Discontinuous Galerkin methods for Friedrichs' systems*

Alexandre Ern[1] and Jean-Luc Guermond[2]

[1] CERMICS, Ecole nationale des ponts et chaussées, Champs sur Marne, 77455 Marne la Vallée Cedex 2, France ern@cermics.enpc.fr
[2] Dept. Math, Texas A&M, College Station, TX 77843-3368, USA (on leave from LIMSI-CNRS, BP 133, 91403, Orsay, France) guermond@math.tamu.edu

Summary. This work presents a unified analysis of Discontinuous Galerkin methods to approximate Friedrichs' systems. A general set of boundary conditions is identified to guarantee existence and uniqueness of solutions to these systems. A formulation enforcing the boundary conditions weakly is proposed. This formulation is the starting point for the construction of Discontinuous Galerkin methods formulated in terms of boundary operators and of interface operators that mildly penalize interface jumps. A general convergence analysis is presented. The setting is subsequently specialized to Friedrichs' systems endowed with a particular 2×2 structure in which some of the unknowns can be eliminated to yield a system of second-order elliptic-like PDE's for the remaining unknowns. A general Discontinuous Galerkin method where the above elimination can be performed in each mesh cell is proposed and analyzed. Finally, details are given for four examples, namely advection–reaction equations, advection–diffusion–reaction equations, the linear elasticity equations in the mixed stress–pressure–displacement form, and the Maxwell equations in the so-called elliptic regime.

1 Introduction

Since their introduction in 1973 by Reed and Hill [20] to simulate neutron transport, Discontinuous Galerkin (DG) methods have sparked extensive interest owing to their flexibility in handling non-matching grids, heterogeneous data, and high-order *hp*-adaptivity. However, the development and analysis of DG methods has followed two somewhat parallel routes depending on whether the PDE is hyperbolic or elliptic.

For hyperbolic PDEs, the first analysis of DG methods in an already rather abstract form was performed by Lesaint and Raviart in 1974 [17, 18] and sub-

* This paper has been published in *Numerical Mathematics and Advanced Applications*, Proceedings of ENUMATH 2005, 6th European Conference on Numerical Mathematics and Advanced Applications, A. Berúdez de Castro, D. Gómez, P. Quintela, P. Salgado (Eds.), Santiago de Compostela, Spain, July 2005, Springer-Verlag Berlin Heidelberg (2006) 79–96.

sequently improved by Johnson *et al.* [16] in 1984. More recently, DG methods for hyperbolic and nearly hyperbolic equations experienced a significant development based on the ideas of numerical fluxes, approximate Riemann solvers, and slope limiters; see, e.g., Cockburn *et al.* [8] and the references therein.

For elliptic PDEs, DG methods originated from the early work of Nitsche on boundary-penalty methods [19] and the use of Interior Penalties (IP) to weakly enforce continuity on the solution or its derivatives across the interfaces between adjoining elements; see, e.g., Babuška [3], Babuška and Zlámal [4], Douglas and Dupont [10], Baker [5], Wheeler [21], and Arnold [1]. DG methods for elliptic problems in mixed form were introduced more recently (see, e.g., Bassi and Rebay [6]) and further extended by Cockburn and Shu [9] leading to the so-called Local Discontinuous Galerkin (LDG) method. The fact that several of the above DG methods (including IP methods) share common features and can be tackled by similar analysis tools called for a unified analysis. A first important step in that direction has been recently accomplished in Arnold *et al.* [2], where it is shown that it is possible to cast many DG methods for the Poisson equation with homogeneous Dirichlet boundary conditions into a single framework amenable to a unified error analysis.

The goal of the present work is to propose a unified analysis of DG methods that goes beyond the traditional hyperbolic/elliptic classification of PDEs. To this purpose, we make systematic use of the theory of Friedrichs' systems [15], i.e., systems of first-order PDE's endowed with a symmetry and a positivity property, to formulate DG methods and to perform the convergence analysis. For brevity, the main theoretical results are stated without proof; see [12, 13, 14] for full detail.[3]

This paper is organized as follows. In §2 we revisit Friedrichs' theory and formulate a set of abstract conditions ensuring well–posedness of the continuous problem while avoiding to invoke traces at the boundary. In §3 we formulate and analyze a general DG method to approximate Friedrichs' systems. The design of the method is based on an operator enforcing boundary conditions weakly and an operator penalizing the jumps of the solution across the mesh interfaces. All the design constraints to be fulfilled by the boundary and the interface operators for the error analysis to hold are stated. Moreover, using integration by parts, the DG method is re-interpreted locally by introducing the concept of element fluxes, thus providing a direct link with engineering practice where approximation schemes are often designed by specifying such fluxes. In §4 we specialize the setting to a particular class of Friedrichs' systems with a 2×2 structure in which some of the unknowns can be eliminated to yield a system of second-order elliptic-like PDE's for the remaining unknowns. For such systems, a general Discontinuous Galerkin method is proposed and analyzed. The key feature of the method is that the unknowns that can be eliminated at the continuous level can also be eliminated at the discrete level by solving local problems. In §5, we apply the theoretical results to

[3] Internal reports available at `cermics.enpc.fr/reports/CERMICS-2005`

advection–reaction equations, advection–diffusion–reaction equations, the linear elasticity equations in the mixed stress–pressure–displacement form, and the Maxwell equations in the so-called elliptic regime. Concluding remarks are reported in §6.

2 Friedrichs' systems

Let Ω be a bounded, open, and connected Lipschitz domain in $\mathbb{R}^d$. We denote by $\mathfrak{D}(\Omega)$ the space of $\mathfrak{C}^\infty$ functions that are compactly supported in Ω. 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 $\mathbb{R}^{m,m}$ such that

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

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

$$\forall k \in \{1,\dots,d\},\ \mathcal{A}^k = (\mathcal{A}^k)^t \ \text{ a.e. in } \Omega, \tag{A3}$$

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

where $\mathcal{I}_m$ is the identity matrix in $\mathbb{R}^{m,m}$. Assumptions (A3) and (A4) are, respectively, the symmetry and the positivity property referred to above.

Set $L = [L^2(\Omega)]^m$. A function z in L is said to have an A-weak derivative in L if the linear form $[\mathfrak{D}(\Omega)]^m \ni \phi \longmapsto -\int_\Omega \sum_{k=1}^d z^t \partial_k(\mathcal{A}^k \phi) \in \mathbb{R}$ is bounded on L. In this case, the function in L that can be associated with the above linear form by means of the Riesz representation theorem is denoted by Az. Clearly, if z is smooth, e.g., $z \in [\mathfrak{C}^1(\overline{\Omega})]^m$, $Az = \sum_{k=1}^d \mathcal{A}^k \partial_k z$. Define the so-called graph space $W = \{z \in L;\, Az \in L\}$ equipped with the graph norm $\|z\|_W = \|Az\|_L + \|z\|_L$. The space W is endowed with a Hilbert structure when equipped with the scalar product $(z,y)_L + (Az, Ay)_L$. Define the operators $T \in \mathcal{L}(W;L)$ and $\tilde{T} \in \mathcal{L}(W;L)$ as

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

Assumption (A4) implies that $T + \tilde{T}$ is L-coercive on L.

Let $f \in L$ and consider the problem of seeking $z \in W$ such that $Tz = f$ in L. 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 $T : V \to L$ is an isomorphism. Let $D \in \mathcal{L}(W;W')$ be the operator defined by

$$\forall (z,y) \in W \times W, \quad \langle Dz, y\rangle_{W',W} = (Tz,y)_L - (z, \tilde{T}y)_L. \tag{2}$$

Let W_0 be the closure of $[\mathfrak{D}(\Omega)]^m$ in W. For every subspace $Z \subset W$, let $Z^\perp$ denote the polar set of Z, i.e., the set of linear forms on W that vanish on Z and use a similar notation for the polar sets of subspaces of W'. A key result concerning the operator D is the following

Lemma 1. *The operator D is self-adjoint. Moreover, the following holds:*

$$\mathrm{Ker}(D) = W_0 \qquad \text{and} \qquad \mathrm{Im}(D) = W_0^\perp. \tag{3}$$

To enforce boundary conditions, a simple approach inspired from Friedrichs' work consists of assuming that there is an operator $M \in \mathcal{L}(W; W')$ such that

$$M \text{ is positive, i.e., } \langle Mz, z\rangle_{W',W} \geq 0 \text{ for all } z \text{ in } W, \tag{M1}$$

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

Then by setting $V = \mathrm{Ker}(D - M)$ and $V^* = \mathrm{Ker}(D + M^*)$ where $M^* \in \mathcal{L}(W; W')$ is the adjoint of M and equipping V and V^* with the graph norm, the following theorem can be proved:

Theorem 1. *Assume* (A1)–(A4) *and* (M1)–(M2). *Then, the restricted operators $T : V \to L$ and $\tilde{T} : V^* \to L$ are isomorphisms.*

The proof of Theorem 1 relies on the following fundamental result, the so-called Banach–Nečas–Babuška (BNB) Theorem, that is restated below for completeness (see, e.g., [11]).

Theorem 2 (BNB). *Let V and L be two Banach spaces, and denote by $\langle\cdot,\cdot\rangle_{L',L}$ the duality pairing between L' and L. Then, $T \in \mathcal{L}(V; L)$ is bijective if and only if*

$$\exists \alpha > 0, \quad \forall w \in V, \qquad \sup_{y \in L' \setminus \{0\}} \frac{\langle y, Tw\rangle_{L',L}}{\|y\|_{L'}} \geq \alpha \|w\|_V, \tag{4}$$

$$\forall y \in L', \quad (\langle y, Tw\rangle_{L',L} = 0,\ \forall w \in V) \implies (y = 0). \tag{5}$$

Remark 1. It is possible to formulate an intrinsic criterion for the bijectivity of the operators T and $\tilde{T}$ that circumvents the somewhat *ad hoc* operator M by introducing the concept of maximal boundary conditions. To this purpose, introduce the cones $C^\pm = \{w \in W;\ \pm\langle Dw, w\rangle_{W',W} \geq 0\}$. Let V and V^* be two subspaces of W such that

$$V \subset C^+ \text{ and } V^* \subset C^-, \tag{V1}$$

$$V = D(V^*)^\perp \text{ and } V^* = D(V)^\perp. \tag{V2}$$

Then, under the assumptions (A1)–(A4) and (V1)–(V2), the conclusions of Theorem 1 still hold. Furthermore, one can prove that if V and V^* are two subspaces of W satisfying (V1)–(V2), then V is maximal in C^+ (there is no $x \in W$ such that $V_x := V + \mathrm{span}(x) \subset C^+$ and V is a proper subspace of V_x) and V^* is maximal in C^- (there is no $y \in W$ such that $V_y^* := V^* + \mathrm{span}(y) \subset C^-$ and V^* is a proper subspace of V_y^*). In this sense, the boundary conditions embodied in V and V^* are maximal.

Owing to Theorem 1, the following problems are well-posed:

$$\text{Seek } z \in V \text{ such that } Tz = f, \tag{6}$$

$$\text{Seek } z^* \in V^* \text{ such that } \tilde{T}z^* = f. \tag{7}$$

The boundary conditions in (6) and (7) are enforced strongly by seeking the solutions in V and V^*, respectively. A key feature of Friedrichs' systems is that it is possible to enforce boundary conditions naturally, thus leading to a suitable framework for developing a DG theory. To see this, introduce the following bilinear forms on $W \times W$,

$$a(z, y) = (Tz, y)_L + \tfrac{1}{2}\langle (M - D)z, y\rangle_{W',W}, \tag{8}$$

$$a^*(z, y) = (\tilde{T}z, y)_L + \tfrac{1}{2}\langle (M^* + D)z, y\rangle_{W',W}. \tag{9}$$

It is clear that a and a^* are in $\mathcal{L}(W \times W; \mathbb{R})$. Consider the following problems:

$$\text{Seek } z \in W \text{ such that } a(z, y) = (f, y)_L,\ \forall y \in W, \tag{10}$$

$$\text{Seek } z^* \in W \text{ such that } a^*(z^*, y) = (f, y)_L,\ \forall y \in W. \tag{11}$$

Contrary to (6) and (7), the boundary conditions in (10) and (11) are weakly enforced. For this reason, problem (10) will constitute our working basis for designing DG methods. The key result of this section is the following

Theorem 3. *Assume* (A1)–(A4) *and* (M1)–(M2). *Then, there is a unique solution to* (10) *(resp.,* (11)*) and this solution solves* (6) *(resp.,* (7)*).*

3 Design and analysis of DG methods

The purpose of this section is to design and analyze a general DG method to approximate the unique solution to (10).

3.1 The discrete setting

Let $\{\mathcal{T}_h\}_{h>0}$ be a family of meshes of Ω. The meshes are assumed to be affine to avoid unnecessary technicalities, i.e., Ω is assumed to be a polyhedron. However, we do not make any assumption on the matching of element interfaces. Let p be a non-negative integer and set

$$P_{h,p} = \{v_h \in L^2(\Omega); \forall K \in \mathcal{T}_h,\ v_h|_K \in \mathbb{P}_p\}, \tag{12}$$

where $\mathbb{P}_p$ denotes the vector space of polynomials with real coefficients and total degree less than or equal to p. Define

$$W_h = [P_{h,p}]^m, \qquad W(h) = [H^1(\Omega)]^m + W_h. \tag{13}$$

We denote by $\mathcal{F}_h^{\mathrm{i}}$ the set of interior faces (or interfaces), i.e., $F \in \mathcal{F}_h^{\mathrm{i}}$ if F is a $(d-1)$-manifold and there are $K_1(F)$, $K_2(F) \in \mathcal{T}_h$ such that $F = K_1(F) \cap K_2(F)$. We denote by $\mathcal{F}_h^{\partial}$ the set of the faces that separate the mesh from the exterior of Ω, i.e., $F \in \mathcal{F}_h^{\partial}$ if F is a $(d-1)$-manifold and there is $K(F) \in \mathcal{T}_h$ such that $F = K(F) \cap \partial\Omega$. Finally, we set $\mathcal{F}_h = \mathcal{F}_h^{\mathrm{i}} \cup \mathcal{F}_h^{\partial}$. Since every function v in $W(h)$ has a (possibly two-valued) trace almost everywhere on $F \in \mathcal{F}_h^{\mathrm{i}}$, it is meaningful to set $v^n(x) = \lim_{\substack{y \to x \\ y \in K_n(F)}} v(y)$, $n \in \{1, 2\}$, for a.e. $x \in F$ and

$$[\![v]\!] = v^1 - v^2, \qquad \{v\} = \tfrac{1}{2}(v^1 + v^2), \qquad \text{a.e. on } F. \tag{14}$$

Nothing that is said hereafter depends on the arbitrariness in the sign of $[\![v]\!]$.

For any measurable subset of Ω, say E, $(\cdot,\cdot)_{L,E}$ denotes the usual L^2-scalar product on E. The same notation is used for scalar- and vector-valued functions. For $K \in \mathcal{T}_h$ (resp., $F \in \mathcal{F}_h$), h_K (resp., h_F) denotes the diameter of K (resp., F). The mesh family $\{\mathcal{T}_h\}_{h>0}$ is assumed to be shape-regular so that the usual inverse and trace inverse inequalities hold on W_h. Henceforth, we use the notation $A \lesssim B$ to represent the inequality $A \leq cB$ where c is independent of h.

3.2 The design of the DG bilinear form

Set $\mathcal{D} = \sum_{k=1}^{d} n_k \mathcal{A}^k$, where $n = (n_1, \ldots, n_d)^t$ is the outward unit normal to Ω, and assume that there is a matrix-valued field $\mathcal{M} : \partial\Omega \longrightarrow \mathbb{R}^{m,m}$ such that for all functions y, w smooth enough (e.g., $y, w \in [H^1(\Omega)]^m$),

$$\langle Dy, w\rangle_{W',W} = \int_{\partial\Omega} w^t \mathcal{D} y, \qquad \langle My, w\rangle_{W',W} = \int_{\partial\Omega} w^t \mathcal{M} y. \tag{15}$$

To enforce boundary conditions weakly, we introduce for all $F \in \mathcal{F}_h^{\partial}$ a linear operator $M_F \in \mathcal{L}([L^2(F)]^m; [L^2(F)]^m)$ such that for all $y, w \in [L^2(F)]^m$,

$$(M_F(y), y)_{L,F} \geq 0, \tag{DG1}$$
$$(\mathcal{M}y = \mathcal{D}y) \implies (M_F(y) = \mathcal{D}y), \tag{DG2}$$
$$|(M_F(y) - \mathcal{D}y, w)_{L,F}| \lesssim |y|_{M,F} \|w\|_{L,F}, \tag{DG3}$$
$$|(M_F(y) + \mathcal{D}y, w)_{L,F}| \lesssim \|y\|_{L,F} |w|_{M,F}, \tag{DG4}$$

where for all $y \in W(h)$, $|y|_M^2 = \sum_{F \in \mathcal{F}_h^{\partial}} |y|_{M,F}^2$ with $|y|_{M,F}^2 = (M_F(y), y)_{L,F}$.

For $K \in \mathcal{T}_h$, define the matrix-valued field $\mathcal{D}_{\partial K} : \partial K \to \mathbb{R}^{m,m}$ as

$$\mathcal{D}_{\partial K}(x) = \sum_{k=1}^{d} n_{K,k} \mathcal{A}^k(x) \quad \text{a.e. on } \partial K, \tag{16}$$

where $n_K = (n_{K,1}, \ldots, n_{K,d})^t$ is the unit outward normal to K on ∂K. We extend the matrix-valued field $\mathcal{D}$ on $\mathcal{F}_h = \mathcal{F}_h^{\mathrm{i}} \cup \mathcal{F}_h^{\partial}$ as follows. On $\mathcal{F}_h^{\partial}$, $\mathcal{D}$ is defined as above. On $\mathcal{F}_h^{\mathrm{i}}$, $\mathcal{D}$ is two-valued and for all $F \in \mathcal{F}_h^{\mathrm{i}}$, its two values are $\mathcal{D}_{\partial K_1(F)}$ and $\mathcal{D}_{\partial K_2(F)}$. Note that $\{\mathcal{D}\} = 0$ a.e. on $\mathcal{F}_h^{\mathrm{i}}$. To control the jumps of functions in W_h across mesh interfaces, we introduce for all $F \in \mathcal{F}_h^{\mathrm{i}}$ a linear operator $S_F \in \mathcal{L}([L^2(F)]^m; [L^2(F)]^m)$ such that for all $y, w \in [L^2(F)]^m$,

$$(S_F(y), y)_{L,F} \geq 0, \tag{DG5}$$

$$|(S_F(y), w)_{L,F}| \lesssim |y|_{S,F} |w|_{S,F}, \tag{DG6}$$

$$|(\mathcal{D}_{\partial K(F)} y, w)_{L,F}| \lesssim |y|_{S,F} \|w\|_{L,F}, \tag{DG7}$$

where $F \subset \partial K(F)$ and where for all $y \in W(h)$, $|y|_S^2 = \sum_{F \in \mathcal{F}_h^{\mathrm{i}}} |y|_{S,F}^2$ with $|y|_{S,F}^2 = (S_F(y), y)_{L,F}$. A simple way of enforcing (DG5)–(DG7) consists of setting $S_F(y) = |\mathcal{D}_{\partial K(F)}| y$.

Introduce the bilinear form a_h such that for all z, y in $W(h)$,

$$\begin{aligned} a_h(z, y) = & \sum_{K \in \mathcal{T}_h} (Tz, y)_{L,K} + \sum_{F \in \mathcal{F}_h^{\partial}} \tfrac{1}{2} (M_F(z) - \mathcal{D}z, y)_{L,F} \\ & - \sum_{F \in \mathcal{F}_h^{\mathrm{i}}} 2(\{\mathcal{D}z\}, \{y\})_{L,F} + \sum_{F \in \mathcal{F}_h^{\mathrm{i}}} (S_F([\![z]\!]), [\![y]\!])_{L,F}. \end{aligned} \tag{17}$$

Observe that owing to (DG2), the second term in the definition of a_h weakly enforces the boundary conditions in a way which is consistent with (8). The purpose of the third term is to ensure that an L-coercivity property holds on W_h. The last term controls the jump of the discrete solution across interfaces. Some user-dependent arbitrariness appears in the second and fourth term through the definition of the operators M_F and S_F. An equivalent definition of the DG bilinear form obtained by integration by parts is the following:

$$\begin{aligned} a_h(z, y) = & \sum_{K \in \mathcal{T}_h} (z, \tilde{T}y)_{L,K} + \sum_{F \in \mathcal{F}_h^{\partial}} \tfrac{1}{2} (M_F(z) + \mathcal{D}z, y)_{L,F} \\ & + \sum_{F \in \mathcal{F}_h^{\mathrm{i}}} \tfrac{1}{2} ([\![\mathcal{D}z]\!], [\![y]\!])_{L,F} + \sum_{F \in \mathcal{F}_h^{\mathrm{i}}} (S_F([\![z]\!]), [\![y]\!])_{L,F}. \end{aligned} \tag{18}$$

3.3 Convergence analysis

An approximation to the solution of (10) is constructed as follows: For $f \in L$,

$$\begin{cases} \text{Seek } z_h \in W_h \text{ such that} \\ a_h(z_h, y_h) = (f, y_h)_L, \quad \forall y_h \in W_h. \end{cases} \tag{19}$$

The error analysis uses the following discrete norms on $W(h)$,

$$\|y\|_{h,A}^2 = \|y\|_L^2 + |y|_J^2 + |y|_M^2 + \sum_{K \in \mathcal{T}_h} h_K \|Ay\|_{L,K}^2, \tag{20}$$

$$\|y\|_{h,\frac{1}{2}}^2 = \|y\|_{h,A}^2 + \sum_{K \in \mathcal{T}_h} [h_K^{-1} \|y\|_{L,K}^2 + \|y\|_{L,\partial K}^2], \tag{21}$$

where for all $y \in W(h)$, $|y|_J^2 = \sum_{F\in\mathcal{F}_h^{\mathrm{i}}} |y|_{J,F}^2$ with $|y|_{J,F} = |[\![y]\!]|_{S,F}$. The convergence analysis is performed in the spirit of Strang's Second Lemma. The main result is the following

Theorem 4. *Let z solve* (10) *and let z_h solve* (19)*. Assume that for all $k \in \{1,\dots,d\}$, $\mathcal{A}^k \in [\mathfrak{C}^{0,\frac{1}{2}}(\overline{\Omega})]^{m,m}$. Then,*

$$\|z - z_h\|_{h,A} \lesssim \inf_{y_h\in W_h} \|z - y_h\|_{h,\frac{1}{2}}, \tag{22}$$

if $z \in [H^1(\Omega)]^m$, and $\lim_{h\to 0} \|z - z_h\|_L = 0$ if $z \in V$ only, assuming there is $\gamma > 0$ such that $[H^{1+\gamma}(\Omega)]^m \cap V$ is dense in V.

Using standard interpolation results on W_h, the above result implies that

$$\|z - z_h\|_{h,A} \lesssim h^{p+\frac{1}{2}} \|z\|_{[H^{p+1}(\Omega)]^m} \tag{23}$$

whenever z is in $[H^{p+1}(\Omega)]^m$. In particular, $\|z-z_h\|_L$ converges to order $h^{p+\frac{1}{2}}$, and if the mesh family $\{\mathcal{T}_h\}_{h>0}$ is quasi-uniform, $(\sum_{K\in\mathcal{T}_h} \|A(z-z_h)\|_{L,K}^2)^{\frac{1}{2}}$ converges to order h^p. These estimates are identical to those that can be obtained by other stabilization methods like Galerkin/Least-Squares, subgrid viscosity, etc.

3.4 Localization and the notion of fluxes

The purpose of this section is to discuss briefly some equivalent formulations of the discrete problem (19) in order to emphasize the link with other formalisms derived previously for DG methods based on the notion of fluxes (see, e.g., Arnold *et al.* [2]). Let $K \in \mathcal{T}_h$. For $v \in W(h)$ and $x \in \partial K$, set $v^i(x) = \lim_{\substack{y\to x\\ y\in K}} v(y)$, $v^e(x) = \lim_{\substack{y\to x\\ y\notin K}} v(y)$ (with $v^e(x) = 0$ if $x \in \partial\Omega$), and

$$[\![v]\!]_{\partial K}(x) = v^i(x) - v^e(x), \quad \{v\}_{\partial K}(x) = \tfrac{1}{2}(v^i(x) + v^e(x)). \tag{24}$$

The *element flux* of a function v on ∂K, say $\phi_{\partial K}(v) \in [L^2(\partial K)]^m$, is defined on a face $F \subset \partial K$ by

$$\phi_{\partial K}(v)|_F = \begin{cases} \frac{1}{2}M_F(v|_F) + \frac{1}{2}\mathcal{D}v, & \text{if } F \subset \partial K^{\partial}, \\ S_F([\![v]\!]_{\partial K}|_F) + \mathcal{D}_{\partial K}\{v\}_{\partial K}, & \text{if } F \subset \partial K^{\mathrm{i}}, \end{cases} \tag{25}$$

where ∂K^{i} denotes that part of ∂K that lies in Ω and ∂K^{∂} that part of ∂K that lies on $\partial\Omega$. The relevance of the notion of flux is clarified by the following

Proposition 1. *The discrete problem* (19) *is equivalent to each of the following two local formulations:*

$$\begin{cases} \text{Seek } z_h \in W_h \text{ such that } \forall K \in \mathcal{T}_h \text{ and } \forall y_h \in [\mathbb{P}_p(K)]^m, \\ (z_h, \tilde{T}y_h)_{L,K} + (\phi_{\partial K}(z_h), y_h)_{L,\partial K} = (f, y_h)_{L,K}, \end{cases} \tag{26}$$

$$\begin{cases} \text{Seek } z_h \in W_h \text{ such that } \forall K \in \mathcal{T}_h \text{ and } \forall y_h \in [\mathbb{P}_p(K)]^m, \\ (Tz_h, y_h)_{L,K} + (\phi_{\partial K}(z_h) - \mathcal{D}_{\partial K} z_h^i, y_h)_{L,\partial K} = (f, y_h)_{L,K}. \end{cases} \tag{27}$$

In engineering practice, approximation schemes such as (26) are often designed by specifying the element fluxes. The above analysis then provides a practical means to assess the properties of the scheme. Indeed, once the element fluxes are given, the boundary operators M_F and the interface operators S_F can be directly retrieved from (25). Then, properties (DG1)–(DG7) provide sufficient conditions for convergence.

Remark 2. The element fluxes are conservative in the sense that for all $F = K_1(F) \cap K_2(F) \in \mathcal{F}_h^{\mathrm{i}}$, $\phi_{\partial K_1(F)}(v) + \phi_{\partial K_2(F)}(v) = 0$ on F. The concept of conservativity as such does not play any role in the present analysis of DG methods. It plays a role when deriving improved L^2-error estimates by using the Aubin–Nitsche lemma; see, e.g., Arnold *et al.* [2] and §4.3.

4 DG approximation of block Friedrichs' systems

In this section the setting is specialized to Friedrichs' systems endowed with a 2×2 block structure in which some of the unknowns can be eliminated to yield a system of elliptic-like PDE's for the remaining unknowns. A general DG method to approximate such systems is proposed and analyzed. The key feature is that the unknowns that can be eliminated at the continuous level can be also eliminated at the discrete level by solving local problems. To achieve this goal we will see that at variance with the DG method formulated in §3, where jumps and boundary values are equally controlled among the unknowns, the boundary values and jumps of the discrete unknowns to be eliminated must no longer be controlled whereas the boundary values and jumps of the remaining discrete unknowns must be controlled with an $\mathcal{O}(h^{-1})$ weight.

4.1 The setting

We now assume that there are two positive integers m_σ and m_u with $m = m_\sigma + m_u$ such that the $(d+1)$ $\mathbb{R}^{m,m}$-valued fields $\mathcal{K}$ and $\{\mathcal{A}^k\}_{1\le k\le d}$ have the following 2×2 block structure:

$$\mathcal{K} = \left[\begin{array}{c:c} \mathcal{K}^{\sigma\sigma} & \mathcal{K}^{\sigma u} \\ \hdashline \mathcal{K}^{u\sigma} & \mathcal{K}^{uu} \end{array}\right], \qquad \mathcal{A}^k = \left[\begin{array}{c:c} 0 & \mathcal{B}^k \\ \hdashline [\mathcal{B}^k]^t & \mathcal{C}^k \end{array}\right], \tag{28}$$

with obvious notation for the blocks of $\mathcal{K}$ and where for all $k \in \{1,\ldots,d\}$, $\mathcal{B}^k$ is an $m_\sigma \times m_u$ matrix field and $\mathcal{C}^k$ is a symmetric $m_u \times m_u$ matrix field. Define the operators $B = \sum_{k=1}^d \mathcal{B}^k \partial_k$, $\tilde{B} = \sum_{k=1}^d [\mathcal{B}^k]^t \partial_k$, and $C = \sum_{k=1}^d \mathcal{C}^k \partial_k$. The two key hypotheses on which the present work is based are the following:

$$\exists k_0 > 0,\ \forall \xi \in \mathbb{R}^{m_\sigma},\ \xi^t \mathcal{K}^{\sigma\sigma} \xi \ge k_0 \|\xi\|^2_{\mathbb{R}^{m_\sigma}} \quad \text{a.e. on } \Omega, \tag{A5}$$

$$\forall k \in \{1,\ldots,d\},\ \text{the } m_\sigma \times m_\sigma \text{ upper-left block of } \mathcal{A}^k \text{ is zero.} \tag{A6}$$

Set $L_\sigma = [L^2(\Omega)]^{m_\sigma}$ and $L_u = [L^2(\Omega)]^{m_u}$. Consider the PDE system $Tz = f$ with $f \in L = L_\sigma \times L_u$ and partition z and f into (z^σ, z^u) and (f^σ, f^u), respectively. Assumption (A5) (which implies that the matrix $\mathcal{K}^{\sigma\sigma}$ is invertible) together with assumption (A6) allow for the elimination of z^σ from the PDE system, yielding $z^\sigma = [\mathcal{K}^{\sigma\sigma}]^{-1}(f^\sigma - \mathcal{K}^{\sigma u} z^u - Bz^u)$, and it comes that z^u solves the following second-order PDE:

$$\begin{aligned} &- \tilde{B}[\mathcal{K}^{\sigma\sigma}]^{-1} B z^u + (C - \tilde{B}[\mathcal{K}^{\sigma\sigma}]^{-1}\mathcal{K}^{\sigma u} - \mathcal{K}^{u\sigma}[\mathcal{K}^{\sigma\sigma}]^{-1}B) z^u \\ &\qquad + (\mathcal{K}^{uu} - \mathcal{K}^{u\sigma}[\mathcal{K}^{\sigma\sigma}]^{-1}\mathcal{K}^{\sigma u}) z^u = f^u - (\mathcal{K}^{u\sigma} + \tilde{B})[\mathcal{K}^{\sigma\sigma}]^{-1} f^\sigma. \end{aligned} \tag{29}$$

The leading order term in this PDE has a very particular structure since the matrices $(\mathcal{B}^k)^t[\mathcal{K}^{\sigma\sigma}]^{-1}\mathcal{B}^k$ are positive semi-definite. Hence, the PDE's covered hereafter are elliptic-like.

4.2 The design of the DG bilinear form

Let p and p_σ be two non-negative integers such that $p-1 \le p_\sigma \le p$. Define the vector spaces

$$U_h = [P_{h,p}]^{m_u}, \qquad \Sigma_h = [P_{h,p_\sigma}]^{m_\sigma}, \qquad W_h = U_h \times \Sigma_h. \tag{30}$$

Consider the DG bilinear form defined in (17) and the discrete problem (19). Partition the discrete unknown into $z_h = (z_h^\sigma, z_h^u)$. We now want to design a DG method in which z_h^σ can be eliminated by solving local problems. It is then readily seen from (26) that this is possible only if the σ-component of the flux $\phi_{\partial K}(z_h)$ solely depends on z_h^u. Owing to (25), it is inferred that the boundary operators M_F and the interface operators S_F must be such that

$$M_F^{\sigma\sigma} = 0 \qquad \text{and} \qquad S_F^{\sigma\sigma} = 0. \tag{31}$$

Let $U(h) = [H^1(\Omega)]^{m_u} + U_h$. We define the mapping $\theta_h^1 : U(h) \longrightarrow \Sigma_h$ such that for all $z^u \in U(h)$ and for all $K \in \mathcal{T}_h$, $\theta_h^1(z^u)|_K$ solves the following problem: For all $q^\sigma \in [\mathbb{P}_{p_\sigma}(K)]^{m_\sigma}$,

$$\begin{aligned} (\mathcal{K}^{\sigma\sigma}\theta_h^1(z^u), q^\sigma)_{L_\sigma,K} = &- (\mathcal{K}^{\sigma u} z^u + Bz^u, q^\sigma)_{L_\sigma,K} \\ &- (\phi_{\partial K}^\sigma(z^u) - \mathcal{D}_{\partial K}^{\sigma u}(z^u)^i, q^\sigma)_{L_\sigma,\partial K}. \end{aligned} \tag{32}$$

Owing to (A5), this problem is well–posed. Similarly, we define the mapping $\theta_h^2 : L_\sigma \longrightarrow \Sigma_h$ such that for all $f^\sigma \in L_\sigma$ and for all $K \in \mathcal{T}_h$, $\theta_h^2(f^\sigma)|_K$ solves the following local problem: For all $q^\sigma \in [\mathbb{P}_{p_\sigma}(K)]^{m_\sigma}$,

$$(\mathcal{K}^{\sigma\sigma}\theta_h^2(f^\sigma), q^\sigma)_{L_\sigma,K} = (f^\sigma, q^\sigma)_{L_\sigma,K}. \tag{33}$$

Finally, define the bilinear form ϕ_h on $U(h) \times U(h)$ by

$$\phi_h(z^u, y^u) = a_h((\theta_h^1(z^u), z^u), (0, y^u)), \tag{34}$$

and the linear form ψ_h on $U(h)$ by $\psi_h(y^u) = a_h((\theta_h^2(f^\sigma), 0), (0, y^u))$. This readily leads to the following

Proposition 2. *If the pair* (z_h^σ, z_h^u) *solves* (19)*, then,*

$$z_h^\sigma = \theta_h^1(z_h^u) + \theta_h^2(f^\sigma), \tag{35}$$

and z_h^u *solves the following problem:*

$$\begin{cases} \text{Seek } z_h^u \in U_h \text{ such that} \\ \phi_h(z_h^u, y_h^u) = (f^u, y_h^u)_{L_u} - \psi_h(y_h^u), \quad \forall y_h^u \in U_h. \end{cases} \tag{36}$$

Conversely, if z_h^u *solves* (36) *and if* z_h^σ *is defined by* (35)*, then the pair* (z_h^σ, z_h^u) *solves* (19)*.*

For the convergence analysis of §4.3 to hold, the boundary operators M_F and the interface operators S_F must comply with certain design criteria that are formulated in [13]. This set of conditions simplifies into the following whenever Dirichlet-type boundary conditions are enforced on z^u: For all $F \in \mathcal{F}_h^\partial$ and for all $y = (y^\sigma, y^u) \in [L^2(F)]^m$, we assume that

$$(\mathcal{M}y - \mathcal{D}y = 0) \implies (M_F(y) - \mathcal{D}y = 0), \tag{LDG1}$$

$$(\mathcal{M}^t y + \mathcal{D}y = 0) \implies (M_F^*(y) + \mathcal{D}y = 0), \tag{LDG2}$$

$$M_F^{\sigma\sigma} = 0, \quad M_F^{\sigma u}(y^u) = -\mathcal{D}^{\sigma u} y^u, \quad M_F^{u\sigma}(y^\sigma) = \mathcal{D}^{u\sigma} y^\sigma, \tag{LDG3}$$

$$M_F^{uu} \text{ is self-adjoint}, \tag{LDG4}$$

$$h_F \|\mathcal{D}^{uu} y^u\|_{L_u,F}^2 + h_F^{-1} \|\mathcal{D}^{\sigma u} y^u\|_{L_\sigma,F}^2 \lesssim |y^u|_{M,F}^2 \lesssim h_F^{-1} \|y^u\|_{L_u,F}^2, \tag{LDG5}$$

where M_F^* denotes the adjoint operator of M_F and where for all $y^u \in U(h)$, $|y^u|_{M,F}^2 = (M_F^{uu}(y^u), y^u)_{L_u,F}$. Similarly, for all $F \in \mathcal{F}_h^{\mathrm{i}}$ and for all $y = (y^\sigma, y^u) \in [L^2(F)]^m$, we assume that

$$S_F^{\sigma\sigma} = 0, \quad S_F^{\sigma u} = 0, \quad S_F^{u\sigma} = 0, \tag{LDG6}$$

$$S_F^{uu} \text{ is self-adjoint}, \tag{LDG7}$$

$$h_F \|\mathcal{D}^{uu} y^u\|_{L_u,F}^2 + h_F^{-1} \|\mathcal{D}^{\sigma u} y^u\|_{L_\sigma,F}^2 \lesssim |y^u|_{S,F}^2 \lesssim h_F^{-1} \|y^u\|_{L_u,F}^2, \tag{LDG8}$$

where for all $y^u \in U(h)$, $|y^u|_{S,F}^2 = (S_F^{uu}(y^u), y^u)_{L_u,F}$.

Remark 3. Assumption (LDG1) is a consistency assumption similar to (DG2). Assumption (LDG2) is an adjoint-consistency assumption needed to obtain an improved error estimate for z_h^u in the L_u-norm. Assumption (LDG3) is suitable to enforce Dirichlet boundary conditions and must be modified if other boundary conditions are considered. In this case, assumption (LDG5) must also be modified: M_F^{uu} no longer scales as h_F^{-1}, but is of order 1.

4.3 Convergence analysis

The error analysis uses the following discrete norms on $W(h)$,

$$\|y\|_{h,A'}^2 = \|y^\sigma\|_{L_\sigma}^2 + \|y^u\|_{L_u}^2 + |y^u|_J^2 + |y^u|_M^2 + \sum_{K\in\mathcal{T}_h} \|By^u\|_{L_\sigma,K}^2, \tag{37}$$

$$\|y\|_{h,1}^2 = \|y\|_{h,A'}^2 + \sum_{K\in\mathcal{T}_h} [h_K^{-2}\|y^u\|_{L_u,K}^2 + h_K^{-1}\|y^u\|_{L_u,\partial K}^2 + h_K\|y^\sigma\|_{L_\sigma,\partial K}^2], \tag{38}$$

where for all $y^u \in U(h)$, $|y^u|_M^2 = \sum_{F\in\mathcal{F}_h^\partial} |y^u|_{M,F}^2$ and $|y^u|_J^2 = \sum_{F\in\mathcal{F}_h^{\rm i}} |[\![y^u]\!]|_{S,F}^2$. The main result is the following

Theorem 5. *Let z solve* (10) *and let z_h solve* (19). *Assume that for all $k \in \{1,\dots,d\}$, $\mathcal{B}^k \in [\mathfrak{C}^{0,1}(\overline{\Omega})]^{m,m}$. Then*

$$\|z - z_h\|_{h,A'} \lesssim \inf_{y_h\in W_h} \|z - y_h\|_{h,1}, \tag{39}$$

if $z \in [H^1(\Omega)]^m$, and $\lim_{h\to 0}(\|z - z_h\|_L^2 + \sum_{K\in\mathcal{T}_h} \|B(z^u - z_h^u)\|_{L_\sigma,L}^2) = 0$ if $z \in V$ only, assuming there is $\gamma > 0$ s.t. $[H^\gamma(\Omega)]^{m_\sigma}\times[H^{1+\gamma}(\Omega)]^{m_u} \cap V$ is dense in V.

Using standard interpolation results on W_h and since $p-1 \le p_\sigma \le p$, the above result implies

$$\|z - z_h\|_{h,A'} \lesssim h^p\big(\|z^\sigma\|_{[H^{p_\sigma+1}(\Omega)]^{m_\sigma}} + \|z^u\|_{[H^{p+1}(\Omega)]^{m_u}}\big). \tag{40}$$

whenever z is in $[H^{p_\sigma+1}(\Omega)]^{m_\sigma}\times[H^{p+1}(\Omega)]^{m_u}$. In particular, $\|z - z_h\|_L$ converges to order h^p and if the mesh family $\{\mathcal{T}_h\}_{h>0}$ is quasi-uniform, $(\sum_{K\in\mathcal{T}_h} \|B(z^u - z_h^u)\|_{L_\sigma,K}^2)^{\frac{1}{2}}$ also converges to order h^p. If $p_\sigma = p$, the L-norm error estimate is suboptimal when compared with that obtained using the DG method analyzed in §3. The reason for this optimality loss is that the interface jumps of the σ-component are no longer controlled to allow for this component to be locally eliminated, and the jumps of the u-component are penalized with an $\mathcal{O}(h^{-1})$ weight. If $p_\sigma = p-1$, the L-norm error estimate is still suboptimal for the u-component, but is optimal for the σ-component.

To derive an optimal error estimate for the u-component in the L_u-norm, we use a duality argument. Let $\psi \in V^*$ solve

$$\tilde{T}\psi = (0, z^u - z_h^u). \tag{41}$$

Assuming the above problem yields elliptic regularity, i.e., $\|\psi^u\|_{[H^2(\Omega)]^{m_u}} + \|\psi^\sigma\|_{[H^1(\Omega)]^{m_\sigma}} \lesssim \|z^u - z_h^u\|_{L_u}$, the main result is the following

Theorem 6. *The following holds:*

$$\|z^u - z_h^u\|_{L_u} \lesssim h \inf_{y_h\in W_h} \|z - y_h\|_{h,1^+}, \tag{42}$$

where $\|y\|_{h,1^+}^2 = \|y\|_{h,1}^2 + \sum_{K\in\mathcal{T}_h} [h_K^2\|y^\sigma\|_{[H^1(K)]^{m_\sigma}}^2 + h_K\|y^\sigma\|_{L_\sigma,\partial K}^2]$. In particular, if $z \in [H^{p_\sigma+1}(\Omega)]^{m_\sigma}\times[H^{p+1}(\Omega)]^{m_u}$, then

$$\|z^u - z_h^u\|_{L_u} \lesssim h^{p+1}\big(\|z^\sigma\|_{[H^{p_\sigma+1}(\Omega)]^{m_\sigma}} + \|z^u\|_{[H^{p+1}(\Omega)]^{m_u}}\big). \tag{43}$$

5 Examples

In this section we apply the methods formulated in §3 and §4 to various Friedrichs' systems encountered in engineering applications. To alleviate notation, an index h indicates that the norm is broken on the mesh elements and $\mathfrak{h}$ denotes the piecewise constant function equal to h_K on each $K \in \mathcal{T}_h$.

5.1 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 there is $\mu_0 > 0$ such that $\mu(x) - \frac{1}{2}\nabla\cdot\beta(x) \geq \mu_0$ a.e. in Ω. Let $f \in L^2(\Omega)$. The PDE

$$\mu u + \beta\cdot\nabla u = f \tag{44}$$

is recast as a Friedrichs' system by setting $m = 1$, $\mathcal{K} = \mu$, and $\mathcal{A}^k = \beta^k$ for $k \in \{1, \ldots, d\}$. The graph space is $W = \{w \in L^2(\Omega);\ \beta\cdot\nabla w \in L^2(\Omega)\}$. To enforce boundary conditions, define $\partial\Omega^\pm = \{x \in \partial\Omega;\ \pm\beta(x)\cdot n(x) > 0\}$, and assume that $\partial\Omega^-$ and $\partial\Omega^+$ are well-separated, i.e., $\mathrm{dist}(\partial\Omega^-, \partial\Omega^+) > 0$. Then, the boundary operator D has the following representation: For all $v, w \in W$,

$$\langle Dv, w\rangle_{W',W} = \int_{\partial\Omega} vw(\beta\cdot n). \tag{45}$$

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

Let $\alpha > 0$ (α can vary from face to face) and for all $F \in \mathcal{F}_h$, set

$$\mathcal{M}_F = |\beta\cdot n| \qquad \text{and} \qquad \mathcal{S}_F = \alpha|\beta\cdot n_F|, \tag{46}$$

where n_F is a unit normal vector to F (the orientation is irrelevant). Then, letting $M_F(v) = \mathcal{M}_F v$ and $S_F(v) = \mathcal{S}_F v$, assumptions (DG1)–(DG7) hold. Hence, if $\beta \in [\mathcal{C}^{0,\frac{1}{2}}(\Omega)]^d$ and the exact solution is smooth enough,

$$\|u - u_h\|_{L^2(\Omega)} + \|\mathfrak{h}^{\frac{1}{2}}\beta\cdot\nabla(u - u_h)\|_{h,L^2(\Omega)} \lesssim h^{p+\frac{1}{2}}\|u\|_{H^{p+1}(\Omega)}. \tag{47}$$

Remark 4. The specific value $\alpha = \frac{1}{2}$ leads to the so-called upwind scheme. This coincidence has lead many authors to believe that DG methods are methods of choice to solve hyperbolic problems. Actually DG methods, as presented herein, are merely stabilization techniques tailored to solve symmetric positive systems of first-order PDEs.

5.2 Advection–diffusion–reaction

Let μ, β, and f be as in §5.1. Then, the PDE $-\Delta u + \beta\cdot\nabla u + \mu u = f$ written in the mixed form

$$\begin{cases} \sigma + \nabla u = 0, \\ \mu u + \nabla{\cdot}\sigma + \beta{\cdot}\nabla u = f, \end{cases} \tag{48}$$

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

$$\mathcal{K} = \left[\begin{array}{c|c} \mathcal{I}_d & 0 \\ \hline 0 & \mu \end{array}\right], \qquad \mathcal{A}^k = \left[\begin{array}{c|c} 0 & e^k \\ \hline (e^k)^t & \beta^k \end{array}\right], \tag{49}$$

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 $W = H(\text{div};\Omega)\times H^1(\Omega)$.

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

$$\langle D(\sigma, u), (\tau, v)\rangle_{W',W} = \langle \sigma{\cdot}n, v\rangle_{-\frac{1}{2},\frac{1}{2}} + \langle \tau{\cdot}n, u\rangle_{-\frac{1}{2},\frac{1}{2}} + \int_{\partial\Omega} (\beta{\cdot}n)uv, \tag{50}$$

where $\langle,\rangle_{-\frac{1}{2},\frac{1}{2}}$ denotes the duality pairing between $H^{-\frac{1}{2}}(\partial\Omega)$ and $H^{\frac{1}{2}}(\partial\Omega)$. Dirichlet boundary conditions are enforced by setting $\langle M(\sigma, u), (\tau, v)\rangle_{W',W} = \langle \sigma{\cdot}n, v\rangle_{-\frac{1}{2},\frac{1}{2}} - \langle \tau{\cdot}n, u\rangle_{-\frac{1}{2},\frac{1}{2}}$, yielding $V = H(\text{div};\Omega)\times H^1_0(\Omega)$. Neumann and Robin boundary conditions can be treated similarly.

Let $\alpha_1 > 0$, $\alpha_2 > 0$, and $\eta > 0$ (these design parameters can vary from face to face) and for all $F \in \mathcal{F}_h$, set

$$\mathcal{M}_F = \left[\begin{array}{c|c} 0 & -n \\ \hline n^t & \eta \end{array}\right], \qquad \mathcal{S}_F = \left[\begin{array}{c|c} \alpha_1 n_F \otimes n_F & 0 \\ \hline 0 & \alpha_2 \end{array}\right]. \tag{51}$$

Then, letting $M_F(\sigma, u) = \mathcal{M}_F(\sigma, u)$ and $S_F(\sigma, u) = \mathcal{S}_F(\sigma, u)$, assumptions (DG1)–(DG7) hold. Hence, if $\beta \in [\mathcal{C}^{0,\frac{1}{2}}(\Omega)]^d$ and the exact solution is smooth enough,

$$\begin{aligned} \|u - u_h\|_{L^2(\Omega)} + \|\sigma - \sigma_h\|_{[L^2(\Omega)]^d} + \|\mathfrak{h}^{\frac{1}{2}}\nabla(u - u_h)\|_{h,[L^2(\Omega)]^d} \\ + \|\mathfrak{h}^{\frac{1}{2}}\nabla{\cdot}(\sigma - \sigma_h)\|_{h,L^2(\Omega)} \lesssim h^{p+\frac{1}{2}}\|(\sigma, u)\|_{[H^{p+1}(\Omega)]^{d+1}}. \end{aligned} \tag{52}$$

The above Friedrichs' system can be equipped with the 2×2 block structure analyzed in §4 by setting $z^\sigma := \sigma$ and $z^u := u$. Take

$$\mathcal{M}_F = \left[\begin{array}{c|c} 0 & -n \\ \hline n^t & \eta h_F^{-1} \end{array}\right], \qquad \mathcal{S}_F = \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & \alpha_2 h_F^{-1} \end{array}\right]. \tag{53}$$

Then, letting $M_F(\sigma, u) = \mathcal{M}_F(\sigma, u)$ and $S_F(\sigma, u) = \mathcal{S}_F(\sigma, u)$, assumptions (LDG1)–(LDG8) hold. If the exact solution is smooth enough,

$$\begin{aligned} \|u - u_h\|_{L^2(\Omega)} + h\|\sigma - \sigma_h\|_{[L^2(\Omega)]^d} + h\|\nabla(u - u_h)\|_{h,[L^2(\Omega)]^d} \\ \lesssim h^{p+1}\|(\sigma, u)\|_{[H^p(\Omega)]^d\times H^{p+1}(\Omega)}. \end{aligned} \tag{54}$$

Remark 5. Other choices for the operators M_F and S_F are possible. In particular, one can show that the DG method of Brezzi *et al.* [7], that of Bassi and Rebay [6], that of Douglas and Dupont [10], and that of Cockburn and Shu [9] fit into the present framework.

5.3 Linear elasticity

Let ς and γ 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 $\sigma = \frac{1}{2}(\nabla u + (\nabla u)^t) + \frac{1}{\gamma}(\nabla{\cdot}u)\mathcal{I}_d$ and $-\nabla{\cdot}\sigma + \varsigma u = f$ can be written in the following mixed stress–pressure–displacement form

$$\begin{cases} \sigma + \pi\mathcal{I}_d - \frac{1}{2}(\nabla u + (\nabla u)^t) = 0, \\ \mathrm{tr}(\sigma) + (d+\gamma)\pi = 0, \\ -\frac{1}{2}\nabla{\cdot}(\sigma + \sigma^t) + \varsigma u = f. \end{cases} \tag{55}$$

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 $1 \le i,j \le d$ and $[ij] = d(j-1)+i$. Then, (55) falls into the category of Friedrichs' systems by setting $m = d^2+1+d$ and

$$\mathcal{K} = \left[\begin{array}{c:c:c} \mathcal{I}_{d^2} & \mathcal{Z} & 0 \\ \hdashline (\mathcal{Z})^t & (d+\gamma) & 0 \\ \hdashline 0 & 0 & \varsigma\mathcal{I}_d \end{array}\right], \qquad \mathcal{A}^k = \left[\begin{array}{c:c:c} 0 & 0 & \mathcal{E}^k \\ \hdashline 0 & 0 & 0 \\ \hdashline (\mathcal{E}^k)^t & 0 & 0 \end{array}\right], \tag{56}$$

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 s.t. for all $(z := (\overline{\sigma},\pi,u), y := (\overline{\tau},\rho,v)) \in W$,

$$\langle Dz, y\rangle_{W',W} = -\langle \tfrac{1}{2}(\tau + \tau^t){\cdot}n, u\rangle_{-\frac{1}{2},\frac{1}{2}} - \langle \tfrac{1}{2}(\sigma + \sigma^t){\cdot}n, v\rangle_{-\frac{1}{2},\frac{1}{2}}. \tag{57}$$

Letting $\langle Mz, y\rangle_{W',W} = \langle \frac{1}{2}(\tau+\tau^t){\cdot}n, u\rangle_{-\frac{1}{2},\frac{1}{2}} - \langle \frac{1}{2}(\sigma+\sigma^t){\cdot}n, v\rangle_{-\frac{1}{2},\frac{1}{2}}$, 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.

Let $\alpha_1 > 0$, $\alpha_2 > 0$, and $\eta > 0$ (these design parameters can vary from face to face) and for all $F \in \mathcal{F}_h$, set

$$\mathcal{M}_F = \left[\begin{array}{c:c:c} 0 & 0 & -\mathcal{H} \\ \hdashline 0 & 0 & 0 \\ \hdashline \mathcal{H}^t & 0 & \eta\mathcal{I}_d \end{array}\right], \qquad \mathcal{S}_F = \left[\begin{array}{c:c:c} \alpha_1\mathcal{H}_F{\cdot}\mathcal{H}_F^t & 0 & 0 \\ \hdashline 0 & 0 & 0 \\ \hdashline 0 & 0 & \alpha_2\mathcal{I}_d \end{array}\right], \tag{58}$$

where $\mathcal{H} = \sum_{k=1}^d n_k\mathcal{E}^k \in \mathbb{R}^{d^2,d}$ and $\mathcal{H}_F$ is defined similarly by substituting n_F to n. Then, letting $M_F(\overline{\sigma},\pi,u) = \mathcal{M}_F(\overline{\sigma},\pi,u)$ and $S_F(\overline{\sigma},\pi,u) = \mathcal{S}_F(\overline{\sigma},\pi,u)$, assumptions (DG1)–(DG7) hold. Hence, if the exact solution is smooth enough,

$$\begin{aligned} &\|u-u_h\|_{[L^2(\Omega)]^d} + \|\pi-\pi_h\|_{L^2(\Omega)} + \|\sigma-\sigma_h\|_{[L^2(\Omega)]^{d,d}} \\ &\quad + \|\mathfrak{h}^{\frac{1}{2}}\nabla(u-u_h)\|_{h,[L^2(\Omega)]^{d,d}} + \|\mathfrak{h}^{\frac{1}{2}}\nabla{\cdot}((\sigma+\sigma^t)-(\sigma_h+\sigma_h^t))\|_{h,[L^2(\Omega)]^d} \\ &\qquad \lesssim h^{p+\frac{1}{2}}\|(\overline{\sigma},\pi,u)\|_{[H^{p+1}(\Omega)]^{d^2+1+d}}. \end{aligned} \tag{59}$$

The above Friedrichs' system can be equipped with the 2×2 block structure analyzed in §4 by setting $z^\sigma := (\overline{\sigma}, \pi)$ and $z^u := u$. Take

$$\mathcal{M}_F = \left[\begin{array}{c:c:c} 0 & 0 & -\mathcal{H} \\ \hdashline 0 & 0 & 0 \\ \hdashline \mathcal{H}^t & 0 & \eta h_F^{-1}\mathcal{I}_d \end{array}\right], \qquad \mathcal{S}_F = \left[\begin{array}{c:c:c} 0 & 0 & 0 \\ \hdashline 0 & 0 & 0 \\ \hdashline 0 & 0 & \alpha_2 h_F^{-1}\mathcal{I}_d \end{array}\right]. \tag{60}$$

Then, letting $M_F(\overline{\sigma}, \pi, u) = \mathcal{M}_F(\overline{\sigma}, \pi, u)$ and $S_F(\overline{\sigma}, \pi, u) = \mathcal{S}_F(\overline{\sigma}, \pi, u)$, assumptions (LDG1)–(LDG8) hold. Hence, if the exact solution is smooth enough,

$$\begin{aligned} &\|u - u_h\|_{[L^2(\Omega)]^d} + h\|\pi - \pi_h\|_{L^2(\Omega)} + h\|\sigma - \sigma_h\|_{[L^2(\Omega)]^{d,d}} \\ &\quad + h\|\nabla(u - u_h)\|_{h,[L^2(\Omega)]^{d,d}} \lesssim h^{p+1}\|(\overline{\sigma}, \pi, u)\|_{[H^p(\Omega)]^{d^2+1}\times[H^{p+1}(\Omega)]^d}. \end{aligned} \tag{61}$$

5.4 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

$$\mu H + \nabla\times E = f, \qquad \sigma E - \nabla\times H = g, \tag{62}$$

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

$$\mathcal{K} = \left[\begin{array}{c:c} \mu\mathcal{I}_3 & 0 \\ \hdashline 0 & \sigma\mathcal{I}_3 \end{array}\right], \qquad \mathcal{A}^k = \left[\begin{array}{c:c} 0 & \mathcal{R}^k \\ \hdashline (\mathcal{R}^k)^t & 0 \end{array}\right], \tag{63}$$

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 $W = H(\mathrm{curl}; \Omega)\times H(\mathrm{curl}; \Omega)$. The boundary operator D is such that for all $(H, E), (h, e) \in W$,

$$\begin{aligned} \langle D(H, E), (h, e)\rangle_{W',W} = (\nabla\times E, h)_{[L^2(\Omega)]^3} - (E, \nabla\times h)_{[L^2(\Omega)]^3} \\ + (H, \nabla\times e)_{[L^2(\Omega)]^3} - (\nabla\times H, e)_{[L^2(\Omega)]^3}. \end{aligned} \tag{64}$$

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

Let $\alpha_1 > 0$, $\alpha_2 > 0$, and $\eta > 0$ (these design parameters can vary from face to face) and for all $F \in \mathcal{F}_h$, set

$$\mathcal{M}_F = \left[\begin{array}{c:c} 0 & -\mathcal{N} \\ \hdashline \mathcal{N}^t & \eta\mathcal{N}^t\mathcal{N} \end{array}\right], \qquad \mathcal{S}_F = \left[\begin{array}{c:c} \alpha_1\mathcal{N}_F^t\mathcal{N}_F & 0 \\ \hdashline 0 & \alpha_2\mathcal{N}_F^t\mathcal{N}_F \end{array}\right], \tag{65}$$

where $\mathcal{N} = \sum_{k=1}^{d} n_k \mathcal{R}^k \in \mathbb{R}^{3,3}$ and $\mathcal{N}_F$ is defined similarly by substituting n_F to n. Then, letting $M_F(H,E) = \mathcal{M}_F(H,E)$ and $S_F(H,E) = \mathcal{S}_F(H,E)$, assumptions (DG1)–(DG7) hold. Hence, if the exact solution is smooth enough,

$$\|E - E_h\|_{[L^2(\Omega)]^3} + \|H - H_h\|_{[L^2(\Omega)]^3} + \|\mathfrak{h}^{\frac{1}{2}} \nabla\times(E - E_h)\|_{h,[L^2(\Omega)]^3} + \|\mathfrak{h}^{\frac{1}{2}} \nabla\times(H - H_h)\|_{h,[L^2(\Omega)]^3} \lesssim h^{p+\frac{1}{2}} \|(H,E)\|_{[H^{p+1}(\Omega)]^6}. \quad (66)$$

The above Friedrichs' system can also be equipped with the 2×2 block structure analyzed in §4 by setting $z^\sigma := H$ and $z^u := E$. Take

$$\mathcal{M}_F = \left[\begin{array}{c|c} 0 & -\mathcal{N} \\ \hline \mathcal{N}^t & \eta h_F^{-1} \mathcal{N}^t \mathcal{N} \end{array}\right], \qquad \mathcal{S}_F = \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & \alpha_2 h_F^{-1} \mathcal{N}_F^t \mathcal{N}_F \end{array}\right]. \quad (67)$$

Then, letting $M_F(H,E) = \mathcal{M}_F(H,E)$ and $S_F(H,E) = \mathcal{S}_F(H,E)$, assumptions (LDG1)–(LDG8) hold. Hence, if the exact solution is smooth enough,

$$\|E - E_h\|_{[L^2(\Omega)]^3} + h\|H - H_h\|_{[L^2(\Omega)]^3} + h\|\nabla\times(E - E_h)\|_{h,[L^2(\Omega)]^3} \lesssim h^{p+1} \|(H,E)\|_{[H^p(\Omega)]^3 \times [H^{p+1}(\Omega)]^3}. \quad (68)$$

6 Concluding remarks

In this paper we have presented a unified analysis of DG methods by making systematic use of Friedrichs' systems. As already pointed out by Friedrichs, such systems go beyond the traditional hyperbolic/elliptic classification of PDE's. Furthermore, DG methods as presented herein appear to be merely stabilization methods where the boundary operators M_F and the interface operators S_F have to be set (tuned) by the user so as to comply with the design criteria (DG1)–(DG7) or (LDG1)–(LDG8).

References

1. D.N. Arnold. An interior penalty finite element method with discontinuous elements. *SIAM J. Numer. Anal.*, 19:742–760, 1982.
2. D.N. Arnold, F. Brezzi, B. Cockburn, and L.D. Marini. Unified analysis of discontinuous Galerkin methods for elliptic problems. *SIAM J. Numer. Anal.*, 39(5):1749–1779, 2001/02.
3. I. Babuška. The finite element method with penalty. *Math. Comp.*, 27:221–228, 1973.
4. I. Babuška and M. Zlámal. Nonconforming elements in the finite element method with penalty. *SIAM J. Numer. Anal.*, 10(5):863–875, 1973.
5. G.A. Baker. Finite element methods for elliptic equations using nonconforming elements. *Math. Comp.*, 31(137):45–59, 1977.

6. F. Bassi and S. Rebay. A high-order accurate discontinuous finite element method for the numerical solution of the compressible Navier-Stokes equations. *J. Comput. Phys.*, 131(2):267–279, 1997.
7. F. Brezzi, M. Manzini, L.D. Marini, P. Pietra, and A. Russo. Discontinous Finite Elements for Diffusion Problems. In *Francesco Brioschi (1824-1897). Convegno di studi matematici*, volume 16 of *Incontri di Studio*, pages 197–217. Istituto Lombardo Accademia di Scienze e Lettere, 1999.
8. B. Cockburn, G.E. Karniadakis, and C.W. Shu. *Discontinuous Galerkin Methods - Theory, Computation and Applications*, volume 11 of *Lecture Notes in Computer Science and Engineering*. Springer, 2000.
9. B. Cockburn and C.W. Shu. The local discontinuous Galerkin method for time-dependent convection-diffusion systems. *SIAM J. Numer. Anal.*, 35:2440–2463, 1998.
10. 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.
11. A. Ern and J.-L. Guermond. *Theory and Practice of Finite Elements*, volume 159 of *Applied Mathematical Sciences*. Springer-Verlag, New York, NY, 2004.
12. A. Ern and J.-L. Guermond. Discontinuous Galerkin methods for Friedrichs' systems. I. General theory. *SIAM J. Numer. Anal.*, 2005. In press.
13. A. Ern and J.-L. Guermond. Discontinuous Galerkin methods for Friedrichs' systems. II. Second-order PDEs. *SIAM J. Numer. Anal.*, 2005. submitted.
14. A. Ern, J.-L. Guermond, and G. Caplain. An intrinsic criterion for the bijectivity of Hilbert operators related to Friedrichs' systems. *Comm. Partial. Differ. Equ.*, 2005. submitted.
15. K.O. Friedrichs. Symmetric positive linear differential equations. *Comm. Pure Appl. Math.*, 11:333–418, 1958.
16. C. Johnson, U. Nävert, and J. Pitkäranta. Finite element methods for linear hyperbolic equations. *Comput. Methods Appl. Mech. Engrg.*, 45:285–312, 1984.
17. P. Lesaint. *Sur la résolution des sytèmes hyperboliques du premier ordre par des méthodes d'éléments finis*. PhD thesis, University of Paris VI, 1975.
18. 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*, pages 89–123. Publication No. 33. Math. Res. Center, Univ. of Wisconsin-Madison, Academic Press, New York, 1974.
19. J. Nitsche. Über ein Variationsprinzip zur Lösung von Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unterworfen sind. *Abh. Math. Sem. Univ. Hamburg*, 36:9–15, 1971.
20. W.H. Reed and T.R. Hill. Triangular mesh methods for the neutron transport equation. Technical Report LA-UR-73-479, Los Alamos Scientific Laboratory, Los Alamos, NM, 1973.
21. M.F. Wheeler. An elliptic collocation-finite element method with interior penalties. *SIAM J. Numer. Anal.*, 15:152–161, 1978.