rm i}}|[\![\nabla v_h]\!]|_{S,F}^2 + \gamma\|\mathfrak{h}^{\frac{1}{2}}Av_h\|_\Omega^2.$$

Collecting the above bounds and using (40), it is inferred that

$$\|\mathfrak{h}^{\frac{1}{2}}Av_h\|_\Omega^2 \lesssim a_h(v_h,\zeta_h) + a(v_h,v_h).$$

Since $\|\!|\zeta_h|\!\| \lesssim \|\!|v_h|\!\|$, the conclusion is straightforward. □

Lemma 3.2 (Continuity)**.** *Define the following norm on $W(h)$,*

$$\|\!|v|\!\|_*^2 = \|\!|v|\!\|^2 + \sum_{T\in\mathcal{T}_h}[h_T^{-1}\|v\|_T^2 + \|v\|_{\partial T}^2]. \tag{45}$$

Then, the following holds:

$$\forall(v,w)\in W(h)\times W(h),\quad a_h(v,w)\lesssim \|\!|v|\!\|_*\|\!|w|\!\|. \tag{46}$$

Proof. We bound the three terms in the right-hand side of (39). For the first term, we obtain using (28),

$$|(v,\tilde{T}w)_\Omega|\lesssim \sum_{T\in\mathcal{T}_h}\|v\|_T(\|w\|_T+\|Aw\|_T)\lesssim \|\!|v|\!\|_*\|\!|w|\!\|.$$

For the second term, (33) yields

$$\sum_{F\in\mathcal{F}_h^\partial}\tfrac{1}{2}|((\mathcal{M}_F+\mathcal{D})v,w)_F|\lesssim \sum_{F\in\mathcal{F}_h^\partial}\|v\|_F|w|_{M,F}\lesssim \|\!|v|\!\|_*\|\!|w|\!\|.$$

The bound on the third term is straightforward. □

Lemma 3.3 (Consistency)**.** *Let z solve* (4) *and let z_h solve* (37)*. If $z\in[H^2(\Omega)]^m$, then,*

$$\forall y_h\in W_h,\quad a_h(z-z_h,y_h)=0. \tag{47}$$

Proof. Since $z\in[H^2(\Omega)]^m$ solves (4), $\mathcal{M}z=\mathcal{D}z$ a.e. on $\partial\Omega$ and $Tz=f$ in L. Assumption (31) yields $\mathcal{M}_F z|_F=\mathcal{D}z|_F$ for all $F\in\mathcal{F}_h^\partial$. Moreover, $[\![\nabla z]\!]_F=0$ for all $F\in\mathcal{F}_h^{\mathrm{i}}$. The conclusion follows readily. □

The above results readily yield the following

Theorem 3.1 (Convergence)**.** *Let z solve* (4) *and let z_h solve* (37)*. Assume $z\in[H^2(\Omega)]^m$. Then, under the assumptions of Lemma* 3.1*,*

$$\|\!|z-z_h|\!\|\lesssim \inf_{v_h\in W_h}\|\!|z-v_h|\!\|_*. \tag{48}$$

Using standard interpolation properties in W_h, it is inferred that

$$\|\!|z-z_h|\!\|\lesssim h^{p+\frac{1}{2}}\|z\|_{[H^{p+1}(\Omega)]^m}, \tag{49}$$

if $z\in[H^{p+1}(\Omega)]^m$. In particular, the method yields $(p+\frac{1}{2})$-order convergence in the L-norm and, provided the mesh family is quasi-uniform, optimal order convergence in the graph norm. These estimates are identical to those that can be obtained by other stabilization methods like Galerkin/Least-Squares, subgrid viscosity, and other methods.

Remark 3.2. When the exact solution is too rough to be in $[H^2(\Omega)]^m$, assuming that $[H^2(\Omega)]^m\cap W$ is dense in W, it can be proven by proceeding as in [10] that $\lim_{h\to 0}\|z-z_h\|_\Omega=0$.

3.3. **Examples**

In this section we apply the theoretical results of §3.2 to the Friedrichs' systems discussed in §2.2. For brevity, proofs are omitted.

3.3.1. *Advection–reaction*

Let $\alpha>0$ and take

$$\mathcal{S}_F=\alpha h_F^2|\beta\cdot n_F|,\qquad \mathcal{M}_F=|\beta\cdot n|. \tag{50}$$

Then, (30)–(35) hold. Hence, if $\beta\in[\mathcal{C}^{0,\frac{1}{2}}(\Omega)]^d$ and the exact solution is smooth enough,

$$\|u-u_h\|_\Omega+\|\mathfrak{h}^{\frac{1}{2}}\beta\cdot\nabla(u-u_h)\|_\Omega\lesssim h^{p+\frac{1}{2}}\|u\|_{H^{p+1}(\Omega)}. \tag{51}$$

3.3.2. *Advection–diffusion–reaction*

Let $\alpha_1 > 0$, $\alpha_2 > 0$, and $\eta > 0$ and take

$$\mathcal{S}_F = h_F^2 \left[\begin{array}{c:c} \alpha_1 n_F \otimes n_F & 0 \\ \hdashline 0 & \alpha_2 \end{array}\right], \qquad \mathcal{M}_F = \left[\begin{array}{c:c} 0 & -n \\ \hdashline n^t & \eta \end{array}\right]. \tag{52}$$

Then, (30)–(35) hold. Hence, if $\beta \in [\mathcal{C}^{0,\frac{1}{2}}(\Omega)]^d$ and the exact solution is smooth enough,

$$\|u - u_h\|_\Omega + \|\mathfrak{h}^{\frac{1}{2}} \nabla(u - u_h)\|_\Omega + \|\sigma - \sigma_h\|_\Omega + \|\mathfrak{h}^{\frac{1}{2}} \nabla\cdot(\sigma - \sigma_h)\|_\Omega \lesssim h^{p+\frac{1}{2}} \|(\sigma, u)\|_{[H^{p+1}(\Omega)]^{d+1}}. \tag{53}$$

3.3.3. *Linear elasticity*

Let $\alpha_1 > 0$, $\alpha_2 > 0$, and $\eta > 0$ and take

$$\mathcal{S}_F = h_F^2 \left[\begin{array}{c:c:c} \alpha_1 \mathcal{H}_F \cdot \mathcal{H}_F^t & 0 & 0 \\ \hdashline 0 & 0 & 0 \\ \hdashline 0 & 0 & \alpha_2 \mathcal{I}_d \end{array}\right], \qquad \mathcal{M}_F = \left[\begin{array}{c:c:c} 0 & 0 & -\mathcal{H} \\ \hdashline 0 & 0 & 0 \\ \hdashline \mathcal{H}^t & 0 & \eta \mathcal{I}_d \end{array}\right], \tag{54}$$

where $\mathcal{H}_F$ is defined similarly to $\mathcal{H}$ by substituting n_F to n. Then, (30)–(35) hold. Hence, if the exact solution is smooth enough,

$$\begin{aligned} &\|u - u_h\|_\Omega + \|\mathfrak{h}^{\frac{1}{2}} \nabla(u - u_h)\|_\Omega + \|p - p_h\|_\Omega \\ &\qquad + \|\sigma - \sigma_h\|_\Omega + \|\mathfrak{h}^{\frac{1}{2}} \nabla\cdot((\sigma + \sigma^t) - (\sigma_h + \sigma_h^t))\|_\Omega \lesssim h^{p+\frac{1}{2}} \|(\overline{\sigma}, p, u)\|_{[H^{p+1}(\Omega)]^{d^2+1+d}}, \end{aligned} \tag{55}$$

where Korn's Second Inequality has been used to simplify the estimate on the graph norm of the displacement.

Remark 3.3. The above finite element method needs to be modified in the incompressible limit; see, e.g., [5] for an analysis of a face penalty stabilized finite element approximation to the Stokes equations.

3.3.4. *Maxwell's equations in the elliptic regime*

Let $\alpha_1 > 0$, $\alpha_2 > 0$, and $\eta > 0$ and take

$$\mathcal{S}_F = h_F^2 \left[\begin{array}{c:c} \alpha_1 \mathcal{N}_F^t \mathcal{N}_F & 0 \\ \hdashline 0 & \alpha_2 \mathcal{N}_F^t \mathcal{N}_F \end{array}\right], \qquad \mathcal{M}_F = \left[\begin{array}{c:c} 0 & -\mathcal{N} \\ \hdashline \mathcal{N}^t & \eta \mathcal{N}^t \mathcal{N} \end{array}\right], \tag{56}$$

where $\mathcal{N}_F$ is defined similarly to $\mathcal{N}$ by substituting n_F to n. Then, (30)–(35) hold. Hence, if the exact solution is smooth enough,

$$\|E - E_h\|_\Omega + \|\mathfrak{h}^{\frac{1}{2}} \nabla\times(E - E_h)\|_\Omega + \|H - H_h\|_\Omega + \|\mathfrak{h}^{\frac{1}{2}} \nabla\times(H - H_h)\|_\Omega \lesssim h^{p+\frac{1}{2}} \|(H, E)\|_{[H^{p+1}(\Omega)]^6}. \tag{57}$$

4. Friedrichs' systems with 2×2 block structure

This section deals with a specific class of Friedrichs' systems endowed with a particular 2×2 block structure such that the dependent variable z in (4) can be partitioned into the form $z = (z^\sigma, z^u)$ and the variable z^σ can be eliminated to yield a system of second-order PDE's for z^u that is of elliptic type. This class of Friedrichs' systems and its approximation by a local DGM was recently analyzed in [11]. The purpose of this section is to design and analyze a SCFEM where only the jumps of the gradient of the z^u-component are penalized. One motivation for this strategy is to substantially reduce the number of nonzero entries in the stiffness matrix, thus alleviating considerably memory requirements.

4.1. The continuous and discrete settings

Let m_σ and m_u be two positive integers such that $m = m_\sigma + m_u$ and assume that for all $k \in \{1,\dots,d\}$, the matrices $\mathcal{A}^k$ have the following structure

$$\mathcal{A}^k = \left[\begin{array}{c|c} 0 & \epsilon^{\frac{1}{2}}\mathcal{B}^k \\ \hline \epsilon^{\frac{1}{2}}(\mathcal{B}^k)^t & \mathcal{C}^k \end{array}\right], \tag{58}$$

where $\mathcal{B}^k$ is $\mathbb{R}^{m_\sigma,m_u}$-valued and $\mathcal{C}^k$ is $\mathbb{R}^{m_u,m_u}$-valued. To handle the case of advection–diffusion equations with dominant advection, we have also included a positive parameter ϵ that is at most of order unity but can take arbitrarily small values. The notation $A \lesssim B$ now means that $A \le cB$ with c positive and independent of h and ϵ. Furthermore, the fields $\mathcal{B}^k$ and $\mathcal{C}^k$ are of order unity. Examples of Friedrichs' systems endowed with the above structure are advection–diffusion–reaction equations (ϵ is the diffusion coefficient, $m_\sigma = d$, and $m_u = 1$), linear elasticity equations ($\epsilon = 1$, $m_\sigma = d^2 + 1$, and $m_u = d$), and the Maxwell equations in the elliptic regime ($\epsilon = 1$, $m_\sigma = 3$, and $m_u = 3$). Owing to (58), the matrix $\mathcal{D}$ is such that

$$\mathcal{D} = \left[\begin{array}{c|c} 0 & \epsilon^{\frac{1}{2}}\mathcal{D}^{\sigma u} \\ \hline \epsilon^{\frac{1}{2}}(\mathcal{D}^{\sigma u})^t & \mathcal{D}^{uu} \end{array}\right], \tag{59}$$

with obvious notation. For simplicity, we restrict ourselves to the case where the boundary conditions can be enforced by taking

$$\mathcal{M} = \left[\begin{array}{c|c} 0 & -\epsilon^{\frac{1}{2}}\mathcal{D}^{\sigma u} \\ \hline \epsilon^{\frac{1}{2}}(\mathcal{D}^{\sigma u})^t & 0 \end{array}\right]. \tag{60}$$

This corresponds to a Dirichlet condition on u both for the advection–diffusion–reaction equation and for the linear elasticity equations, while it enforces the condition $E\times n = 0$ for the Maxwell equations in the elliptic regime.

To enforce boundary conditions weakly, we introduce for all $F \in \mathcal{F}_h^\partial$ a matrix-valued field $\mathcal{M}_F$ such that

$$\mathcal{M}_F = \left[\begin{array}{c|c} 0 & -\epsilon^{\frac{1}{2}}\mathcal{D}^{\sigma u} \\ \hline \epsilon^{\frac{1}{2}}(\mathcal{D}^{\sigma u})^t & \mathcal{M}_F^{uu} \end{array}\right]. \tag{61}$$

We still assume that the consistency condition (31) holds. However, instead of (30), (32), and (33), we now assume that $\mathcal{M}_F^{uu}$ is symmetric and that

$$\tfrac{\epsilon}{h_F}((\mathcal{D}^{\sigma u})^t\mathcal{D}^{\sigma u})^{\frac{1}{2}} + |\mathcal{D}^{uu}| \lesssim \mathcal{M}_F^{uu} \lesssim (1 + \tfrac{\epsilon}{h_F})\mathcal{I}_{m_u}. \tag{62}$$

If $\epsilon = 1$, this yields $\frac{1}{h_F}((\mathcal{D}^{\sigma u})^t\mathcal{D}^{\sigma u})^{\frac{1}{2}} \lesssim \mathcal{M}_F^{uu} \lesssim \frac{1}{h_F}\mathcal{I}_{m_u}$, while if $\epsilon \ll h$, (62) implies that $\mathcal{M}_F^{uu}$ and $\mathcal{D}^{uu}$ satisfy (30), (32), and (33).

To penalize the jumps of the gradient of the z^u-component only, we introduce for all $F \in \mathcal{F}_h^{\mathrm{i}}$ a matrix-valued field $\mathcal{S}_F$ such that

$$\mathcal{S}_F = \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & \mathcal{S}_F^{uu} \end{array}\right]. \tag{63}$$

Instead of (34) and (35), we now assume that $\mathcal{S}_F^{uu}$ is symmetric and that

$$\tfrac{\epsilon}{h_F}((\mathcal{D}_F^{\sigma u})^t\mathcal{D}_F^{\sigma u})^{\frac{1}{2}} + |\mathcal{D}_F^{uu}| \lesssim \tfrac{1}{h_F^2}\mathcal{S}_F^{uu} \lesssim (1 + \tfrac{\epsilon}{h_F})\mathcal{I}_{m_u}. \tag{64}$$

If $\epsilon = 1$, this yields $h_F((\mathcal{D}_F^{\sigma u})^t\mathcal{D}_F^{\sigma u})^{\frac{1}{2}} \lesssim \mathcal{S}_F^{uu} \lesssim h_F\mathcal{I}_{m_u}$, while if $\epsilon \ll h$, conditions (34) and (35) are recovered.

Owing to the above setting, the bilinear form a_h defined by (36) becomes

$$a_h(v,w) = (\mathcal{K}v,w)_\Omega + \epsilon^{\frac12}(Bv^u,w^\sigma)_\Omega + \epsilon^{\frac12}(\tilde{B}v^\sigma,w^u)_\Omega + (Cv^u,w^u)_\Omega \\ + \sum_{F\in\mathcal{F}_h^\partial}[-\epsilon^{\frac12}(\mathcal{D}^{\sigma u}v^u,w^\sigma)_F + \tfrac12((\mathcal{M}_F^{uu}-\mathcal{D}^{uu})v^u,w^u)_F] + \sum_{F\in\mathcal{F}_h^{\mathrm{i}}}(\mathcal{S}_F^{uu}[\![\nabla v^u]\!],[\![\nabla w^u]\!])_F, \tag{65}$$

where $B=\sum_{k=1}^d \mathcal{B}^k\partial_k$, $\tilde{B}=\sum_{k=1}^d(\mathcal{B}^k)^t\partial_k$, and $C=\sum_{k=1}^d\mathcal{C}^k\partial_k$. The discrete problem is still (37) with the discrete space W_h unchanged.

4.2. **Convergence analysis**

To perform the error analysis we introduce the following norm on $W(h)$,

$$|\!|\!| v |\!|\!|^2 = \|v\|_\Omega^2 + \sum_{F\in\mathcal{F}_h^\partial}|v^u|_{M,F}^2 + \sum_{F\in\mathcal{F}_h^{\mathrm{i}}}|[\![\nabla v^u]\!]|_{S,F}^2 + \|\epsilon^{\frac12}Bv^u\|_\Omega^2 + \|\mathfrak{h}^{\frac12}Cv^u\|_\Omega^2, \tag{66}$$

with the semi-norms $|v^u|_{M,F}=(\mathcal{M}_F^{uu}v^u,v^u)_F^{\frac12}$ and $|[\![\nabla v^u]\!]|_{S,F}=(\mathcal{S}_F^{uu}[\![\nabla v^u]\!],[\![\nabla v^u]\!])_F^{\frac12}$.

Lemma 4.1 (Stability). *Assume that for all $k\in\{1,\dots,d\}$, $\mathcal{B}^k\in[\mathcal{C}^{0,1}(\Omega)]^{m_\sigma,m_u}$ and $\mathcal{C}^k\in[\mathcal{C}^{0,1}(\Omega)]^{m_u,m_u}$, and that*

$$\forall T\in\mathcal{T}_h,\ \forall w_h\in W_h,\quad \|Cw_h^u\|_T\lesssim\|Bw_h^u\|_T. \tag{67}$$

Then, the following holds:

$$\forall v_h\in W_h,\qquad \sup_{w_h\in W_h\setminus\{0\}}\frac{a_h(v_h,w_h)}{|\!|\!| w_h|\!|\!|}\gtrsim |\!|\!| v_h|\!|\!|. \tag{68}$$

Proof. Let $v_h\in W_h$. It is clear that the coercivity property (40) now becomes

$$a_h(v_h,v_h)\gtrsim\|v_h\|_\Omega^2+\sum_{F\in\mathcal{F}_h^\partial}|v_h^u|_{M,F}^2+\sum_{F\in\mathcal{F}_h^{\mathrm{i}}}|[\![\nabla v_h^u]\!]|_{S,F}^2.$$

(i) Take $\zeta_h=\pi_h(0,\mathfrak{h}\overline{C}v_h^u)$. Proceeding as in the proof of Lemma 3.1, it is inferred that

$$\|\zeta_h\|_T\lesssim\min(\|v_h^u\|_{\Delta_1(T)},h_T^{\frac12}\|\mathfrak{h}^{\frac12}Cv_h^u\|_{\Delta_1(T)}+h_T^{\frac12}\|v_h^u\|_{\Delta_1(T)}).$$

Observe that

$$\|\mathfrak{h}^{\frac12}Cv_h^u\|_\Omega^2 = a_h(v_h,\zeta_h)-(\mathcal{K}v_h,\zeta_h)_\Omega-(\epsilon^{\frac12}\tilde{B}v_h^\sigma,\zeta_h^u)_\Omega-\sum_{F\in\mathcal{F}_h^\partial}\tfrac12((\mathcal{M}_F^{uu}-\mathcal{D}^{uu})v_h^u,\zeta_h^u)_F \\ -\sum_{F\in\mathcal{F}_h^{\mathrm{i}}}(\mathcal{S}_F^{uu}[\![\nabla v_h^u]\!],[\![\nabla\zeta_h^u]\!])_F+(Cv_h^u,\mathfrak{h}Cv_h^u-\zeta_h^u)_\Omega := a_h(v_h,\zeta_h)+R_1+R_2+R_3+R_4+R_5.$$

The term R_1 is controlled as in the proof of Lemma 3.1. The same is possible for the term R_5 since (64) (lower bound) implies that $h_F^2|\mathcal{D}_F^{uu}|\lesssim\mathcal{S}_F^{uu}$. Furthermore, owing to (29) and (67),

$$\begin{aligned}(\epsilon^{\frac12}\tilde{B}v_h^\sigma,\zeta_h^u)_T&\lesssim\epsilon^{\frac12}\|v_h^\sigma\|_T\|\overline{C}v_h^u\|_{\Delta_1(T)}\\&\lesssim\epsilon^{\frac12}\|v_h^\sigma\|_T(\|(C-\overline{C})v_h^u\|_{\Delta_1(T)}+\|Cv_h^u\|_{\Delta_1(T)})\\&\lesssim\epsilon^{\frac12}\|v_h^\sigma\|_T(\|v_h^u\|_{\Delta_1(T)}+\|Bv_h^u\|_{\Delta_1(T)}),\end{aligned}$$

owing to (67), whence it follows that

$$|R_2| \lesssim \|v_h\|_\Omega^2 + \gamma\|\epsilon^{\frac{1}{2}} B v_h^u\|_\Omega^2,$$

where γ can be chosen as small as needed. To bound R_3, observe that since $|\mathcal{D}^{uu}| \lesssim \mathcal{M}_F^{uu}$ and $\mathcal{M}_F^{uu}$ is positive,

$$|((\mathcal{M}_F^{uu} - \mathcal{D}^{uu})v_h^u, \zeta_h^u)_F| \lesssim (\mathcal{M}_F^{uu} v_h^u, \zeta_h^u)_F \lesssim |v_h^u|_{M,F} |\zeta_h^u|_{M,F}.$$

Using (62) (upper bound) leads to

$$|\zeta_h^u|_{M,F}^2 \lesssim (\zeta_h^u, \zeta_h^u)_F + \tfrac{\epsilon}{h_F}(\zeta_h^u, \zeta_h^u)_F.$$

The first term is bound as in the proof of Lemma 3.1. Moreover, using (29), (43), and (67) to bound the second term eventually leads to

$$|\zeta_h^u|_{M,F}^2 \lesssim \|v_h^u\|_{\Delta_3(F)}^2 + \|\mathfrak{h}^{\frac{1}{2}} C v_h^u\|_{\Delta_3(F)}^2 + \|\epsilon^{\frac{1}{2}} B v_h^u\|_{\Delta_3(F)}^2,$$

with $\Delta_3(F) = \{T \in \mathcal{T}_h;\ T \cap F \neq \emptyset\}$, whence it follows that

$$|R_3| \lesssim \|v_h\|_\Omega^2 + \sum_{F \in \mathcal{F}_h^\partial} |v_h^u|_{M,F}^2 + \gamma\|\mathfrak{h}^{\frac{1}{2}} C v_h^u\|_\Omega^2 + \gamma\|\epsilon^{\frac{1}{2}} B v_h^u\|_\Omega^2.$$

Finally, to bound R_4 use the positivity of $\mathcal{S}_F^{uu}$ and (64) (upper bound) and proceed as above to infer

$$|R_4| \lesssim \|v_h\|_\Omega^2 + \sum_{F \in \mathcal{F}_h^{\mathrm{i}}} |[\![\nabla v_h^u]\!]|_{S,F}^2 + \gamma\|\mathfrak{h}^{\frac{1}{2}} C v_h^u\|_\Omega^2 + \gamma\|\epsilon^{\frac{1}{2}} B v_h^u\|_\Omega^2.$$

Collecting the above bounds yields

$$\|\mathfrak{h}^{\frac{1}{2}} C v_h^u\|_\Omega^2 \lesssim a_h(v_h, \zeta_h) + a_h(v_h, v_h) + \gamma\|\epsilon^{\frac{1}{2}} B v_h^u\|_\Omega^2. \tag{69}$$

(ii) Take $\xi_h = \pi_h(\epsilon^{\frac{1}{2}} \overline{B} v_h^u, 0)$ so that

$$\|\xi_h\|_T \lesssim \|v_h^u\|_{\Delta_1(T)} + \|\epsilon^{\frac{1}{2}} B v_h^u\|_{\Delta_1(T)},$$

since the fields $\mathcal{B}^k$ are Lipschitz and $\epsilon \le 1$ by assumption. Observe that

$$\begin{aligned}\|\epsilon^{\frac{1}{2}} B v_h^u\|_\Omega^2 &= a_h(v_h, \xi_h) - (\mathcal{K} v_h, \xi_h)_\Omega + \sum_{F \in \mathcal{F}_h^\partial} (\epsilon^{\frac{1}{2}} \mathcal{D}^{\sigma u} v_h^u, \xi_h^\sigma)_F + (\epsilon^{\frac{1}{2}} B v_h^u, \epsilon^{\frac{1}{2}} B v_h^u - \xi_h^\sigma)_\Omega \\ &:= a_h(v_h, \zeta_h) + R_1 + R_2 + R_3.\end{aligned}$$

It is clear that

$$|R_1| \lesssim \|v_h^u\|_\Omega^2 + \gamma\|\epsilon^{\frac{1}{2}} B v_h^u\|_\Omega^2.$$

Furthermore, since $\frac{\epsilon}{h_F}((\mathcal{D}^{\sigma u})^t \mathcal{D}^{\sigma u})^{\frac{1}{2}} \lesssim \mathcal{M}_F^{uu}$ owing to (62) (lower bound), it is inferred that

$$(\epsilon^{\frac{1}{2}} \mathcal{D}^{\sigma u} v_h^u, \xi_h^\sigma)_F \lesssim \|(\tfrac{\epsilon}{h_F})^{\frac{1}{2}} \mathcal{D}^{\sigma u} v_h^u\|_F \|\xi_h\|_{\mathcal{T}(F)} \lesssim |v_h^u|_{M,F} \|\xi_h\|_{\mathcal{T}(F)},$$

whence it follows that

$$|R_2| \lesssim \|v_h^u\|_\Omega^2 + \sum_{F \in \mathcal{F}_h^\partial} |v_h^u|_{M,F}^2 + \gamma\|\epsilon^{\frac{1}{2}} B v_h^u\|_\Omega^2.$$

Finally, proceeding as in the proof of Lemma 3.1 and using $((\mathcal{D}_F^{\sigma u})^t\mathcal{D}_F^{\sigma u})^{\frac{1}{2}} \lesssim (\epsilon h_F)^{-1}\mathcal{S}_F^{uu}$ from (64) (lower bound) yields

$$|R_3| \lesssim \|v_h^u\|_\Omega^2 + \sum_{F\in\mathcal{F}_h^{\mathrm{i}}} |[\![\nabla v_h^u]\!]|_{S,F}^2 + \gamma\|\epsilon^{\frac{1}{2}}Bv_h^u\|_\Omega^2.$$

Collecting the above bounds leads to

$$\|\epsilon^{\frac{1}{2}}Bv_h^u\|_\Omega^2 \lesssim a(v_h,\xi_h) + a_h(v_h,v_h). \tag{70}$$

(iii) The bounds (69) and (70) readily imply

$$|\!|\!|v_h|\!|\!|^2 \lesssim a(v_h,\zeta_h) + a(v_h,\xi_h) + a_h(v_h,v_h).$$

Since the fields $\mathcal{C}^k$ are Lipschitz, assumption (67) implies $\|\overline{C}v_h^u\|_T \lesssim \|Bv_h^u\|_T + \|v_h^u\|_T$. Hence, owing to (62) (upper bound) and proceeding as above yields

$$|\zeta_h^u|_{M,F} \lesssim \|v_h^u\|_{\Delta_3(F)} + \|\epsilon^{\frac{1}{2}}Bv_h^u\|_{\Delta_3(F)} + \|\mathfrak{h}^{\frac{1}{2}}Cv_h^u\|_{\Delta_3(F)}.$$

Proceeding similarly to bound the other terms in $|\!|\!|\zeta_h|\!|\!|$, it is inferred that

$$|\!|\!|\zeta_h|\!|\!| \lesssim |\!|\!|v_h|\!|\!|.$$

Similarly, $|\!|\!|\xi_h|\!|\!| \lesssim |\!|\!|v_h|\!|\!|$, and (68) follows readily. □

Lemma 4.2 (Continuity). *Define the following norm on $W(h)$,*

$$|\!|\!|v|\!|\!|_*^2 = |\!|\!|v|\!|\!|^2 + \sum_{T\in\mathcal{T}_h} [(1+\tfrac{\epsilon}{h_T})h_T^{-1}\|v^u\|_T^2 + (1+\tfrac{\epsilon}{h_T})\|v^u\|_{\partial T}^2 + h_T\|v^\sigma\|_{\partial T}^2]. \tag{71}$$

Then, the following holds:

$$\forall (v,w_h)\in W(h)\times W_h, \quad a_h(v,w_h) \lesssim |\!|\!|v|\!|\!|_*|\!|\!|w_h|\!|\!|. \tag{72}$$

Proof. The idea is to bound the three terms in the right-hand side (39) making use of the block structure under consideration. Owing to (28) and the symmetry of $\mathcal{C}^k$,

$$\begin{aligned}|(v,\tilde{T}w_h)_T| &\lesssim \|v\|_T\|w_h\|_T + \|v^\sigma\|_T\|\epsilon^{\frac{1}{2}}Bw_h^u\|_T + |(v^u,\epsilon^{\frac{1}{2}}\tilde{B}w_h^\sigma)| + |(v^u,Cw_h^u)_T|\\ &\lesssim \|v\|_T\|w_h\|_T + \|v^\sigma\|_T\|\epsilon^{\frac{1}{2}}Bw_h^u\|_T + \epsilon^{\frac{1}{2}}h_T^{-1}\|v^u\|_T\|w_h^\sigma\|_T + h_T^{-\frac{1}{2}}\|v^u\|_T\|\mathfrak{h}^{\frac{1}{2}}Cw_h^u\|_T.\end{aligned}$$

Hence,

$$|(v,\tilde{T}w_h)_\Omega| \lesssim |\!|\!|v|\!|\!|_*|\!|\!|w_h|\!|\!|.$$

Furthermore,

$$((\mathcal{M}_F+\mathcal{D})v,w_h)_F = 2(\epsilon^{\frac{1}{2}}(\mathcal{D}^{\sigma u})^t v^\sigma, w_h^u)_F + ((\mathcal{M}_F^{uu}+\mathcal{D}^{uu})v^u,w_h^u)_F.$$

Using (62) yields

$$|((\mathcal{M}_F+\mathcal{D})v,w_h)_F| \lesssim h_F^{\frac{1}{2}}\|v^\sigma\|_F|w_h^u|_{M,F} + (1+\tfrac{\epsilon}{h_T})^{\frac{1}{2}}\|v^u\|_F|w_h^u|_{M,F}.$$

Hence,

$$\sum_{F\in\mathcal{F}_h^\partial} \tfrac{1}{2}|((\mathcal{M}_F+\mathcal{D})v,w_h)_F| \lesssim |\!|\!|v|\!|\!|_*|\!|\!|w_h|\!|\!|.$$

Finally, the bound on the third term is straightforward. □

Since a consistency result analogous to Lemma 3.3 holds, the following convergence theorem is readily inferred from Lemma 4.1 and Lemma 4.2.

Theorem 4.1 (Convergence)**.** *Let z solve* (4) *and let z_h solve* (37)*. Assume that $z = (z^\sigma, z^u) \in [H^1(\Omega)]^{m_\sigma} \times [H^2(\Omega)]^{m_u}$. Then, under the assumptions of Lemma* 4.1*,*

$$\|z - z_h\| \lesssim \inf_{v_h \in W_h} \|z - v_h\|_*. \tag{73}$$

Corollary 4.1. *If $\epsilon \sim 1$ and if $z = (z^\sigma, z^u) \in [H^p(\Omega)]^{m_\sigma} \times [H^{p+1}(\Omega)]^{m_u}$, then*

$$\|z - z_h\| \lesssim h^p \|z\|_{[H^p(\Omega)]^{m_\sigma} \times [H^{p+1}(\Omega)]^{m_u}}. \tag{74}$$

If $\epsilon \ll h$ and if $z = (z^\sigma, z^u) \in [H^{p+1}(\Omega)]^m$, then

$$\|z - z_h\| \lesssim h^{p+\frac{1}{2}} \|z\|_{[H^{p+1}(\Omega)]^m}. \tag{75}$$

Estimate (74) yields that the error $\|z^u - z_h^u\|_\Omega$ converges to order p, which is suboptimal. This estimate can be improved by using the duality argument introduced in [11] for Friedrichs' systems. Consider the following continuous dual problem: letting $V^* = \mathrm{Ker}(D + M^*)$ where $M^* \in \mathcal{L}(W; W')$ is the adjoint of the operator M,

$$\text{Seek } \psi \in V^* \text{ such that } \tilde{T}\psi = (0, z^u - z_h^u) \text{ in } L. \tag{76}$$

Assume the following (elliptic) regularity result:

$$\|\psi^\sigma\|_{[H^1(\Omega)]^{m_\sigma}} + \|\psi^u\|_{[H^2(\Omega)]^{m_u}} \lesssim \|z^u - z_h^u\|_\Omega. \tag{77}$$

Lemma 4.3. *Under the above hypotheses,*

$$\forall v \in W(h), \quad a_h(v, \psi) = (v^u, z^u - z_h^u)_\Omega. \tag{78}$$

Proof. The identity results from (39). Since ψ solves (76), $(v, \tilde{T}\psi)_\Omega = (v^u, z^u - z_h^u)_\Omega$. Moreover, since $\psi \in V^*$ and owing to the particular structure of $\mathcal{M}$ and $\mathcal{M}_F$, it is clear that $(\mathcal{M}_F^t + \mathcal{D})\psi = 0$. Hence,

$$\sum_{F \in \mathcal{F}_h^\partial} \tfrac{1}{2}((\mathcal{M}_F + \mathcal{D})v, \psi)_F = 0.$$

Finally, the last term in (39) vanishes because $\psi^u \in [H^2(\Omega)]^{m_u}$. □

Lemma 4.4. *The following holds:*

$$\forall (v = (0, v^u), w) \in W(h) \times W(h), \quad a_h((0, v^u), w) \lesssim \|(0, v^u)\| \, \|w\|_*. \tag{79}$$

Proof. The proof is similar to that of Lemma 4.2 except that we use (65) instead of (39). □

Theorem 4.2. *In the above framework, the following holds:*

$$\|z^u - z_h^u\|_\Omega \lesssim h\|z - z_h\| + h \inf_{(q_h^\sigma, 0) \in W_h} \left(\sum_{F \in \mathcal{F}_h^\partial} h_F \|z^\sigma - q_h^\sigma\|_F^2 + \|z^\sigma - q_h^\sigma\|_{T(F)}^2 \right)^{\frac{1}{2}}. \tag{80}$$

Hence, if $z \in [H^p(\Omega)]^{m_\sigma} \times [H^{p+1}(\Omega)]^{m_u}$, then

$$\|z^u - z_h^u\|_\Omega \lesssim h^{p+1} \|z\|_{[H^p(\Omega)]^{m_\sigma} \times [H^{p+1}(\Omega)]^{m_u}}. \tag{81}$$

Proof. Owing to (78),
$$\|z^u - z_h^u\|_\Omega^2 = a_h(z - z_h, \psi) = a_h(z - z_h, \psi - w_h),$$
where w_h is arbitrary in W_h. Hence,

$$\begin{aligned}\|z^u - z_h^u\|_\Omega^2 &= a_h((0, z^u - z_h^u), \psi - w_h) + (\mathcal{K}(z^\sigma - z_h^\sigma, 0), \psi - w_h)_\Omega + \epsilon^{\frac{1}{2}}(\tilde{B}(z^\sigma - z_h^\sigma), \psi^u - w_h^u)_\Omega \\ &:= T_1 + T_2 + T_3.\end{aligned}$$

Owing to (79),
$$|T_1| \lesssim \|\!|z - z_h|\!\|\,\|\!|\psi - w_h|\!\|_*,$$
and clearly, $|T_2| \lesssim \|\!|z - z_h|\!\|\,\|\!|\psi - w_h|\!\|_*$. Integrating by parts and using (62) (lower bound) yields

$$\begin{aligned}|T_3| &\lesssim \|z^\sigma - z_h^\sigma\|_\Omega \|\epsilon^{\frac{1}{2}} B(\psi^u - w_h^u)\|_\Omega + \sum_{F \in \mathcal{F}_h^\partial} h_F^{\frac{1}{2}} \|z^\sigma - z_h^\sigma\|_F |\psi^u - w_h^u|_{M,F} \\ &\lesssim \|\!|z - z_h|\!\|\,\|\!|\psi - w_h|\!\|_* + \left(\sum_{F \in \mathcal{F}_h^\partial} h_F \|z^\sigma - z_h^\sigma\|_F^2\right)^{\frac{1}{2}} \|\!|\psi - w_h|\!\|_*.\end{aligned}$$

Using (29) yields for all $(q_h^\sigma, 0) \in W_h$,

$$|T_3| \lesssim \|\!|z - z_h|\!\|\,\|\!|\psi - w_h|\!\|_* + \left(\sum_{F \in \mathcal{F}_h^\partial} h_F \|z^\sigma - q_h^\sigma\|_F^2 + \|z^\sigma - q_h^\sigma\|_{T(F)}^2\right)^{\frac{1}{2}} \|\!|\psi - w_h|\!\|_*,$$

where $T(F)$ is the mesh element of which F is a face. Using (77) and classical interpolation results yields $\|\!|\psi - w_h|\!\|_* \lesssim h\|z^u - z_h^u\|_\Omega$. The conclusion is straightforward. □

Remark 4.1. Estimate (74) also yields that the error $\|z^\sigma - z_h^\sigma\|_\Omega$ converges to order p, which is suboptimal. Optimality for both the σ- and u-components can be recovered by considering polynomial interpolation of order $(p-1)$ for the σ-component, but this procedure can make the implementation more cumbersome. Moreover, numerical experiments on structured and unstructured meshes for smooth solutions indicate that $\|z^\sigma - z_h^\sigma\|_\Omega$ often converges to optimal order when considering equal-order interpolation for the σ- and u-components.

4.3. **Examples**

In this section we apply the theoretical results of §4.2 to the three Friedrichs' systems endowed with the 2×2 block structure discussed in §4.1.

4.3.1. *Advection–diffusion–reaction*

Set $z^\sigma = \sigma$ and $z^u = u$. Clearly (67) holds since $|\beta\cdot\nabla w_h^u| \lesssim \|\nabla w_h^u\|$. Let $\alpha > 0$ and take

$$\mathcal{S}_F^{uu} = \alpha h_F^2(|\beta\cdot n_F| + \tfrac{\epsilon}{h_F}), \qquad \mathcal{M}_F^{uu} = |\beta\cdot n| + \tfrac{\epsilon}{h_F}. \tag{82}$$

The boundary operator $\mathcal{M}_F^{uu}$ is designed in such a way that the boundary operator relevant to the pure advection–reaction limit is recovered as $\epsilon \to 0$. If $\epsilon \sim 1$, $\beta \in [\mathcal{C}^{0,1}(\Omega)]^d$, and the exact solution is smooth enough,
$$\|u - u_h\|_\Omega + h\|\nabla(u - u_h)\|_\Omega + h\|\sigma - \sigma_h\|_\Omega \lesssim h^{p+1}\|(\sigma, u)\|_{[H^p(\Omega)]^d \times H^{p+1}(\Omega)}, \tag{83}$$
and if $\epsilon \ll h$ and the exact solution is smooth enough,

$$\|u - u_h\|_\Omega + \|\mathfrak{h}^{\frac{1}{2}}\beta\cdot\nabla(u - u_h)\|_\Omega + \|\sigma - \sigma_h\|_\Omega \lesssim h^{p+\frac{1}{2}}\|(\sigma, u)\|_{[H^{p+1}(\Omega)]^{d+1}}. \tag{84}$$

Comparing with the estimate (53), we observe that the optimal convergence of $\|\nabla\cdot(\sigma-\sigma_h)\|_\Omega$ is lost and that $\|\sigma-\sigma_h\|_\Omega$ converges only to order p (instead of $p+\frac{1}{2}$).

4.3.2. *Linear elasticity*

Set $z^\sigma=(\overline{\sigma},p)$ and $z^u=u$. Clearly (67) holds since $C=0$. Let $\alpha>0$ and $\eta>0$ and take

$$\mathcal{S}_F^{uu}=\alpha h_F\mathcal{I}_d,\qquad \mathcal{M}_F^{uu}=\eta h_F^{-1}\mathcal{I}_d. \tag{85}$$

Then, if the exact solution is smooth enough,

$$\|u-u_h\|_\Omega+h\|\nabla(u-u_h)\|_\Omega+h\|p-p_h\|_\Omega+h\|\sigma-\sigma_h\|_\Omega\lesssim h^{p+1}\|(\overline{\sigma},p,u)\|_{[H^p(\Omega)]^{d^2+1}\times[H^{p+1}(\Omega)]^d}, \tag{86}$$

Comparing with the estimate (55), we observe that the optimal convergence of $\|\nabla\cdot(\sigma-\sigma_h)\|_\Omega$ is lost and that $\|\sigma-\sigma_h\|_\Omega$ and $\|p-p_h\|_\Omega$ converge only to order p (instead of $p+\frac{1}{2}$).

4.3.3. *Maxwell's equations in the elliptic regime*

Set $z^\sigma=H$ and $z^u=E$. Clearly (67) holds since $C=0$. Let $\alpha>0$ and $\eta>0$ and take

$$\mathcal{S}_F^{uu}=\alpha h_F\mathcal{N}_F^t\mathcal{N}_F,\qquad \mathcal{M}_F^{uu}=\eta h_F^{-1}\mathcal{N}^t\mathcal{N}. \tag{87}$$

Then, if the exact solution is smooth enough,

$$\|E-E_h\|_\Omega+h\|\nabla\times(E-E_h)\|_\Omega+h\|H-H_h\|_\Omega\lesssim h^{p+1}\|(H,E)\|_{[H^p(\Omega)]^3\times[H^{p+1}(\Omega)]^3}. \tag{88}$$

Comparing with the estimate (57), we observe that the optimal convergence of $\|\nabla\times(H-H_h)\|_\Omega$ is lost and that $\|H-H_h\|_\Omega$ converges only to order p (instead of $p+\frac{1}{2}$).

5. Numerical results

All the numerical experiments are carried out using FreeFem++ [14]. We first consider test cases with analytical solutions to illustrate the convergence analysis and then test cases with rough solutions to illustrate how the present finite element method is suitable to control oscillations. The stabilization parameter for $\mathcal{M}_F$ is set to 1 and those for $\mathcal{S}_F$ to 10^{-2}. Although a systematic investigation to optimize the values of jump penalty parameters goes beyond the present scope, we observe that setting them to 10^{-2} leads to a fairly efficient choice for two-dimensional problems and polynomials orders up to 2; see, e.g., [6] for further discussion.

5.1. **Convergence rates for smooth solutions**

We consider the four examples of Friedrichs' systems presented in §2.2. The data and right-hand side are chosen to yield the following exact solutions on the unit square:

- Advection–reaction: $\mu=1$, $\beta=(1,0)^t$, $u(x,y)=\arctan(\frac{(y-0.5)}{0.1})\exp(-\mu x)$, and homogeneous Dirichlet boundary conditions enforced on the line $\{x=0\}$.
- Advection–diffusion–reaction: $\mu=1$, $\beta=(1,0)^t$, $u(x,y)=\sin(\pi x)\sin(\pi y)$, and homogeneous Dirichlet boundary conditions enforced on u.
- Linear elasticity: $\gamma_1=\gamma_2=1$, $u_1(x,y)=u_2(x,y)=\sin(\pi x)\sin(\pi y)$, and homogeneous Dirichlet boundary conditions enforced on u.
- Maxwell's equations (two-dimensional setting): $\mu=1$, $\sigma=1$, $E(x,y)=\sin(2\pi x)\sin(2\pi y)$, $H(x,y)=2\pi(\sin(2\pi x)\cos(2\pi y),\sin(2\pi y)\cos(2\pi x))^t$, and homogeneous Dirichlet boundary conditions enforced on E.

h	$h^{\frac{1}{2}}\times(51)$	$h^{\frac{1}{2}}\times(53)$	$h^{\frac{1}{2}}\times(55)$	$h^{\frac{1}{2}}\times(57)$	(83)	(86)	(88)
2^{-3}	3.2e-2	1.4e-1	4.2e-1	1.6e-0	3.4e-2	7.8e-2	2.0e-1
2^{-4}	7.7e-3	3.7e-2	1.2e-1	2.9e-1	8.4e-3	1.5e-2	3.6e-2
2^{-5}	1.8e-3	8.5e-3	2.3e-2	6.1e-2	1.7e-3	3.1e-3	7.5e-3
2^{-6}	3.9e-4	2.1e-3	6.4e-3	1.5e-2	4.8e-4	8.4e-4	1.8e-3
2^{-7}	1.1e-4	5.1e-4	1.6e-4	3.8e-3	1.1e-4	2.0e-4	4.6e-4

TABLE 1. Convergence results for $p = 1$ on unstructured meshes. The number in the first row refers to the equation number of the corresponding estimate

h	$h^{\frac{1}{2}}\times(51)$	$h^{\frac{1}{2}}\times(53)$	$h^{\frac{1}{2}}\times(55)$	$h^{\frac{1}{2}}\times(57)$	(83)	(86)	(88)
2^{-3}	8.5e-3	1.2e-1	4.2e-1	1.5e-0	3.2e-2	8.2e-2	1.6e-1
2^{-4}	1.8e-3	1.7e-2	6.2e-2	2.1e-1	6.2e-3	1.2e-2	2.3e-2
2^{-5}	3.2e-4	2.7e-3	9.4e-3	3.0e-2	1.3e-3	2.0e-3	4.0e-3
2^{-6}	5.6e-5	4.2e-4	1.5e-3	4.5e-3	2.6e-4	4.1e-4	7.7e-4
2^{-7}	9.7e-6	7.1e-5	2.5e-4	6.9e-4	5.6e-5	8.9e-5	1.7e-4

TABLE 2. Convergence results for $p = 1$ on structured meshes. The number in the first row refers to the equation number of the corresponding estimate

h	$h^{\frac{1}{2}}\times(51)$	$h^{\frac{1}{2}}\times(53)$	$h^{\frac{1}{2}}\times(55)$	$h^{\frac{1}{2}}\times(57)$	(83)	(86)	(88)
2^{-3}	5.3e-3	1.6e-2	4.9e-2	2.5e-1	5.4e-3	1.1e-2	4.6e-2
2^{-4}	2.2e-3	1.9e-3	5.5e-3	2.9e-2	7.1e-4	1.5e-3	6.3e-3
2^{-5}	9.4e-5	2.1e-4	6.3e-4	3.4e-3	9.0e-5	2.0e-4	7.2e-4
2^{-6}	2.0e-5	2.7e-5	8.5e-5	4.2e-4	1.1e-5	2.4e-5	8.7e-5

TABLE 3. Convergence results for $p = 2$ on unstructured meshes. The number in the first row refers to the equation number of the corresponding estimate

Tables 1 and 2 present convergence results obtained with $p = 1$ and the two stabilization techniques discussed in §3 and §4. To better appreciate the convergence orders, the errors associated with the method designed in §3 have been multiplied by a factor $h^{\frac{1}{2}}$ so that all the errors in Tables 1 and 2 should theoretically be divided by a factor of 4 from one mesh to the next finer one. All the errors obtained on unstructured meshes (Table 1) match theoretical predictions. When working on structured meshes (Table 2), a super-convergence phenomenon by a factor of $h^{\frac{1}{2}}$ is observed for the estimates derived in §3. This can be linked to the fact that when using uniform meshes in one space dimension, the stabilization parameter can be chosen to yield a finite difference scheme of higher order on a 5-point stencil.

Tables 3 and 4 present convergence results obtained with $p = 2$ and the two stabilization techniques discussed in §3 and §4. All the errors in Tables 3 and 4 should theoretically be divided by a factor of 8 from one mesh to the next finer one. On unstructured meshes (Table 3), numerical results match theoretical predictions. For the advection–reaction equation, the overall convergence order is correct, despite some irregularities on coarser meshes. On structured meshes (Table 4), numerical results also match theoretical predictions. The super-convergence phenomenon is much less pronounced than for $p = 1$.

h	$h^{\frac{1}{2}}\times(51)$	$h^{\frac{1}{2}}\times(53)$	$h^{\frac{1}{2}}\times(55)$	$h^{\frac{1}{2}}\times(57)$	(83)	(86)	(88)
2^{-3}	2.2e-3	1.5e-2	8.8e-2	2.6e-1	8.8e-3	1.6e-2	5.7e-2
2^{-4}	2.2e-4	1.8e-3	1.1e-2	3.2e-2	1.2e-3	2.4e-3	9.2e-3
2^{-5}	1.9e-5	2.1e-4	1.3e-3	3.9e-3	1.5e-4	3.0e-4	1.3e-3
2^{-6}	1.8e-6	2.5e-5	1.5e-4	4.6e-4	1.7e-5	3.6e-5	1.6e-4

TABLE 4. Convergence results for $p = 2$ on structured meshes. The number in the first row refers to the equation number of the corresponding estimate

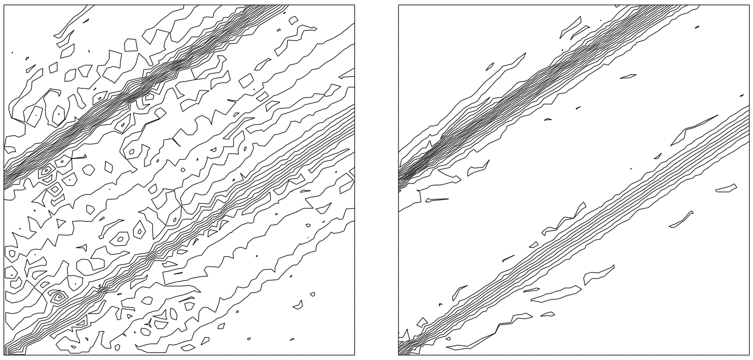

FIGURE 1. Advection–reaction: approximate solution obtained with CFEM without (left) and with (right) stabilization

5.2. Controlling oscillations in rough solutions

For the four Friedrichs' systems, we now consider geometries and data leading to rough solutions producing oscillations if approximated by a CFEM without stabilization. The test cases are the following:

- Advection–reaction: Ω is the unit square, $\mu = 0$, $\beta = (\frac{3}{4}, \frac{1}{2})^t$, $u(0, y) = \arctan(\frac{(y-0.5)}{0.01})$, and $u(x, 0) = 0$. Observe that the inflow data is discontinuous at the origin.
- Advection–diffusion–reaction: Ω is an L–shaped domain, homogeneous Dirichlet boundary conditions are enforced, $\mu = 0$, $\beta = (1, 0)^t$, $\epsilon = 1$, and $f = 100\exp(-100((x - 0.5)^2 + (y - 0.5)^2))$.
- Linear elasticity: Ω is an L–shaped domain, homogeneous Dirichlet boundary conditions are enforced on the displacement, $\gamma_1 = \gamma_2 = 1$, $f_1 = 50\exp(-50((x - 0.5)^2 + (y - 0.5)^2))$, and $f_2 = 0$.
- Maxwell's equation in the diffusive regime: Ω is the unit square, homogeneous Dirichlet boundary conditions are enforced on the electric field, $\mu = 1$, $\sigma = 1$, $f = 750\exp(-750((x - 0.5)^2 + (y - 0.5)^2))$, and $g = 0$.

Figure 1 compares the approximate solution obtained with CFEM without (left) and with (right) stabilization for the advection–reaction equation. As expected, global oscillations are eliminated by the SCFEM; as for all stabilized finite element methods, spurious oscillations remain in the vicinity of layers.

Figure 2 compares the approximate solution obtained with CFEM without stabilization (left), with stabilization on σ and u (center), and with stabilization on u only (right) for the advection–diffusion–reaction equation. The σ-components computed with CFEM without stabilization exhibit some oscillations in the lower part of the domain. Furthermore, owing to the sharp variations in the data f near the point $(\frac{1}{2}, \frac{1}{2})$ yielding insufficient regularity in the σ-components, the SCFEM with stabilization on u produces slightly better results than the SCFEM with stabilization on σ and u.

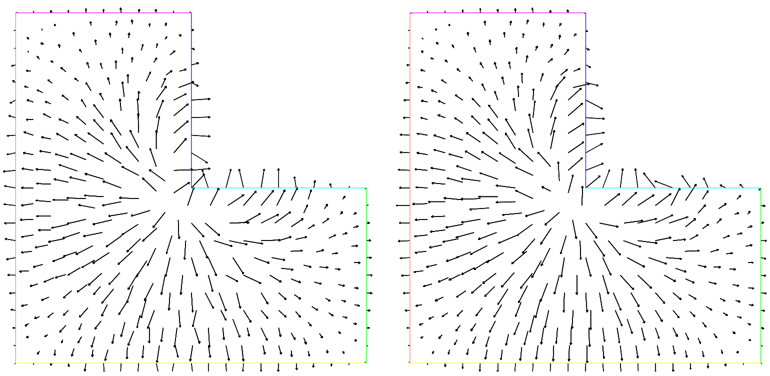

FIGURE 2. Advection–diffusion–reaction: approximate solution (σ_1, σ_2) obtained with CFEM without stabilization (left), with stabilization on σ and u (center), and with stabilization on u only (right)

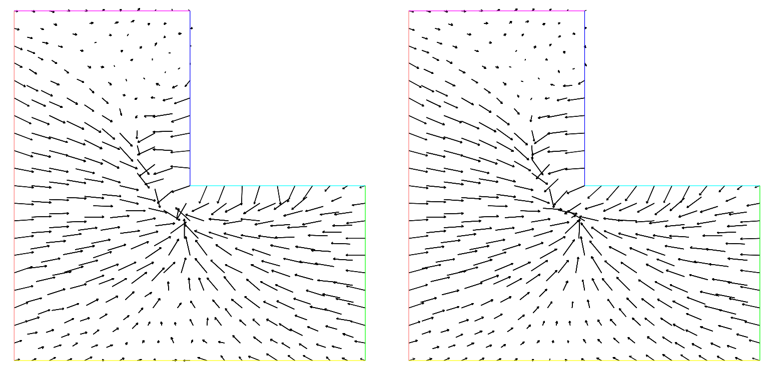

FIGURE 3. Linear elasticity: approximate solution $(\sigma_{11}, \sigma_{12})$ obtained with CFEM without stabilization (left), with stabilization on σ and u (center), and with stabilization on u only (right)

Figure 3 compares the approximate solution obtained with CFEM without stabilization (left), with stabilization on σ and u (center), and with stabilization on u only (right) for the linear elasticity equations. The σ-components computed with CFEM without stabilization exhibit some oscillations in the lower part of the domain. Furthermore, the two versions of the SCFEM produce similar results since the data f has now smoother variations than that considered in the previous case.

Figure 4 compares the approximate solution obtained with CFEM without stabilization (left), with stabilization on H and E (center), and with stabilization on E only (right) for Maxwell's equations in the elliptic regime. The magnetic field produced by the CFEM without stabilization is polluted by oscillations, while the two versions of the SCFEM yield similar and acceptable results.

6. Conclusion

In this paper we have shown that the SCFEM provides a viable alternative to approximate Friedrichs' systems with respect to existing methods such as the DGM. The convergence analysis yields similar error estimates,

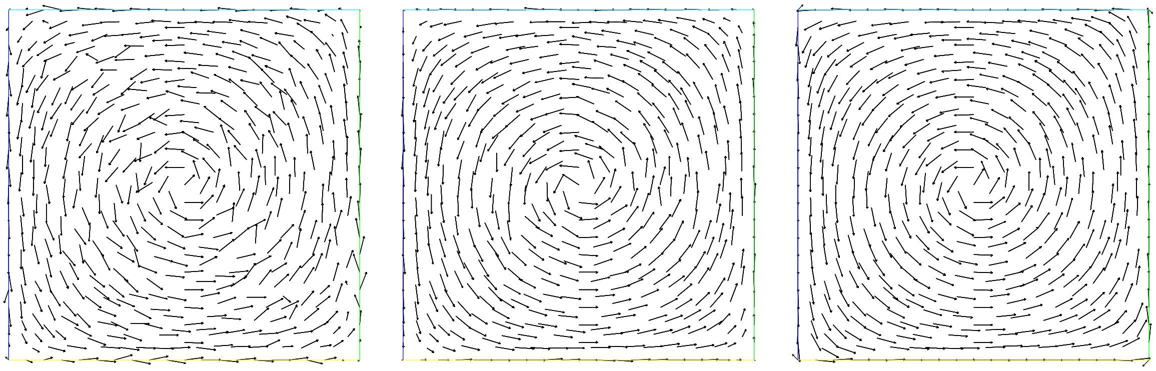

FIGURE 4. Maxwell's equations in the elliptic regime: approximate solution (H_1, H_2) obtained with CFEM without stabilization (left), with stabilization on H and E (center), and with stabilization on E only (right)

while for low-order elements, the SCFEM yields a significant reduction of the number of nonzero entries in the stiffness matrix compared to the DGM. For elliptic-like PDE's, the mixed form has to be considered. Two stabilization strategies have been proposed yielding different convergence orders for the primal variable and its flux. The choice between the two strategies can be driven by the regularity of the exact solution and cost considerations since the demand on memory is much lighter when the flux is not stabilized.

References

[1] I. Babuška. The finite element method with penalty. *Math. Comp.*, 27:221–228, 1973.

[2] I. Babuška and M. Zlámal. Nonconforming elements in the finite element method with penalty. *SIAM J. Numer. Anal.*, 10(5):863–875, 1973.

[3] E. Burman. A unified analysis for conforming and nonconforming stabilized finite element methods using interior penalty. *SIAM J. Numer. Anal.*, 2005. In press.

[4] E. Burman and P. Hansbo. Edge stabilization for Galerkin approximations of convection–diffusion–reaction problems. *Comput. Methods Appl. Mech. Engrg.*, 193:1437–1453, 2004.

[5] E. Burman and P. Hansbo. Edge stabilization for Galerkin approximations of the generalized Stokes' problem. *Comput. Methods Appl. Mech. Engrg.*, 2005. In press.

[6] E. Burman and B. Stamm. Discontinuous and continuous finite element methods with interior penalty for hyperbolic problems. *J. Numer. Math*, 2005. submitted (EPFL-IACS report 17.2005).

[7] 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.

[8] L. El Alaoui and A. Ern. Residual and hierarchical a posteriori estimates for nonconforming mixed finite element methods. *M2AN Math. Model. Numer. Anal.*, 38(6):903–929, 2004.

[9] A. Ern and J.-L. Guermond. *Theory and Practice of Finite Elements*, volume 159 of *Applied Mathematical Sciences*. Springer-Verlag, New York, NY, 2004.

[10] A. Ern and J.-L. Guermond. Discontinuous Galerkin methods for Friedrichs' systems. I. General theory. *SIAM J. Numer. Anal.*, 2005. In press.

[11] A. Ern and J.-L. Guermond. Discontinuous Galerkin methods for Friedrichs' systems. II. Second-order PDEs. *SIAM J. Numer. Anal.*, 2005. submitted (internal report available at `cermics.enpc.fr/reports/CERMICS-2005`).

[12] A. Ern and J.-L. Guermond. Evaluation of the condition number in linear systems arising in finite element approximations. *M2AN Math. Model. Numer. Anal.*, 2005. In press.

[13] K.O. Friedrichs. Symmetric positive linear differential equations. *Comm. Pure Appl. Math.*, 11:333–418, 1958.

[14] F. Hecht and O. Pironneau. *FreeFEM++ Manual*. Laboratoire Jacques-Louis Lions, University Paris VI, 2005.

[15] R.H.W. Hoppe and B. Wohlmuth. Element-oriented and edge-oriented local error estimators for non-conforming finite element methods. *M2AN Math. Model. Numer. Anal.*, 30:237–263, 1996.

[16] M. Jensen. *Discontinuous Galerkin Methods for Friedrichs Systems with Irregular Solutions*. PhD thesis, University of Oxford, 2004.

[17] C. Johnson and J. Pitkäranta. An analysis of the discontinuous Galerkin method for a scalar hyperbolic equation. *Math. Comp.*, 46(173):1–26, 1986.

[18] O. Karakashian and F. Pascal. A-posteriori error estimates for a discontinuous Galerkin approximation of second order elliptic problems. *SIAM J. Numer. Anal.*, 41(6):2374–2399, 2003.

[19] P. Lesaint. *Sur la résolution des systèmes hyperboliques du premier ordre par des méthodes d'éléments finis.* PhD thesis, University of Paris VI, 1975.

[20] P. Lesaint and P.-A. Raviart. On a finite element method for solving the neutron transport equation. In C. de Boors, editor, *Mathematical aspects of Finite Elements in Partial Differential Equations*, pages 89–123. Academic Press, 1974.