

# Stabilization mechanisms in Discontinuous Galerkin finite element methods

*F. Brezzi<sup>1,2</sup>, B. Cockburn<sup>3</sup>, L.D. Marini<sup>1,2</sup>, and E. Süli<sup>4</sup>*

## Abstract

In this paper we propose a new general framework for the construction and the analysis of Discontinuous Galerkin (DG) methods which reveals a basic mechanism, responsible for certain distinctive stability properties of DG methods. We show that this mechanism is common to apparently unrelated stabilizations, including jump penalty, upwinding, and Hughes–Franca type residual-based stabilizations.

## 1 Introduction

The aim of this paper is to introduce a new perspective on the construction of Discontinuous Galerkin (DG) methods for linear partial differential equations. Here, we present DG methods as methods that define an approximation by means of a variational formulation that tries to enforce the partial differential equation together with the boundary and continuity conditions satisfied by the exact solution. In this way, the DG method establishes a linear relationship between the residual of the approximation inside each element and its jumps across interelement boundaries. We then uncover a simple and general approach devised to ensure the desired stability properties of the resulting methods. In doing so, we identify the ingredients which enforce those properties; the introduction of such ingredients is what is commonly called *stabilization* of the method. This approach allows us to realize that, in the context of discontinuous Galerkin methods, the use of jump

---

<sup>1</sup>Dipartimento di Matematica, Università di Pavia, Via Ferrata 1, 27100 Pavia, Italy

<sup>2</sup>IMATI del CNR, Via Ferrata 1, 27100 Pavia, Italy

<sup>3</sup>School of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455, U.S.A.

<sup>4</sup>University of Oxford, Computing Laboratory, Wolfson Building, Parks Road, Oxford OX1 3QD, United Kingdom

penalties, upwinding and Hughes–Franca type residual-based stabilizations are all different forms of the same mechanism.

Let us briefly illustrate this approach on the model problem of Poisson’s equation with homogeneous Dirichlet boundary condition.

Given a domain  $\Omega$  in  $\mathbb{R}^2$  which, for simplicity, we assume for the moment to be a convex polygon, and an element  $f$  in  $L^2(\Omega)$ , the problem reads: *Find  $u$  in  $H_0^1(\Omega)$  such that*

$$-\Delta u = f \quad \text{in } \Omega, \quad (1.1)$$

$$u = 0 \quad \text{on } \partial\Omega. \quad (1.2)$$

First of all, well-known regularity results tell us that the unique solution  $u$  of (1.1)-(1.2) belongs to  $H^2(\Omega) \cap H_0^1(\Omega)$ . Let us consider now a regular sequence  $\{\mathcal{T}_h\}_h$  of decompositions of  $\Omega$  into triangles  $T$  (see e.g. [15]), and let  $\mathcal{E}_h$  be the union of the edges of the triangles in  $\mathcal{T}_h$ . We can then consider the space

$$H^2(\mathcal{T}_h) := \{v \in L^2(\Omega) \text{ such that } v|_T \in H^2(T) \quad \forall T \in \mathcal{T}_h\} \quad (1.3)$$

and look, a priori, for a solution (which we still denote by  $u$ ) that belongs to  $H^2(\mathcal{T}_h)$ . This is very reasonable, given that we shall approximate  $u$  by means of a discontinuous piecewise polynomial function  $u_h$ . To do so, however, Problem (1.1)-(1.2) will have to be written in a different form: it will amount to requiring that  $u$  satisfies  $-\Delta u = f$  in each  $T$  and that, simultaneously, the jumps of  $u$  and  $\partial u / \partial \mathbf{n}$  vanish on the interelement boundaries. To express this in a more precise way we need to introduce suitable notational conventions.

We use the notation employed in [2]. Let  $e$  be an interior edge shared by elements  $T_1$  and  $T_2$ . Define the unit normal vectors  $\mathbf{n}^1$  and  $\mathbf{n}^2$  on  $e$  pointing exterior to  $T_1$  and  $T_2$ , respectively. For a function  $\varphi$ , piecewise smooth on  $\mathcal{T}_h$ , with  $\varphi^i := \varphi|_{T_i}$  we define

$$\{\varphi\} = \frac{1}{2}(\varphi^1 + \varphi^2), \quad \llbracket \varphi \rrbracket = \varphi^1 \mathbf{n}^1 + \varphi^2 \mathbf{n}^2 \quad \text{on } e \in \mathcal{E}_h^\circ, \quad (1.4)$$

where  $\mathcal{E}_h^\circ$  is the set of interior edges  $e$ . For a vector-valued function  $\boldsymbol{\tau}$ , piecewise smooth on  $\mathcal{T}_h$ , with analogous meanings for  $\boldsymbol{\tau}^1$  and  $\boldsymbol{\tau}^2$ , we define

$$\{\boldsymbol{\tau}\} = \frac{1}{2}(\boldsymbol{\tau}^1 + \boldsymbol{\tau}^2), \quad \llbracket \boldsymbol{\tau} \rrbracket = \boldsymbol{\tau}^1 \cdot \mathbf{n}^1 + \boldsymbol{\tau}^2 \cdot \mathbf{n}^2 \quad \text{on } e \in \mathcal{E}_h^\circ. \quad (1.5)$$

Notice that the jump  $\llbracket \varphi \rrbracket$  of the scalar function  $\varphi$  across  $e \in \mathcal{E}_h^\circ$  is a vector parallel to the normal to  $e$ , and the jump  $\llbracket \boldsymbol{\tau} \rrbracket$  of the vector function  $\boldsymbol{\tau}$  is a scalar quantity. The advantage of these definitions is that they do not depend on the ordering that is assigned to the elements  $T_i$ . For  $e \in \mathcal{E}_h^\partial$ , the set of boundary edges, we let

$$\llbracket \varphi \rrbracket = \varphi \mathbf{n}, \quad \{\boldsymbol{\tau}\} = \boldsymbol{\tau} \quad \text{on } e \in \mathcal{E}_h^\partial. \quad (1.6)$$

The quantities  $\{\varphi\}$  or  $\llbracket \boldsymbol{\tau} \rrbracket$  on boundary edges are defined analogously.

We can now rewrite problem (1.1)-(1.2) as

$$-\Delta u = f \quad \text{in each } T \in \mathcal{T}_h, \quad (1.7)$$

$$\llbracket u \rrbracket|_e = 0 \quad \text{on each } e \in \mathcal{E}_h, \quad (1.8)$$

$$\llbracket \nabla u \rrbracket|_e = 0 \quad \text{on each } e \in \mathcal{E}_h^\circ. \quad (1.9)$$

Next, we introduce a variational formulation in which each of the three equations above is thought to have the same relevance, and is, therefore, treated in the same fashion. To do so, we assume that we are given three linear operators  $B_0$ ,  $\mathbf{B}_1$  and  $B_2$  from  $\mathbf{H}^2(\mathcal{T}_h)$  to  $L^2(\Omega)$ ,  $(L^2(\mathcal{E}_h))^2$  and  $L^2(\mathcal{E}_h^\circ)$ , respectively, and we consider the problem: *Find*  $u \in \mathbf{H}^2(\mathcal{T}_h)$  *such that*

$$(-\Delta u - f, B_0 v)_{\mathcal{T}_h} + (\llbracket u \rrbracket, \mathbf{B}_1 v)_{\mathcal{E}_h} + (\llbracket \nabla u \rrbracket, B_2 v)_{\mathcal{E}_h^\circ} = 0 \quad \forall v \in \mathbf{H}^2(\mathcal{T}_h), \quad (1.10)$$

with obvious meanings of the symbols. Note that, as  $\llbracket u \rrbracket$  is a vector, then  $\mathbf{B}_1 v$  has to be a vector as well; however, as  $\llbracket u \rrbracket$  is directed along the normal, then only the normal component of  $\mathbf{B}_1$  will be relevant in the method.

It is not difficult to see (and this will be done in more detail in the next section) that if we choose  $B_0 v = v$  and  $B_2 v = \{v\}$ , integrate the first term in (1.10) by parts, and discretize the resulting identity by *continuous* finite elements, we obtain the usual conforming finite element formulation.

Now, observe that the solution  $u$  to our original problem (1.1)-(1.2) is always a solution to problem (1.10). Hence, if we ensure uniqueness of solution to the latter problem, such a solution will be nothing but the solution to the original problem. Let us consider a simple example. A general class of admissible choices will be discussed in the next section. Assume that we take:

$$B_0 v = v, \quad \mathbf{B}_1 v = \{\nabla v\} \quad \text{and} \quad B_2 v = \{v\}. \quad (1.11)$$

Inserting (1.11) into (1.10), integrating the first term by parts, and manipulating terms (in a way that will be made clear in the next section) we obtain

$$(\nabla u, \nabla v)_{\mathcal{T}_h} + (\llbracket u \rrbracket, \{\nabla v\})_{\mathcal{E}_h} - (\llbracket v \rrbracket, \{\nabla u\})_{\mathcal{E}_h} = (f, v)_{\mathcal{T}_h} \quad \forall v \in \mathbf{H}^2(\mathcal{T}_h). \quad (1.12)$$

Now, (1.12) is precisely the weak formulation which the Baumann–Oden method [7] is based upon. It is not difficult to see that (1.12) has a unique solution (and this will be discussed in more detail in Section 2).

Notice, however, that uniqueness may be lost under discretization. Indeed, if we select a finite-dimensional subspace  $W_h$  of  $\mathbf{H}^2(\mathcal{T}_h)$  consisting of discontinuous piecewise polynomial functions defined on  $\mathcal{T}_h$ , restrict the bilinear form on the left-hand side of (1.12) to  $W_h \times W_h$  and set  $f = 0$  on the

right, the resulting homogeneous problem does not, in general, have the trivial solution as its unique solution in  $W_h$ : to be more precise, if, for instance,  $W_h$  consists of piecewise linear functions, the solution of the discretized problem will be unique if, and only if, there exists an internal vertex shared by an odd number of triangles (see e.g. [8]).

We can avoid this degenerate situation by taking, for  $s_1 > 0$ ,

$$B_0 v = v, \quad \mathbf{B}_1 v = \{\nabla v\} + s_1 \llbracket v \rrbracket \quad \text{and} \quad B_2 v = \{v\}. \quad (1.13)$$

Indeed, in this case, the discretized problem will be: *Find  $u_h \in W_h$  such that*

$$(\nabla u_h, \nabla v)_{\mathcal{T}_h} + (\llbracket u_h \rrbracket, \{\nabla v\} + s_1 \llbracket v \rrbracket)_{\mathcal{E}_h} - (\llbracket v \rrbracket, \{\nabla u_h\})_{\mathcal{E}_h} = (f, v)_{\mathcal{T}_h}, \quad (1.14)$$

for all  $v \in W_h$ . Thus, when  $f = 0$ , we get, after selecting  $v = u_h$ ,

$$(\nabla u_h, \nabla u_h)_{\mathcal{T}_h} + s_1 (\llbracket u_h \rrbracket, \llbracket u_h \rrbracket)_{\mathcal{E}_h} = 0.$$

This immediately implies that  $u_h = 0$  and hence (1.14) has a unique solution in  $W_h$ , provided  $s_1 > 0$ .

We note that changing the definition of the operator  $\mathbf{B}_1$  from (1.11) to (1.13) is equivalent to supplementing the variational formulation (1.12) by the term

$$s_1 (\llbracket u \rrbracket, \llbracket v \rrbracket)_{\mathcal{E}_h},$$

which is commonly considered to be a penalization of the jumps. Since this added term enhances the stability of the method, it is natural to consider it as *stabilization* of the finite element method that would have resulted from the weak formulation (1.12). We also note that from the point of view proposed here, the jump  $\llbracket u \rrbracket$  is nothing but the residual associated with the second of our equations, that is, (1.8). Thus, the above stabilization is considered to be a residual-based stabilization.

This immediately suggests that residual-based stabilizations can be obtained which *are based on supplementing the bilinear form by terms containing the (piecewise) inner product of one of the three residuals and of a suitable operator acting on the test function*. In our case, this amounts to properly defining the operators  $B_0$ ,  $\mathbf{B}_1$  and  $B_2$  so that the resulting weak formulation contains terms of the type

$$s_0 (\Delta u + f, \Delta v)_{\mathcal{T}_h} \quad \text{or} \quad s_1 (\llbracket u \rrbracket, \llbracket v \rrbracket)_{\mathcal{E}_h} \quad \text{or} \quad s_2 (\llbracket \nabla u \rrbracket, \llbracket \nabla v \rrbracket)_{\mathcal{E}_h^o}, \quad (1.15)$$

where the  $s_j$ , for  $j = 0, 1, 2$ , are suitable coefficients that may (and will often do) depend on the local mesh-size. In this paper, we show that this approach allows us to view all the classical stabilization techniques for DG methods as different incarnations of a single mechanism.

The outline of the paper is as follows. In the next section we shall deal with linear second-order elliptic operators (in divergence form). We shall consider both primal and mixed formulations, and we will highlight the relations between the present approach and the more classical flux formulations. In the third section we shall deal with linear first-order hyperbolic equations. A remark will briefly address the case of strongly advection-dominated advection-diffusion equations.

## 2 Elliptic equations

Let  $\Omega$  be a bounded open polygonal domain in  $\mathbb{R}^2$ , and let  $\alpha$  be a smooth function in  $\Omega$  such that there exist constants  $\alpha_0$  and  $\alpha_1$  with

$$0 < \alpha_0 \leq \alpha(x) \leq \alpha_1 \quad \forall x \in \Omega. \quad (2.1)$$

We consider the differential operator

$$\mathcal{A}v := -\operatorname{div}(\alpha \nabla v). \quad (2.2)$$

It is well-known that for every  $f \in L^2(\Omega)$  there exists a unique solution  $u$  in  $H_0^1(\Omega) \cap H^2(\Omega)$  of the problem

$$\mathcal{A}u = f \text{ in } \Omega, \quad u = 0 \text{ on } \partial\Omega. \quad (2.3)$$

It is also well-known that on defining  $\boldsymbol{\sigma} = \boldsymbol{\sigma}(u)$  by

$$\boldsymbol{\sigma} := \alpha \nabla u \quad (2.4)$$

we have that  $\boldsymbol{\sigma} \in H^1(\Omega)$  and therefore the normal component of  $\boldsymbol{\sigma}$  across any straight line intersecting  $\Omega$  is continuous. If now  $\mathcal{T}_h$  is a decomposition of  $\Omega$  into triangles, we consider again the space  $H^2(\mathcal{T}_h)$  as defined in (??)

**Remark.** The convexity assumption on  $\Omega$  is clearly much stronger than necessary, and could surely be weakened. For instance, assuming that  $\Omega$  is just a bounded open polygon we have that it exists an  $s > 3/2$  such that  $u \in H_0^1(\Omega) \cap H^s(\Omega)$ . We can therefore consider, for instance, the space

$$\mathcal{H} := \{v \in L^2(\Omega) : v|_T \in H^s(T) \quad \forall T \in \mathcal{T}_h\}. \quad (2.5)$$

We could also take more general cases, as the one of *mixed boundary conditions*, changing from Dirichlet to Neumann at a vertex, or, with some additional work, possibly using Besov spaces, other more general conditions. However this would be well beyond the aims of this paper, that is concentrated on the meaning of stabilizing mechanisms for discontinuous piecewise polynomial approximations. ■

## 2.1 Primal formulations

It is well-known that a function  $u \in H_{\mathcal{A}}^1(\mathcal{T}_h)$  will coincide with the solution of (2.3) if, and only if,

$$\mathcal{A}u = f \quad \text{in each } T \in \mathcal{T}_h, \quad (2.6)$$

$$\llbracket u \rrbracket|_e = \mathbf{0} \quad \text{on each } e \in \mathcal{E}_h, \quad (2.7)$$

$$\llbracket \boldsymbol{\sigma}(u) \rrbracket|_e = 0 \quad \text{on each } e \in \mathcal{E}_h^\circ. \quad (2.8)$$

Now, as in the previous section, we select three operators  $B_0$ ,  $\mathbf{B}_1$ , and  $B_2$  defined on  $H_{\mathcal{A}}^1(\mathcal{T}_h)$ , and consider the variational problem: Find  $u \in H_{\mathcal{A}}^1(\mathcal{T}_h)$  such that

$$(\mathcal{A}u - f, B_0v)_{\mathcal{T}_h} + (\llbracket u \rrbracket, \mathbf{B}_1v)_{\mathcal{E}_h} + (\llbracket \boldsymbol{\sigma}(u) \rrbracket, B_2v)_{\mathcal{E}_h^\circ} = 0 \quad \forall v \in H_{\mathcal{A}}^1(\mathcal{T}_h), \quad (2.9)$$

where we used the notation

$$(f, g)_{\mathcal{T}_h} := \sum_{T \in \mathcal{T}_h} \int_T f g \, dx \quad (2.10)$$

together with

$$(\mathbf{f}, \mathbf{g})_{\mathcal{E}_h} := \sum_{e \in \mathcal{E}_h} \int_e \mathbf{f} \cdot \mathbf{g} \, ds \quad \text{and} \quad (f, g)_{\mathcal{E}_h^\circ} := \sum_{e \in \mathcal{E}_h^\circ} \int_e f g \, ds. \quad (2.11)$$

It is interesting to note that (2.9) can be seen as *a way of enforcing a linear relation* (depending on the choice of the *B-operators*:  $B_0$ ,  $\mathbf{B}_1$  and  $B_2$ ) *among the three residuals of the three equations (2.6), (2.7) and (2.8)*. Since a similar property also holds for all DG methods defined by means of numerical traces, see [16], this seems to be a very reasonable framework for the study of a wide class of DG methods. Now, setting

$$D_0 := \prod_{T \in \mathcal{T}_h} C_0^\infty(T), \quad \mathbf{D}_1 := \prod_{e \in \mathcal{E}_h} C_0^\infty(e) \mathbf{n}_e, \quad D_2 := \prod_{e \in \mathcal{E}_h^\circ} C_0^\infty(e), \quad (2.12)$$

where  $\mathbf{n}_e$  is a unit normal to  $e$  (its orientation is immaterial, as it does not change the space  $\mathbf{D}_1$ ), we have the following *sufficient condition* for the uniqueness of the solution of (2.9).

**Theorem 1** *Assume that:*

$$\forall \varphi \in D_0 \exists v \in H_{\mathcal{A}}^1(\mathcal{T}_h) \text{ such that } B_0v = \varphi, \quad \mathbf{B}_1v = \mathbf{0}, \quad B_2v = 0, \quad (2.13)$$

$$\forall \boldsymbol{\psi} \in \mathbf{D}_1 \exists v \in H_{\mathcal{A}}^1(\mathcal{T}_h) \text{ such that } (\mathbf{B}_1v - \boldsymbol{\psi}) \cdot \mathbf{n} = 0, \quad B_2v = 0, \quad (2.14)$$

$$\forall \chi \in D_2 \exists v \in H_{\mathcal{A}}^1(\mathcal{T}_h) \text{ such that } B_2 v = \chi. \quad (2.15)$$

Then problem (2.9) has a unique solution that coincides with the solution of (2.3).

**Proof.** The proof is an easy exercise. First, one uses the density of  $C_0^\infty(T)$  in  $L^2(T)$  (for each  $T \in \mathcal{T}_h$ ) to obtain that  $\mathcal{A}u = f$  in each  $T$ . Hence the first term in (2.9) vanishes. Next, we use the density of  $C_0^\infty(e)$  in  $L^2(e)$  (for each  $e \in \mathcal{E}_h$ ) to obtain that  $\llbracket u \rrbracket = \mathbf{0}$  on each  $e \in \mathcal{E}_h$ . Hence the second term in (2.9) also vanishes. Finally, we use the density of  $C_0^\infty(e)$  in  $L^2(e)$  (for each  $e \in \mathcal{E}_h^\circ$ ) to obtain that  $\llbracket \boldsymbol{\sigma}(u) \rrbracket = 0$  on each  $e \in \mathcal{E}_h^\circ$ . Thus we deduce that any solution of (2.9) has to satisfy (2.6)-(2.8), and hence it has to coincide with the (unique) solution of (2.3). It is clear that the same conclusion could also be reached by exchanging the order of the three operators  $B_0$ ,  $\mathbf{B}_1$ , and  $B_2$ . ■

It is clear that the condition of Theorem 1 is just a sufficient condition, and it is far from being necessary. For instance, it is less than obvious how to apply it in order to deduce the (obvious) uniqueness of the solution of (2.9) with the choice

$$B_0 v = \mathcal{A}v, \quad \mathbf{B}_1 v = \llbracket v \rrbracket \quad \text{and} \quad B_2 v = \llbracket \boldsymbol{\sigma}(v) \rrbracket.$$

On the other hand, if one starts by taking  $B_0 v = v$ , then condition (2.13) holds trivially. Moreover, one can pick up almost any combination of boundary operators  $\llbracket v \rrbracket$ ,  $\{v\}$ ,  $\llbracket \alpha \nabla v \rrbracket$ , and  $\{\alpha \nabla v\}$ , and easily ensure properties (2.14) and (2.15). Indeed, we have the following obvious proposition.

**Proposition 1** *For every triangle  $T$ , for every edge  $e$  of  $T$  and for every  $\phi \in C_0^\infty(e)$  there exist two functions  $v_1$  and  $v_2$  in  $H^2(T)$  such that*

$$v_{1|e} = \phi, \quad (\alpha \nabla v_1 \cdot \mathbf{n})|_e = 0 \text{ and } v_{2|e} = 0, \quad (\alpha \nabla v_2 \cdot \mathbf{n})|_e = \phi. \quad (2.16)$$

It transpires from Theorem 1, however, that one should *not* take any of the operators  $B_j$  to be identically zero, because this would amount to *omitting the corresponding equation* in (2.6)-(2.8).

Let us discuss now in more detail what happens in the case of the first basic choice,  $B_0 v = v$ . In this case, one can integrate the first term in (2.9) by parts to obtain

$$\sum_{T \in \mathcal{T}_h} \int_T \mathcal{A}u v \, dx = \sum_{T \in \mathcal{T}_h} \int_T \alpha \nabla u \cdot \nabla v \, dx - \sum_{T \in \mathcal{T}_h} \int_{\partial T} v \boldsymbol{\sigma}(u) \cdot \mathbf{n} \, ds. \quad (2.17)$$

Recalling the following identity (see [2]) which holds for vectors  $\boldsymbol{\tau}$  and scalars  $\varphi$ , piecewise smooth on  $\mathcal{T}_h$ ,

$$\sum_{T \in \mathcal{T}_h} \int_{\partial T} (\boldsymbol{\tau} \cdot \mathbf{n}) \varphi \, ds = \sum_{e \in \mathcal{E}_h} \int_e \{\boldsymbol{\tau}\} \cdot \llbracket \varphi \rrbracket \, ds + \sum_{e \in \mathcal{E}_h^\circ} \int_e \llbracket \boldsymbol{\tau} \rrbracket \{\varphi\} \, ds, \quad (2.18)$$

and setting

$$a_h(u, v) := \sum_{T \in \mathcal{T}_h} \int_T \alpha \nabla u \cdot \nabla v \, dx, \quad (2.19)$$

we can write (2.17) as

$$(\mathcal{A}u, B_0v)_{\mathcal{T}_h} = a_h(u, v) - (\{\boldsymbol{\sigma}(u)\}, \llbracket v \rrbracket)_{\mathcal{E}_h} - (\llbracket \boldsymbol{\sigma}(u) \rrbracket, \{v\})_{\mathcal{E}_h^\circ}. \quad (2.20)$$

This suggests an obvious choice for  $B_2v$ , which is, in fact, the preferred one in most DG methods (even though they are not presented in this way!), namely

$$B_2v := \{v\}, \quad (2.21)$$

which transforms (2.9) into

$$a_h(u, v) - (\{\boldsymbol{\sigma}(u)\}, \llbracket v \rrbracket)_{\mathcal{E}_h} + (\llbracket u \rrbracket, \mathbf{B}_1v)_{\mathcal{E}_h} = (f, v) \quad \forall v \in H_{\mathcal{A}}^1(\mathcal{T}_h). \quad (2.22)$$

It is interesting to note that, if we restrict (2.22) to a subspace consisting of *continuous* functions (for instance, if we discretize by using continuous piecewise polynomials), then both  $\llbracket v \rrbracket$  and  $\llbracket u \rrbracket$  vanish, and we recover, as a particular case, the usual *conforming* finite element approximation. Notice also that, from this new perspective, we can consider conforming finite element methods as ones that enforce a linear relation between the residual inside the element and the *jump* in the normal component of the gradient across interelement boundaries, since only the operators  $B_0$  and  $B_2$  are active.

Now, the (most common) choice  $\mathbf{B}_1v = -\{\boldsymbol{\sigma}(v)\}$ , giving rise to the classical (nonstabilized) IP method ([20], [4], [28], [1]):

$$a_h(u, v) - (\{\boldsymbol{\sigma}(u)\}, \llbracket v \rrbracket)_{\mathcal{E}_h} - (\llbracket u \rrbracket, \{\boldsymbol{\sigma}(v)\})_{\mathcal{E}_h} = (f, v) \quad \forall v \in H_{\mathcal{A}}^1(\mathcal{T}_h), \quad (2.23)$$

is often motivated by the desire *to render the discrete formulation symmetric*, but it corresponds in fact to the *necessity of having a nonzero  $\mathbf{B}_1$* , in view of Theorem 1. The same can be said of the choice  $\mathbf{B}_1v = \{\boldsymbol{\sigma}(v)\}$ , corresponding to the Baumann–Oden method (cf. [7]):

$$a_h(u, v) - (\{\boldsymbol{\sigma}(u)\}, \llbracket v \rrbracket)_{\mathcal{E}_h} + (\llbracket u \rrbracket, \{\boldsymbol{\sigma}(v)\})_{\mathcal{E}_h} = (f, v) \quad \forall v \in H_{\mathcal{A}}^1(\mathcal{T}_h), \quad (2.24)$$

which is often motivated by the ease of obtaining stability estimates. On the other hand, the choice  $\mathbf{B}_1 \equiv \mathbf{0}$ , made for instance in [27] and [19], leads to

$$a_h(u, v) - (\{\boldsymbol{\sigma}(u)\}, \llbracket v \rrbracket)_{\mathcal{E}_h} = (f, v) \quad \forall v \in H_{\mathcal{A}}^1(\mathcal{T}_h); \quad (2.25)$$

this will not satisfy the assumptions of Theorem 1. In this case, the subsequent addition of a term of the form  $c_h(\llbracket u \rrbracket, \llbracket v \rrbracket)_{\mathcal{E}_h}$  (with a suitable choice of the coefficient  $c_h$ ) should not then be seen as *stabilization*, but rather as a primary choice of  $\mathbf{B}_1 v = c_h \llbracket v \rrbracket$ . The fact that the same term could be seen as a stabilization in one case, and not in another case is quite a subtle (and somewhat fuzzy) philosophical issue. The attitude we have adopted here is that it is helpful to distinguish between terms which are added in order to ensure uniqueness of solution to the infinite-dimensional problem (2.9), and terms which are further added in order to provide uniqueness of solution to its discretization and ensure stability of the discrete problem.

Indeed, once the  $B$ -operators have been chosen in such a way that problem (2.9) has a unique solution, and once a finite-dimensional subspace  $W_h$  of  $H_{\mathcal{A}}^1(\mathcal{T}_h)$  has been selected, it is natural to consider the discretized problem: *Find  $u_h \in W_h$  such that*

$$(\mathcal{A}u_h - f, B_0 v)_{\mathcal{T}_h} + (\llbracket u_h \rrbracket, \mathbf{B}_1 v)_{\mathcal{E}_h} + (\llbracket \boldsymbol{\sigma}(u_h) \rrbracket, B_2 v)_{\mathcal{E}_h} = 0 \quad \forall v \in W_h. \quad (2.26)$$

However, as we have already noted in the Introduction in connection with (1.12), it will not be generally true that the discretized problem (2.26) inherits uniqueness of solution from its infinite-dimensional counterpart, and even when it does, the discretization may fail to be stable in a norm suitable for proving convergence and optimal error estimates. Hence, additional stabilization may be needed in order to recover stability in a convenient norm.

Stability properties are generally proved in the norm  $\|\!\| \cdot \|\!\|$  defined by

$$\|\!\|v\|\!\|^2 := \|v\|_{0, \mathcal{T}_h}^2 + \sum_{e \in \mathcal{E}_h} \frac{1}{|e|} \|\llbracket v \rrbracket\|_{0, e}^2. \quad (2.27)$$

To stabilize the problem without losing consistency one could (and very often *does*, in practice) add further *residual-dependent stabilizing terms*. As we have seen, this corresponds in our case to the addition of terms having, *for instance*, the form

$$s_0 (\mathcal{A}u - f, \mathcal{A}v)_{\mathcal{T}_h} \quad \text{or} \quad s_1 (\llbracket u \rrbracket, \llbracket v \rrbracket)_{\mathcal{E}_h} \quad \text{or} \quad s_2 (\llbracket \alpha \nabla u \rrbracket, \llbracket \alpha \nabla v \rrbracket)_{\mathcal{E}_h}, \quad (2.28)$$

where the  $s_j$ , for  $j = 0, 1, 2$ , are suitable coefficients which, in general, depend on the local mesh-size. The first term would correspond to the most classical

Hughes–Franca stabilization (see e.g. [24], [21]), and is not normally used in this type of “smooth” problem. The second one is the preferred choice in the literature, while the third, used for instance in [13], is less popular, in spite of the fact that it may prove convenient in certain situations. The convenience of the second term in (2.28) is highlighted by the following property (see e.g. [2]). Assume that the space  $W_h$  is a subspace of the space of discontinuous piecewise polynomials of degree  $k \geq 1$ . Then, there exists a constant  $\gamma = \gamma(k, \theta_0)$ , depending only on  $k$  and on the minimum angle in  $\mathcal{T}_h$ , such that

$$(\llbracket v \rrbracket, \{\nabla v\})_{\mathcal{E}_h} \leq \gamma \|\nabla v\|_{0, \mathcal{T}_h} \left( \sum_{e \in \mathcal{E}_h} \frac{1}{|e|} \|\llbracket v \rrbracket\|_{0, e}^2 \right)^{1/2}. \quad (2.29)$$

With the use of (2.29) and similar properties it is a straightforward matter to establish stability for a number of possible choices of the  $B$ -operators once an appropriate stabilizing term (such as  $s_1(\llbracket u \rrbracket, \llbracket v \rrbracket)_{\mathcal{E}_h}$ , with a suitable choice of  $s_1$ ) has been introduced. It is not our intention to enter into an exercise of this kind here. Instead, we confine ourselves to listing, in Table 1, the correspondence between certain methods and the associated operators  $B_0$ ,  $\mathbf{B}_1$ ,  $B_2$ . There, and in some of the following tables, the notation, say,  $(\llbracket u \rrbracket \equiv 0)$  is used to specify that the definition of  $\mathbf{B}_1 v$  would be meaningless, as it would be multiplied by zero.

TABLE 1  
*Some methods for the model equation*

Method	$B_0 v$	$\mathbf{B}_1 v$	$B_2 v$
classical $C^1$ -conforming	$v$	$(\llbracket u \rrbracket \equiv 0)$	$(\llbracket \boldsymbol{\sigma}(u) \rrbracket \equiv 0)$
classical $C^0$ -conforming	$v$	$(\llbracket u \rrbracket \equiv 0)$	$v$
IP [20]	$v$	$(\llbracket u \rrbracket \equiv 0)$	$v + s_2 \llbracket \alpha \nabla v \rrbracket$
B.O. [7]	$v$	$\{\alpha \nabla v\}$	$\{v\}$
NIPG [26]	$v$	$\{\alpha \nabla v\} + s_1 \llbracket v \rrbracket$	$\{v\}$
IP [1, 4, 28]	$v$	$-\{\alpha \nabla v\} + s_1 \llbracket v \rrbracket$	$\{v\}$
D.S.W. [19]	$v$	$s_1 \llbracket v \rrbracket$	$\{v\}$

A possible stabilization, similar to  $s_1(\llbracket u \rrbracket, \llbracket v \rrbracket)_{\mathcal{E}_h}$ , emerged in the paper [5], and more clearly in the subsequent analysis performed in [10] and [2]. The basic idea is to construct a *lifting* of jumps from the interelement boundaries into  $\Omega$ . Roughly speaking, this is done (at the discrete level) by choosing a finite element space  $\Sigma_h$  consisting of discontinuous piecewise polynomial

vector-functions, and then defining, for each piecewise smooth function  $v$ , the *lifting of its jumps*,  $R(\llbracket v \rrbracket) \in \Sigma_h$ , as the unique solution, in  $\Sigma_h$ , of

$$(R(\llbracket v \rrbracket), \boldsymbol{\tau})_{\mathcal{T}_h} = -(\llbracket v \rrbracket, \{\boldsymbol{\tau}\})_{\mathcal{E}_h} \quad \forall \boldsymbol{\tau} \in \Sigma_h. \quad (2.30)$$

The presence of the minus sign in (2.30), whose historical roots originate in [5], stems from the desire to ensure that  $R(\llbracket v \rrbracket)$  resembles  $-\nabla v$  (as would be natural for Darcy's law) rather than  $\nabla v$ . We refer, for example, to [2] for further details and variants of this idea. What is relevant here is the possibility to stabilize (2.9) by adding a term of the form

$$s_{11}(R(\llbracket u \rrbracket), R(\llbracket v \rrbracket))_{\mathcal{T}_h}. \quad (2.31)$$

As shown in [10], [2], and [11] (in different contexts, and with increasing generality) one has, in general,

$$\|R(\llbracket v \rrbracket)\|_{\mathcal{T}_h}^2 \simeq \sum_{e \in \mathcal{E}_h} \frac{1}{|e|} \|\llbracket v \rrbracket\|_{0,e}^2 \quad (2.32)$$

or possibly

$$\|\nabla v\|_{0,\mathcal{T}_h}^2 + \|R(\llbracket v \rrbracket)\|_{\mathcal{T}_h}^2 \simeq \sum_{e \in \mathcal{E}_h} \frac{1}{|e|} \|\llbracket v \rrbracket\|_{0,e}^2, \quad (2.33)$$

when  $\Sigma_h$  is insufficiently rich. Hence, in practice, the addition of (2.31) is completely equivalent to the common *jump stabilization*.

Without entering the zoo of possible choices for the  $B$ -operators, we confine ourselves to noting that, in general, once one has a reasonably stable method, in a suitable norm such as (2.27), for the majority of finite element discretizations the proof of convergence and respective error bounds will easily emerge (see, for instance, [2]).

## 2.2 Mixed formulations

In this section, we consider mixed formulations of (2.3). Thus, instead of (2.6)-(2.8), the equations are now

$$\alpha^{-1} \boldsymbol{\sigma} = \nabla u \quad \text{in each } T \in \mathcal{T}_h, \quad (2.34)$$

$$-\operatorname{div} \boldsymbol{\sigma} = f \quad \text{in each } T \in \mathcal{T}_h, \quad (2.35)$$

$$\llbracket u \rrbracket|_e = 0 \quad \text{on each } e \in \mathcal{E}_h, \quad (2.36)$$

$$\llbracket \boldsymbol{\sigma} \rrbracket|_e = 0 \quad \text{on each } e \in \mathcal{E}_h^\circ. \quad (2.37)$$

The natural extension of (2.9) is now, roughly speaking,

$$(\alpha^{-1} \boldsymbol{\sigma} - \nabla u, \mathbf{B}_{00} \boldsymbol{\tau})_{\mathcal{T}_h} + (\llbracket u \rrbracket, \mathbf{B}_{01} \boldsymbol{\tau})_{\mathcal{E}_h} + (\llbracket \boldsymbol{\sigma} \rrbracket, B_{02} \boldsymbol{\tau})_{\mathcal{E}_h^\circ} = 0 \quad \forall \boldsymbol{\tau}, \quad (2.38)$$

$$(-\operatorname{div} \boldsymbol{\sigma} - f, B_{10} v)_{\mathcal{T}_h} + (\llbracket u \rrbracket, \mathbf{B}_{11} v)_{\mathcal{E}_h} + (\llbracket \boldsymbol{\sigma} \rrbracket, B_{12} v)_{\mathcal{E}_h^\circ} = 0 \quad \forall v. \quad (2.39)$$

In order to state a theorem similar to Theorem 1, we let

$$H^1(\mathcal{T}_h) := \{v \in L^2(\Omega) \mid v|_K \in H^1(T) \quad \forall T \in \mathcal{T}_h\}, \quad (2.40)$$

and, for  $s > 2$ ,

$$H^s(\text{div}; \mathcal{T}_h) := \{\boldsymbol{\tau} \in (L^s(\Omega))^2 \mid \text{div}(\boldsymbol{\tau}|_K) \in L^1(T) \quad \forall T \in \mathcal{T}_h\} \quad (2.41)$$

be the pair of spaces where the solution  $(u, \boldsymbol{\sigma})$  of (2.38), (2.39) is sought (and through which the test-functions  $(v, \boldsymbol{\tau})$  vary).

**Theorem 2** *Assume that the following conditions are satisfied.*

- For every  $\boldsymbol{\varphi}$  in  $(D_0)^2$  there exists a  $\boldsymbol{\tau} \in H^s(\text{div}; \mathcal{T}_h)$  such that  $\mathbf{B}_{00}\boldsymbol{\tau} = \boldsymbol{\varphi}$  and  $\mathbf{B}_{01}\boldsymbol{\tau} = \mathbf{0}$ ,  $B_{02}\boldsymbol{\tau} = 0$ .
- For every  $\varphi$  in  $D_0$  there exists a  $v \in H^1(\mathcal{T}_h)$  such that  $B_{10}v = \varphi$  and  $\mathbf{B}_{11}v = \mathbf{0}$ ,  $B_{12}v = 0$ .
- For every  $\boldsymbol{\psi} \in D_1$  at least one of the following two properties holds:
  - i) there exists  $\boldsymbol{\tau} \in H^s(\text{div}; \mathcal{T}_h)$  with  $(\mathbf{B}_{01}\boldsymbol{\tau} - \boldsymbol{\psi}) \cdot \mathbf{n} = 0$  and  $B_{02}\boldsymbol{\tau} = 0$ ;
  - ii) there exists  $v \in H^1(\mathcal{T}_h)$  with  $(\mathbf{B}_{11}v - \boldsymbol{\psi}) \cdot \mathbf{n} = 0$  and  $B_{12}v = 0$ .
- For every  $\chi \in D_2$  at least one of the following two properties holds:
  - i) there exists  $\boldsymbol{\tau} \in H^s(\text{div}; \mathcal{T}_h)$  with  $B_{02}\boldsymbol{\tau} = \chi$ ;
  - ii) there exists  $v \in H^1(\mathcal{T}_h)$  with  $B_{12}v = \chi$ .

*Then (2.38)–(2.39) has a unique solution.*

The proof is an elementary adaptation of the proof of Theorem 1, and is left to the reader.

As in the previous section, we remark that the *natural* choices  $\mathbf{B}_{00}\boldsymbol{\tau} = \boldsymbol{\tau}$  and  $B_{10}v = v$  immediately ensure that the first two requirements of Theorem 2 are satisfied. Moreover, properties of the kind stated in Proposition 1 will, with any reasonable choice of the remaining operators  $B_{ij}$  in terms of averages and/or jumps of  $v$  and  $\boldsymbol{\tau}$ , guarantee the remaining properties required in Theorem 2. We point out however that we can allow either  $\mathbf{B}_{01}$  or  $\mathbf{B}_{11}$  to be zero, but not both: otherwise the equation (2.36) will not be taken into account. Similarly, one of  $B_{02}$  and  $B_{12}$  can be zero, but not both: otherwise the equation (2.37) will not be taken into account. In a sense we may then say that (2.38) enforces a linear relation between the residual of equation (2.34) and the residual of *at least one* of the two jump equations (2.36)–(2.37), while (2.39) enforces a linear relation between (2.35) and *at least the other one* of the two jump equations (2.36)–(2.37).

Next, let us discuss some possible choices for the operators  $B_{ij}$ . In view of the above discussion, the *plain vanilla choice* seems to be

$$\mathbf{B}_{00}\boldsymbol{\tau} = \boldsymbol{\tau}, \quad \mathbf{B}_{01}\boldsymbol{\tau} = \{\boldsymbol{\tau}\}, \quad B_{02}\boldsymbol{\tau} = 0,$$

and

$$B_{10}v = v, \quad \mathbf{B}_{11}v = \mathbf{0}, \quad B_{12}v = \{v\},$$

which gives

$$(\alpha^{-1}\boldsymbol{\sigma} - \nabla u, \boldsymbol{\tau})_{\mathcal{T}_h} + (\llbracket u \rrbracket, \{\boldsymbol{\tau}\})_{\mathcal{E}_h} = 0 \quad \forall \boldsymbol{\tau} \in \mathbf{H}^s(\text{div}; \mathcal{T}_h), \quad (2.42)$$

$$(-\text{div}\boldsymbol{\sigma} - f, v)_{\mathcal{T}_h} + (\llbracket \boldsymbol{\sigma} \rrbracket, \{v\})_{\mathcal{E}_h^\circ} = 0 \quad \forall v \in \mathbf{H}^1(\mathcal{T}_h), \quad (2.43)$$

and which corresponds to the original formulation of [5]. The uniqueness of solution to (2.42)-(2.43) emerges directly from Theorem 2. It is interesting to note that (2.18) easily implies

$$(-\text{div}\boldsymbol{\sigma}, v)_{\mathcal{T}_h} + (\llbracket \boldsymbol{\sigma} \rrbracket, \{v\})_{\mathcal{E}_h^\circ} = (\boldsymbol{\sigma}, \nabla v) - (\llbracket v \rrbracket, \{\boldsymbol{\sigma}\})_{\mathcal{E}_h}, \quad (2.44)$$

so that (2.43) (on changing the sign of the whole equation) can be written as

$$-(\boldsymbol{\sigma}, \nabla v)_{\mathcal{T}_h} + (\llbracket v \rrbracket, \{\boldsymbol{\sigma}\})_{\mathcal{E}_h} + (f, v) = 0 \quad \forall v \in \mathbf{H}^1(\mathcal{T}_h), \quad (2.45)$$

rendering the problem (2.42)-(2.45) *symmetric*. However, as it has been shown in [10], the discretization of (2.42)-(2.43) (whether symmetrized, or not) does not generally give rise to a discrete problem that has a unique solution. Hence, once again, we encounter the need for supplementing the discrete problem with additional terms. Before doing so, however, we point out that we may consider the possibility of restricting (2.42)-(2.43) to subspaces consisting of continuous functions, as we did for primal formulations in the previous section. Here, in fact, we have several possibilities. Let us consider subspaces  $V_h \subset \mathbf{H}^1(\mathcal{T}_h)$  and  $\Sigma_h \subset \mathbf{H}^s(\text{div}; \mathcal{T}_h)$ , and the corresponding restricted problem: *Find*  $(u_h, \boldsymbol{\sigma}_h) \in V_h \times \Sigma_h$  *such that*

$$(\alpha^{-1}\boldsymbol{\sigma}_h - \nabla u_h, \boldsymbol{\tau})_{\mathcal{T}_h} + (\llbracket u_h \rrbracket, \{\boldsymbol{\tau}\})_{\mathcal{E}_h} = 0 \quad \forall \boldsymbol{\tau} \in \Sigma_h, \quad (2.46)$$

$$(-\text{div}\boldsymbol{\sigma}_h - f, v)_{\mathcal{T}_h} + (\llbracket \boldsymbol{\sigma}_h \rrbracket, \{v\})_{\mathcal{E}_h^\circ} = 0 \quad \forall v \in V_h, \quad (2.47)$$

where (2.47) is still equivalent to

$$-(\boldsymbol{\sigma}_h, \nabla v)_{\mathcal{T}_h} + (\llbracket v \rrbracket, \{\boldsymbol{\sigma}_h\})_{\mathcal{E}_h} + (f, v) = 0 \quad \forall v \in V_h. \quad (2.48)$$

We remark that, on taking  $\Sigma_h \subset H(\text{div}; \Omega)$ , integrating by parts the term containing  $\nabla u_h$  in (2.46), and using (2.18) we get, rather easily,

$$(\alpha^{-1}\boldsymbol{\sigma}_h, \boldsymbol{\tau})_{\mathcal{T}_h} + (u_h, \text{div}\boldsymbol{\tau})_{\mathcal{T}_h} = 0 \quad \forall \boldsymbol{\tau} \in \Sigma_h. \quad (2.49)$$

This then, when combined with (2.47) (where now the second term vanishes) gives one of the two classical *conforming mixed formulations* for (2.3) (actually, the most common one). On the other hand, taking instead  $V_h \subset H_0^1(\Omega)$ , the terms containing  $\llbracket u_h \rrbracket$  and  $\llbracket v \rrbracket$  vanish from (2.46) and (2.48), respectively, and we obtain *the other classical conforming mixed formulation*. As is well-known, both classical mixed formulations may still lack stability (and even uniqueness of solution) unless the subspaces are properly chosen, or some form of stabilization is introduced (see, e.g., [9]).

Let us now return to the more general case when both  $V_h$  and  $\Sigma_h$  consist of discontinuous piecewise polynomial functions. If we assume that

$$\nabla_h(V_h) \subset \Sigma_h, \quad (2.50)$$

where  $\nabla_h$  is the gradient taken element by element, and, for simplicity, that  $\alpha$  is piecewise constant, we can use the operator  $R$  defined in (2.30) and write (2.46) as

$$\boldsymbol{\sigma}_h = [\alpha(\nabla u_h + R(\llbracket u_h \rrbracket))]_{\mathcal{T}_h}. \quad (2.51)$$

Inserting (2.51) into (2.48) and using (2.30) once again, we obtain

$$(\alpha(\nabla u_h + R(\llbracket u_h \rrbracket)), \nabla v + R(\llbracket v \rrbracket))_{\mathcal{T}_h} = (f, v) \quad \forall v \in V_h. \quad (2.52)$$

Unfortunately, despite its attractive appearance, (2.52) is unstable, and the matrix of the resulting linear system can even become singular on certain grids (see [10], where it is shown that a nontrivial solution of  $\nabla_h u_h + R(\llbracket u_h \rrbracket) = 0$  can be constructed). Hence, some stabilization is in order. Having four residuals, we can think of residual-dependent stabilizing terms of the form

$$s_{e1}(\alpha^{-1}\boldsymbol{\sigma}_h - \nabla u_h, \boldsymbol{\tau})_{\mathcal{T}_h}, \quad s_{e2}(\alpha^{-1}\boldsymbol{\sigma}_h - \nabla u_h, \nabla v)_{\mathcal{T}_h}, \quad s_{e3}(\operatorname{div}\boldsymbol{\sigma}_h, \operatorname{div}\boldsymbol{\tau})_{\mathcal{T}_h}, \quad (2.53)$$

(depending on the equations inside the elements) or

$$s_1(\llbracket u_h \rrbracket, \llbracket v \rrbracket)_{\mathcal{E}_h}, \quad s_2(\llbracket \boldsymbol{\sigma}_h \rrbracket, \llbracket \boldsymbol{\tau} \rrbracket)_{\mathcal{E}_h^\circ}, \quad (2.54)$$

(depending on the jumps). Indeed, there are several other possibilities for stabilization.

At any rate, once suitable stabilizing terms have been chosen, the proof of stability is, generally, not too difficult. A basic inequality (see e.g. [2]) that extends (2.29) is the following: if  $V_h$  is included in the space of discontinuous piecewise polynomials of degree  $\leq k$ , and  $\Sigma_h$  is included in the space of (vector valued) discontinuous piecewise polynomials of degree  $\leq m$ , then there exists a constant  $\gamma = \gamma(k, m, \theta_0)$ , depending only on  $k$ ,  $m$ , and the

minimum angle  $\theta_0$  of the decomposition, such that for all  $v \in V_h$  and for all  $\boldsymbol{\tau} \in \Sigma_h$  we have

$$(\llbracket v \rrbracket, \{\boldsymbol{\tau}\})_{\mathcal{E}_h} \leq \gamma \|\boldsymbol{\tau}\|_{0,\Omega} \left( \sum_{e \in \mathcal{E}_h} \frac{1}{|e|} \|\llbracket v \rrbracket\|_{0,e}^2 \right)^{1/2}. \quad (2.55)$$

Other similar inequalities can also be easily proved. Then, we take  $\boldsymbol{\tau} = \boldsymbol{\sigma}_h$  in the first equation,  $v = \pm u_h$  (the choice of sign being a matter of convenience) in the second equation, and sum up. We repeatedly apply Cauchy's inequality  $2ab \leq \varepsilon a^2 + b^2/\varepsilon$ , valid for any positive  $\varepsilon$ , and the stability result follows. There are obviously less trivial cases, but the proofs of the straightforward ones proceed in precisely this manner. Of course, it would be quite interesting to explore, both theoretically and computationally, some of the possibilities which have not yet been considered in the literature. Instead of pursuing that avenue, here we confine ourselves to discussing some simple examples.

Possibly the most standard idea is to take

$$\begin{aligned} (\alpha^{-1} \boldsymbol{\sigma}_h - \nabla u_h, \boldsymbol{\tau})_{\mathcal{T}_h} + (\llbracket u_h \rrbracket, \{\boldsymbol{\tau}\})_{\mathcal{E}_h} &= 0 \quad \forall \boldsymbol{\tau} \in \Sigma_h, \\ (-\operatorname{div} \boldsymbol{\sigma}_h - f, v)_{\mathcal{T}_h} + s_1 (\llbracket u_h \rrbracket, \llbracket v \rrbracket)_{\mathcal{E}_h} + (\llbracket \boldsymbol{\sigma}_h \rrbracket, \{v\})_{\mathcal{E}_h^\circ} &= 0 \quad \forall v \in V_h, \end{aligned}$$

where the second equation can obviously be written in the symmetric form

$$(-\boldsymbol{\sigma}_h, \nabla v)_{\mathcal{T}_h} + (\llbracket v \rrbracket, \{\boldsymbol{\sigma}_h\})_{\mathcal{E}_h} - s_1 (\llbracket u_h \rrbracket, \llbracket v \rrbracket)_{\mathcal{E}_h} + (f, v) = 0 \quad \forall v \in V_h.$$

This is nothing but a particular case of the LDG method [18]. The proof of stability is immediate provided that we assume (2.50) and select  $s_1 = c/h$  with  $c$  large enough. It is interesting to note that this same scheme can also be obtained with particular choices of the operators  $B_{ij}$ , namely by taking

$$\mathbf{B}_{00} \boldsymbol{\tau} = \boldsymbol{\tau}, \quad \mathbf{B}_{01} \boldsymbol{\tau} = \{\boldsymbol{\tau}\}, \quad B_{02} \boldsymbol{\tau} = 0,$$

and

$$B_{10} v = v, \quad \mathbf{B}_{11} v = s_1 \llbracket v \rrbracket, \quad \text{and} \quad B_{12} v = \{v\}.$$

Our second example is an interesting variation on the DG method just considered; it does not use stabilization based on the jumps but on the residual of the first equation, namely

$$(\alpha^{-1} \boldsymbol{\sigma}_h - \nabla u_h, \boldsymbol{\tau})_{\mathcal{T}_h} + (\llbracket u_h \rrbracket, \{\boldsymbol{\tau}\})_{\mathcal{E}_h} = 0 \quad (2.56)$$

$$(-\boldsymbol{\sigma}_h, \nabla v)_{\mathcal{T}_h} + (\llbracket v \rrbracket, \{\boldsymbol{\sigma}_h\})_{\mathcal{E}_h} + s_0 (\boldsymbol{\sigma} - \alpha \nabla u, \nabla v)_{\mathcal{T}_h} + (f, v) = 0 \quad (2.57)$$

for all  $(\boldsymbol{\tau}, v) \in \Sigma_h \times V_h$ , in the spirit of [25] and [11]. Stability follows easily whenever  $0 < s_0 < 4$ .

It is remarkable (and somehow surprising) that this method too could be obtained with a particular choice of the operators  $B_{ij}$ . Indeed, on assuming that (2.50) holds, and supposing again, for simplicity, that  $\alpha$  is piecewise constant on  $\mathcal{T}_h$ , we can take

$$\mathbf{B}_{00}\boldsymbol{\tau} = \boldsymbol{\tau}, \quad \mathbf{B}_{01}\boldsymbol{\tau} = \{\boldsymbol{\tau}\}, \quad B_{02}\boldsymbol{\tau} = 0,$$

and

$$B_{10}v = v, \quad \mathbf{B}_{11}v = s_0\{\alpha\nabla v\}, \quad \text{and} \quad B_{12}v = \{v\}.$$

After the usual passage to the symmetric form, we obtain

$$\begin{aligned} (\alpha^{-1}\boldsymbol{\sigma}_h - \nabla u_h, \boldsymbol{\tau})_{\mathcal{T}_h} + (\llbracket u_h \rrbracket, \{\boldsymbol{\tau}\})_{\mathcal{E}_h} &= 0 \quad \forall \boldsymbol{\tau} \in \Sigma_h, \\ (-\boldsymbol{\sigma}_h, \nabla v)_{\mathcal{T}_h} + (\llbracket v \rrbracket, \{\boldsymbol{\sigma}_h\})_{\mathcal{E}_h} - s_0(\llbracket u_h \rrbracket, \{\alpha\nabla v\})_{\mathcal{E}_h} + (f, v) &= 0 \quad \forall v \in V_h. \end{aligned}$$

At this point, using (2.50), we can take  $\boldsymbol{\tau} = \alpha\nabla v$  in the first equation, to obtain

$$-(\llbracket u_h \rrbracket, \{\alpha\nabla v\})_{\mathcal{E}_h} = (\alpha^{-1}\boldsymbol{\sigma}_h - \nabla u_h, \alpha\nabla v)_{\mathcal{T}_h}$$

which, when inserted into the second equation, exactly reproduces (2.56)–(2.57).

However, there is a third point of view on this method. Indeed, we can again use (2.51) and substitute it into (2.57) to obtain

$$(\alpha(\nabla u_h + R(\llbracket u_h \rrbracket)), \nabla v + R(\llbracket v \rrbracket))_{\mathcal{T}_h} - s_0(\alpha R(\llbracket u_h \rrbracket), \nabla v)_{\mathcal{T}_h} = (f, v) \quad \forall v \in V_h. \quad (2.58)$$

It is interesting to note now that for  $s_0 = 2$  we obtain the Baumann–Oden method (2.24) stabilized with the addition of the term  $(\alpha R(\llbracket u_h \rrbracket), R(\llbracket v \rrbracket))_{\mathcal{T}_h}$ . Similarly, with  $s_0 = 1$  we obtain the stabilization of the method of Dawson–Sun–Wheeler (2.25) with the addition of the same term. Clearly, the choice  $s_0 = 0$  can also be seen as (insufficient) stabilization of the IP method (2.23), — always with the same term.

With quite similar arguments, we find the equivalence between the stabilization proposed by Hughes and Masud [25]:

$$\begin{aligned} (\alpha^{-1}\boldsymbol{\sigma} - \nabla u, \boldsymbol{\tau})_{\mathcal{T}_h} + (\llbracket u \rrbracket, \{\boldsymbol{\tau}\})_{\mathcal{E}_h} - s_e(\alpha^{-1}\boldsymbol{\sigma} - \nabla u, \boldsymbol{\tau})_{\mathcal{T}_h} &= 0 \quad \forall \boldsymbol{\tau}, \\ (-\operatorname{div}\boldsymbol{\sigma} - f, v)_{\mathcal{T}_h} + (\llbracket \boldsymbol{\sigma} \rrbracket, \{v\})_{\mathcal{E}_h} - s_e(\boldsymbol{\sigma} - \alpha\nabla u, \nabla v)_{\mathcal{T}_h} &= 0 \quad \forall v, \end{aligned}$$

and the following choice for the  $B_{ij}$ :

$$\mathbf{B}_{00}\boldsymbol{\tau} = \boldsymbol{\tau} - s_e\boldsymbol{\tau}, \quad \mathbf{B}_{01}\boldsymbol{\tau} = \{\boldsymbol{\tau}\}, \quad B_{02}\boldsymbol{\tau} = 0,$$

and

$$B_{10}v = v, \quad \mathbf{B}_{11}v = \frac{s_e}{1 - s_e}\{\alpha\nabla v\}, \quad \text{and} \quad B_{12}v = \{v\},$$

Note that, as shown in [11], this is also equivalent to stabilizing the Bassi–Rebay formulation (2.52) with the addition of the stabilizing term

$$\frac{s_e}{1-s_e}(\alpha R(\llbracket u_h \rrbracket), R(\llbracket v \rrbracket))_{\mathcal{T}_h},$$

showing that stability holds in this case for  $0 < s_e < 1$ .

For many further variants of the idea of stabilizing with the residuals (2.53), see [25] and [11].

To conclude this section, we show that most DG methods defined in terms of numerical fluxes (cf. Table 2) can be easily recast in the present formulation. To this end, let us recall that those methods have the following form: in each element  $T$  we set

$$(\alpha^{-1}\boldsymbol{\sigma}, \boldsymbol{\tau})_T = -(u, \operatorname{div} \boldsymbol{\tau})_T + (\widehat{u}_T, \boldsymbol{\tau} \cdot \mathbf{n}_T)_{\partial T}, \quad (2.59)$$

$$(\boldsymbol{\sigma}, \nabla v)_T = (f, v)_T + (v, \widehat{\boldsymbol{\sigma}}_T \cdot \mathbf{n}_T)_{\partial T}. \quad (2.60)$$

Different choices of the numerical fluxes  $\widehat{u}_T$  and  $\widehat{\boldsymbol{\sigma}}_T$  in the above formulae give rise to different DG methods (as shown in [2]: see Table 2, quoted from there; we refer to [2] for those bits of notation which have not been specified here).

After straightforward integrations by parts in (2.59) and in (2.60) we obtain

$$(\alpha^{-1}\boldsymbol{\sigma} - \nabla u, \boldsymbol{\tau})_T = (\widehat{u}_T - u, \boldsymbol{\tau} \cdot \mathbf{n}_T)_{\partial T}, \quad (2.61)$$

$$(-\operatorname{div} \boldsymbol{\sigma} - f, v)_T = (v, (\widehat{\boldsymbol{\sigma}}_T - \boldsymbol{\sigma}) \cdot \mathbf{n}_T)_{\partial T}. \quad (2.62)$$

We immediately see that if we define the numerical fluxes  $\widehat{u}_T$  and  $\widehat{\boldsymbol{\sigma}}_T$  in such a way that the quantities  $\widehat{u}_T - u$  and  $\widehat{\boldsymbol{\sigma}}_T - \boldsymbol{\sigma}$  depend only on the jumps  $\llbracket u \rrbracket$  and  $\llbracket \boldsymbol{\sigma} \rrbracket$ , then (2.61) and (2.62) establish a linear relation between the four residuals of the equations under consideration. In fact, almost all known DG methods can be obtained in this way, as was shown in [2].

Let us now see if it is possible to identify the  $B$ -operators associated with those methods. Summing over all the elements  $T$  of the triangulation  $\mathcal{T}_h$ , and using (2.18) we get

$$\begin{aligned} (\alpha^{-1}\boldsymbol{\sigma} - \nabla u, \boldsymbol{\tau})_{\mathcal{T}_h} &= (\{\widehat{u} - u\}, \llbracket \boldsymbol{\tau} \rrbracket)_{\mathcal{E}_h^\circ} + (\llbracket \widehat{u} - u \rrbracket, \{\boldsymbol{\tau}\})_{\mathcal{E}_h}, \\ (-\operatorname{div} \boldsymbol{\sigma} - f, v)_{\mathcal{T}_h} &= (\{v\}, \llbracket \widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma} \rrbracket)_{\mathcal{E}_h^\circ} + (\llbracket v \rrbracket, \{\widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}\})_{\mathcal{E}_h}. \end{aligned}$$

It is now clear that if we set

$$\mathbf{B}_{00}\boldsymbol{\tau} = \boldsymbol{\tau} \quad \text{and} \quad B_{10}v = v,$$

TABLE 2  
The numerical fluxes of some DG methods.

Method	$\widehat{\boldsymbol{\sigma}}_T$	$\widehat{u}_T$
B.R. [5]	$\{\boldsymbol{\sigma}\}$	$\{u\}$
LDG [18]	$\{\boldsymbol{\sigma}\} + \boldsymbol{\beta}[\![\boldsymbol{\sigma}]\!] - s_1[\![u]\!]$	$\{u\} - \boldsymbol{\beta} \cdot [\![u]\!]$
C.C.P.S. [14]	$\{\boldsymbol{\sigma}\} + \boldsymbol{\beta}[\![\boldsymbol{\sigma}]\!] - s_1[\![u]\!]$	$\{u\} - \boldsymbol{\beta} \cdot [\![u]\!] - s_2[\![\boldsymbol{\sigma}]\!]$
B.M.M.P.R.[10]	$\{\boldsymbol{\sigma}\} - \alpha^r([\![u]\!])$	$\{u\}$
IP [1, 4, 28]	$\{\alpha \nabla u\} - s_1[\![u]\!]$	$\{u\}$
B.R.M.P.S. [6]	$\{\alpha \nabla u\} - \alpha^r([\![u]\!])$	$\{u\}$
B.O. [7]	$\{\alpha \nabla u\}$	$\{u\} + \mathbf{n}_T \cdot [\![u]\!]$
NIPG [26]	$\{\alpha \nabla u\} - s_1[\![u]\!]$	$\{u\} + \mathbf{n}_T \cdot [\![u]\!]$
B.Z. [3]	$-s_1[\![u]\!]$	$(u _T) _{\partial T}$
B.M.M.P.R.[10]	$-\alpha^r([\![u]\!])$	$(u _T) _{\partial T}$

we obtain the following relation between the  $B$ -operators and the numerical fluxes

$$-([\![u]\!], \mathbf{B}_{01}\boldsymbol{\tau})_{\mathcal{E}_h} - ([\![\boldsymbol{\sigma}]\!], B_{02}\boldsymbol{\tau})_{\mathcal{E}_h^\circ} = (\{\widehat{u} - u\}, [\![\boldsymbol{\tau}]\!])_{\mathcal{E}_h^\circ} + ([\![\widehat{u} - u]\!], \{\boldsymbol{\tau}\})_{\mathcal{E}_h}, \quad (2.63)$$

$$-([\![u]\!], \mathbf{B}_{11}v)_{\mathcal{E}_h} - ([\![\boldsymbol{\sigma}]\!], B_{12}v)_{\mathcal{E}_h^\circ} = (\{v\}, [\![\widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}]\!])_{\mathcal{E}_h^\circ} + ([\![v]\!], \{\widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}\})_{\mathcal{E}_h}. \quad (2.64)$$

It is now a simple matter to identify the  $B$ -operators associated with each of the DG methods. For example, for the Bassi–Rebay method [5], we have

$$\widehat{u} = \{u\} \quad \widehat{\boldsymbol{\sigma}} = \{\boldsymbol{\sigma}\},$$

and so

$$\begin{aligned} -([\![u]\!], \mathbf{B}_{01}\boldsymbol{\tau})_{\mathcal{E}_h} - ([\![\boldsymbol{\sigma}]\!], B_{02}\boldsymbol{\tau})_{\mathcal{E}_h^\circ} &= -([\![u]\!], \{\boldsymbol{\tau}\})_{\mathcal{E}_h}, \\ -([\![u]\!], \mathbf{B}_{11}v)_{\mathcal{E}_h} - ([\![\boldsymbol{\sigma}]\!], B_{12}v)_{\mathcal{E}_h^\circ} &= -(\{v\}, [\![\boldsymbol{\sigma}]\!])_{\mathcal{E}_h^\circ}. \end{aligned}$$

Hence,

$$\mathbf{B}_{01}\boldsymbol{\tau} = \{\boldsymbol{\tau}\}, \quad B_{02}\boldsymbol{\tau} = 0, \quad \mathbf{B}_{11}v = \mathbf{0}, \quad \text{and} \quad B_{12}v = \{v\},$$

TABLE 3  
The  $B$ -operators for some DG methods.

Method	$\mathbf{B}_{00}\boldsymbol{\tau}$	$\mathbf{B}_{01}\boldsymbol{\tau}$	$B_{02}\boldsymbol{\tau}$	$B_{10}v$	$\mathbf{B}_{11}v$	$B_{12}v$
conforming mixed I	$\boldsymbol{\tau}$	$\{\boldsymbol{\tau}\}$	$([\boldsymbol{\sigma}] \equiv 0)$	$v$	$\mathbf{0}$	$([\boldsymbol{\sigma}] \equiv 0)$
conforming mixed II	$\boldsymbol{\tau}$	$([u] \equiv 0)$	0	$v$	$([u] \equiv 0)$	$v$
IP [20]	$\boldsymbol{\tau}$	$([u] \equiv 0)$	0	$v$	$([u] \equiv 0)$	$v + s_2[\alpha \nabla v]$
B.R. [5]	$\boldsymbol{\tau}$	$\{\boldsymbol{\tau}\}$	0	$v$	$\mathbf{0}$	$\{v\}$
LDG [18]	$\boldsymbol{\tau}$	$\{\boldsymbol{\tau}\} + \boldsymbol{\beta}[\boldsymbol{\tau}]$	0	$v$	$s_1[v]$	$\{v\} - \boldsymbol{\beta} \cdot [v]$
C.C.P.S. [14]	$\boldsymbol{\tau}$	$\{\boldsymbol{\tau}\} + \boldsymbol{\beta}[\boldsymbol{\tau}]$	$s_2[\boldsymbol{\tau}]$	$v$	$s_1[v]$	$\{v\} - \boldsymbol{\beta} \cdot [v]$
IP [1, 4, 28]	$\boldsymbol{\tau}$	$\{\boldsymbol{\tau}\}$	0	$v$	$s_1[v]$	$\{v\}$
B.O. [7]	$\boldsymbol{\tau}$	$-\{\boldsymbol{\tau}\}$	0	$v$	$\mathbf{0}$	$\{v\}$
NIPG [26]	$\boldsymbol{\tau}$	$-\{\boldsymbol{\tau}\}$	0	$v$	$s_1[v]$	$\{v\}$
DSW [26]	$\boldsymbol{\tau}$	$\mathbf{0}$	0	$v$	$s_1[v]$	$\{v\}$
(2.56)-(2.57)	$\boldsymbol{\tau}$	$\{\boldsymbol{\tau}\}$	0	$v$	$s_0\{\alpha \nabla v\}$	$\{v\}$
H.M. [25]	$\boldsymbol{\tau} - s_e \boldsymbol{\tau}$	$\{\boldsymbol{\tau}\}$	0	$v$	$\frac{s_e}{1-s_e}\{\alpha \nabla v\}$	$\{v\}$

which is our original *plain vanilla* choice, giving rise to (2.42)-(2.43).

We end the section by summarizing, in Table 3, all these results. It is interesting to note that the superpenalty methods considered in [3] and in [10] do not fall within the present framework, because they have an *inconsistent* numerical flux for  $\boldsymbol{\sigma}$ , and this in turn makes it impossible to find operators  $\mathbf{B}_{11}$  and  $B_{12}$  which satisfy the equation (2.64). We also note that it is the operators  $\mathbf{B}_{11}$  and  $B_{02}$  that carry all the burden of stabilizing DG methods by means of the jumps.

### 3 Hyperbolic equations

Let, again,  $\Omega$  be a bounded polygonal domain in  $\mathbb{R}^2$ , and let the advective velocity field  $\boldsymbol{\beta} = (\beta_1, \beta_2)^T$  be a vector-valued function defined on  $\overline{\Omega}$  with  $\beta_i \in C^1(\overline{\Omega})$ ,  $i = 1, 2$ . We define the *inflow* and *outflow* parts of  $\Gamma = \partial\Omega$  in the usual fashion:

$$\begin{aligned}\Gamma_- &= \{x \in \Gamma : \boldsymbol{\beta}(x) \cdot \mathbf{n}(x) < 0\} = \text{inflow}, \\ \Gamma_+ &= \{x \in \Gamma : \boldsymbol{\beta}(x) \cdot \mathbf{n}(x) > 0\} = \text{outflow},\end{aligned}$$

where  $\mathbf{n}(x)$  denotes the unit outward normal vector to  $\Gamma$  at  $x \in \Gamma$ .

Let  $\gamma \in C(\overline{\Omega})$ ,  $f \in L^2(\Omega)$ . As a model problem we will consider the hyperbolic boundary value problem

$$\begin{aligned}\mathcal{L}u &\equiv \operatorname{div}(\boldsymbol{\beta}u) + \gamma u = f && \text{in } \Omega, \\ u &= 0 && \text{on } \Gamma_-.\end{aligned}\tag{3.1}$$

We shall assume the existence of a positive constant  $\gamma_0$  such that

$$\gamma(x) + \frac{1}{2} \operatorname{div} \beta(x) \geq \gamma_0 \quad \text{for all } x \in \Omega. \quad (3.2)$$

Consider now a decomposition  $\mathcal{T}_h$  as before, and introduce the space

$$\mathbf{H}_{\mathcal{L}}(\mathcal{T}_h) := \{v \in L^2(\Omega) \mid \mathcal{L}v|_K \in L^2(T) \quad \forall T \in \mathcal{T}_h\}. \quad (3.3)$$

For the sake of simplicity we assume that there exists a constant  $\kappa > 0$  such that  $|\beta \cdot \mathbf{n}| \geq \kappa$  on each edge of  $\mathcal{E}_h$ . The latter condition can be omitted, but then a number of details become more complicated. Note that the above assumption implies, among other things, that the traces on  $\mathcal{E}_h$  of functions in  $\mathbf{H}_{\mathcal{L}}(\mathcal{T}_h)$  are well defined and belong to  $L^2(\mathcal{E}_h)$ .

A function  $u \in \mathbf{H}_{\mathcal{L}}(\mathcal{T}_h)$  will coincide with the solution  $u$  of (3.1) if, and only if,

$$\begin{aligned} \mathcal{L}u &= f && \text{in each } K \in \mathcal{T}_h, \\ \llbracket \beta u \rrbracket|_e &= 0 && \text{in each } e \in \mathcal{E}_h^\circ, \\ \{\beta u\}|_e &= 0 && \text{in each } e \in \mathcal{E}_h^-. \end{aligned}$$

With the same reasoning as before, we then consider the variational problem: *Find  $u \in \mathbf{H}_{\mathcal{L}}(\mathcal{T}_h)$  such that*

$$(\mathcal{L}u - f, B_0v)_{\mathcal{T}_h} + (\llbracket \beta u \rrbracket, B_2v)_{\mathcal{E}_h^\circ} + (\{\beta u\}, \mathbf{B}_3v)_{\mathcal{E}_h^-} = 0 \quad \forall v \in \mathbf{H}_{\mathcal{L}}(\mathcal{T}_h). \quad (3.4)$$

Again, on setting

$$\mathbf{D}_3 := \prod_{e \in \mathcal{E}_h^-} C_0^\infty(e) \mathbf{n}_e \quad (3.5)$$

we have the analogue of Theorems 1 and 2, telling us that we cannot take any of the  $B_j$  to be zero.

**Theorem 3** *Assume that:*

- for each  $\varphi \in D_0$  there exists a  $v \in \mathbf{H}_{\mathcal{L}}(\mathcal{T}_h)$  such that  $B_0v = \varphi$  and  $B_2v = 0$ ,  $\mathbf{B}_3v \cdot \mathbf{n} = 0$ ;
- for every  $\chi \in D_2$  there exists a  $v \in \mathbf{H}_{\mathcal{L}}(\mathcal{T}_h)$  such that  $B_2v = \chi$  and  $\mathbf{B}_3v \cdot \mathbf{n} = 0$ ;
- for every  $\zeta \in \mathbf{D}_3$  there exists a  $v \in \mathbf{H}_{\mathcal{L}}(\mathcal{T}_h)$  such that  $(\mathbf{B}_3v - \zeta) \cdot \mathbf{n} = 0$ .

Then problem (3.4) has a unique solution that coincides with the solution of (3.1).

If we now choose  $B_0v = v$  and integrate by parts we have

$$(\mathcal{L}u, B_0v)_{\mathcal{T}_h} = \sum_{T \in \mathcal{T}_h} \left( \int_T (-u(\boldsymbol{\beta} \cdot \nabla v) + \gamma uv) dx + \int_{\partial T} (\boldsymbol{\beta} \cdot \mathbf{n}) u v ds \right), \quad (3.6)$$

and applying (2.18) we deduce that

$$\sum_{T \in \mathcal{T}_h} \int_{\partial T} (\boldsymbol{\beta} \cdot \mathbf{n}) u v ds = (\{\boldsymbol{\beta}u\}, \llbracket v \rrbracket)_{\mathcal{E}_h} + (\llbracket \boldsymbol{\beta}u \rrbracket, \{v\})_{\mathcal{E}_h^\circ}. \quad (3.7)$$

Setting

$$a_h(u, v) = \sum_{T \in \mathcal{T}_h} \int_T (-u(\boldsymbol{\beta} \cdot \nabla v) + \gamma uv) dx, \quad (3.8)$$

equation (3.4) becomes

$$\begin{aligned} a_h(u, v) + (\{\boldsymbol{\beta}u\}, \llbracket v \rrbracket)_{\mathcal{E}_h} + (\llbracket \boldsymbol{\beta}u \rrbracket, \{v\})_{\mathcal{E}_h^\circ} + (\llbracket \boldsymbol{\beta}u \rrbracket, B_2v)_{\mathcal{E}_h^\circ} + (\{\boldsymbol{\beta}u\}, \mathbf{B}_3v)_{\mathcal{E}_h^-} \\ = (f, v)_{\mathcal{T}_h} \quad \forall v \in \mathbf{H}^1(\mathcal{T}_h). \end{aligned} \quad (3.9)$$

At this point it is difficult to resist the temptation of choosing  $B_2v = -\{v\}$  which gives

$$a_h(u, v) + (\{\boldsymbol{\beta}u\}, \llbracket v \rrbracket)_{\mathcal{E}_h} + (\{\boldsymbol{\beta}u\}, \mathbf{B}_3v)_{\mathcal{E}_h^-} = (f, v)_{\mathcal{T}_h} \quad \forall v \in \mathbf{H}^1(\mathcal{T}_h). \quad (3.10)$$

Denoting now  $\mathcal{E}_h^\circ \cup \mathcal{E}_h^+$  by  $\mathcal{E}_h^*$  and choosing  $\mathbf{B}_3v := -\llbracket v \rrbracket$  we finally obtain:

$$a_h(u, v) + (\{\boldsymbol{\beta}u\}, \llbracket v \rrbracket)_{\mathcal{E}_h^*} = (f, v)_{\mathcal{T}_h} \quad \forall v \in \mathbf{H}^1(\mathcal{T}_h). \quad (3.11)$$

The formulation (3.11) is stable, but only in the  $L^2(\Omega)$ -norm. The practical consequences of this can be detrimental: discontinuities in the data may trigger large, nonphysical oscillations in the numerical solution. We need therefore to apply some form of *stabilization*. The most common practice (which, surprisingly enough, is often considered as *not being a stabilization*) is the so called *upwinding*: on every internal edge  $e$ , common to the triangles  $T^1$  and  $T^2$ , one usually substitutes the average  $\{\boldsymbol{\beta}u\}$  which appears in (3.11) by the *upwind value* of  $\boldsymbol{\beta}u$ , defined as

$$\{\boldsymbol{\beta}u\}_{\text{upw}} = \begin{cases} \boldsymbol{\beta}u^1 & \text{if } \boldsymbol{\beta} \cdot \mathbf{n}^1 > 0 \\ \boldsymbol{\beta}u^2 & \text{if } \boldsymbol{\beta} \cdot \mathbf{n}^1 < 0 \\ \boldsymbol{\beta}\{u\} & \text{if } \boldsymbol{\beta} \cdot \mathbf{n}^1 = 0. \end{cases} \quad (3.12)$$

On the edges  $e \in \mathcal{E}_h^+$  we had already  $\{\boldsymbol{\beta}u\} = \boldsymbol{\beta}u$  (the only value available) and we leave it unchanged.

As  $\{\boldsymbol{\beta}u\}_{\text{upw}}$ , in (3.11), is multiplied by  $\llbracket v \rrbracket$ , which is directed as the normal  $\mathbf{n}$  to  $e$ , it is clear that only the *normal component* of  $\{\boldsymbol{\beta}u\}_{\text{upw}}$  will feature in the scheme.

On the other hand, it is a simple matter to check that, if  $\mathbf{n}$  is normal to  $e$ , then  $\{\boldsymbol{\beta}u\}_{\text{upw}} \cdot \mathbf{n}$  can also be written as

$$\{\boldsymbol{\beta}u\}_{\text{upw}} \cdot \mathbf{n} = (\{\boldsymbol{\beta}u\} + c^* \llbracket u \rrbracket) \cdot \mathbf{n} \quad (3.13)$$

where  $\{\boldsymbol{\beta}u\}$  is again the usual average and  $c^*$  is given by

$$c^* = |\boldsymbol{\beta} \cdot \mathbf{n}|/2. \quad (3.14)$$

It is also known (see e.g. [16] or [12]) that we could still achieve stability in a norm that is stronger than  $\|\cdot\|_{0,\Omega}$  if we replace the upwind average  $\{\boldsymbol{\beta}u\}_{\text{upw}}$  by  $\{\boldsymbol{\beta}u\} + c_e \llbracket u \rrbracket$ , provided that  $c_e$  is a nonnegative function chosen on each  $e$  in such a way that

$$c_e \geq \theta_0 |\boldsymbol{\beta} \cdot \mathbf{n}|, \quad (3.15)$$

with  $\theta_0$  a positive constant independent of  $e$  and  $h$ . Note that this is equivalent to stabilizing (3.11) with

$$\sum_{e \in \mathcal{E}_h^\circ} c_e \int_e \llbracket u \rrbracket \cdot \llbracket v \rrbracket ds, \quad (3.16)$$

which is a residual-based stabilization. It is interesting to note that the addition of (3.16) can be interpreted as just *a different choice of  $B_2$* .

It is also interesting to note that (3.11) could equally well be stabilized by adding terms of the type

$$(\mathcal{L}u - f, \mathcal{L}v)_{\mathcal{T}_h} \text{ or } (\mathcal{L}u - f, \boldsymbol{\beta} \cdot \nabla v)_{\mathcal{T}_h}; \quad (3.17)$$

see, for example, [22].

Finally, we would like to point out that, if we had a convection-dominated equation of the type

$$\varepsilon \mathcal{A}u + \mathcal{L}u = f, \quad (3.18)$$

the use of (3.16) for the convective part would provide a sufficient amount of jump stabilization for the elliptic part as well (at least when  $\varepsilon$  is small compared to  $\boldsymbol{\beta} \cdot \mathbf{n}/h$ ); see [23] and, for the time-dependent case, [17, 18].

## 4 Conclusions

We proposed a new framework for designing and analyzing DG methods. Although we confined ourselves to a few simple model problems, the framework can immediately be adapted to much more complicated cases.

The basic idea is to consider the equation(s) inside each element and the equations for the jumps (where we enforce the continuity requirements from one element to another) as *basic equations*, all having the same ‘dignity’. Then, the equations that we use in our methods are linear combinations of the residuals of the *basic equations*, weighted by suitable operators applied to the test functions.

Changing the weighting operators results in a galaxy of possible methods where one can easily recognize essentially all DG methods proposed and used so far in the literature, as well as many others that have not been studied so far.

Our equations (= combinations of residuals) can be written at the continuous level, that is, using infinite-dimensional spaces of functions that are a priori discontinuous from one element to another, and have the usual regularity properties inside each element. Suitable conditions, which are very easy to check, ensure that the resulting equations have a unique solution.

The uniqueness of solution to the infinite-dimensional problem does not (rather obviously) imply that any finite element discretization will be stable, or even uniquely solvable. In general, to ensure stability one has to make, in a suitable sense, a *stronger choice* of the weighting operators. Typically this includes some additional term in the weighting operators that generates a sort of *least square term* for one (or more) of the basic equations.

Surprisingly enough, using least squares terms involving the equations inside the elements (in the spirit of Hughes–Franca stabilizations) or involving the jump equations (as in traditional DG methods) provide identical or very similar effects. Hence we can conceive this as *one* basic stabilization mechanism.

However, one may incorporate into the discretization these stronger choices of the weighting operators (as was already done at the infinite-dimensional level at the very beginning of the process of construction of the method). From this perspective, then, it is not easy to decide whether such additional terms should or should not be viewed as *stabilization*. This subtle philosophical question has been left open.

## References

- [1] D.N. ARNOLD, *An interior penalty finite element method with discontinuous element*, SIAM J. Numer. Anal., **19** (1982), pp. 742–760.
- [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** (2002), pp. 1749–1779.
- [3] I. BABUŠKA AND M. ZLÁMAL, *Nonconforming elements in the finite element method with penalty*, SIAM J. Numer. Anal., **10** (1973), pp. 863–875.
- [4] G.A. BAKER, *Finite element methods for elliptic equations using nonconforming elements*, Math Comp., **31** (1977), pp. 45–59.
- [5] 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** (1997), pp. 267–279.
- [6] F. BASSI, S. REBAY, G. MARIOTTI, S. PEDINOTTI, AND M. SAVINI, *A high-order accurate discontinuous finite element method for inviscid and viscous turbomachinery flows*, in 2nd European Conference on Turbomachinery Fluid Dynamics and Thermodynamics, R. Decuyper and G. Dibelius, eds., Antwerpen, Belgium, March 5–7 1997, Technologisch Instituut, pp. 99–108.
- [7] C. BAUMANN AND J. ODEN, *A discontinuous hp-finite element method for convection-diffusion problems*, Comput. Methods Appl. Mech. Engrg., **175** (1999), pp. 311–341.
- [8] F. BREZZI, B. COCKBURN, L.D. MARINI, AND E. SÜLI, *Paper in preparation*.
- [9] F. BREZZI AND M. FORTIN, *Mixed and hybrid finite element methods*, Springer-Verlag, New York, 1991.
- [10] F. BREZZI, G. MANZINI, L.D. MARINI, P. PIETRA, AND A. RUSSO, *Discontinuous Galerkin approximations for elliptic problems*, Numer. Methods Partial Differential Equations., **16** (2000), pp. 365–378.
- [11] F. BREZZI, T.J.R. HUGHES, L.D. MARINI, AND A. MASUD, *Mixed Discontinuous Galerkin methods for Darcy flow*, to appear in J. Sci. Comput..

- [12] F. BREZZI, L.D. MARINI, AND E. SÜLI, *Discontinuous Galerkin methods for first-order hyperbolic problems*, Math. Models Methods Appl. Sci. **14**, (12), (2004).
- [13] E. BURMAN AND P. HANSBO, *Edge stabilization for Galerkin approximations of convection-diffusion-reaction problems*, to appear in Comput. Methods Appl. Mech. Engrg..
- [14] P. CASTILLO, B. COCKBURN, I. PERUGIA, AND D. SCHÖTZAU, *An a priori error analysis of the local discontinuous Galerkin method for elliptic problems*, SIAM J. Numer. Anal., **38** (2000), pp. 1676–1706.
- [15] P. CIARLET, *The finite element method for elliptic problems*, North-Holland, Amsterdam, 1978.
- [16] B. COCKBURN, *Discontinuous Galerkin Methods*, ZAMM Z. Angew. Math. Mech., **83** (2003), pp. 731–754.
- [17] B. COCKBURN AND C. DAWSON, *Some extensions of the local discontinuous Galerkin method for convection-diffusion equations in multidimensions*, J.R. Whiteman, Ed., in The Proceedings of the Conference on the Mathematics of Finite Elements and Applications: MAFELAP X, Elsevier (2000), pp. 225–238.
- [18] B. COCKBURN AND C.W. SHU, *The local discontinuous Galerkin finite element method for convection-diffusion systems*, SIAM J. Numer. Anal., **35** (1998), pp. 2440–2463.
- [19] C.N. DAWSON, S. SUN, AND M.F. WHEELER, *Compatible algorithms for coupled flow and transport*. to appear in Comput. Meth. Appl. Mech. Engrg..
- [20] J. DOUGLAS, JR. AND T. DUPONT, *Interior penalty procedures for elliptic and parabolic Galerkin methods*, Lecture Notes in Physics, **58**, Springer-Verlag, Berlin, 1976.
- [21] L.P. FRANCA, S.L. FREY AND T.J.R. HUGHES, *Stabilized Finite Element Methods: I. Application to the Advective-Diffusive Model*, Comput. Methods Appl. Mech. Engrg., **95** (1992), pp. 253–276.
- [22] P. HOUSTON, M. JENSEN, AND E. SÜLI, *hp-Discontinuous Galerkin finite element methods with least-squares stabilization*, Journal of Scientific Computing, **17** (2002), pp. 1–26.

- [23] P. HOUSTON, CH. SCHWAB, AND E. SÜLI, *Discontinuous hp-finite element methods for advection-diffusion-reaction problems*, SIAM J. of Numer. Anal., **39** (2002), pp. 2133–2163.
- [24] T.J.R. HUGHES AND L.P. FRANCA, *A new finite element formulation for computational fluid dynamics: VII. The Stokes problem with various well-posed boundary conditions: symmetric formulations that converge for all velocity/pressure spaces*, Comput. Methods Appl. Mech. Engrg., **65** (1987), pp. 85–96.
- [25] A. MASUD AND T.J.R. HUGHES, *A stabilized mixed finite element method for Darcy flow*, Comput. Methods Appl. Mech. Engrg., **191** (2002), pp. 4341–4370.
- [26] B. RIVIÈRE, M. WHEELER, AND V. GIRAULT, *Improved energy estimates for interior penalty, constrained and discontinuous Galerkin methods for elliptic problems. Part I*, Comp. Geosci., **3** (1999), pp. 337–360.
- [27] S. SUN, *Discontinuous Galerkin Methods for Reactive Transport in Porous Media*. Ph.D. thesis, The University of Texas at Austin, 2003.
- [28] M.F. WHEELER, *An elliptic collocation-finite element method with interior penalties*, SIAM J. Numer. Anal., **15** (1978), pp. 152–161.