ESAIM: M2AN
M2AN, Vol. 41, N° 1, 2007, pp. 55–76
DOI: 10.1051/m2an:2007007

ESAIM: Mathematical Modelling and Numerical Analysis
www.edpsciences.org/m2an

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. Stability is achieved by penalizing the jumps across mesh interfaces of the normal derivative 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 and reducing the number of nonzero entries in the stiffness matrix is also proposed and analyzed. Finally, numerical results are presented to illustrate the theoretical analysis.

Mathematics Subject Classification. 65N30, 65N12, 74S05, 78M10, 76R99, 35F15.

Received: December 21, 2005.

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 [18], has made considerable progress in the last decades; see, *e.g.*, Jensen's thesis [21]. Recently, the theory was revisited by Ern and Guermond [13] 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 [24–26] where the discontinuous Galerkin method (DGM) was analyzed. The convergence estimate was subsequently improved by Johnson and Pitkäranta [22] and Falk and Richter [17], and more recently a thorough systematic analysis was proposed by Ern and Guermond [13, 14]. 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.

Keywords and phrases. Finite elements, interior penalty, stabilization methods, Friedrichs' systems, first-order PDE's.

[1] Department of Mathematics, École Polytechnique Fédérale de Lausanne, Switzerland. erik.burman@epfl.ch

[2] CERMICS, École des ponts, ParisTech, Champs sur Marne, 77455 Marne la Vallée Cedex 2, France.
ern@cermics.enpc.fr

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], Baker [3], and Douglas and Dupont [10] on interior penalty methods for elliptic problems, the analysis of face penalty finite element methods has been recently extended to advection-diffusion equations [5,6]. 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 (assuming that each vertex is shared on the average by six triangles and that the number of triangles is twice the number of vertices), 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, when working with first-order finite elements, and this number scales as 46, 100, and 288 times the number of mesh vertices for CFEM, SCFEM, and DGM, respectively, when working with second-order finite elements. Another technique for solving systems of first-order PDE's was proposed in [9]. This is a least-squares technique that results in a symmetric system at the price of a squared condition number.

The goal of this work is to generalize the face penalty technique of [5,6] in order to approximate satisfactorily Friedrichs' systems using continuous finite elements. In Section 2 the main results on Friedrichs' systems derived in [13, 14] are briefly restated. The reader familiar with this material can directly jump to Section 3 where the SCFEM with face penalty is designed and analyzed. In Section 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 can reduce the condition number of the stiffness matrix from a multiple of h^{-2} to a multiple of h^{-1} (see, *e.g.* [16]), and it can be the only viable formulation whenever complex constitutive laws such as those of viscoelastic fluids are considered (see, *e.g.* [4]). Finally, in Section 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

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

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

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

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

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

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 $(v, w) \in W \times W$,

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

The operator D is self-adjoint and is a boundary operator in the sense that $[\mathfrak{D}(\Omega)]^m \subset \mathrm{Ker}(D)$.

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 $T : V \to L$ 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

$$\langle Mw, w\rangle_{W',W} \geq 0 \text{ for all } w \text{ in } W, \tag{M1}$$
$$W = \mathrm{Ker}(D - M) + \mathrm{Ker}(D + M). \tag{M2}$$

Assumption (M2) implies that $\mathrm{Ker}(D) = \mathrm{Ker}(M)$ so that M is also a boundary operator. For all $(v, w) \in W \times W$, let

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

In this framework, the main result proven in [13] is the following:

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

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

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 $n = (n_1, \ldots, n_d)^t$ 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 Friedrichs' systems considered 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 [15] for more details.

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 the jumps across mesh interfaces of the normal derivative are penalized for some components of the discrete solution. The main results are Theorem 3.1 and estimate (30) which yield a suboptimal estimate (of order $\frac{1}{2}$) for the L^2-norm and an optimal estimate for the graph norm if the mesh is quasi-uniform.

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 $A \leq cB$ 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. For $F \in \mathcal{F}_h^{\partial}$, let $T(F)$ denote the mesh element of which F is a face. 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 jump of its normal derivative as

$$[\![\nabla v]\!]_F = (\nabla v|_{T_1(F)} - \nabla v|_{T_2(F)}) \cdot n_F. \tag{7}$$

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

$$V_h = \{v_h \in \mathfrak{C}^0(\Omega);\ \forall T \in \mathcal{T}_h,\ v_h|_T \in \mathbb{P}_p\}, \tag{8}$$

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 $T \in \mathcal{T}_h$,

$$\|\nabla v_h\|_T \lesssim h_T^{-1} \|v_h\|_T, \tag{9}$$

$$\|v_h\|_F \lesssim h_T^{-\frac{1}{2}} \|v_h\|_T, \quad \forall F \subset \partial T. \tag{10}$$

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 $v, w \in [L^2(F)]^m$,

$$0 \leq \mathcal{M}_F \leq \mathcal{I}_m, \tag{11}$$

$$(\mathcal{M} v = \mathcal{D} v) \implies (\mathcal{M}_F v = \mathcal{D} v), \tag{12}$$

$$|((\mathcal{M}_F - \mathcal{D})v, w)_{L,F}| \lesssim |v|_{M,F} \|w\|_F, \tag{13}$$

$$|((\mathcal{M}_F + \mathcal{D})v, w)_{L,F}| \lesssim \|v\|_F |w|_{M,F}, \tag{14}$$

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 normal derivative 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

$$\mathcal{S}_F \text{ is symmetric,} \tag{15}$$

$$h_F^2 |\mathcal{D}_F| \lesssim \mathcal{S}_F \lesssim h_F^2 \mathcal{I}_m, \tag{16}$$

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

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

$$a_h(v,w) = (Tv, w)_\Omega + \sum_{F \in \mathcal{F}_h^\partial} \tfrac{1}{2}((\mathcal{M}_F - \mathcal{D})v, w)_F + \sum_{F \in \mathcal{F}_h^{\mathrm{i}}} (\mathcal{S}_F [\![\nabla v]\!], [\![\nabla w]\!])_F. \tag{17}$$

Then, to approximate the solution z of (4), the following problem is considered:

$$\text{Find } z_h \in W_h \text{ such that } a_h(z_h, y_h) = (f, y_h)_\Omega, \ \forall y_h \in W_h. \tag{18}$$

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 [13]. 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 $W(h)$,

$$|\!|\!| v |\!|\!|^2 = \|v\|_\Omega^2 + \sum_{F \in \mathcal{F}_h^\partial} |v|_{M,F}^2 + \sum_{F \in \mathcal{F}_h^{\mathrm{i}}} |[\![\nabla v]\!]|_{S,F}^2 + \|\mathfrak{h}^{\frac{1}{2}} A v\|_\Omega^2. \tag{19}$$

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

$$a_h(v,w) = (v, \tilde{T} w)_\Omega + \sum_{F \in \mathcal{F}_h^\partial} \tfrac{1}{2}((\mathcal{M}_F + \mathcal{D})v, w)_F + \sum_{F \in \mathcal{F}_h^{\mathrm{i}}} (\mathcal{S}_F [\![\nabla v]\!], [\![\nabla w]\!])_F. \tag{20}$$

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

$$a_h(v_h, v_h) \gtrsim \|v_h\|_\Omega^2 + \sum_{F \in \mathcal{F}_h^\partial} |v_h|_{M,F}^2 + \sum_{F \in \mathcal{F}_h^{\mathrm{i}}} |[\![\nabla v_h]\!]|_{S,F}^2, \tag{21}$$

which shows that the bilinear form a_h is at least L-coercive on W_h. To control the last term in (19), a sharper stability result is needed. This is the purpose of the following lemma. The proof combines the ideas of [5, 6] for the SCFEM approximation of scalar transport equations and those of [13] for the DGM approximation of Friedrichs' systems.

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:*

$$\forall v_h \in W_h, \qquad \sup_{w_h \in W_h \setminus \{0\}} \frac{a_h(v_h, w_h)}{|\!|\!| w_h |\!|\!|} \gtrsim |\!|\!| v_h |\!|\!|. \tag{22}$$

Proof. Let $v_h \in W_h$. Owing to (21), the first three terms in the norm $|\!|\!| v_h |\!|\!|$ defined by (19) are already controlled, so that it only remains to control the last term.

(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

$$\|\zeta_h'\|_T \lesssim \min(\|v_h\|_T, h_T^{\frac{1}{2}}\|\mathfrak{h}^{\frac{1}{2}}Av_h\|_T + h_T^{\frac{1}{2}}\|v_h\|_T). \tag{23}$$

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 [5, 11, 12, 20, 23]:

$$\|\pi_h w_h\|_T \lesssim \|w_h\|_{\Delta_1(T)}, \tag{24}$$

$$\|w_h - \pi_h w_h\|_T \lesssim \sum_{F\in\Delta_2(T)} h_F^{\frac{1}{2}}\|[\![w_h]\!]\|_F, \tag{25}$$

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 (9), (10), (11) (upper bound), (16) (upper bound), (23), and (24), it is inferred that

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

(ii) Observe that

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

We now bound the remainder terms R_1 to R_4. Using (23) (first bound) and (24) yields

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

Using (13), (10), (23) (second bound), (24), 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 (16) (upper bound), (9), (10), (24), (23) (second bound), and Young's inequality yields

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

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 (9) yields

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

Using (25) yields

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

For all $F \in \mathcal{F}_h^{\mathrm{i}}$, using the continuity of $\mathfrak{h}$ it is inferred that

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

Now, (16) (lower bound) is used to control $\|\mathfrak{h}[\![Av_h]\!]\|_F$. Indeed, since v_h and the fields $\mathcal{A}^k$ are continuous,

$$\begin{aligned}\|\mathfrak{h}[\![Av_h]\!]\|_F^2 \lesssim h_F^2([\![Av_h]\!],[\![Av_h]\!])_F &= h_F^2(\mathcal{D}_F[\![\nabla v_h]\!],\mathcal{D}_F[\![\nabla v_h]\!])_F \\ &\lesssim h_F^2(|\mathcal{D}_F|[\![\nabla v_h]\!],[\![\nabla v_h]\!])_F \lesssim |[\![\nabla v_h]\!]|_{S,F}^2.\end{aligned}$$

This yields

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

Collecting the above bounds, using (21), and taking γ small enough leads to

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

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

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

$$|\!|\!|v|\!|\!|_*^2 = |\!|\!|v|\!|\!|^2 + \sum_{T\in\mathcal{T}_h} [h_T^{-1}\|v\|_T^2 + \|v\|_{\partial T}^2]. \tag{26}$$

Then, the following holds:

$$\forall (v,w) \in W(h)\times W(h), \quad a_h(v,w) \lesssim |\!|\!|v|\!|\!|_* |\!|\!|w|\!|\!|. \tag{27}$$

Proof. We bound the three terms in the right-hand side of (20). For the first term,

$$|(v,\tilde{T}w)_\Omega| \lesssim \|v\|_\Omega \|w\|_\Omega + \sum_{T\in\mathcal{T}_h} h_T^{-\frac{1}{2}} \|v\|_T h_T^{\frac{1}{2}} \|Aw\|_T \lesssim |\!|\!|v|\!|\!|_* |\!|\!|w|\!|\!|.$$

For the second term, (14) 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* (18)*. If $z \in [H^2(\Omega)]^m$, then,*

$$\forall y_h \in W_h, \quad a_h(z - z_h, y_h) = 0. \tag{28}$$

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 (12) 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 yield the following:

Theorem 3.1 (convergence)**.** *Let z solve* (4) *and let z_h solve* (18)*. Assume $z \in [H^2(\Omega)]^m$. Then, under the assumption of Lemma* 3.1*,*

$$|||z - z_h||| \lesssim \inf_{v_h \in W_h} |||z - v_h|||_*. \tag{29}$$

Using standard interpolation properties in W_h, it is inferred that

$$|||z - z_h||| \lesssim h^{p+\frac{1}{2}} \|z\|_{[H^{p+1}(\Omega)]^m}, \tag{30}$$

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 obtained with other stabilization methods like Galerkin/Least-Squares, subgrid viscosity, or DGM.

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 [13] that $\lim_{h\to 0} \|z - z_h\|_\Omega = 0$.

3.3. **Examples**

In this section we apply the theoretical results of Section 3.2 to the four examples of Friedrichs' systems for which the approximation by DGM is discussed in [13, 14]. For brevity, proofs are omitted.

3.3.1. *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

$$\mu u + \beta\cdot\nabla u = f \tag{31}$$

falls into the category of Friedrichs' systems by setting $m = 1$, $\mathcal{K} = \mu$ and $\mathcal{A}^k = \beta^k$ for $k \in \{1, \ldots, d\}$. The graph space is $W = \{w \in L^2(\Omega);\ \beta\cdot\nabla w \in L^2(\Omega)\}$. 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, an admissible boundary condition is to enforce homogeneous Dirichlet conditions at the inflow boundary [13]. The boundary operators D and M admit the representation (5)–(6) with

$$\mathcal{D} = \beta\cdot n, \qquad \mathcal{M} = |\beta\cdot n|. \tag{32}$$

Let $\alpha > 0$ and take

$$\mathcal{S}_F = \alpha h_F^2 |\beta\cdot n_F|, \qquad \mathcal{M}_F = |\beta\cdot n_F|. \tag{33}$$

Then, (11)–(16) hold. Hence, if $\beta \in [\mathcal{C}^{0,\frac{1}{2}}(\Omega)]^d$ and the exact solution is smooth enough,

$$\|u - u_h\|_\Omega + \|\mathfrak{h}^{\frac{1}{2}}\beta\cdot\nabla(u - u_h)\|_\Omega \lesssim h^{p+\frac{1}{2}} \|u\|_{H^{p+1}(\Omega)}. \tag{34}$$

3.3.2. *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

$$\begin{cases} \sigma + \nabla u = 0, \\ \mu u + \nabla\cdot\sigma + \beta\cdot\nabla u = f, \end{cases} \tag{35}$$

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

$$\mathcal{K} = \left[\begin{array}{c:c} \mathcal{I}_d & 0 \\ \hdashline 0 & \mu \end{array}\right], \qquad \mathcal{A}^k = \left[\begin{array}{c:c} 0 & e^k \\ \hdashline (e^k)^t & \beta^k \end{array}\right], \tag{36}$$

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 $W = H(\mathrm{div};\Omega)\times H^1(\Omega)$. An admissible boundary condition is to enforce a Dirichlet condition on u (Neumann

and Robin boundary conditions can be treated as well; see [13]). Then, the boundary operators D and M admit the representation (5)–(6) with

$$\mathcal{D} = \left[\begin{array}{c|c} 0 & n \\ \hline n^t & \beta\cdot n \end{array}\right], \qquad \mathcal{M} = \left[\begin{array}{c|c} 0 & -n \\ \hline n^t & 0 \end{array}\right]. \tag{37}$$

Remark 3.3. Using a Poincaré inequality, one can show that well-posedness still holds if $\mu(x) - \frac{1}{2}\nabla\cdot\beta(x) \geq 0$ *a.e.* in Ω.

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

$$\mathcal{S}_F = h_F^2 \left[\begin{array}{c|c} \alpha_1 n_F\otimes n_F & 0 \\ \hline 0 & \alpha_2 \end{array}\right], \qquad \mathcal{M}_F = \left[\begin{array}{c|c} 0 & -n_F \\ \hline n_F^t & \eta \end{array}\right]. \tag{38}$$

Then, (11)–(16) hold. Hence, if $\beta \in [\mathcal{C}^{0,\frac{1}{2}}(\Omega)]^d$ and the exact solution is smooth enough,

$$\|u-u_h\|_\Omega + \|\mathfrak{h}^{\frac{1}{2}}\nabla(u-u_h)\|_\Omega + \|\sigma-\sigma_h\|_\Omega + \|\mathfrak{h}^{\frac{1}{2}}\nabla\cdot(\sigma-\sigma_h)\|_\Omega \lesssim h^{p+\frac{1}{2}}\|(\sigma,u)\|_{[H^{p+1}(\Omega)]^{d+1}}. \tag{39}$$

3.3.3. *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

$$\left\{\begin{array}{l} \sigma + p\mathcal{I}_d - \frac{1}{2}(\nabla u + (\nabla u)^t) = 0, \\ \mathrm{tr}(\sigma) + (d+\gamma_1)p = 0, \\ -\frac{1}{2}\nabla\cdot(\sigma+\sigma^t) + \gamma_2 u = f. \end{array}\right. \tag{40}$$

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, (40) falls into the category of Friedrichs' systems by setting $m = d^2 + 1 + d$ and

$$\mathcal{K} = \left[\begin{array}{c|c|c} \mathcal{I}_{d^2} & \mathcal{Z} & 0 \\ \hline (\mathcal{Z})^t & (d+\gamma_1) & 0 \\ \hline 0 & 0 & \gamma_2\mathcal{I}_d \end{array}\right], \qquad \mathcal{A}^k = \left[\begin{array}{c|c|c} 0 & 0 & \mathcal{E}^k \\ \hline 0 & 0 & 0 \\ \hline (\mathcal{E}^k)^t & 0 & 0 \end{array}\right], \tag{41}$$

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 $H_{\overline{\sigma}} = \{\overline{\sigma} \in [L^2(\Omega)]^{d^2}; \nabla\cdot(\sigma+\sigma^t) \in [L^2(\Omega)]^d\}$. An admissible boundary condition is to enforce a Dirichlet condition on u. Then, the boundary operators D and M admit the representation (5)–(6) with

$$\mathcal{D} = \left[\begin{array}{c|c|c} 0 & 0 & \mathcal{H} \\ \hline 0 & 0 & 0 \\ \hline \mathcal{H}^t & 0 & 0 \end{array}\right], \qquad \mathcal{M} = \left[\begin{array}{c|c|c} 0 & 0 & -\mathcal{H} \\ \hline 0 & 0 & 0 \\ \hline \mathcal{H}^t & 0 & 0 \end{array}\right], \tag{42}$$

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 3.4. Using a Korn inequality, one can show that well-posedness still holds if $\gamma_2 \geq 0$ *a.e.* in Ω.

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

$$\mathcal{S}_F = h_F^2 \left[\begin{array}{c|c|c} \alpha_1\mathcal{H}_F\cdot\mathcal{H}_F^t & 0 & 0 \\ \hline 0 & 0 & 0 \\ \hline 0 & 0 & \alpha_2\mathcal{I}_d \end{array}\right], \qquad \mathcal{M}_F = \left[\begin{array}{c|c|c} 0 & 0 & -\mathcal{H}_F \\ \hline 0 & 0 & 0 \\ \hline \mathcal{H}_F^t & 0 & \eta\mathcal{I}_d \end{array}\right], \tag{43}$$

where $\mathcal{H}_F$ is defined similarly to $\mathcal{H}$ by substituting n_F to n. Then, (11)–(16) hold. Hence, if the exact solution is smooth enough,

$$\|u-u_h\|_\Omega + \|\mathfrak{h}^{\frac12}\nabla(u-u_h)\|_\Omega + \|p-p_h\|_\Omega + \|\sigma-\sigma_h\|_\Omega + \|\mathfrak{h}^{\frac12}\nabla\cdot((\sigma+\sigma^t)-(\sigma_h+\sigma_h^t))\|_\Omega \lesssim h^{p+\frac12}\|(\overline{\sigma},p,u)\|_{[H^{p+1}(\Omega)]^{d^2+1+d}}, \quad (44)$$

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

Remark 3.5. Numerical experiments indicate that the above formulation is stable in the incompressible limit as $\gamma_1 \to 0$. However the method becomes more sensitive to the choice of the stabilization parameters. A thorough analysis of the limit case goes beyond the present scope. A SCFEM for the Stokes equations similar to the one proposed here is analyzed in [7].

3.3.4. *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

$$\begin{cases} \mu H + \nabla\times E = f, \\ \sigma E - \nabla\times H = g, \end{cases} \quad (45)$$

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

$$\mathcal{K} = \left[\begin{array}{c|c} \mu\mathcal{I}_3 & 0 \\ \hline 0 & \sigma\mathcal{I}_3 \end{array}\right], \qquad \mathcal{A}^k = \left[\begin{array}{c|c} 0 & \mathcal{R}^k \\ \hline (\mathcal{R}^k)^t & 0 \end{array}\right], \quad (46)$$

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 $W = H(\mathrm{curl};\Omega)\times H(\mathrm{curl};\Omega)$. An admissible boundary condition is to enforce a Dirichlet condition on the tangential component of the electric field. Then, the boundary operators D and M admit the representation (5)–(6) with

$$\mathcal{D} = \left[\begin{array}{c|c} 0 & \mathcal{N} \\ \hline \mathcal{N}^t & 0 \end{array}\right], \qquad \mathcal{M} = \left[\begin{array}{c|c} 0 & -\mathcal{N} \\ \hline \mathcal{N}^t & 0 \end{array}\right], \quad (47)$$

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 $\xi\in\mathbb{R}^3$.

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

$$\mathcal{S}_F = h_F^2\left[\begin{array}{c|c} \alpha_1\mathcal{N}_F^t\mathcal{N}_F & 0 \\ \hline 0 & \alpha_2\mathcal{N}_F^t\mathcal{N}_F \end{array}\right], \qquad \mathcal{M}_F = \left[\begin{array}{c|c} 0 & -\mathcal{N}_F \\ \hline \mathcal{N}_F^t & \eta\mathcal{N}_F^t\mathcal{N}_F \end{array}\right], \quad (48)$$

where $\mathcal{N}_F$ is defined similarly to $\mathcal{N}$ by substituting n_F to n. Then, (11)–(16) hold. Hence, if the exact solution is smooth enough,

$$\|E-E_h\|_\Omega + \|\mathfrak{h}^{\frac12}\nabla\times(E-E_h)\|_\Omega + \|H-H_h\|_\Omega + \|\mathfrak{h}^{\frac12}\nabla\times(H-H_h)\|_\Omega \lesssim h^{p+\frac12}\|(H,E)\|_{[H^{p+1}(\Omega)]^6}. \quad (49)$$

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 [14]. The purpose of this section is to design and analyze a SCFEM where only the jumps of the normal derivative of the z^u-component are penalized. The motivation for using this type of stabilization is to substantially reduce the number of nonzero entries in

the stiffness matrix, thus alleviating considerably memory requirements. The main results are Theorem 4.1 (along with Cor. 4.1) and Theorem 4.2. The key difference with the analysis of Section 3 is that only the graph norm of the u-component (instead of the full graph norm weighted by $\mathfrak{h}^{\frac{1}{2}}$) is controlled. Moreover, an optimal L^2-error estimate for the u-component is derived using elliptic regularity and a duality argument. Furthermore, a singular perturbation of the error estimate is included in the analysis to recover optimal error estimates when the terms associated with elliptic behavior are actually dominated by other first-order derivatives, *e.g.*, for advection-dominated advection-diffusion problems.

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

$$\mathcal{A}^k = \left[\begin{array}{c|c} 0 & \epsilon^{\frac{1}{2}}\mathcal{B}^k \\ \hline \epsilon^{\frac{1}{2}}(\mathcal{B}^k)^t & \mathcal{C}^k \end{array}\right], \tag{50}$$

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 (50), the matrix $\mathcal{D}$ is such that

$$\mathcal{D} = \left[\begin{array}{c|c} 0 & \epsilon^{\frac{1}{2}}\mathcal{D}^{\sigma u} \\ \hline \epsilon^{\frac{1}{2}}(\mathcal{D}^{\sigma u})^t & \mathcal{D}^{uu} \end{array}\right], \tag{51}$$

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

$$\mathcal{M} = \left[\begin{array}{c|c} 0 & -\epsilon^{\frac{1}{2}}\mathcal{D}^{\sigma u} \\ \hline \epsilon^{\frac{1}{2}}(\mathcal{D}^{\sigma u})^t & 0 \end{array}\right]. \tag{52}$$

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

$$\mathcal{M}_F = \left[\begin{array}{c|c} 0 & -\epsilon^{\frac{1}{2}}\mathcal{D}^{\sigma u} \\ \hline \epsilon^{\frac{1}{2}}(\mathcal{D}^{\sigma u})^t & \mathcal{M}_F^{uu} \end{array}\right]. \tag{53}$$

We still assume that the consistency condition (12) holds. However, instead of (11), (13), and (14), we now assume that $\mathcal{M}_F^{uu}$ is symmetric and that

$$\tfrac{\epsilon}{h_F}((\mathcal{D}^{\sigma u})^t\mathcal{D}^{\sigma u})^{\frac{1}{2}} + |\mathcal{D}^{uu}| \lesssim \mathcal{M}_F^{uu} \lesssim (1 + \tfrac{\epsilon}{h_F})\mathcal{I}_{m_u}. \tag{54}$$

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 $\epsilon \ll h$, (54) implies that $\mathcal{M}_F^{uu}$ and $\mathcal{D}^{uu}$ satisfy (11), (13), and (14).

To penalize the jumps of the normal derivative 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

$$\mathcal{S}_F = \left[\begin{array}{c:c} 0 & 0 \\ \hdashline 0 & \mathcal{S}_F^{uu} \end{array}\right]. \tag{55}$$

Instead of (15) and (16), we now assume that $\mathcal{S}_F^{uu}$ is symmetric and that

$$\tfrac{\epsilon}{h_F}((\mathcal{D}_F^{\sigma u})^t\mathcal{D}_F^{\sigma u})^{\frac12} + |\mathcal{D}_F^{uu}| \lesssim \tfrac{1}{h_F^2}\mathcal{S}_F^{uu} \lesssim (1+\tfrac{\epsilon}{h_F})\mathcal{I}_{m_u}. \tag{56}$$

If $\epsilon = 1$, this yields $h_F((\mathcal{D}_F^{\sigma u})^t\mathcal{D}_F^{\sigma u})^{\frac12} \lesssim \mathcal{S}_F^{uu} \lesssim h_F\mathcal{I}_{m_u}$, while if $\epsilon \ll h$, (56) implies that $\mathcal{S}_F^{uu}$ and $\mathcal{D}^{uu}$ satisfy (16).

Owing to the above setting, the bilinear form a_h defined by (17) becomes

$$\begin{aligned} a_h(v,w) = {} & (\mathcal{K}v,w)_\Omega + \epsilon^{\frac12}(Bv^u,w^\sigma)_\Omega + \epsilon^{\frac12}(\tilde{B}v^\sigma,w^u)_\Omega + (Cv^u,w^u)_\Omega \\ & + \sum_{F\in\mathcal{F}_h^\partial}[-\epsilon^{\frac12}(\mathcal{D}^{\sigma u}v^u,w^\sigma)_F + \tfrac12((\mathcal{M}_F^{uu}-\mathcal{D}^{uu})v^u,w^u)_F] + \sum_{F\in\mathcal{F}_h^{\mathrm{i}}}(\mathcal{S}_F^{uu}[\![\nabla v^u]\!],[\![\nabla w^u]\!])_F, \end{aligned} \tag{57}$$

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 (18) with the discrete space W_h unchanged, *i.e.*, equal-order interpolation is used for the u- and the σ-components.

4.2. **Convergence analysis**

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

$$|\!|\!|v|\!|\!|^2 = \|v\|_\Omega^2 + \sum_{F\in\mathcal{F}_h^\partial}|v^u|_{M,F}^2 + \sum_{F\in\mathcal{F}_h^{\mathrm{i}}}|[\![\nabla v^u]\!]|_{S,F}^2 + \|\epsilon^{\frac12}Bv^u\|_\Omega^2 + \|\mathfrak{h}^{\frac12}Cv^u\|_\Omega^2, \tag{58}$$

with the semi-norms $|v^u|_{M,F} = (\mathcal{M}_F^{uu}v^u,v^u)_F^{\frac12}$ and $|[\![\nabla v^u]\!]|_{S,F} = (\mathcal{S}_F^{uu}[\![\nabla v^u]\!],[\![\nabla v^u]\!])_F^{\frac12}$. The triple norm $|\!|\!|\cdot|\!|\!|$ with which we want to control the error is substantially different from that used in Section 3 and defined by (19). Indeed, the full graph norm weighted by $\mathfrak{h}^{\frac12}$ is not present in (58); instead, we now want to control the B-derivatives of the u-component weighted by $\epsilon^{\frac12}$ (*e.g.*, the diffusive flux for an advection-diffusion problem) and the C-derivatives of the u-component weighted by $\mathfrak{h}^{\frac12}$ (*e.g.*, the advective derivative weighted by the same factor). Furthermore, only the jumps of the gradient of the u-component are controlled.

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*

$$\forall T\in\mathcal{T}_h,\ \forall w_h\in W_h,\quad \|Cw_h^u\|_T \lesssim \|Bw_h^u\|_T. \tag{59}$$

Then, the following holds:

$$\forall v_h\in W_h,\qquad \sup_{w_h\in W_h\setminus\{0\}}\frac{a_h(v_h,w_h)}{|\!|\!|w_h|\!|\!|} \gtrsim |\!|\!|v_h|\!|\!|. \tag{60}$$

Proof. Let $v_h\in W_h$. It is clear that the coercivity property (21) 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 $\zeta_h = \pi_h(0,\mathfrak{h}\overline{C}v_h^u)$ where $\overline{C}$ is defined similarly to $\overline{A}$ in the proof of Lemma 3.1. Let us first prove that $|\!|\!|\zeta_h|\!|\!| \lesssim |\!|\!|v_h|\!|\!|$. Proceeding as in the proof of Lemma 3.1, it is inferred that for all $T\in\mathcal{T}_h$,

$$\|\zeta_h\|_T \lesssim \min(\|v_h^u\|_{\Delta_1(T)}, h_T^{\frac12}\|\mathfrak{h}^{\frac12}Cv_h^u\|_{\Delta_1(T)} + h_T^{\frac12}\|v_h^u\|_{\Delta_1(T)}), \tag{61}$$

so that $\|\zeta_h\|_\Omega \lesssim \|v_h\|_\Omega \le |\!|\!| v_h |\!|\!|$. Furthermore, for all $F \in \mathcal{F}_h^\partial$, using (54) (upper bound) leads to $|\zeta_h^u|_{M,F}^2 \lesssim \|\zeta_h^u\|_F^2 + \frac{\epsilon}{h_F}\|\zeta_h^u\|_F^2$, and the first term is bounded using (10) and (61) by

$$\|\zeta_h^u\|_F^2 \lesssim \|v_h^u\|_{\Delta_1(T(F))}^2 + \|\mathfrak{h}^{\frac12} C v_h^u\|_{\Delta_1(T(F))}^2.$$

Moreover, using (10), (24), (59), and (61) yields

$$\tfrac{\epsilon}{h_F}\|\zeta_h^u\|_F^2 \lesssim \epsilon\|\overline{C} v_h^u\|_{\Delta_1(T(F))}^2 \lesssim \|(C-\overline{C})v_h^u\|_{\Delta_1(T(F))}^2 + \epsilon\|C v_h^u\|_{\Delta_1(T(F))}^2 \lesssim \|v_h^u\|_{\Delta_1(T(F))}^2 + \|\epsilon^{\frac12} B v_h^u\|_{\Delta_1(T(F))}^2,$$

since $\epsilon \le 1$. Hence,

$$\sum_{F\in\mathcal{F}_h^\partial} |\zeta_h^u|_{M,F}^2 \lesssim |\!|\!| v_h |\!|\!|^2.$$

Similarly, using (56),

$$\sum_{F\in\mathcal{F}_h^{\mathrm{i}}} |[\![\nabla \zeta_h^u]\!]|_{S,F}^2 \lesssim |\!|\!| v_h |\!|\!|^2.$$

Moreover, using (9), (59), and the fact that $\epsilon \le 1$ yields

$$\|\epsilon^{\frac12} B\zeta_h^u\|_\Omega \lesssim \|\epsilon^{\frac12}\overline{C} v_h^u\|_\Omega \lesssim \|(C-\overline{C})v_h^u\|_\Omega + \|\epsilon^{\frac12} C v_h^u\|_\Omega \lesssim \|v_h^u\|_\Omega + \|\epsilon^{\frac12} B v_h^u\|_\Omega \lesssim |\!|\!| v_h |\!|\!|.$$

Proceeding similarly yields $\|\mathfrak{h}^{\frac12} C\zeta_h^u\|_\Omega \lesssim \|v_h^u\|_\Omega + \|\mathfrak{h}^{\frac12} C v_h^u\|_\Omega \lesssim |\!|\!| v_h |\!|\!|$. Collecting the above bounds, it is finally inferred that

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

Now, observe that

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

The term R_1 is controlled as in the proof of Lemma 3.1. The same is possible for the term R_5 since (56) (lower bound) implies that $h_F^2|\mathcal{D}_F^{uu}| \lesssim \mathcal{S}_F^{uu}$. Furthermore, owing to (10) and (59),

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

whence it follows that

$$|R_2| \lesssim \|v_h\|_\Omega^2 + \gamma\|\epsilon^{\frac12} B v_h^u\|_\Omega^2,$$

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,

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

and proceed similarly to bound R_4. Collecting the above bounds yields

$$\|\mathfrak{h}^{\frac12} C v_h^u\|_\Omega^2 \lesssim a_h(v_h,\zeta_h) + a_h(v_h,v_h) + \gamma\|\epsilon^{\frac12} B v_h^u\|_\Omega^2. \tag{62}$$

(ii) Take $\xi_h = \pi_h(\epsilon^{\frac{1}{2}}\overline{B}v_h^u, 0)$. Proceeding as above leads to

$$|\!|\!|\xi_h|\!|\!| \lesssim |\!|\!|v_h|\!|\!|.$$

Now, observe that

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

It is clear that

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

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 (54) (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

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

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 (56) (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

$$\|\epsilon^{\frac{1}{2}}Bv_h^u\|_\Omega^2 \lesssim a_h(v_h,\xi_h) + a_h(v_h,v_h). \tag{63}$$

(iii) The bounds (62) and (63) readily imply

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

and the conclusion results from the fact that $|\!|\!|\zeta_h|\!|\!| \lesssim |\!|\!|v_h|\!|\!|$ and $|\!|\!|\xi_h|\!|\!| \lesssim |\!|\!|v_h|\!|\!|$. □

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

$$|\!|\!|v|\!|\!|_*^2 = |\!|\!|v|\!|\!|^2 + \sum_{T\in\mathcal{T}_h}[(1+\tfrac{\epsilon}{h_T})h_T^{-1}\|v^u\|_T^2 + (1+\tfrac{\epsilon}{h_T})\|v^u\|_{\partial T}^2 + h_T\|v^\sigma\|_{\partial T}^2]. \tag{64}$$

Then, the following holds:

$$\forall(v,w_h)\in W(h)\times W_h,\quad a_h(v,w_h) \lesssim |\!|\!|v|\!|\!|_*|\!|\!|w_h|\!|\!|. \tag{65}$$

Proof. The idea is to bound the three terms in the right-hand side (20) making use of the block structure under consideration. Owing to (9) and the symmetry of $\mathcal{C}^k$, for all $T\in\mathcal{T}_h$,

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

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 (54) 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 if the exact solution is smooth enough, the following convergence theorem is readily inferred from Lemmas 4.1 and 4.2.

Theorem 4.1 (convergence)**.** *Let z solve* (4) *and let z_h solve* (18)*. 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*,*

$$|\!|\!| z-z_h |\!|\!| \lesssim \inf_{v_h\in W_h} |\!|\!| z-v_h |\!|\!|_*. \tag{66}$$

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*

$$|\!|\!| z-z_h |\!|\!| \lesssim h^p \|z\|_{[H^p(\Omega)]^{m_\sigma}\times[H^{p+1}(\Omega)]^{m_u}}. \tag{67}$$

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

$$|\!|\!| z-z_h |\!|\!| \lesssim h^{p+\frac{1}{2}} \|z\|_{[H^{p+1}(\Omega)]^m}. \tag{68}$$

Estimate (67) yields optimal convergence order for the B-directional derivative of the error if $\epsilon\sim 1$, whereas if $\epsilon\ll h$, estimate (68) yields optimal convergence order for the C-directional derivative of the error if the mesh family is quasi-uniform. When $\epsilon\sim 1$, estimate (67) 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 Aubin-Nitsche duality argument introduced in [14] 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,

$$\text{Find } \psi\in V^* \text{ such that } \tilde{T}\psi = (0, z^u-z_h^u) \text{ in } L. \tag{69}$$

Assume the following (elliptic) regularity result:

$$\|\psi^\sigma\|_{[H^1(\Omega)]^{m_\sigma}} + \|\psi^u\|_{[H^2(\Omega)]^{m_u}} \lesssim \|z^u-z_h^u\|_\Omega. \tag{70}$$

Lemma 4.3. *Under the above hypotheses,*

$$\forall v\in W(h),\quad a_h(v,\psi) = (v^u, z^u-z_h^u)_\Omega. \tag{71}$$

Proof. The identity results from (20). Since ψ solves (69), $(v,\tilde{T}\psi)_\Omega = (v^u, z^u-z_h^u)_\Omega$. 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 (20) vanishes because $\psi^u\in[H^2(\Omega)]^{m_u}$. □

Lemma 4.4. *The following holds:*

$$\forall (v=(0,v^u),w)\in W(h)\times W(h),\quad a_h((0,v^u),w)\lesssim |\!|\!|(0,v^u)|\!|\!|\,|\!|\!|w|\!|\!|_*. \tag{72}$$

Proof. The proof is similar to that of Lemma 4.2 except that we use (57) instead of (20). □

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

$$\|z^u-z_h^u\|_\Omega \lesssim h|\!|\!|z-z_h|\!|\!| + h\inf_{(q_h^\sigma,0)\in W_h}\left(\sum_{F\in\mathcal{F}_h^\partial} h_F\|z^\sigma-q_h^\sigma\|_F^2+\|z^\sigma-q_h^\sigma\|_{T(F)}^2\right)^{\frac12}. \tag{73}$$

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

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

Proof. Owing to (71) and Lemma 3.3,

$$\|z^u-z_h^u\|_\Omega^2 = a_h(z-z_h,\psi)=a_h(z-z_h,\psi-w_h),$$

where w_h is arbitrary in W_h. Hence,

$$\begin{aligned}\|z^u-z_h^u\|_\Omega^2 &= a_h((0,z^u-z_h^u),\psi-w_h)+(\mathcal{K}(z^\sigma-z_h^\sigma,0),\psi-w_h)_\Omega+\epsilon^{\frac12}(\tilde B(z^\sigma-z_h^\sigma),\psi^u-w_h^u)_\Omega\\ &:= T_1+T_2+T_3.\end{aligned}$$

Owing to (72),

$$|T_1|\lesssim |\!|\!|z-z_h|\!|\!|\,|\!|\!|\psi-w_h|\!|\!|_*,$$

and clearly, $|T_2|\lesssim |\!|\!|z-z_h|\!|\!|\,|\!|\!|\psi-w_h|\!|\!|_*$. Integrating by parts and using (54) (lower bound) yields

$$\begin{aligned}|T_3| &\lesssim \|z^\sigma-z_h^\sigma\|_\Omega\|\epsilon^{\frac12}B(\psi^u-w_h^u)\|_\Omega+\sum_{F\in\mathcal{F}_h^\partial} h_F^{\frac12}\|z^\sigma-z_h^\sigma\|_F|\psi^u-w_h^u|_{M,F}\\ &\lesssim |\!|\!|z-z_h|\!|\!|\,|\!|\!|\psi-w_h|\!|\!|_*+\left(\sum_{F\in\mathcal{F}_h^\partial} h_F\|z^\sigma-z_h^\sigma\|_F^2\right)^{\frac12}|\!|\!|\psi-w_h|\!|\!|_*.\end{aligned}$$

Let $(q_h^\sigma,0)$ be arbitrary in W_h. Using (10) and triangle inequalities yields

$$\begin{aligned}\|z^\sigma-z_h^\sigma\|_F &\le \|z^\sigma-q_h^\sigma\|_F+\|z_h^\sigma-q_h^\sigma\|_F\lesssim \|z^\sigma-q_h^\sigma\|_F+h_F^{-\frac12}\|z_h^\sigma-q_h^\sigma\|_{T(F)}\\ &\le \|z^\sigma-q_h^\sigma\|_F+h_F^{-\frac12}\|z^\sigma-q_h^\sigma\|_{T(F)}+h_F^{-\frac12}\|z^\sigma-z_h^\sigma\|_{T(F)}.\end{aligned}$$

Hence,

$$|T_3|\lesssim |\!|\!|z-z_h|\!|\!|\,|\!|\!|\psi-w_h|\!|\!|_*+\left(\sum_{F\in\mathcal{F}_h^\partial} h_F\|z^\sigma-q_h^\sigma\|_F^2+\|z^\sigma-q_h^\sigma\|_{T(F)}^2\right)^{\frac12}|\!|\!|\psi-w_h|\!|\!|_*.$$

Using (70) and classical interpolation results yields $\inf_{w_h\in W_h}|\!|\!|\psi-w_h|\!|\!|_*\lesssim h\|z^u-z_h^u\|_\Omega$. The conclusion is straightforward. □

Remark 4.1. Estimate (67) 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 Section 4.2 to the three Friedrichs' systems endowed with the 2×2 block structure discussed in Section 4.1.

4.3.1. *Advection-diffusion-reaction*

Set $z^\sigma = \sigma$ and $z^u = u$. Clearly (59) holds since $|\beta{\cdot}\nabla w_h^u| \lesssim \|\nabla w_h^u\|$. Let $\alpha > 0$ and take

$$\mathcal{S}_F^{uu} = \alpha\, h_F^2(|\beta{\cdot}n_F| + \tfrac{\epsilon}{h_F}), \qquad \mathcal{M}_F^{uu} = |\beta{\cdot}n_F| + \tfrac{\epsilon}{h_F}. \tag{75}$$

Observe that the design of $\mathcal{M}_F^{uu}$ is such 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,

$$\|u - u_h\|_\Omega + h\|\nabla(u - u_h)\|_\Omega + h\|\sigma - \sigma_h\|_\Omega \lesssim h^{p+1}\|(\sigma, u)\|_{[H^p(\Omega)]^d \times H^{p+1}(\Omega)}. \tag{76}$$

Comparing (76) with (39), 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 $p + \frac{1}{2}$). Furthermore, if $\epsilon \ll h$ and the exact solution is smooth enough,

$$\|u - u_h\|_\Omega + \|\mathfrak{h}^{\frac{1}{2}}\beta{\cdot}\nabla(u - u_h)\|_\Omega + \|\sigma - \sigma_h\|_\Omega \lesssim h^{p+\frac{1}{2}}\|(\sigma, u)\|_{[H^{p+1}(\Omega)]^{d+1}}. \tag{77}$$

4.3.2. *Linear elasticity*

Set $z^\sigma = (\overline{\sigma}, p)$ and $z^u = u$. Clearly (59) holds since $C = 0$. Let $\alpha > 0$ and $\eta > 0$ and take

$$\mathcal{S}_F^{uu} = \alpha\, h_F\, \mathcal{I}_d, \qquad \mathcal{M}_F^{uu} = \eta\, h_F^{-1}\, \mathcal{I}_d. \tag{78}$$

Then, if the exact solution is smooth enough,

$$\|u - u_h\|_\Omega + h\|\nabla(u - u_h)\|_\Omega + h\|p - p_h\|_\Omega + h\|\sigma - \sigma_h\|_\Omega \lesssim h^{p+1}\|(\overline{\sigma}, p, u)\|_{[H^p(\Omega)]^{d^2+1} \times [H^{p+1}(\Omega)]^d}. \tag{79}$$

Comparing (79) with (44), we observe that the optimal convergence of the divergence of the symmetric part of σ is lost and that $\|\sigma - \sigma_h\|_\Omega$ and $\|p - p_h\|_\Omega$ converge only to order p (instead of $p + \frac{1}{2}$).

Remark 4.2. Numerical experiments indicate that the above formulation is unstable in the incompressible limit. To obtain a stable formulation, a penalty on the jumps of the normal derivative of the discrete pressure has to be included, yielding a SCFEM similar to that proposed in [4] for the three-field Stokes problem. A similar modification is analyzed in [15] for the DGM.

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

Set $z^\sigma = H$ and $z^u = E$. Clearly (59) holds since $C = 0$. Let $\alpha > 0$ and $\eta > 0$ and take

$$\mathcal{S}_F^{uu} = \alpha\, h_F\, \mathcal{N}_F^t\, \mathcal{N}_F, \qquad \mathcal{M}_F^{uu} = \eta\, h_F^{-1}\, \mathcal{N}_F^t\, \mathcal{N}_F. \tag{80}$$

Then, if the exact solution is smooth enough,

$$\|E - E_h\|_\Omega + h\|\nabla{\times}(E - E_h)\|_\Omega + h\|H - H_h\|_\Omega \lesssim h^{p+1}\|(H, E)\|_{[H^p(\Omega)]^3 \times [H^{p+1}(\Omega)]^3}. \tag{81}$$

Comparing with the estimate (49), 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 $p + \frac{1}{2}$).

TABLE 1. Convergence results for $p = 1$ and the SCFEM analyzed in Section 3.

h	$h^{\frac{1}{2}}\times(34)$	$\mathrm{O}(h^{\xi})$	$h^{\frac{1}{2}}\times(39)$	$\mathrm{O}(h^{\xi})$	$h^{\frac{1}{2}}\times(44)$	$\mathrm{O}(h^{\xi})$	$h^{\frac{1}{2}}\times(49)$	$\mathrm{O}(h^{\xi})$
2^{-3}	3.2e–2	-	1.4e–1	-	4.2e–1	-	1.6e–0	-
2^{-4}	7.7e–3	2.1	3.7e–2	1.9	1.2e–1	1.8	2.9e–1	2.5
2^{-5}	1.8e–3	2.2	8.5e–3	2.1	2.3e–2	2.4	6.1e–2	2.2
2^{-6}	3.9e–4	2.2	2.1e–3	2.0	6.4e–3	1.8	1.5e–2	2.0
2^{-7}	1.1e–4	1.8	5.1e–4	2.0	1.6e–3	2.0	3.8e–3	2.0

5. Numerical results

All the numerical experiments are carried out using FreeFem++ [19]. 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 polynomial orders up to 2; see, *e.g.*, [8] for further discussion on the optimal choice of penalty parameters. When the solution is rough and for problems having a 2×2 block structure, the SCFEM analyzed in Section 4 appears to be more robust with respect to the choice of stabilization parameters than the SCFEM analyzed in Section 3.

5.1. Convergence rates for smooth solutions

We consider the four examples of Friedrichs' systems discussed above. 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 are 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 are 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 are 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 are enforced on E.

Tables 1–4 present convergence results on unstructured meshes obtained with $p = 1$ and $p = 2$ for the two stabilization techniques discussed in Sections 3 and 4. The number in the first row of the tables refers to the equation number of the estimate, and the columns labeled $\mathrm{O}(h^{\xi})$ indicate convergence orders. All the convergence orders match theoretical predictions. For the advection-reaction equation and $p = 2$, the overall convergence order is correct, despite some irregularities on coarser meshes. All the experiments have been repeated on structured meshes leading to similar results. When working with structured meshes, a super-convergence phenomenon by a factor of $h^{\frac{1}{2}}$ is observed for $p = 1$ and the estimates derived in Section 3. This observation 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.

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.

TABLE 2. Convergence results for $p = 1$ and the SCFEM analyzed in Section 4.

h	(76)	$O(h^\xi)$	(79)	$O(h^\xi)$	(81)	$O(h^\xi)$
2^{-3}	3.4e–2	-	7.8e–2	-	2.0e–1	-
2^{-4}	8.4e–3	2.0	1.5e–2	2.4	3.6e–2	2.5
2^{-5}	1.7e–3	2.3	3.1e–3	2.3	7.5e–3	2.3
2^{-6}	4.8e–4	1.8	8.4e–4	1.9	1.8e–3	2.1
2^{-7}	1.1e–4	2.1	2.0e–4	2.1	4.6e–4	2.0

TABLE 3. Convergence results for $p = 2$ and the SCFEM analyzed in Section 3.

h	$h^{\frac{1}{2}}\times$(34)	$O(h^\xi)$	$h^{\frac{1}{2}}\times$(39)	$O(h^\xi)$	$h^{\frac{1}{2}}\times$(44)	$O(h^\xi)$	$h^{\frac{1}{2}}\times$(49)	$O(h^\xi)$
2^{-3}	5.3e–3	-	1.6e–2	-	4.9e–2	-	2.5e–1	-
2^{-4}	2.2e–3	1.3	1.9e–3	3.1	5.5e–3	3.2	2.9e–2	3.1
2^{-5}	9.4e–5	4.5	2.1e–4	3.2	6.3e–4	3.1	3.4e–3	3.1
2^{-6}	2.0e–5	2.2	2.7e–5	3.0	8.5e–5	2.9	4.2e–4	3.0

TABLE 4. Convergence results for $p = 2$ and the SCFEM analyzed in Section 4.

h	(76)	$O(h^\xi)$	(79)	$O(h^\xi)$	(81)	$O(h^\xi)$
2^{-3}	5.4e–3	-	1.1e–2	-	4.6e–2	-
2^{-4}	7.1e–4	2.9	1.5e–3	2.9	6.3e–3	2.9
2^{-5}	9.0e–5	3.0	2.0e–4	2.9	7.2e–4	3.1
2^{-6}	1.1e–5	3.0	2.4e–5	3.1	8.7e–5	3.0

- Advection-diffusion-reaction: Ω is an L-shaped domain, homogeneous Dirichlet boundary conditions are enforced, $\mu = 0$, $\beta = (1, 0)^t$, $\epsilon = 1$, and $f = 10\exp(-100((x - 0.5)^2 + (y - 0.5)^2))$.
- Linear elasticity: Ω is an L-shaped domain, homogeneous Dirichlet boundary conditions are enforced on the displacement, $\gamma_1 = \gamma_2 = 1$, $f_1 = 10\exp(-(x - 0.5)^2 - (y - 0.5)^2)$, 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 solution computed without stabilization exhibits oscillations, while oscillations are essentially eliminated by the SCFEM. 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-component, the SCFEM with stabilization on u produces slightly better results than the SCFEM with stabilization on σ and u.

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. As in the previous case, the solution computed without stabilization exhibits oscillations, while oscillations are essentially eliminated by the SCFEM. Furthermore, the two versions of the SCFEM produce similar results since the data f has smoother variations than in the previous case. For the method analyzed in Section 3, the stabilization parameters for the displacements and the stresses have to be chosen separately. In the present case, we took $\alpha_1 = 10^{-3}$ and $\alpha_2 = 0.05$ in (43). For the method analyzed in Section 4, choosing $\alpha \in [10^{-1}, 10^{-3}]$ in (78) yields satisfactory results.

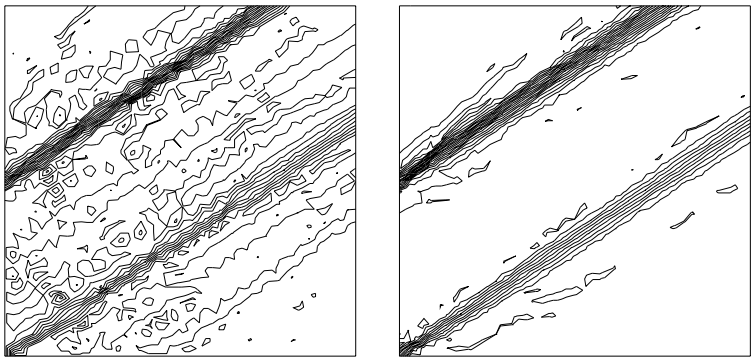

FIGURE 1. Advection-reaction: approximate solution obtained with CFEM without (left) and with (right) stabilization.

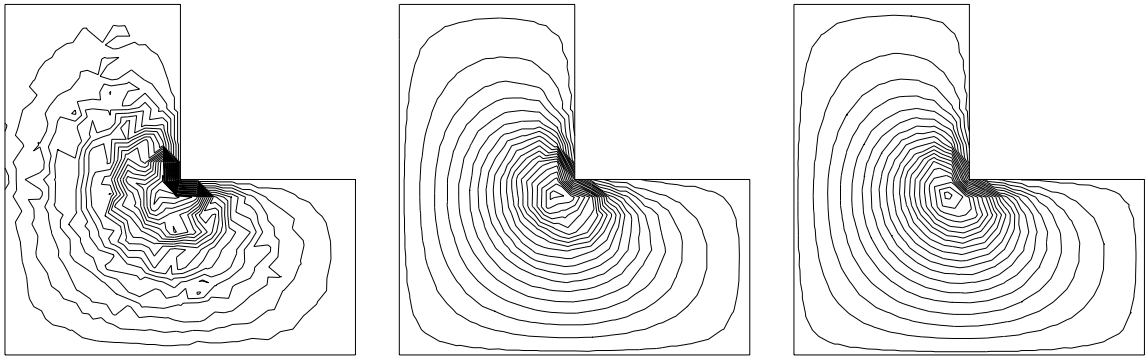

FIGURE 2. Advection-diffusion-reaction: approximate solution u obtained with CFEM without stabilization (left), with stabilization on σ and u (center), and with stabilization on u only (right).

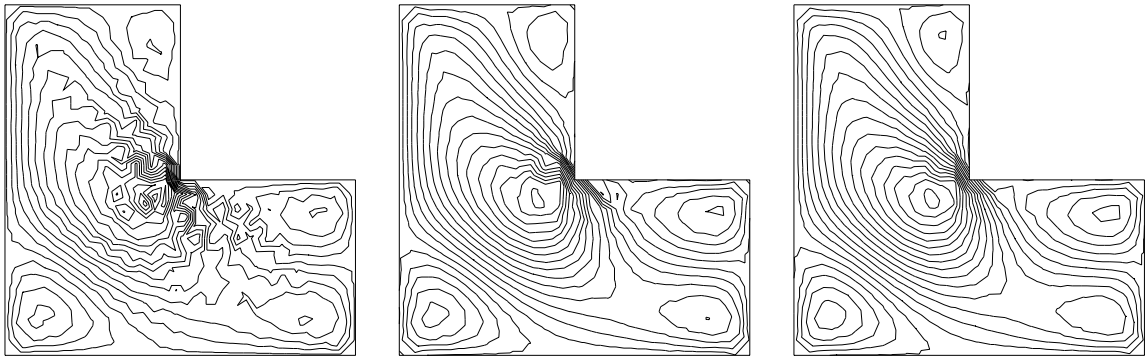

FIGURE 3. Linear elasticity: approximate solution, second displacement component u_2 obtained with CFEM without stabilization (left), with stabilization on σ and u (center), and with stabilization on u only (right).

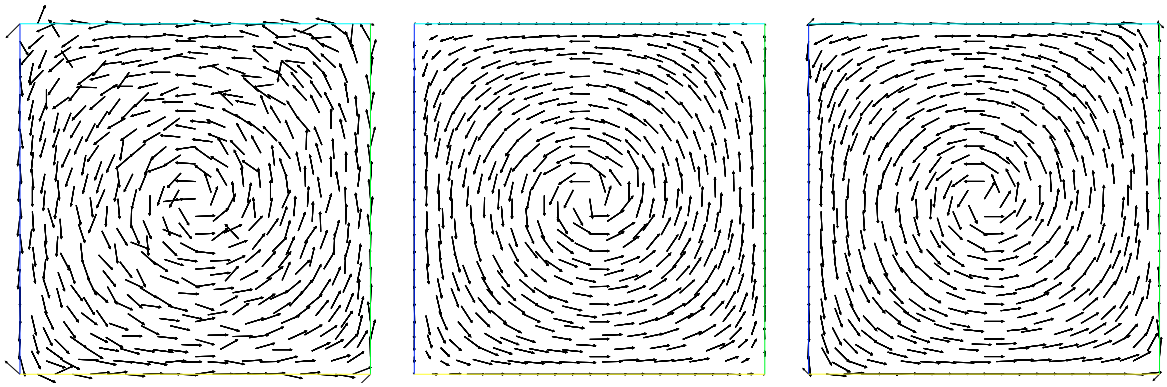

FIGURE 4. Maxwell's equations in the elliptic regime: approximate solution (H_1, H_2) obtained with CFEM without stabilization (left), with stabilization on H and E (center), and with stabilization on E only (right).

Finally, 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

The theoretical analysis and the numerical experiments presented in this work have shown that Friedrichs' systems can be satisfactorily approximated by stabilized continuous finite elements. For elliptic-like PDE's, the mixed form is considered. Two stabilizations of independent interest are proposed, yielding different convergence orders for the primal variable and its flux. The choice between the two stabilizations is driven by the regularity of the exact solution and cost considerations since the demand on memory is much lighter when only the primal variable is stabilized.

References

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

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

[3] G.A. Baker, Finite element methods for elliptic equations using nonconforming elements. *Math. Comp.* **31** (1977) 45–59.

[4] A. Bonito and E. Burman, A face penalty method for the three fields Stokes equation arising from Oldroyd-B viscoelastic flows, in *Numerical Mathematics and Advanced Applications, ENUMATH Conf. Proc.*, Springer (2006).

[5] E. Burman, A unified analysis for conforming and nonconforming stabilized finite element methods using interior penalty. *SIAM J. Numer. Anal.* **43** (2005) 2012–2033.

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

[7] E. Burman and P. Hansbo, Edge stabilization for the generalized Stokes problem: a continuous interior penalty method. *Comput. Methods Appl. Mech. Engrg.* **195** (2006) 2393–2410.

[8] 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).

[9] Z. Cai, T.A. Manteuffel, S.F. McCormick and S.V. Parter. First-order system least squares (FOSLS) for planar linear elasticity: Pure traction problem. *SIAM J. Numer. Anal.* **35** (1998) 320–335.

[10] J. Douglas, Jr., and T. Dupont, *Interior Penalty Procedures for Elliptic and Parabolic Galerkin Methods*. Lect. Notes Phys. **58**, Springer-Verlag, Berlin (1976).

[11] L. El Alaoui and A. Ern, Residual and hierarchical *a posteriori* estimates for nonconforming mixed finite element methods. *ESAIM: M2AN* **38** (2004) 903–929.

[12] A. Ern and J.-L. Guermond, *Theory and Practice of Finite Elements*. Appl. Math. Sci. **159**, Springer-Verlag, New York, NY (2004).
[13] A. Ern and J.-L. Guermond, Discontinuous Galerkin methods for Friedrichs' systems. I. General theory. *SIAM J. Numer. Anal.* **44** (2006) 753–778.
[14] A. Ern and J.-L. Guermond, Discontinuous Galerkin methods for Friedrichs' systems. II. Second-order PDEs. *SIAM J. Numer. Anal.* **44** (2006) 2363–2388.
[15] A. Ern and J.-L. Guermond, Discontinuous Galerkin methods for Friedrichs' systems. III. Multi-field theories with partial coercivity. *SIAM J. Numer. Anal.* (2006) Submitted (CERMICS report 2006–320).
[16] A. Ern and J.-L. Guermond, Evaluation of the condition number in linear systems arising in finite element approximations. *ESAIM: M2AN* **40** (2006) 29–48.
[17] R.S. Falk and G.R. Richter, Explicit finite element methods for symmetric hyperbolic equations. *SIAM J. Numer. Anal.* **36** (1999) 935–952.
[18] K.O. Friedrichs, Symmetric positive linear differential equations. *Comm. Pure Appl. Math.* **11** (1958) 333–418.
[19] F. Hecht and O. Pironneau, *FreeFEM++ Manual*. Laboratoire Jacques-Louis Lions, University Paris VI (2005).
[20] R.H.W. Hoppe and B. Wohlmuth, Element-oriented and edge-oriented local error estimators for non-conforming finite element methods. *RAIRO Math. Model. Anal. Numer.* **30** (1996) 237–263.
[21] M. Jensen, *Discontinuous Galerkin Methods for Friedrichs Systems with Irregular Solutions*. Ph.D. thesis, University of Oxford (2004).
[22] C. Johnson and J. Pitkäranta, An analysis of the discontinuous Galerkin method for a scalar hyperbolic equation. *Math. Comp.* **46** (1986) 1–26.
[23] O. Karakashian and F. Pascal, *A-posteriori* error estimates for a discontinuous Galerkin approximation of second order elliptic problems. *SIAM J. Numer. Anal.* **41** (2003) 2374–2399.
[24] P. Lesaint, Finite element methods for symmetric hyperbolic equations. *Numer. Math.* **21** (1973/74) 244–255.
[25] P. Lesaint, *Sur la résolution des systèmes hyperboliques du premier ordre par des méthodes d'éléments finis*. Ph.D. thesis, University of Paris VI, France (1975).
[26] P. Lesaint and P.-A. Raviart. On a finite element method for solving the neutron transport equation, in *Mathematical Aspects of Finite Elements in Partial Differential Equations*, C. de Boors Ed., Academic Press (1974) 89–123.