IMA Journal of Numerical Analysis (2001) **21**, 165–197

Subgrid stabilization of Galerkin approximations of linear monotone operators

J.-L. GUERMOND[†]

LIMSI (CNRS-UPR 3152), BP 133, 91403, Orsay, France

[Received 16 September 1998 and in revised form 17 January 2000]

This paper presents a stabilized Galerkin technique for approximating monotone linear operators in a Hilbert space. The key idea consists in introducing an approximation space that is broken up into resolved scales and subgrid scales so that the bilinear form associated with the problem satisfies a uniform inf-sup condition with respect to this decomposition. An optimal Galerkin approximation is obtained by introducing an artificial diffusion on the subgrid scales.

Keywords: linear first-order PDEs; non-coercive linear operator; monotone operators; hierarchical finite elements; stabilization; subgrid modelling; artificial viscosity.

1. Introduction

This paper presents a stabilized Galerkin technique for approximating non-coercive monotone linear operators in a Hilbert space. More precisely, let $X \subset V \subset L$ be three Hilbert spaces with dense and continuous embedding. The inner product of L is denoted by $(\cdot,\cdot)_L$. This paper deals with the approximation of the following abstract problem:

$$\text{For } f \in L,\ \text{find } u \in X,\ \forall v \in X \qquad a(u,v) + \epsilon d(u,v) = (f,v)_L, \tag{1.1}$$

where $\epsilon \geqslant 0$, $d \in \mathcal{L}(X \times X; \mathbb{R})$ is X-coercive and the bilinear form a is in $\mathcal{L}(V \times L; \mathbb{R})$ and satisfies

$$\begin{cases} \exists c > 0 & \inf\limits_{u\in V} \sup\limits_{v\in L} \dfrac{a(u,v)}{\|u\|_V \|v\|_L} \geqslant c \\ \forall v \in L & (v \neq 0) \Rightarrow \left(\sup\limits_{u\in V} \dfrac{a(u,v)}{\|u\|_V} \neq 0 \right). \end{cases} \tag{1.2}$$

By defining the operator $A : D(A) = V \longrightarrow L$ so that $(Au,v)_L = a(u,v)$, these conditions are equivalent to stating that A is bijective. Furthermore, we assume that d is associated with an unbounded operator $D : D(D) \subset X \longrightarrow L$ so that $(Du,v)_L = d(u,v)$. In practice, the situation we shall study corresponds to A being a first-order differential operator and D being a coercive second-order differential operator (think of $D = -\Delta : H^2(\Omega) \cap H_0^1(\Omega) \longrightarrow L^2(\Omega)$).

[†]Email: guermond@limsi.fr

When $\epsilon = 0$, problem (1.1) reduces to the following.

$$\text{For } f \in L, \quad \text{find } u \in V,\ \forall v \in L \qquad a(u, v) = (f, v)_L. \tag{1.3}$$

This problem can be solved efficiently by considering its least square formulation; namely,

$$\text{For } f \in L \quad \text{find } u \in V,\ \forall v \in V \qquad (Au, Av)_L = (f, Av)_L.$$

Thanks to the first inequality in (1.2), which is equivalent to $\|Aw\|_L \geqslant c\|w\|_V$ for all $w \in V$, the bilinear form $(Au, Av)_L$ is V-coercive; hence, this formulation lends itself quite efficiently to approximation by means of the conforming Galerkin technique.

In general, the situation is a little bit more complex, since ϵ is not zero but may be arbitrarily small. As a result, coercivity may not be strong enough for the Galerkin approximation to work properly. Consequently, the least square technique may seem to be a good alternative to solve (1.1) also. However, since the domain of $A+\epsilon D$ controls second derivatives, conformity requires the least square method to work with $\mathcal{C}^1$ finite elements or with the scalar product of the dual of X; namely X'. One way to avoid this difficulty is to use the Galerkin/least square technique (see e.g. Brooks & Hughes, 1982; Hughes *et al.*, 1989). This approach consists of a linear combination of the Galerkin and the least square formulations. More precisely, by denoting $A_\epsilon = A + \epsilon D$, it consists of the following.

$$\begin{cases} \text{For } f \in L,\ \text{find } u \in X,\ \text{so that } \forall v \in X \\ a(u, v) + \epsilon d(u, v) + \delta(\epsilon, h) \displaystyle\sum_{T \in \mathcal{T}_h} (A_\epsilon u, A_\epsilon v)_{L,T} = (f, v)_L + \delta(\epsilon, h) \sum_{T \in \mathcal{T}_h} (f, A_\epsilon v)_{L,T}, \end{cases}$$

where $\cup_{T \in \mathcal{T}_h} T$ is a triangulation, $(\cdot, \cdot)_{L,T}$ is the restriction of the L-scalar product to the element T, and the coefficient $\delta(\epsilon, h)$ is chosen as follows

$$\begin{cases} \delta(\epsilon, h) = h, & \text{if } \epsilon \leqslant h \\ 0 \leqslant \delta(\epsilon, h) \leqslant h, & \text{if } h \leqslant \epsilon \leqslant 2h \\ \delta(\epsilon, h) = 0, & \text{if } 2h \leqslant \epsilon. \end{cases}$$

This method is quite popular and works quite well. However, there are two problems:

1. There is a tuning coefficient that depends on the presence or the absence of the coercive operator. The tuning is easily controllable in academic situations, but is a tricky task for realistic problems (think of variable nonlinear viscosity or degenerate elliptic operators, etc).
2. To the best of the author's knowledge, the least square and Galerkin least square methods cannot be generalized to time-dependent problems without using discontinuous space–time finite elements.

The objective of the present paper is to propose a method that has the following features. First, it has the same stability and approximation properties as the least square and Galerkin least square methods for problem (1.1) but has *no tuning coefficient* that depends on ϵ (see Remark 3.2 below). Second, it can be very easily generalized to approximate linear contraction semi-groups by using standard finite element techniques.

The theory developed herein is based on two principles:

(i) Since (1.2) guarantees the solution of problem (1.3) to be stable in the norm of V, we introduce three approximation spaces $X_h = X_H \oplus X_h^H$ so that the triplet (X_h, X_H, a) satisfies a discrete inf-sup condition similar to (1.2). We refer to X_H as the resolved scale space and to X_h^H as the subgrid scale space. The discrete inf-sup condition in question permits the resolved scales of the approximate solution to be controlled in the graph norm of A.
(ii) The subgrid scales can be controlled, in turn, by means of a small artificial diffusion mechanism; the control being provided by a simple energy argument.

The outline of the paper is as follows. In §2 we concentrate on model problem (1.3). Stability and quasi-optimal convergence results are proved. The results of §2 are generalised in §3 to the case of (1.1). The theory presented in §2 and §3 relies on a uniform inverse inequality (2.9) that is true for finite elements provided the mesh underlying the approximation is quasi-uniform. This constraint being too strong for practical purposes, since it *a priori* excludes local refinement and mesh adaptation, the theory is generalized in §4 to the case of nonuniform meshes by using a local version of the inverse inequality. Examples of applications of the present theory are shown in §5. Some of the results presented in this paper were announced in Guermond (1999b,c).

2. Approximation of a model problem on uniform meshes

In this section, we concentrate on model problem (1.3) and we think of A as a first-order differential operator.

2.1 *A model problem*

Let L be a real separable Hilbert space and V be a dense subspace continuously embedded in L. Hereafter we identify L and its dual so that we are in the following classical situation: $V \subset L \equiv L' \subset V'$.

We introduce a continuous bilinear form $a : V \times L \longrightarrow \mathbb{R}$, and we assume that a is monotone; that is

$$\forall u \in V \qquad a(u,u) \geqslant 0. \tag{2.1}$$

We introduce the symmetric part $a_s : V \times V \longrightarrow \mathbb{R}$ of a as follows

$$\forall (u,v) \in V \times V \qquad a_s(u,v) = \frac{1}{2}(a(u,v) + a(v,u)). \tag{2.2}$$

It is clear that for all u in V we have $a(u,u) = a_s(u,u) \geqslant 0$; as a result, a_s is a symmetric monotone bilinear form. Hereafter we shall make use of the following classical property:

LEMMA 2.1 Let E be a vector space and $x : E \times E \longrightarrow \mathbb{R}$ be a symmetric monotone bilinear form then

$$\forall (u,v) \in E \times E \qquad x(u,v) \leqslant x(u,u)^{1/2} x(v,v)^{1/2}.$$

Furthermore, we assume that there is $c > 0$ so that

$$\begin{cases} \forall u \in V & \sup_{v \in L} \dfrac{a(u,v)}{\|v\|_L} \geqslant c\|u\|_V^2, \\ \forall v \in L & (v \neq 0) \Rightarrow \left(\sup_{u \in V} \dfrac{a(u,v)}{\|u\|_V} > 0 \right). \end{cases} \tag{2.3}$$

This condition is equivalent to stating that the problem

$$\begin{cases} \text{For } f \in L, \text{ find } u \in V \text{ so that} \\ a(u,v) = (f,v) \qquad \forall v \in L, \end{cases} \tag{2.4}$$

has a unique solution. More precisely we have the following result.

THEOREM 2.1 Problem (2.4) has a unique solution and this solution satisfies

$$c_1\|u\|_V \leqslant \|f\|_L. \tag{2.5}$$

Proof. This is a consequence of (2.3) together with a classical characterization of bijective linear operators in reflexive Banach spaces, cf. e.g. Brezis (1983, pp 29–31). □

REMARK 2.1 When A is a first-order differential operator, problem (2.4) is essentially a Petrov–Galerkin problem; that is, the solution space and the test space are different. The failure of discrete Galerkin techniques to approximate properly this problem is rooted in this basic fact. In general, the first inequality in (2.3) is not satisfied (uniformly with respect to the mesh size) at the discrete level.

2.2 *The discrete setting*

To build a discrete approximation of u, we introduce X_H and X_h, two finite-dimensional subspaces of V. The indices H and h denote two positive parameters that tend to zero. In the practical applications described in §5 we have $h \approx H/2$.

The space X_H is assumed to have the following approximation property: there is W, a dense subspace of V, and there are $k > 0$ and $c > 0$ so that, for all $v \in W$

$$\inf_{w_H \in X_H} \|v - w_H\|_L + H\|v - w_H\|_V \leqslant cH^{k+1}\|v\|_W. \tag{2.6}$$

From now on, c denotes a generic constant that does not depend on (H, h) and the value of which may change in different occurrences.

The couple (X_H, X_h) is assumed to satisfy the following discrete inf-sup condition: there is $c_a > 0$, independent of (H, h), such that

$$\forall v_H \in X_H \qquad \sup_{\phi_h \in X_h} \frac{a(v_H, \phi_h)}{\|\phi_h\|_L} \geqslant c_a\|v_H\|_V. \tag{2.7}$$

Furthermore, we assume that $X_H \subset X_h$, and there is a linear projection operator $P_H : X_h \longrightarrow X_H$ that is stable with respect to the L-norm:

$$\exists c > 0,\ \forall (H,h),\ \forall v_h \in X_h \qquad \|P_H v_h\|_L \leqslant c\|v_h\|_L. \tag{2.8}$$

For further references, we denote $X_h^H = (1 - P_H)X_h$, and for all v_h in X_h we set $v_H = P_H v_h$ and $v_h^H = v_h - v_H$.

Since X_h is a finite-dimensional normed vector space, we assume that the following inverse inequality holds:

$$\forall v_h \in V \qquad \|v_h\|_V \leqslant c_i H^{-1} \|v_h\|_L. \tag{2.9}$$

REMARK 2.2 It is shown in §5 that it is possible in general to find couples (X_h, X_H) satisfying the discrete condition above, where X_h can be broken up as follows: $X_h = X_H \oplus X_h^H$, the decomposition being L-stable. We refer to X_H as the resolved scale space and to X_h^H as the subgrid scale space. This decomposition can be formally interpreted as follows. For instance, assume that a is associated with a first-order differential operator and assume that a finite element piecewise linear approximation of u is sought. The action of the differential operator on any piecewise linear function generates discontinuities at the interfaces of the finite elements. These discontinuities contain very small Fourier modes that cannot be captured when tested against piecewise linear test functions; as a consequence, the Galerkin technique is in general suboptimal for this class of problems (unless the test space is finely tuned, see Baiocchi *et al.*, 1993). On the other hand, optimality can be recovered if subgrid scale test functions (i.e. those of X_h^H) are added to the conventional test space composed of piecewise linear functions (i.e. X_H; the resolved scales). The inf-sup condition (2.7) is the discrete counterpart of the first inequality in (1.2); it warrants the resolved scales of the approximate solution to be stable in the norm of V (i.e. to be free of spurious numerical wiggles).

REMARK 2.3 In the case of a finite element approximation, (2.9) holds uniformly if A is a first-order differential operator, the mesh is quasi-uniform, and $c_1 h \leqslant H \leqslant c_2 h$. The quasi-uniformity constraint is quite stringent since it *a priori* excludes local refinement and mesh adaptation. The present theory will be extended to nonuniform meshes in §4. The second constraint is equivalent to assuming that the dimension of X_H is a fraction of that of X_h. We shall see in §5 that in applications we have $H = h$ or $H = 2h$.

2.3 *The discrete problem*

Having introduced subgrid scales to control the resolved scales of the approximate solution, we are left with the problem of controlling also subgrid scales. To this purpose, we introduce an artificial diffusion mechanism; that is, we define a bilinear form $b_h : X_h^H \times X_h^H \longrightarrow \mathbb{R}$ that satisfies the following continuity and coercivity properties: there is $c_B > 0$ such that

$$\begin{cases} b_h(v_h^H, v_h^H) \geqslant H \|v_h^H\|_b^2, \\ b_h(v_h^H, w_h^H) \leqslant c_B H \|v_h^H\|_b \|w_h^H\|_b, \end{cases} \tag{2.10}$$

where the norm $\|\cdot\|_b$ is such that there are two constants $c_{e1} > 0$ and $c_{e2} > 0$ so that

$$\forall v_h^H \in X_h^H \qquad c_{e1} \|v_h^H\|_V \leqslant \|v_h^H\|_b \leqslant c_{e2} H^{-1} \|v_h^H\|_L. \tag{2.11}$$

EXAMPLE 2.1 Let $((\cdot, \cdot))_V$ denote the inner product in V. The simplest choice for b_h is $b_h(v_h^H, w_h^H) = H((v_h^H, w_h^H))_V$.

EXAMPLE 2.2 Let X be a dense subspace of V, continuously embedded in V, so that $\|v_h^H\|_X \leqslant c_{e2}H^{-1}\|v_h^H\|_L$ for all v_h^H in X_h^H; then, by assuming that $X_h \subset X$ and by denoting $((\cdot,\cdot))_X$ the inner product in X, one can set $b_h(v_h^H, w_h^H) = H((v_h^H, w_h^H))_X$.

EXAMPLE 2.3 For the scalar advection equation $\beta\cdot\nabla u = f$ in Ω, with suitable assumptions on the vector field β, we have $L = \mathrm{L}^2(\Omega)$ and $V = \{v \in \mathrm{L}^2(\Omega) \mid \beta\cdot\nabla v \in \mathrm{L}^2(\Omega),\ v_{|\Gamma^-} = 0\}$, where Γ^- is the inflow boundary (see Azerad & Pousin, 1996, or Bardos, 1970, for technical details on this problem). Since $\mathrm{H}^1(\Omega) \subset V$, the following two definitions are possible for b_h:

$$b_h(v_h^H, w_h^H) = H\int_\Omega v_h^H w_h^H + \begin{cases} H\int_\Omega(\beta\cdot\nabla v_h^H)(\beta\cdot\nabla w_h^H), \\ H\int_\Omega(\nabla v_h^H)\cdot(\nabla w_h^H). \end{cases} \tag{2.12}$$

The second model may be helpful in practice to avoid cross-wind oscillations when approximating very stiff problems.

LEMMA 2.2 There is $c_b > 0$ such that

$$\forall(H,h),\ \forall v_h^H \in X_h^H \qquad \sup_{w_h\in X_h}\frac{b_h(v_h^H, w_h^H)}{\|w_h\|_L} \leqslant c_b\|v_h^H\|_b. \tag{2.13}$$

Proof. The inverse stability property (2.11) together with the stability hypotheses (2.8) and (2.10) yields

$$\begin{aligned} b_h(v_h^H, w_h^H) &\leqslant c_B H\|v_h^H\|_b\|w_h^H\|_b \\ &\leqslant c_B c_{e2}\|v_h^H\|_b\|w_h^H\|_L \\ &\leqslant c_B c_{e2}\|v_h^H\|_b\|(1-P_H)w_h\|_L \\ &\leqslant c_B c_{e2}\|1-P_H\|\,\|v_h^H\|_b\|w_h\|_L. \end{aligned}$$

The proof is complete. □

The discrete problem we consider hereafter is

$$\begin{cases} \text{Find } u_h \text{ in } X_h \text{ such that} \\ a(u_h, v_h) + b_h(u_h^H, v_h^H) = (f, v_h) \qquad \forall v_h \in X_h. \end{cases} \tag{2.14}$$

PROPOSITION 2.1 The discrete problem (2.14) has a unique solution.

Proof. Since the problem is set in a finite-dimensional vector space, it is sufficient to prove an *a priori* bound on u_h. By using u_h as a test function in (2.14) we obtain

$$\begin{aligned} H\|u_h^H\|_b^2 &\leqslant \|f\|_L\|u_h\|_L \\ &\leqslant c\|f\|_L\|u_h\|_V \\ &\leqslant c\|f\|_L(\|u_h^H\|_V + \|u_H\|_V) \\ &\leqslant c\|f\|_L(c_{e1}^{-1}\|u_h^H\|_b + \|u_H\|_V). \end{aligned}$$

An *a priori* bound on $\|u_H\|_V$ is provided by the discrete inf-sup condition (2.7),

$$\begin{aligned} c_a\|u_H\|_V &\leqslant \sup_{\phi_h\in X_h} \frac{a(u_H,\phi_h)}{\|\phi_h\|_L} \\ &\leqslant \sup_{\phi_h\in X_h} \frac{f - a(u_h^H,\phi_h) - b_h(u_h^H,\phi_h^H)}{\|\phi_h\|_L} \\ &\leqslant \|f\|_L + \|a\|\|u_h^H\|_V + c_b\|u_h^H\|_b \\ &\leqslant \|f\|_L + c_{e1}^{-1}\|a\|\|u_h^H\|_b + c_b\|u_h^H\|_b \\ &\leqslant \|f\|_L + c\|u_h^H\|_b. \end{aligned}$$

As a result, by substituting this bound into the previous inequality, we have

$$\begin{aligned} H\|u_h^H\|_b^2 &\leqslant c\|f\|_L(\|f\|_L + \|u_h^H\|_b) \\ &\leqslant c\Big(1+\frac{c}{2H}\Big)\|f\|_L^2 + \frac{1}{2}H\|u_h^H\|_b^2, \end{aligned}$$

from which we infer

$$\|u_h^H\|_b + \|u_H\|_V \leqslant c(H)\|f\|_L,$$

where $c(H)$ depends continuously on H. This completes the proof. □

REMARK 2.4 The basic principles of the proposed technique can be summarized as follows: introduce subgrid scales to capture the discontinuities generated by the differential operator when acting on the approximate solution, and control the subgrid scales by an artificial viscosity. The goal of the present paper is to show that a quasi-optimal Galerkin approximation of problem (1.3) can be built by combining these two ideas. The notion of scale separation and artificial dissipation of subgrid scales is rooted in many works: e.g. subgrid modelling (Smagorinsky, 1963; Germano *et al.*, 1991), the nonlinear Galerkin method (Foias *et al.*, 1988; Marion & Temam, 1990), and the stabilizing property of bubble functions (Arnold *et al.*, 1984; Brezzi *et al.*, 1992; Baiocchi *et al.*, 1993; Crouzeix & Raviart, 1973).

2.4 *Error analysis*

The main convergence result of this section is

THEOREM 2.2 The discrete solution u_h of (2.14) satisfies

$$\begin{cases} a_s(u-u_h,u-u_h)^{1/2} \leqslant c \inf\limits_{w_H\in X_H} \big[H^{-1/2}\|u-w_H\|_L + H^{1/2}\|u-w_H\|_V\big], \\ \|u-u_h\|_V + \|u_h^H\|_b \leqslant c \inf\limits_{w_H\in X_H} \big[H^{-1}\|u-w_H\|_L + \|u-w_H\|_V\big]. \end{cases} \tag{2.15}$$

Proof. Let us introduce some notation. Let w_H be an arbitrary element in X_H; we set $\eta_h = u - w_H$, and $e_h = w_H - u_h$. Note that we have $u - u_h = \eta_h + e_h$.

The equation that controls e_h is obtained by subtracting (2.14) from (2.4) where the test functions span X_h:

$$\forall v_h \in X_h \qquad a(e_h, v_h) - b_h(u_h^H, v_h^H) = -a(\eta_h, v_h).$$

Since X_H is invariant under the projection P_H and P_H is linear, we infer

$$\begin{aligned} u_h^H &= u_h - P_H u_h \\ &= u_h - w_H - P_H(u_h - w_H) \\ &= -e_h + P_H e_h \\ &= -e_h^H. \end{aligned}$$

As a result, the equation that controls e_h can be recast into the form

$$\forall v_h \in X_h \qquad a(e_h, v_h) + b_h(e_h^H, v_h^H) = -a(\eta_h, v_h).$$

By taking e_h as a test function and by using the coercivity property (2.10) we obtain

$$a_s(e_h, e_h) + H\|e_h^H\|_b^2 \leqslant -a(\eta_h, e_h).$$

We control the right-hand side of the inequality above by proceeding as follows.

$$\begin{aligned} a_s(e_h, e_h) + H\|e_h^H\|_b^2 &\leqslant -a(\eta_h, e_h) \\ &\leqslant a(e_h, \eta_h) - 2a_s(e_h, \eta_h) \\ &\leqslant \|a\|\,\|e_h\|_V\|\eta_h\|_L + \gamma a_s(e_h, e_h) + \frac{1}{\gamma}a_s(\eta_h, \eta_h), \end{aligned} \tag{2.16}$$

where we have used Lemma 2.1 and the inequality $2xy \leqslant \gamma x^2 + y^2/\gamma$ which is valid for any positive constant γ. This constant will be chosen to meet our needs. Hereafter, γ denotes a generic constant that can be chosen as small as needed and c_γ is a constant that depends on γ; the values of γ and c_γ may change at different occurrences.

To obtain a control on $\|e_h\|_V$ we use the discrete inf-sup condition (2.7),

$$\begin{aligned} c_a\|e_H\|_V &\leqslant \sup_{\phi_h \in X_h} \frac{a(e_H, \phi_h)}{\|\phi_h\|_L} \\ &\leqslant \sup_{\phi_h \in X_h} \frac{-a(\eta_h, \phi_h) - a(e_h^H, \phi_h) - b_h(e_h^H, \phi_h^H)}{\|\phi_h\|_L} \\ &\leqslant \|a\|(\|\eta_h\|_V + \|e_h^H\|_V) + c_b\|e_h^H\|_b \\ &\leqslant c(\|\eta_h\|_V + \|e_h^H\|_b). \end{aligned}$$

By using this bound and a triangle inequality, we obtain $\|e_h\|_V \leqslant c(\|\eta_h\|_V + \|e_h^H\|_b)$. By substituting this bound into (2.16), we have

$$\begin{aligned} (1-\gamma)a_s(e_h, e_h) + H\|e_h^H\|_b^2 &\leqslant c(\|\eta_h\|_V + \|e_h^H\|_b)\|\eta_h\|_L + \frac{1}{\gamma}a_s(\eta_h, \eta_h), \\ &\leqslant c_\gamma\|\eta_h\|_V\|\eta_h\|_L + \gamma H\|e_h^H\|_b^2 + c'_\gamma H^{-1}\|\eta_h\|_L^2. \end{aligned}$$

By choosing $\gamma = 1/2$, we obtain

$$a_s(e_h, e_h) + H\|e_h^H\|_b^2 \leqslant cH^{-1}(\|\eta_h\|_L^2 + H^2\|\eta_h\|_V^2).$$

As a result we infer

$$a_s(u - u_h, u - u_h) + H(\|u - u_h\|_V^2 + \|u_h^H\|_b^2) \leqslant cH^{-1} \inf_{w_H \in X_H} [\|u - w_H\|_L^2 + H^2\|u - w_H\|_V^2].$$

The proof is complete. □

COROLLARY 2.1 If u, the solution of (2.4), is in W, the discrete solution u_h of (2.14) satisfies

$$\begin{cases} a_s(u - u_h, u - u_h)^{1/2} \leqslant cH^{k+1/2}\|u\|_W, \\ \|u - u_h\|_V + \|u_h^H\|_b \leqslant cH^k\|u\|_W. \end{cases} \tag{2.17}$$

REMARK 2.5 The bound (2.17) is optimal in the norm of V. On the other hand, if a_s is L-coercive, (2.17) is not optimal in the norm of L: a factor $H^{1/2}$ is missing. Actually, it can be shown, by proceeding is in Zhou (1997) or Guermond (1999a), that optimality can be recovered if the mesh underlying the approximation space X_h satisfies special geometric properties.

REMARK 2.6 The estimate (2.17) is identical to the one that could be obtained by applying the Galerkin least square method to the present problem (see Johnson *et al.*, 1984, or Hughes *et al.*, 1989).

EXAMPLE 2.4 In the case of a convection problem, $\beta \cdot \nabla u = f$ (under some reasonable assumptions on β), we have $\|\cdot\|_L = \|\cdot\|_0$ and $\|\cdot\|_V = \|\cdot\|_0 + \|\beta \cdot \nabla \cdot\|_0$. For finite element approximations, the convergence result reads $\|u - u_h\|_V \leqslant cH^k\|u\|_{k+1}$, which is optimal; see Guermond (1999a) or §5 for examples of admissible $\mathbb{P}_1$ and $\mathbb{P}_2$ finite elements.

2.5 *A possible improvement in the definition of b_h*

The definition of the bilinear form b_h can be sharpened if further assumptions on a are made. Let us assume that $a = a_0 + a_1$ where the two bilinear forms a_0 and a_1 have the following continuity properties:

$$\forall (u, v) \in V \times L \qquad a_0(u, v) \leqslant c_0 a_s(u, u)^{1/2}\|v\|_L, \tag{2.18}$$

$$\forall (u, v) \in V \times L \qquad a_1(u, v) \leqslant c_1|u|_V\|v\|_L, \tag{2.19}$$

where $|\cdot|_V$ is a semi-norm in V such that

$$\forall u \in V \qquad \|u\|_V \leqslant c(a_s(u, u)^{1/2} + |u|_V). \tag{2.20}$$

Furthermore, we assume that a_1, X_h, and X_H satisfy the following property: There are $c_{a1} > 0$, $c_\delta \geqslant 0$, independent of (H, h), such that

$$\forall u_h \in X_h \qquad \sup_{v_h \in X_h} \frac{a_1(u_H, v_h)}{\|v_h\|_L} \geqslant c_{a1}|u_H|_V - c_\delta[a_s(u_h, u_h)^{1/2} + |u_h^H|_V]. \tag{2.21}$$

REMARK 2.7 Note that (2.21) is weaker that (2.7); actually, only $|u_H|_V$ needs to be controlled by the inf-sup inequality, since $a_s(u_h, u_h)$ is already controlled by the monotonicity of a, and $|u_h^H|_V$ by the coercivity of b_h.

Now we weaken slightly the definition of b_h as follows:

$$\forall (v_h^H, w_h^H) \in X_h^H \times X_h^H \qquad \begin{cases} b_h(v_h^H, v_h^H) \geqslant H|v_h^H|_b^2, \\ b_h(v_h^H, w_h^H) \leqslant c_B H|v_h^H|_b|w_h^H|_b, \end{cases} \tag{2.22}$$

where the semi-norm $|\cdot|_b$ is such that there are $c_{e1} > 0$ and $c_{e2} > 0$ such that

$$\forall v_h^H \in X_h^H \qquad c_{e1}|v_h^H|_V \leqslant |v_h^H|_b \leqslant c_{e2} H^{-1} \|v_h^H\|_L. \tag{2.23}$$

EXAMPLE 2.5 This situation corresponds to equations like $u + \beta \cdot \nabla u = f$ in Ω. Assuming $\|\operatorname{div} \beta\|_{0,\infty} < 2$, the bilinear form a is $L^2(\Omega)$-coercive. Then, instead of using $b_h(v_h^H, w_h^H) = H \int_\Omega v_h^H w_h^H + (\beta \cdot \nabla v_h^H)(\beta \cdot \nabla w_h^H)$, one can use one of the following definitions

$$\forall (v_h^H, w_h^H) \in X_h^H \times X_h^H \qquad b_h(v_h^H, w_h^H) = \begin{cases} H \int_\Omega (\beta \cdot \nabla v_h^H)(\beta \cdot \nabla w_h^H), \\ H \int_\Omega (\nabla v_h^H) \cdot (\nabla w_h^H). \end{cases} \tag{2.24}$$

The main interest of the first alternative definition is that the artificial dissipation is zero in the regions where β is zero and it does not introduce cross-wind diffusion. In both models, the stabilizing terms are expected to be small in the regions where the gradient of the solution is small. In other words, unlike models (2.12), models (2.24) put artificial diffusion only where it is needed.

THEOREM 2.3 If u, the solution of (2.4), is in W, then the discrete solution u_h of (2.14) satisfies

$$a_s(u - u_h, u - u_h)^{1/2} + H^{1/2}|u - u_h|_V \leqslant c H^{k+1/2} \|u\|_W. \tag{2.25}$$

Proof. The proof is almost identical to that of Theorem 2.2. By taking e_h as a test function and by using the coercivity property (2.22), we obtain

$$a_s(e_h, e_h) + H|e_h^H|_b^2 \leqslant a_1(e_h, \eta_h) + a_0(e_h, \eta_h) - 2a_s(e_h, \eta_h).$$

Each term on the right-hand side is bounded from above as follows:

$$a_1(e_h, \eta_h) \leqslant c|e_h|_V \|\eta_h\|_L,$$

$$\begin{aligned} a_0(e_h, \eta_h) &\leqslant c a_s(e_h, e_h)^{1/2} \|\eta_h\|_L \\ &\leqslant \gamma a_s(e_h, e_h) + c_\gamma \|\eta_h\|_L^2, \end{aligned}$$

$$\begin{aligned} -2a_s(e_h, \eta_h) &\leqslant \gamma a_s(e_h, e_h) + c_\gamma a_s(\eta_h, \eta_h) \\ &\leqslant \gamma a_s(e_h, e_h) + c_\gamma \|\eta_h\|_V \|\eta_h\|_L. \end{aligned}$$

By inserting these bounds into the inequality above, we obtain

$$a_s(e_h, e_h) + H|e_h^H|_b^2 \leqslant c(|e_h|_V \|\eta_h\|_L + \|\eta_h\|_V \|\eta_h\|_L).$$

The control on the remaining term $|e_h|_V$ is obtained by means of the weakened inf-sup inequality (2.21):

$$\begin{aligned} c_a|e_H|_V &\leqslant \sup_{v_h \in X_h} \frac{a_1(e_H, v_h)}{\|v_h\|_L} + c_\delta[a_s(e_h, e_h)^{1/2} + |e_h^H|_V] \\ &\leqslant \sup_{v_h \in X_h} \frac{-a_0(e_h, v_h) - a_1(e_h^H, v_h) - b_h(e_h^H, v_h^H) - a(\eta_h, v_h)}{\|v_h\|_L} \\ &\quad + c_\delta[a_s(e_h, e_h)^{1/2} + |e_h^H|_V] \\ &\leqslant c[a_s(e_h, e_h)^{1/2} + |e_h^H|_b + \|\eta_h\|_V]. \end{aligned}$$

Hence, we have

$$\begin{aligned} |e_h|_V \|\eta_h\|_L &\leqslant c(a_s(e_h, e_h)^{1/2} + |e_h^H|_b + \|\eta_h\|_V)\|\eta_h\|_L \\ &\leqslant \gamma[a_s(e_h, e_h) + H|e_h^H|_b^2] + c_\gamma[H^{-1}\|\eta_h\|_L^2 + \|\eta_h\|_V \|\eta_h\|_L]. \end{aligned}$$

Finally we obtain

$$a_s(e_h, e_h) + H|e_h^H|_b^2 \leqslant c[H^{-1}\|\eta_h\|_L^2 + H\|\eta_h\|_V^2].$$

The rest of the proof is evident. □

3. The full problem

In this section we return to the original problem (1.1). We shall treat this situation as a perturbation of the previous one. We shall think of the operator D as a (possibly degenerate) elliptic second-order differential operator.

3.1 *The abstract framework*

In addition to the two Hilbert spaces, L and V, already defined, we introduce a new Hilbert space X that is dense and continuously embedded in L. Hereafter, we make the identifications

$$V \subset L \equiv L' \subset V' \quad \text{and} \quad X \subset L \equiv L' \subset X'.$$

We introduce a continuous bilinear form $d : X \times X \longrightarrow \mathbb{R}$, and we assume that there is a semi-norm $|\cdot|_X$ in X such that $d(u, v) \leqslant c_d|u|_X|v|_X$ for all u and v in X. In practice, d can be a degenerate elliptic operator. We also assume that $a + d$ is coercive with respect to the semi-norm $|\cdot|_X$; that is,

$$\forall v \in V \cap X \qquad |v|_X^2 \leqslant a_s(v, v) + d_s(v, v) = a(v, v) + d(v, v). \tag{3.1}$$

We shall now consider the following problem:

$$\text{For } f \in L, \text{ find } u \in V \cap X, \quad \forall v \in V \cap X \quad a(u, v) + \epsilon d(u, v) = (f, v), \tag{3.2}$$

where ϵ is a positive real number which may be arbitrarily small. Hereafter, we assume that ϵ is bounded from above by a constant; say $\epsilon \leqslant 1/2$. This hypothesis means only that the problem has been properly normalized. The analysis of this problem is quite difficult in general (see Bardos, 1970, for an introduction to this type of equations). One suitable tool to treat this class of problem consists in the viscosity method (see Barles, 1994, for an introduction to this method) but we shall not dwell on this matter. Actually, the hypotheses assumed up to now are not sufficient to ensure that a solution exists, even if $a + d$ is fully X-coercive. We propose the following counterexample. Let $\Omega =]0, 1[^2$ and consider the following problem

$$\begin{cases} u + \dfrac{\partial u}{\partial x} - \epsilon \dfrac{\partial}{\partial y}\left(\left(x - \dfrac{1}{2}\right)^+ \dfrac{\partial u}{\partial y}\right) = x + 1, \\ u = 0, \quad \text{on} \quad \{0\}\times]0, 1[, \\ u = 0, \quad \text{on} \quad]1/2, 1[\times\{0, 1\}. \end{cases}$$

Let

$$V = \left\{ v \in \mathrm{L}^2(\Omega) \;\middle|\; \frac{\partial v}{\partial x} \in \mathrm{L}^2(\Omega),\ v_{\{0\}\times]0,1[} = 0 \right\}$$

and

$$X = \left\{ v \in \mathrm{L}^2(\Omega) \;\middle|\; H\big(x - \tfrac{1}{2}\big)\frac{\partial v}{\partial y} \in \mathrm{L}^2(\Omega),\ v_{]1/2,1[\times\{0,1\}} = 0 \right\}.$$

For the bilinear forms a and d we have

$$a(u, v) = \int_\Omega uv + v\frac{\partial u}{\partial x}$$

and

$$d(u, v) = \int_\Omega H\big(x - \tfrac{1}{2}\big)\frac{\partial u}{\partial y}\frac{\partial v}{\partial y}.$$

It is clear that $a + d$ is X-coercive, hence uniqueness is ensured, but it can be quite easily shown that the problem considered has no solution in $X \cap V$.

To guarantee that problem (3.2) is well posed it would be sufficient to assume that X is continuously embedded in V and $a+d$ is X-coercive, but we shall not make this hypothesis for the time being. We shall only assume hereafter that problem (3.2) has a unique solution in $V \cap X$.

3.2 *The discrete problem*

Let X_H and X_h be two finite-dimensional subspaces of $V \cap X$. The two spaces X_h and X_H are assumed to satisfy the same hypotheses as in §2; namely, hypotheses (2.6), (2.7), (2.8), (2.9), (2.10), and (2.11). Moreover, we assume that X_h satisfies the following inverse inequality: there is $\mu(H) > 0$ such that

$$\forall v_h \in X_h \qquad \|v_h\|_X \leqslant \mu(H)^{-1}\|v_h\|_L. \tag{3.3}$$

REMARK 3.1 $\mu(H) \sim H$ when d is associated with a second-order differential operator, and $\mu(H) \sim H^2$ when d is associated with a fourth-order differential operator.

The discrete problem we consider hereafter consists in finding u_h in X_h such that

$$\forall v_h \in X_h \qquad a(u_h, v_h) + \epsilon d(u_h, v_h) + b_h(u_h^H, v_h^H) = (f, v_h). \tag{3.4}$$

PROPOSITION 3.1 Problem (3.4) has a unique solution.

Proof. The proof is quite similar to that of Proposition 2.1. Let us prove an *a priori* bound on u_h. By using u_h as a test function in (3.4) we obtain

$$\begin{aligned} a_s(u_h, u_h) + \epsilon d_s(u_h, u_h) + H\|u_h^H\|_b^2 &\leqslant \|f\|_L \|u_h\|_L \\ &\leqslant c\|f\|_L (c_{e1}^{-1}\|u_h^H\|_b + \|u_H\|_V). \end{aligned}$$

An *a priori* bound on $\|u_H\|_V$ is provided by the discrete inf-sup condition (2.7).

$$\begin{aligned} c_a\|u_H\|_V &\leqslant \sup_{\phi_h \in X_h} \frac{a(u_H, \phi_h)}{\|\phi_h\|_L} \\ &\leqslant \sup_{\phi_h \in X_h} \frac{f - a(u_h^H, \phi_h) - b_h(u_h^H, \phi_h^H) - \epsilon d(u_h, \phi_h)}{\|\phi_h\|_L} \\ &\leqslant \|f\|_L + \|a\|\,\|u_h^H\|_V + c_b\|u_h^H\|_b + \epsilon c_d |u_h|_X \sup_{\phi_h \in X_h} \frac{|\phi_h|_X}{\|\phi_h\|_L} \\ &\leqslant \|f\|_L + c_{e1}^{-1}\|a\|\,\|u_h^H\|_b + c_B c_{e2}\|u_h^H\|_b + \epsilon c_d \mu(H)^{-1}|u_h|_X \\ &\leqslant \|f\|_L + c\|u_h^H\|_b + c'\epsilon\mu(H)^{-1}|u_h|_X. \end{aligned}$$

As a result, by substituting this bound into the previous inequality and by using the inequality (3.1) we obtain

$$\begin{aligned} \epsilon|u_h|_X^2 + H\|u_h^H\|_b^2 &\leqslant c\|f\|_L(\|f\|_L + \|u_h^H\|_b + \epsilon\mu(H)^{-1}|u_h|_X), \\ &\leqslant c\|f\|_L^2\left(1 + \frac{c}{2H} + \frac{c\epsilon}{2\mu(H)^2}\right) + \frac{H}{2}\|u_h^H\|_b^2 + \frac{\epsilon}{2}|u_h|_X^2, \end{aligned}$$

from which we infer

$$\|u_h^H\|_b + \|u_H\|_V \leqslant c(H, \mu(H))\|f\|_L.$$

This *a priori* bound (uniform in ϵ) proves uniqueness of the solution to problem (3.4); since X_h is finite-dimensional, this bound also proves existence. □

3.3 *Error analysis*

The main result of this section consists in the following.

THEOREM 3.1 The discrete solution of (3.4) satisfies

$$\begin{aligned} a_s(u - u_h, u - u_h)^{1/2} + \epsilon^{1/2}|u - u_h|_X \leqslant c \inf_{w_H \in X_H} [&H^{1/2}\|u - w_H\|_V \\ &+ \epsilon^{1/2}\|u - w_H\|_X + \kappa H^{-1/2}\|u - w_H\|_L], \end{aligned} \tag{3.5}$$

$$\|u-u_h\|_V \leqslant c\kappa \inf_{w_H\in X_H} [\|u-w_H\|_V + \epsilon^{1/2}H^{-1/2}\|u-u_H\|_X + \kappa H^{-1}\|u-w_H\|_L], \tag{3.6}$$

where we have set $\kappa = 1+\epsilon^{1/2}H^{1/2}\mu(H)^{-1}$.

Proof. We proceed as in the proof of Theorem 2.2. Let w_H be an arbitrary element in X_H, and let us set $\eta_h = u - w_H$ and $e_h = w_H - u_h$.

The equation that controls e_h is obtained by subtracting (3.4) from (3.2) with the test functions spanning X_h:

$$\forall v_h \in X_h \qquad a(e_h,v_h)+\epsilon d(e_h,v_h)+b_h(e_h^H,v_h^H) = -a(\eta_h,v_h)-\epsilon d(\eta_h,v_h),$$

where we have used $u_h^H = -e_h^H$, since X_H is invariant under the projection P_H and P_H is linear.

By taking e_h as a test function and by using the coercivity properties (2.10) and (3.1) we obtain

$$(1-\epsilon)a_s(e_h,e_h)+\epsilon|e_h|_X^2 + H\|e_h^H\|_b^2 \leqslant -a(\eta_h,e_h)-\epsilon d(\eta_h,e_h). \tag{3.7}$$

Now, we have to control the right-hand side of the inequality above. First, we find a bound by proceeding as follows.

$$\begin{aligned}
-a(\eta_h,e_h)-\epsilon d(\eta_h,e_h) &\leqslant a(e_h,\eta_h) - 2a_s(e_h,\eta_h) + \epsilon c_d|\eta_h|_X|e_h|_X \\
&\leqslant \|a\|\,\|e_h\|_V\|\eta_h\|_L + \gamma a_s(e_h,e_h) + \frac{1}{\gamma}a_s(\eta_h,\eta_h) + c\epsilon|\eta_h|_X|e_h|_X \\
&\leqslant \|a\|\,\|e_h\|_V\|\eta_h\|_L + \gamma a_s(e_h,e_h) + \gamma\epsilon|e_h|_X^2 \\
&\quad + c_\gamma\|\eta_h\|_V\|\eta_h\|_L + c'_\gamma\epsilon|\eta_h|_X^2.
\end{aligned}$$

By choosing $\gamma = (1-\epsilon)/2$, one obtains

$$a_s(e_h,e_h)+\epsilon|e_h|_X^2 + H\|e_h^H\|_b^2 \leqslant c\|e_h\|_V\|\eta_h\|_L + c'(\|\eta_h\|_L\|\eta_h\|_V + \epsilon|\eta_h|_X^2). \tag{3.8}$$

The control on $\|e_H\|_V$ is provided by the discrete inf-sup condition (2.7),

$$\begin{aligned}
c_a\|e_H\|_V &\leqslant \sup_{\phi_h\in X_h}\frac{a(e_H,\phi_h)}{\|\phi_h\|_L}, \\
&\leqslant \sup_{\phi_h\in X_h}\frac{-a(\eta_h,\phi_h)-a(e_h^H,\phi_h)-\epsilon d(e_h,v_h)-\epsilon d(\eta_h,v_h)-b_h(e_h^H,\phi_h^H)}{\|\phi_h\|_L}, \\
&\leqslant \|a\|(\|\eta_h\|_V+\|e_h^H\|_V) + c_b\|e_h^H\|_b + \epsilon c_d(|e_h|_X+|\eta_h|_X)\sup_{\phi_h\in X_h}\frac{|\phi_h|_X}{\|\phi_h\|_L}, \\
&\leqslant c(\|\eta_h\|_V + \epsilon\mu(H)^{-1}|\eta_h|_X + \|e_h^H\|_b + \epsilon\mu(H)^{-1}|e_h|_X).
\end{aligned}$$

The triangle inequality together with (2.11) yields $\|e_h\|_V \leqslant c(\|e_H\|_V + \|e_h^H\|_b)$. By substituting this bound into the inequality (3.8) we have

$$a_s(e_h, e_h) + \epsilon|e_h|_X^2 + H\|e_h^H\|_b^2 \leqslant c(\|\eta_h\|_L\|\eta_h\|_V + \epsilon|\eta_h|_X^2 + \epsilon\mu(H)^{-1}\|\eta_h\|_L|\eta_h|_X) + c'\|\eta_h\|_L(\|e_h^H\|_b + \mu(H)^{-1}\epsilon|e_h|_X)$$

which finally yields

$$a_s(e_h, e_h) + \epsilon\|e_h\|_X^2 + H\|e_h^H\|_b^2 \leqslant c[\|\eta_h\|_L^2(H^{-1} + \epsilon\mu(H)^{-2}) + H\|\eta_h\|_V^2 + \epsilon|\eta_h|_X^2].$$

The final result is a consequence of this inequality together with the definition $u - u_h = e_h + \eta_h$ and the *a priori* bound on $\|e_H\|_V$ provided by the discrete inf-sup condition. □

EXAMPLE 3.1 Let us assume that u, the solution of (3.2), is in W, and

$$\inf_{w_H \in X_H} \|u - u_H\|_L + H(\|u - u_H\|_V + \|u - u_H\|_X) \leqslant cH^{k+1}\|u\|_W.$$

Assume also that $\mu(H) \sim H$ and $\epsilon = \mathcal{O}(H)$ (which is the case of interest in practice). These hypotheses are satisfied if problem (3.2) corresponds to a second-order PDE and finite elements are used. Then

$$a_s(u - u_H, u - u_H) + \epsilon^{1/2}|u - u_H|_X + H^{1/2}\|u - u_H\|_V \leqslant c(u)H^{k+1/2}.$$

This bound is optimal in the norm of V.

The case $H = \mathcal{O}(\epsilon)$ can be treated without relying on the discrete inf-sup condition if we assume that X is continuously embedded in V.

THEOREM 3.2 Assume that X is continuously embedded in V, $|\cdot|_X \sim \|\cdot\|_X$ and $H \sim \mu(H)$. The solution of problem (3.4) satisfies

$$a_s(u - u_h, u - u_h)^{1/2} + \epsilon^{1/2}\|u - u_h\|_X \leqslant c \inf_{w_H \in X_H} [H^{-1/2}\|u - w_H\|_L + H^{1/2}\|u - w_H\|_V + \epsilon^{1/2}\|u - w_H\|_X]. \tag{3.9}$$

$$\|u - u_h\|_V \leqslant c \inf_{w_H \in X_H} [H^{-1}\|u - w_H\|_L + \|u - w_H\|_X] \tag{3.10}$$

Proof. Assume first that $\epsilon \leqslant H$. By proceeding exactly as in the proof of Theorem 3.1, we obtain

$$a_s(e_h, e_h) + \epsilon\|e_h\|_X^2 + H\|e_h^H\|_b^2 \leqslant c[\|\eta_h\|_L^2(H^{-1} + \epsilon H^{-2}) + H\|\eta_h\|_V^2 + \epsilon|\eta_h|_X^2] \leqslant c[H^{-1}\|\eta_h\|_L^2 + H\|\eta_h\|_V^2 + \epsilon|\eta_h|_X^2].$$

Now assume that $H \leqslant \epsilon$. By proceeding as in the proof of Theorem 3.1 we obtain

$$\begin{aligned}
a_s(e_h, e_h) + \epsilon\|e_h\|_X^2 &\leqslant a(e_h, \eta_h) - 2a_s(e_h, \eta_h) - \epsilon d(\eta_h, e_h) \\
&\leqslant \|a\|\,\|e_h\|_V\|\eta_h\|_L + \gamma a_s(e_h, e_h) + c_\gamma a_s(\eta_h, \eta_h) + \epsilon c_d\|\eta_h\|_X\|e_h\|_X \\
&\leqslant \gamma a_s(e_h, e_h) + c[\|e_h\|_X(\|\eta_h\|_L + \epsilon\|\eta_h\|_X) + c_\gamma\|\eta_h\|_V\|\eta_h\|_L] \\
&\leqslant \gamma a_s(e_h, e_h) + \gamma\epsilon\|e_h\|_X^2 + c_\gamma[\epsilon^{-1}\|\eta_h\|_L^2 + \epsilon\|\eta_h\|_X^2] \\
&\leqslant \gamma a_s(e_h, e_h) + \gamma\epsilon\|e_h\|_X^2 + c_\gamma[H^{-1}\|\eta_h\|_L^2 + \epsilon\|\eta_h\|_X^2].
\end{aligned}$$

By choosing $\gamma = 1/2$, this bounds yields

$$a_s(e_h, e_h)^{1/2} + \epsilon^{1/2}\|e_h\|_X \leqslant c[H^{-1/2}\|\eta_h\|_L + \epsilon^{1/2}\|\eta_h\|_X].$$

The estimate (3.9) is an easy consequence of this bound and the previous one.

Now, let us prove the estimate (3.10). Assume first that $\epsilon \leqslant H$. The discrete inf-sup condition (2.7) yields

$$\|e_h\|_V \leqslant c(\|\eta_h\|_V + \epsilon H^{-1}\|\eta_h\|_X + \|e_h^H\|_b + \epsilon H^{-1}\|e_h\|_X).$$

As a result,

$$\begin{aligned}
\max(\epsilon, H)\|e_h\|_V^2 &\leqslant H\|e_h\|_V^2, \\
&\leqslant c[H\|\eta_h\|_V^2 + \epsilon^2 H^{-1}\|\eta_h\|_X^2 + H\|e_h^H\|_b^2 + \epsilon^2 H^{-1}\|e_h\|_X^2] \\
&\leqslant c[H\|\eta_h\|_V^2 + \epsilon\|\eta_h\|_X^2 + H\|e_h^H\|_b^2 + \epsilon\|e_h\|_X^2] \\
&\leqslant c[H^{-1}\|\eta_h\|_L^2 + H\|\eta_h\|_V^2 + \epsilon\|\eta_h\|_X^2] \\
&\leqslant c\max(\epsilon, H)[H^{-2}\|\eta_h\|_L^2 + \|\eta_h\|_X^2].
\end{aligned}$$

This bound yields

$$\|e_h\|_V \leqslant c[H^{-1}\|\eta_h\|_L + \|\eta_h\|_X].$$

Now, let us assume $H \leqslant \epsilon$,

$$\begin{aligned}
\max(\epsilon, H)\|e_h\|_V^2 &\leqslant c\epsilon\|e_h\|_X^2 \\
&\leqslant c[H^{-1}\|\eta_h\|_L^2 + \epsilon\|\eta_h\|_X^2] \\
&\leqslant c\max(\epsilon, H)[H^{-2}\|\eta_h\|_L^2 + \|\eta_h\|_X^2].
\end{aligned}$$

From this bound we obtain again

$$\|e_h\|_V \leqslant c[H^{-1}\|\eta_h\|_L + \|\eta_h\|_X].$$

The proof is complete. □

REMARK 3.2 Note that the bound (3.10) is uniform with respect to ϵ. As claimed in the introduction, this result is obtained without having to tune a stabilizing coefficient with respect to ϵ as in the Galerkin least square method.

4. Approximation on nonuniform meshes

The main drawback of the theory presented in §2 and §3 is that it relies on the inverse inequality (2.9). In finite element frameworks, this property is true provided the mesh underlying the approximation is quasi-uniform (see e.g. Girault & Raviart, 1986, p 103). This constraint may be too strong for practical purposes since it *a priori* excludes local refinement and mesh adaptation. The goal of this section is to generalize §2 and §3 to the case of nonuniform meshes by using a local version of (2.9).

4.1 *The model problem*

Our model problem is of the same type as (3.2); however, we sharpen slightly the hypotheses as follows.

We assume that $a \in \mathcal{L}(V \times L; \mathbb{R})$ and a is monotone. Furthermore, we assume the following decomposition $a = a_0 + a_1$ where the bilinear forms a_0 and a_1 have the following continuity properties:

$$a_0(v, w) \leqslant c_0 a_s(v, v)^{1/2} \|w\|_L, \tag{4.1}$$

$$a_1(v, w) \leqslant c_1 |v|_V \|w\|_L, \tag{4.2}$$

where $|\cdot|_V$ is a semi-norm in V such that

$$\exists c > 0,\ \forall v \in V \qquad \|v\|_V \leqslant c(a_s(v, v)^{1/2} + |v|_V). \tag{4.3}$$

The bilinear form d is in $\mathcal{L}(X \times X; \mathbb{R})$ and satisfies the following properties: There are $c \geqslant 0$, and a semi-norm $|\cdot|_X$ such that

$$\forall v \in X \qquad |v|_X^2 \leqslant a(v, v) + d(v, v), \tag{4.4}$$

$$\forall (v, w) \in X^2 \qquad d(v, w) \leqslant c|v|_X |w|_X. \tag{4.5}$$

For the sake of simplicity we shall assume hereafter that X is continuously embedded in V. The problem for which we want to build an approximate solution is: For f in L,

$$\begin{cases} \text{find } u \in X \text{ such that,} \\ a(u, v) + \epsilon d(u, v) = (f, v) \qquad \forall v \in X. \end{cases} \tag{4.6}$$

We shall assume hereafter that this problem has a unique solution.

4.2 *The discrete setting*

To approximate problem (4.6), we shall use $X_H \subset X_h \subset X$ two finite-dimensional subspaces of X. We assume also that there are two sequences $(T_i)_{i=1,\ldots,I(H)}$ and $(H_i)_{i=1,\ldots,I(H)}$ that satisfy the following properties:

For every subspace of L involved in the theory developed hereafter, say Y, and for every semi-norm in Y that we shall use, say $|\cdot|_Y$, we have

$$\forall v \in Y \qquad |v|_Y = \left(\sum_i |v|_{Y,T_i}^2 \right)^{1/2}. \tag{4.7}$$

We assume also that all the bilinear forms involved hereafter, say $x \in \mathcal{L}(Y \times Z)$, satisfy the following property

$$\forall (y, z) \in Y \times Z \qquad x(y, z) = \sum_i x_i(y, z). \tag{4.8}$$

EXAMPLE 4.1 If $a(v, w) = \int_\Omega vw + (\beta \cdot \nabla v)w$, with suitable assumptions on the vector field β, we have $L = \mathrm{L}^2(\Omega)$ and $V = \{v \in \mathrm{L}^2(\Omega) \mid \beta \cdot \nabla v \in \mathrm{L}^2(\Omega),\ v_{|\Gamma^-} = 0\}$. Moreover, by assuming that X_H is a finite element space based on a triangulation $(T_i)_{i=1,\dots,I(H)}$ (T_i being tetrahedrons, hexahedrons, etc), we have $|v|_{L,T_i} = \|v\|_{0,T_i}$. Moreover,

$$a_{0,i}(v, w) = \int_{T_i} vw, \qquad a_{1,i}(v, w) = \int_{T_i} (\beta \cdot \nabla v)w,$$

$$a_{s,i}(v, w) = \int_{T_i} (1 - \operatorname{div} \beta/2)vw + \int_{\Gamma^+ \cap T_i} vw(\beta \cdot n)/2,$$

and

$$|v|_{V,T_i} = \left(\int_{T_i} (\beta \cdot \nabla v)^2 \right)^{1/2}.$$

The approximation space X_H satisfies the following local interpolation property: There is W, a dense subspace of V, and there are $k > 0$ and $c > 0$ such that, for all $v \in W$

$$\forall T_i \qquad \inf_{w_H \in X_H} \|v - w_H\|_{L,T_i} + H_i \|v - w_H\|_{X,T_i} \leqslant c H_i^{k+1} \|v\|_{W,T_i}. \tag{4.9}$$

Furthermore, we assume that there is a linear projection operator $P_H : X_h \longrightarrow X_H$ that satisfies the following local L-stability:

$$\exists c > 0,\ \forall (H, h),\ \forall v_h \in X_h,\ \forall T_i \qquad \|P_H v_h\|_{L,T_i} \leqslant c \|v_h\|_{L,T_i}. \tag{4.10}$$

We shall hereafter denote $X_h^H = (1 - P_H)X_h$, and for all v_h in X_h we set $v_H = P_H v_h$ and $v_h^H = v_h - v_H$.

Concerning the bilinear forms a, we assume the following local continuity properties:

$$\forall (v, w) \in V^2,\ \forall T_i \qquad a_{0,i}(v, w) \leqslant c_0 a_{s,i}(v, v)^{1/2} \|w\|_{L,T_i}, \tag{4.11}$$

$$\forall (v, w) \in V^2,\ \forall T_i \qquad a_{1,i}(v, w) \leqslant c_1 |v|_{V,T_i} \|w\|_{L,T_i}, \tag{4.12}$$

and

$$\forall v \in V,\ \forall T_i \qquad \|v\|_{V,T_i} \leqslant c(a_{s,i}(v, v)^{1/2} + |v|_{V,T_i}). \tag{4.13}$$

We assume now that there is a subspace $\oplus_{i=1}^{I(H)} X_h(T_i)$ of X_h such that the bilinear form a_1 and the couple (X_H, X_h) satisfy the following local, discrete, inf-sup condition: there are $c_a > 0$, $c_\delta \geqslant 0$, independent of (H, h), such that

$$\forall v_h \in X_h,\ \forall T_i \qquad \sup_{\phi_h \in X_h(T_i)} \frac{a_{1,i}(v_H, \phi_h)}{\|\phi_h\|_{L,T_i}} \geqslant c_a |v_H|_{V,T_i} - c_\delta a_{s,i}(v_h, v_h)^{1/2}. \tag{4.14}$$

Since X is continuously embedded in V, we assume

$$\exists c > 0, \forall v \in X, \ \forall T_i \qquad |v|_{V,T_i} \leqslant c|v|_{X,T_i}. \tag{4.15}$$

Moreover, the finite-dimensional space X_h is assumed to satisfy the following inverse stability property:

$$\forall v_h \in X_h, \ \forall T_i \qquad |v_h|_{V,T_i} + |v_h|_{X,T_i} \leqslant cH_i^{-1}\|v_h\|_{L,T_i}. \tag{4.16}$$

REMARK 4.1 In practice, this hypothesis means that, in terms of PDEs, X and V are domains of first-order differential operators: think of $X = H_0^1(\Omega)$ and $V = \{v \in \mathrm{L}^2(\Omega) \mid \beta\cdot\nabla v \in \mathrm{L}^2(\Omega), \ v_{|\Gamma^-} = 0\}$.

Now we introduce the stabilizing bilinear form b_h, and we assume that it satisfies the following properties:

$$\forall (v_h^H, w_h^H) \in X_h^H, \ \forall T_i \qquad \begin{cases} b_{h,i}(v_h^H, v_h^H) \geqslant H_i|v_h^H|_{b,T_i}^2, \\ b_{h,i}(v_h^H, w_h^H) \leqslant c_B H_i |v_h^H|_{b,T_i}|w_h^H|_{b,T_i}, \end{cases} \tag{4.17}$$

and the semi-norm $|\cdot|_b$ is such that there are two constants $c_{e1} > 0$ and $c_{e2} > 0$ such that

$$\forall v_h^H \in X_h^H, \ \forall T_i \qquad c_{e1}|v_h^H|_{V,T_i} \leqslant |v_h^H|_{b,T_i} \leqslant c_{e2}H_i^{-1}\|v_h^H\|_{L,T_i}, \tag{4.18}$$

Owing to (4.10), (4.17), and (4.18) we infer

LEMMA 4.1 There is $c_b > 0$ such that

$$\forall (H,h), \ \forall v_h^H \in X_h^H, \ \forall T_i \qquad \sup_{w_h \in X_h(T_i)} \frac{b_{h,i}(v_h^H, w_h^H)}{\|w_h\|_{L,T_i}} \leqslant c_b \|v_h^H\|_{b,T_i}. \tag{4.19}$$

The discrete problem consists in the following:

$$\begin{cases} \text{Find } u_h \text{ in } X_h \text{ such that} \\ a(u_h, v_h) + \epsilon d(u_h, v_h) + b_h(u_h^H, v_h^H) = (f, v_h) \qquad \forall v_h \in X_h. \end{cases} \tag{4.20}$$

4.3 *The error analysis*

The main convergence result of this section is summarized as follows:

THEOREM 4.1 The discrete solution to (4.20) satisfies the bounds:

$$a_s(u-u_h, u-u_h)^{1/2} + \epsilon^{1/2}|u-u_h|_X \leqslant c\left[\inf_{w_H \in X_H} \sum_i [H_i^{-1}\|u-w_H\|_{L,T_i}^2 + H_i\|u-w_H\|_{V,T_i}^2 + \epsilon\|u-w_H\|_{X,T_i}^2]\right]^{1/2}. \tag{4.21}$$

$$\left[\sum_i \max(H_i,\epsilon)|u-u_h|_{V,T_i}^2\right]^{1/2} \leqslant c\left[\inf_{w_H \in X_H} \sum_i \max(H_i,\epsilon)[H_i^{-2}\|u-w_H\|_{L,T_i}^2 + \|u-w_H\|_{X,T_i}^2]\right]^{1/2}. \tag{4.22}$$

Proof. By using the same notation as in Theorem 2.2, the equation that controls e_h is obtained by subtracting (4.20) from (4.6) with the test functions spanning X_h:

$$\forall v_h \in X_h \qquad a(e_h, v_h) + \epsilon d(e_h, v_h) + b_h(e_h^H, v_h^H) = -a(\eta_h, v_h) - \epsilon d(\eta_h, v_h),$$

where we have used $u_h^H = -e_h^H$.

By taking e_h as a test function and by using the coercivity properties (4.17) and (4.4) we obtain

$$(1-\epsilon)a_s(e_h, e_h) + \epsilon \sum_i |e_h|^2_{X,T_i} + \sum_i H_i |e_h^H|^2_{b,T_i} \leqslant a_1(e_h, \eta_h) + a_0(e_h, \eta_h) \\ -2a_s(e_h, \eta_h) - \epsilon d(\eta_h, e_h).$$

We derive bounds from above for the last three terms of the right-hand side as follows:

$$\begin{aligned} a_0(e_h, \eta_h) &\leqslant c a_s(e_h, e_h)^{1/2} \|\eta_h\|_L \\ &\leqslant \gamma a_s(e_h, e_h) + c_\gamma \sum_i \|\eta_h\|^2_{L,T_i}, \end{aligned}$$

$$\begin{aligned} -2a_s(e_h, \eta_h) &\leqslant \gamma a_s(e_h, e_h) + c_\gamma a_s(\eta_h, \eta_h) \\ &\leqslant \gamma a_s(e_h, e_h) + c_\gamma \sum_i \|\eta_h\|_{V,T_i} \|\eta_h\|_{L,T_i}, \end{aligned}$$

$$\begin{aligned} -\epsilon d(\eta_h, e_h) &\leqslant c \sum_i \epsilon |\eta_h|_{X,T_i} |e_h|_{X,T_i} \\ &\leqslant \gamma \sum_i \epsilon |e_h|^2_{X,T_i} + c_\gamma \sum_i \epsilon \|\eta_h\|^2_{X,T_i}. \end{aligned}$$

By inserting these bounds into the inequality above, we obtain

$$a_s(e_h, e_h) + \epsilon \sum_i |e_h|^2_{X,T_i} + \sum_i H_i |e_h^H|^2_{b,T_i} \leqslant c a_1(e_h, \eta_h) \\ + c' \sum_i \Big[\|\eta_h\|_{V,T_i} \|\eta_h\|_{L,T_i} + \epsilon \|\eta_h\|^2_{X,T_i} \Big]. \tag{4.23}$$

To control the remaining term, $a_1(e_h, \eta_h)$, we proceed as follows:

$$\begin{aligned} a_1(e_h, \eta_h) &\leqslant c \sum_i |e_h|_{V,T_i} \|\eta_h\|_{L,T_i} \\ &\leqslant c \sum_{\{i \mid \epsilon \geqslant H_i\}} |e_h|_{X,T_i} \|\eta_h\|_{L,T_i} + c' \sum_{\{i \mid \epsilon < H_i\}} |e_h|_{V,T_i} \|\eta_h\|_{L,T_i} \\ &\leqslant \gamma \sum_{\{i \mid \epsilon \geqslant H_i\}} \epsilon |e_h|^2_{X,T_i} + c_\gamma \sum_{\{i \mid \epsilon \geqslant H_i\}} H_i^{-1} \|\eta_h\|^2_{L,T_i} \\ &\quad + c' \sum_{\{i \mid \epsilon < H_i\}} |e_h|_{V,T_i} \|\eta_h\|_{L,T_i}. \end{aligned}$$

The most critical term is $|e_h|_{V,T_i}$ for T_i such that $\epsilon < H_i$. This term is controlled by means of the local discrete inf-sup inequality.

$$\begin{aligned}
c_a|e_H|_{V,T_i} &\leqslant \sup_{v_h\in X_h(T_i)} \frac{a_{1,i}(e_H, v_h)}{\|v_h\|_{L,T_i}} + c_\delta a_{s,i}(e_h, e_h)^{1/2} \\
&\leqslant \sup_{v_h\in X_h(T_i)} \frac{-a_{0,i}(e_h, v_h) - a_{1,i}(e_h^H, v_h) - \epsilon d_i(e_h, v_h) - b_{h,i}(e_h^H, v_h^H)}{\|v_h\|_{L,T_i}} \\
&\quad + \sup_{v_h\in X_h(T_i)} \frac{-a_i(\eta_h, v_h) - \epsilon d_i(\eta_h, v_h)}{\|v_h\|_{L,T_i}} + c_\delta a_{s,i}(e_h, e_h)^{1/2} \\
&\leqslant c[a_{s,i}(e_h, e_h)^{1/2} + H_i^{-1}\epsilon|e_h|_{X,T_i} + |e_h^H|_{b,T_i} \\
&\quad + \|\eta_h\|_{V,T_i} + H_i^{-1}\epsilon\|\eta_h\|_{X,T_i}].
\end{aligned}$$

Hence, we have

$$\begin{aligned}
\sum_{\{i|\epsilon<H_i\}} |e_h|_{V,T_i}\|\eta_h\|_{L,T_i} &\leqslant c \sum_{\{i|\epsilon<H_i\}} [a_{s,i}(e_h, e_h)^{1/2} + H_i^{-1}\epsilon|e_h|_{X,T_i} + |e_h^H|_{b,T_i} \\
&\quad + \|\eta_h\|_{V,T_i} + H_i^{-1}\epsilon\|\eta_h\|_{X,T_i}]\|\eta_h\|_{L,T_i} \\
&\leqslant \gamma \sum_i [a_{s,i}(e_h, e_h) + \epsilon|e_h|^2_{X,T_i} + |e_h^H|^2_{b,T_i}] \\
&\quad + c_\gamma \sum_{\{i|\epsilon<H_i\}} [H_i^{-2}\epsilon\|\eta_h\|^2_{L,T_i} + H_i^{-1}\|\eta_h\|^2_{L,T_i} \\
&\quad + \|\eta_h\|_{V,T_i}\|\eta_h\|_{L,T_i} + \epsilon\|\eta_h\|^2_{X,T_i}] \\
&\leqslant \gamma \sum_i [a_{s,i}(e_h, e_h) + \epsilon|e_h|^2_{X,T_i} + |e_h^H|^2_{b,T_i}] \\
&\quad + c_\gamma \sum_{\{i|\epsilon<H_i\}} [H_i^{-1}\|\eta_h\|^2_{L,T_i} + H_i\|\eta_h\|^2_{V,T_i} + \epsilon\|\eta_h\|^2_{X,T_i}].
\end{aligned}$$

Finally we obtain

$$\begin{aligned}
a_s(e_h, e_h) + \epsilon|e_h|^2_X + \sum_i H_i|e_h^H|^2_{b,T_i} &\leqslant c\sum_i [H_i^{-1}\|\eta_h\|^2_{L,T_i} + H_i\|\eta_h\|^2_{V,T_i} \\
&\quad + \epsilon\|\eta_h\|^2_{X,T_i}],
\end{aligned}$$

from which we infer a bound on $a(u-u_h, u-u_h)^{1/2} + \epsilon^{1/2}|u-u_h|_X$.

To obtain an error estimate in the semi-norm $|\cdot|_V$ let us recall that, if $\epsilon < H_i$, the discrete inf-sup condition provides us with the bound

$$\begin{aligned}
\max(H_i, \epsilon)|e_H|^2_{V,T_i} &\leqslant c[H_i a_{s,i}(e_h, e_h) + \epsilon|e_h|^2_{X,T_i} + H_i|e_h^H|^2_{b,T_i} \\
&\quad + H_i\|\eta_h\|^2_{V,T_i} + \epsilon\|\eta_h\|^2_{X,T_i}],
\end{aligned}$$

whereas if $\epsilon \geqslant H_i$ we can use

$$\max(H_i, \epsilon)|e_h|^2_{V,T_i} \leqslant c\epsilon|e_h|^2_{X,T_i}.$$

By combining these two bounds we obtain

$$\sum_i \max(H_i,\epsilon)|e_H|^2_{V,T_i} \leqslant c\sum_i [H_i^{-1}\|\eta_h\|^2_{L,T_i} + H_i\|\eta_h\|^2_{V,T_i} + \epsilon\|\eta_h\|^2_{X,T_i}]$$
$$\leqslant c\sum_i \max(H_i,\epsilon)[H_i^{-2}\|\eta_h\|^2_{L,T_i} + \|\eta_h\|^2_{X,T_i}].$$

The proof is complete. □

COROLLARY 4.1 If u is in W, the following error estimates hold:

$$a_s(u-u_h,u-u_h)^{1/2} + \epsilon^{1/2}|u-u_h|_X \leqslant c\left[\sum_i (H_i^{2k+1} + \epsilon H_i^{2k})\|u\|^2_{W,T_i}\right]^{1/2} \tag{4.24}$$

$$\left[\sum_i \max(H_i,\epsilon)|u-u_h|^2_{V,T_i}\right]^{1/2} \leqslant c\left[\sum_i \max(H_i,\epsilon)H_i^{2k}\|u\|^2_{W,T_i}\right]^{1/2}. \tag{4.25}$$

REMARK 4.2 The error estimate in the V norm is quasi-optimal.

5. Examples

5.1 *Preliminaries*

Let Ω be an open bounded connected subset of $\mathbb{R}^d$. Having in mind general second-order PDEs dominated by a linear first-order differential operator, we consider a sequence of d matrices $(A^k)_{k=1,d}$ such that $A^k : \Omega \longrightarrow \mathcal{M}_m(\mathbb{R})$, where m is a strictly positive integer. We set $\beta = (A^1,\ldots,A^d)$ and for a smooth function $u : \Omega \longrightarrow \mathbb{R}^m$, we conventionally denote by $\beta\cdot\nabla u$ the function $\beta\cdot\nabla u : \Omega \longrightarrow \mathbb{R}^m$ such that

$$1 \leqslant i \leqslant m \qquad (\beta\cdot\nabla u)_i = \sum_{k=1}^{d}\sum_{j=1}^{m} A^k_{ij}\frac{\partial u_j}{\partial x_k}.$$

For a smooth function $v : \Omega \longrightarrow \mathbb{R}^m$, we define $v\cdot(\beta\cdot\nabla u) = \sum_{i=1}^m v_i(\beta\cdot\nabla u)_i$, and we set the notation $|u|_{1,\beta} = \left[\int_\Omega (\beta\cdot\nabla u)\cdot(\beta\cdot\nabla u)\right]^{1/2}$. We are now concerned with bilinear forms involving terms of the following type $\int_\Omega v\cdot(\beta\cdot\nabla u)$.

EXAMPLE 5.1 Let us consider the scalar advection problem in $\Omega \subset \mathbb{R}^d$

$$\begin{cases} \mu u + \beta\cdot\nabla u = f \\ u_{|\Gamma^-} = 0, \end{cases}$$

where we assume $\mu - \frac{1}{2}\mathrm{div}\,\beta \geqslant \mu_0 > 0$. We set $m = 1$,

$$A^k_{11} = \beta_k \qquad \text{for } 1 \leqslant k \leqslant d,$$

$a_0(u, v) = (\mu u, v)_{0,\Omega}$, and $a_1(u, v) = (\beta \cdot \nabla u, v)_{0,\Omega}$. Furthermore, we define

$$L = \mathrm{L}^2(\Omega),$$
$$V = \{v \in \mathrm{L}^2(\Omega) \mid \beta \cdot \nabla v \in \mathrm{L}^2(\Omega),\ v_{|\Gamma^-} = 0\},$$

and $|u|_V = |u|_{1,\beta}$. It is clear that hypotheses (2.18), (2.19), (2.20), and (2.21) are satisfied.

EXAMPLE 5.2 Consider the following Darcy problem in $\Omega \subset \mathbb{R}^d$.

$$\begin{cases} u + \nabla p = f \\ \operatorname{div} u = g \\ p_{|\Gamma} = 0. \end{cases}$$

This problem can be put within the framework defined above by setting $m = d + 1$ and

$$\begin{aligned} A_{ij}^k &= 0, && \text{if } 1 \leqslant i \leqslant m-1, && 1 \leqslant j \leqslant m-1, \\ A_{ij}^k &= \delta_{i,k}, && \text{if } 1 \leqslant i \leqslant m-1, && j = m, \\ A_{ij}^k &= \delta_{j,k}, && \text{if } i = m, && 1 \leqslant j \leqslant m-1, \\ A_{ij}^k &= 0, && \text{if } i = m, && j = m, \end{aligned}$$

where $\delta_{i,k}$ is the Kronecker symbol. Let us also set $a_0((u, p), (v, q)) = (u, v)_{0,\Omega}$ and $a_1((u, p), (v, q)) = (q, \operatorname{div} u)_{0,\Omega} + (\nabla p, v)_{0,\Omega}$. It is clear that, owing to the definition of the generalized vector field β, we have $a_1((u, p), (v, q)) = (\beta \cdot \nabla(u, p), (v, q))_{0,\Omega}$. Furthermore, we define

$$L = \mathrm{L}^2(\Omega)^d \times \mathrm{L}^2(\Omega),$$
$$V = \{v \in \mathrm{L}^2(\Omega) \mid \operatorname{div} v \in \mathrm{L}^2(\Omega)\} \times \mathrm{H}_0^1(\Omega),$$

and $|(u, p)|_V = |(u, p)|_{1,\beta}$. A simple computation yields

$$|(u, p)|_V = (\|\operatorname{div} u\|_{0,\Omega}^2 + \|\nabla p\|_{0,\Omega}^2)^{1/2}.$$

It is clear that hypotheses (2.18), (2.19), (2.20), and (2.21) are satisfied. For instance, hypothesis (2.20) is a simple consequence of the relation $a_s((u, p), (u, p)) = a_0((u, p), (u, p)) = \|u\|_{0,\Omega}^2$, together with the definition of the semi-norm $|\cdot|_V$ and the Poincaré inequality for the pressure.

EXAMPLE 5.3 Let $\Omega \subset \mathbb{R}^3$ and consider the simplified Maxwell equations in Ω:

$$\begin{cases} E + \nabla \times B = f \\ B - \nabla \times E = g \\ E \times n_{|\Gamma} = 0. \end{cases}$$

To put this problem in our classification, we set

$$\begin{aligned} A_{ij}^k &= 0, && \text{if } 1 \leqslant i \leqslant 3,\ 1 \leqslant j \leqslant 3, \\ A_{ij}^k &= \epsilon_{i,(j-3),k}, && \text{if } 1 \leqslant i \leqslant 3,\ 4 \leqslant j \leqslant 6, \\ A_{ij}^k &= -\epsilon_{(i-3),j,k}, && \text{if } 4 \leqslant i \leqslant 6,\ 1 \leqslant j \leqslant 3, \\ A_{ij}^k &= 0, && \text{if } 4 \leqslant i \leqslant 6,\ 4 \leqslant j \leqslant 6, \end{aligned}$$

where $\epsilon_{i,j,k}$ is the Lévy–Chivita tensor. By denoting $u = (E, B)$, it is clear that $\beta\cdot\nabla u = (-\mathrm{rot}\, B, \mathrm{rot}\, E)$. We introduce the Hilbert spaces

$$L = \mathrm{L}^2(\Omega)^3 \times \mathrm{L}^2(\Omega)^3,$$
$$V = \{E \in \mathrm{L}^2(\Omega)^3 \mid \nabla\times E \in \mathrm{L}^2(\Omega)^3,\ E \times n_{|\Gamma} = 0\} \times \{B \in \mathrm{L}^2(\Omega)^3 \mid \nabla\times B \in \mathrm{L}^2(\Omega)^3\},$$

and we consider the following bilinear form $a : V \times L \longrightarrow \mathbb{R}$ such that

$$a(u, v) = \int_\Omega u\cdot v + v\cdot(\beta\cdot\nabla u). \tag{5.1}$$

We have the natural decomposition $a = a_0 + a_1$ with $a_0((E, B), (e, b)) = (E, e)_{0,\Omega} + (B, b)_{0,\Omega}$ and $a_1((E, B), (e, b)) = \int_\Omega (e, b)\cdot(\beta\cdot\nabla(E, B))$. A simple computation shows that $|(E, B)|_V = (\|\nabla\times E\|^2_{0,\Omega} + \|\nabla\times B\|^2_{0,\Omega})^{1/2}$. The hypotheses (2.18), (2.19), (2.20), and (2.21) are simple consequence of the relation

$$a_s((E, B), (E, B)) = a_0((E, B), (E, B)) = \|E\|^2_{0,\Omega} + \|B\|^2_{0,\Omega}$$

together with the definition of the semi-norm $|\cdot|_V$.

5.2 $\mathbb{P}_1$ *and* $\mathbb{P}_2$ *interpolations*

We describe in this section four admissible discrete settings. For the sake of simplicity, we assume hereafter that Ω is a $\mathbb{R}^d$-polyhedron and $\mathcal{T}_H$ is a regular triangulation of Ω composed of affine simplexes, (T_H). The reference simplex is denoted by $\hat{T}$ and $F_H : T_H \longrightarrow \hat{T}$ is the one-to-one affine mapping that maps T_H onto $\hat{T}$.

$\mathbb{P}_1$*/bubble interpolation*

To build a $\mathbb{P}_1$ interpolation space we define X_H as follows

$$X_H = \{v_H \in H^1(\Omega)^m \mid v_{H|T_H} \in \mathbb{P}_1(T_H)^m,\ \forall T_H \in \mathcal{T}_H\}. \tag{5.2}$$

To build a simple subgrid space X_h^H we proceed as follows. Let $\hat{\psi}$ be in $H_0^1(\hat{T})$ with $0 \leqslant \hat{\psi} \leqslant 1$; $\hat{\psi}$ is hereafter referred to as the bubble function (cf. e.g. Arnold *et al.*, 1984, or Crouzeix & Raviart, 1973). By denoting $\psi_h = \hat{\psi}(F_H)$, we define $X_h^T(T_H) = [\mathrm{span}(\psi_h)]^m$ for all T_H in $\mathcal{T}_H$, and we set

$$X_h^H = \oplus_{T_H} X_h^H(T_H). \tag{5.3}$$

The couple (X_H, X_h) is hereafter referred to as the $\mathbb{P}_1$/bubble approximation space.

$\mathbb{P}_2$*/bubble interpolation*

Another possibility that we shall also consider consists in defining X_H as being the $\mathbb{P}_2$ finite element space (conformal in $H^1(\Omega)^m$) associated with the triangulation $\mathcal{T}_H$:

$$X_H = \{v_H \in H^1(\Omega)^m \mid v_{H|T_H} \in \mathbb{P}_2(T_H)^m,\ \forall T_H \in \mathcal{T}_H\}. \tag{5.4}$$

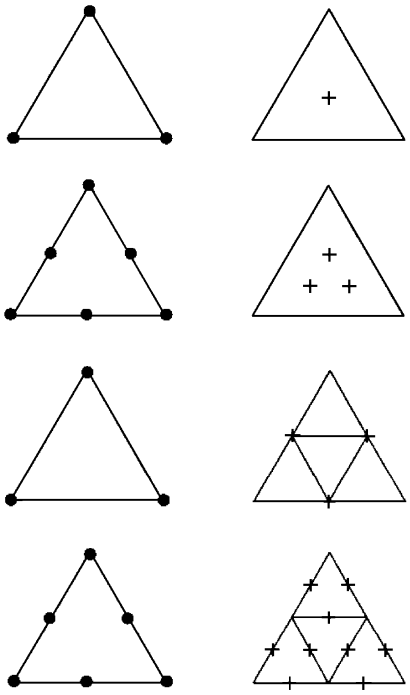

FIG. 1. Four examples of finite elements; (left) resolved scale finite element; (right) subgrid scale finite element. From top to bottom: $\mathbb{P}_1$/bubble finite element; $\mathbb{P}_2$/bubble finite element; two-level $\mathbb{P}_1$ finite element; two-level $\mathbb{P}_2$ finite element.

To build the subgrid scale space we introduce $\hat{\psi}_1, \ldots, \hat{\psi}_{d+1}$, a family of $d+1$ linearly independent, real-valued functions in $H_0^1(\hat{T})$. Let $\hat{a}_1, \ldots, \hat{a}_{d+1}$ be the nodes of the reference simplex $\hat{T}$. Let R_{ij} be the symmetry of $\hat{T}$ such that $R_{ij}(\hat{a}_i) = \hat{a}_j$ and $R_{ij}(\hat{a}_l) = \hat{a}_l$ if $l \notin \{i, j\}$. Now, we assume that the functions $(\hat{\psi}_i)_{i=1,\ldots,d+1}$ satisfy the following symmetry properties

$$\begin{cases} \hat{\psi}_i(R_{ij}) = \hat{\psi}_j, \\ \hat{\psi}_i(R_{jl}) = \hat{\psi}_i, \text{ if } i \notin \{j, l\}. \end{cases} \tag{5.5}$$

We denote $\psi_{i,h} = \hat{\psi}_i(F_H)$ for $1 \leqslant i \leqslant d+1$, we set $X_h^H(T_H) = [\text{span}(\psi_{1,h}, \ldots, \psi_{d+1,h})]^m$, and we finally define

$$X_h^H = \oplus_{T_H} X_h^H(T_H). \tag{5.6}$$

The couple (X_H, X_h) is referred to as the $\mathbb{P}_2$/bubble approximation space.

Two-level $\mathbb{P}_1$ interpolation

The two settings described above are not really two-level approximation spaces since X_H and X_h^H are defined on the same mesh; in some sense, for these two cases $h = H$. We

propose now an alternative approach that is valid in 2D (though it can can be extended to 3D). From each triangle $T_H \in \mathcal{T}_H$, we create 4 new triangles by connecting the middle of the 3 edges of T_H. Let us put $h = H/2$ and denote by $\mathcal{T}_h$ the resulting new triangulation. For each macro-triangle T_H we denote by $\mathbb{P}$ the set of continuous functions on T_H that are piecewise $\mathbb{P}_1$ on each sub triangle of T_H and vanish at the three vertices of T_H. Now we set

$$X_h^H = \{v_h^H \in H^1(\Omega)^m \mid v_{h|T_H}^H \in \mathbb{P}^m, \ \forall T_H \in \mathcal{T}_H\}. \tag{5.7}$$

It is clear that X_h has the following simple characterisation:

$$X_h = \{v_h \in H^1(\Omega)^m \mid v_{h|T_h} \in \mathbb{P}_1(T_h)^m, \ \forall T_h \in \mathcal{T}_h\}. \tag{5.8}$$

We shall call the couple (X_H, X_h) the two-level $\mathbb{P}_1$ approximation.

Two-level $\mathbb{P}_2$ interpolation

Now we build the $\mathbb{P}_2$ extension of the two-level $\mathbb{P}_1$ setting. We again set $h = H/2$, and we denote by $\mathcal{T}_h$ the triangulation that is obtained by dividing each triangle of $\mathcal{T}_H$ into four sub-triangles. For each triangle T_h we denote by ψ_1, ψ_2, ψ_3 the three $\mathbb{P}_2$ nodal functions associated with the middle of each edge of T_h. We define the subgrid scale space by

$$X_h^H = \{v_h^H \in H^1(\Omega)^m \mid v_{h|T_h}^H \in \operatorname{span}(\psi_1, \psi_2, \psi_3)^m, \ \forall T_h \in \mathcal{T}_h\}. \tag{5.9}$$

X_h has the following simple characterization:

$$X_h = \{v_h \in H^1(\Omega)^m \mid v_{h|T_h} \in \mathbb{P}_2(T_h)^m, \ \forall T_h \in \mathcal{T}_h\}. \tag{5.10}$$

The couple (X_H, X_h) is called the two-level $\mathbb{P}_2$ approximation. The finite elements associated with the four settings defined above are shown in Fig. 1.

The inf-sup condition

It is shown in Guermond (1999a) that for the four $\mathbb{P}_1$ and $\mathbb{P}_2$ interpolation spaces defined above, the decomposition $X_h = X_H \oplus X_h^H$ is L^2-stable. Furthermore, to localize the inf-sup condition (2.7), we introduce new definitions. For the $\mathbb{P}_1$/bubble and $\mathbb{P}_2$/bubble frameworks we set $W(T_H) = T_H$ and $Y_h(T_H) = X_h^H(T_H)$. For the other two-level settings, we set $W(T_H) = \{T'_H \in \mathcal{T}_H \mid T'_H \cap T_H \neq \emptyset\}$ and $Y_h(T_H) = \{v_h \in X_h \mid \operatorname{supp}(v_h) \subset W(T_H)\}$. By reasoning as in Guermond (1999a), the following results can be proved:

LEMMA 5.1 If β is piecewise constant on each simplex T_H of $\mathcal{T}_H$, there is $c_\beta > 0$ independent of (H, h), such that

$$\forall u_H \in X_H, \ \forall T_H \qquad \sup_{v_h \in Y_h(T_H)} \frac{\int_{W(T_H)} v_h \cdot (\beta \cdot \nabla u_H)}{\|v_h\|_{0,W(T_H)}} \geqslant c_\beta |u_H|_{1,\beta,W(T_H)}. \tag{5.11}$$

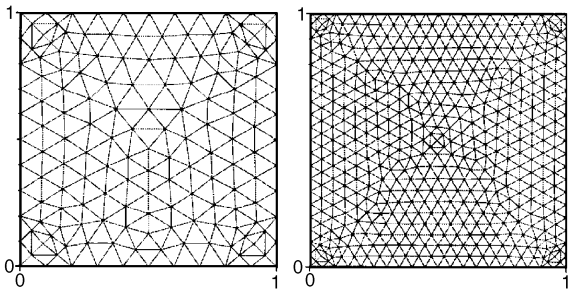

FIG. 2. Two two-level meshes used for tests: left, $h \approx 1/10$, and right, $h \approx 1/20$.

COROLLARY 5.1 If β is in $C^1(\overline{\Omega}; \mathcal{M}_m(\mathbb{R})^d)$, there are $c_\beta > 0$ and $c_\delta \geqslant 0$, both independent of (H, h), such that

$$\forall u_H \in X_H,\ \forall T_H \qquad \sup_{v_h \in Y_h(T_H)} \frac{\int_{W(T_H)} (\beta \cdot \nabla u_H) v_h}{\|v_h\|_{0,W(T_H)}} \geqslant c_\beta |u_H|_{1,\beta,W(T_H)} - c_\delta \|u_H\|_{0,W(T_H)}. \tag{5.12}$$

REMARK 5.1 Note that for the three model problems considered, the four finite element frameworks presented above satisfy all the hypotheses of §4. Hence, the present formulation allows for solving the Maxwell-like problem $a(u, v) = (f, v)$ with $\mathbb{P}_1$ or $\mathbb{P}_2$ finite elements in a quasi-optimal way.

5.3 *Example 1: an advection equation*

To illustrate the method proposed in this paper we apply it to the following 2D problem

$$\begin{cases} \partial_y u = -8\pi \sin(8\pi y) & \text{in } \Omega =]0, 1[^2 \\ u_{|y=0} = 1, \end{cases}$$

where $u = \cos(8\pi y)$ is the exact solution. We tested the two-level $\mathbb{P}_1$ and two-level $\mathbb{P}_2$ frameworks described above. Owing to Lemma 5.1, it is clear that the theory developed in this paper applies. The artificial viscosity is introduced by means of the bilinear form

$$b_h(v_h^H, w_h^H) = c_b \sum_{T_h \in \mathcal{T}_h} \text{meas}(T_h)^{1/2} \int_{T_h} \nabla v_h^H \cdot \nabla w_h^H.$$

To give an idea of the coarseness of the meshes that we use, we have plotted them in Fig. 2. The $\mathbb{P}_2$ calculations are performed on the coarse mesh on the left ($H \approx 1/5$, $h \approx 1/10$), whereas the $\mathbb{P}_1$ calculations are performed on the mesh on the right ($H \approx 1/10$, $h \approx 1/20$).

The results of the $\mathbb{P}_1$ approximation are plotted in Fig. 3 and those of the $\mathbb{P}_2$ approximation are plotted in Fig. 4. In both cases, isovalue contours are shown at the top of the figure and the projection of the solution in the plane $x = 0$ is shown at the bottom. It

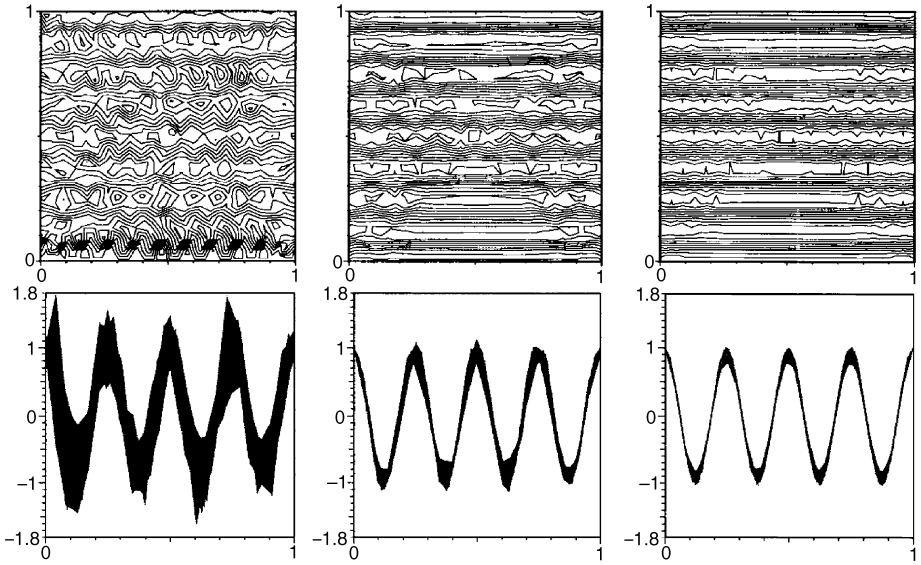

FIG. 3. 2D advection equation: $\partial_y u(x, y) = -8\pi \sin(8\pi y)$; $\mathbb{P}_1$ approximation with $h \approx 1/20$. Isovalue of solution (top) and projection of solution on plane $x = 0$ (bottom). Left, Galerkin solution; centre, two-level stabilized $\mathbb{P}_1$ solution; right, $\mathbb{P}_1$ interpolate of exact solution.

is clear that the Galerkin solution is plagued by spurious oscillations in both cases whereas the stabilized solution behaves correctly. Note that these tests are quite demanding since for the $\mathbb{P}_1$ approximation $\Lambda/H \approx 5$ and for the $\mathbb{P}_2$ approximation $\Lambda/H \approx 2{\cdot}5$, where $\Lambda = 0{\cdot}25$ is the wavelength of the solution. In both cases the stabilizing parameter c_b of the subgrid viscosity is set to $0{\cdot}1$.

5.4 *Example 2: an advection–diffusion equation*

To further illustrate the method, we apply it to the following 2D advection–diffusion problem:

$$\begin{cases} \partial_y u - \nu \nabla^2 u = 0 & \text{in } \Omega =]0, 1[^2 \\ u_{|y=0} = 0, \quad u_{|y=1} = 0; \end{cases}$$

where $u = (\exp(y/\nu) - 1)/(\exp(1/\nu) - 1)$ is the exact solution with $\nu = 0{\cdot}002$. The two-level mesh that we use is composed of 952 elements and 517 nodes and the mesh size h is of order $1/20$. This mesh is depicted in Fig. 5(top left). A 3D rendering of the $\mathbb{P}_1$ Galerkin solution is plotted in Fig. 5(top centre). The projection of this solution in the plane $x = 0$ is shown in Fig. 5(top right). Spurious numerical wiggles are clearly apparent throughout the domain. The projection in the plane $x = 0$ of the $\mathbb{P}_1$ interpolate of the exact solution is plotted in Fig. 5(bottom right).

The subgrid stabilized solution is shown in Fig. 5(bottom left). As expected, all the spurious wiggles have been smoothed out except in the region of the boundary layer where

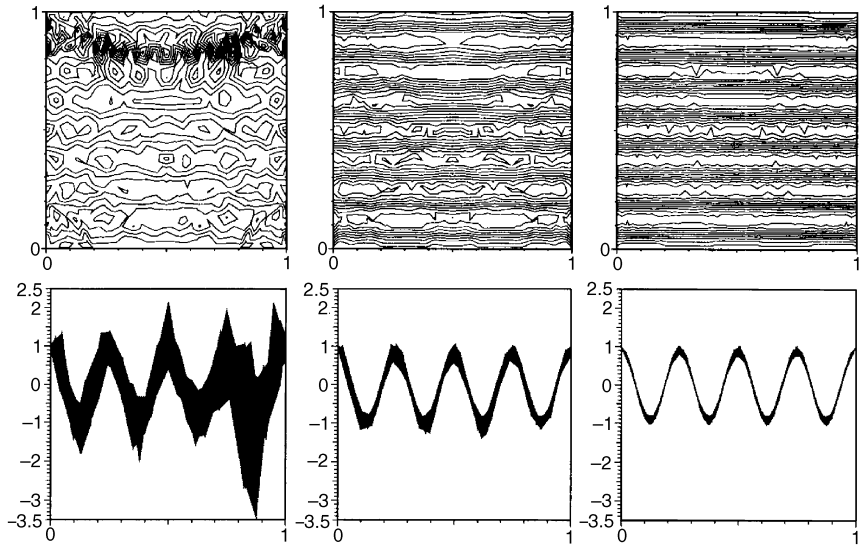

FIG. 4. 2D advection equation: $\partial_y u(x, y) = 8\pi \sin(8\pi y)$; $\mathbb{P}_2$ approximation with $h \approx 1/10$. Isovalue of solution (top) and projection of solution on plane $x = 0$ (bottom). Left, Galerkin solution; centre, $\mathbb{P}_1$ interpolate of two-level stabilized $\mathbb{P}_2$ solution; right, $\mathbb{P}_1$ interpolate of exact solution.

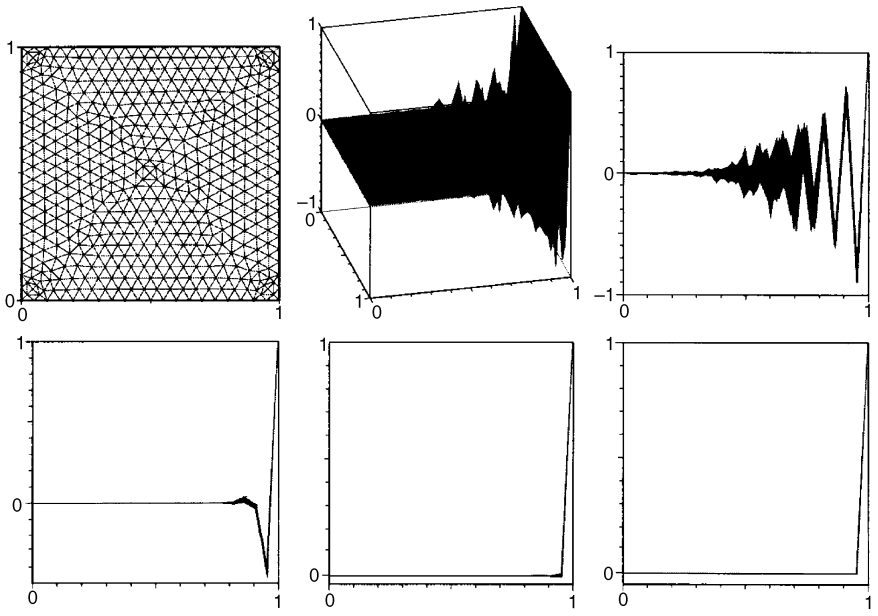

FIG. 5. Boundary layer problem: $\partial_y u - 0{\cdot}002\nabla^2 u = 0$. Top left, finite element mesh; top centre, 3D rendering of Galerkin solution; top right, projection on plane $x = 0$ of Galerkin solution; bottom left, projection on plane $x = 0$ of subgrid viscosity solution; bottom centre, projection on plane $x = 0$ of subgrid viscosity solution + shock capturing; bottom right, projection on plane $x = 0$ of $\mathbb{P}_1$ interpolate of exact solution.

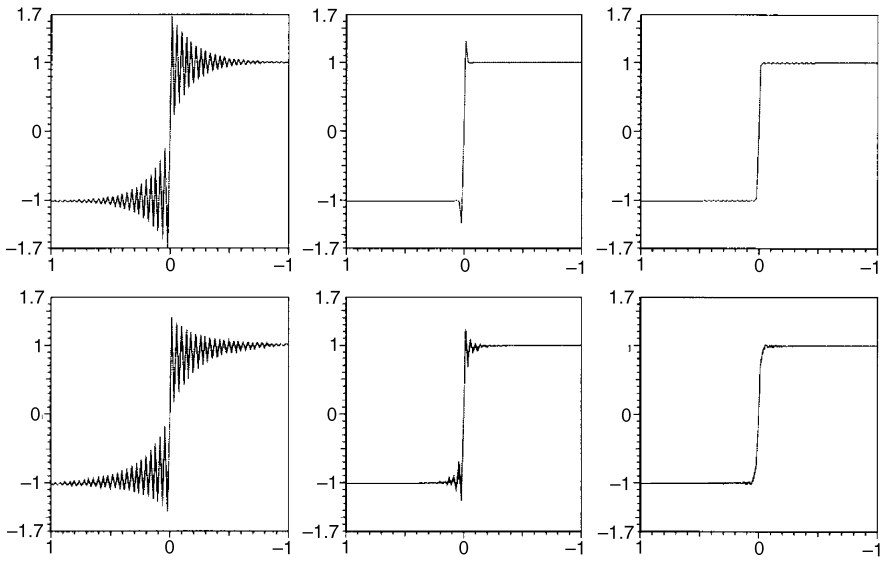

FIG. 6. The Bürgers problem: $u\partial_y(u) - \nu\nabla^2 u = 0$. Top, projection in plane $x = 0$ of $\mathbb{P}_1$ approximation with $h \approx 1/50$ and $H = 1/25$; bottom, $\mathbb{P}_2$ approximation with $h \approx 1/25$ and $H = 2/25$. Left, Galerkin solution; centre, two-level stabilized solution; right, two-level stabilized solution + two-level shock capturing.

the solution is rough. The remaining localized oscillations are linked to Gibbs' phenomena as explained in Maday *et al.* (1993). To eliminate these unwelcome modes, we introduce a subgrid shock capturing form as follows:

$$c_h(u_h^H, v_h, w_h) = c_{sc} \sum_{T_h \in \mathcal{T}_h} \text{meas}(T_h)^{1/2} \int_{T_h} |u_h^H| (\nabla v_h \cdot \nabla w_h).$$

Recall that $u_h^H = (1 - P_H)$ is the subgrid scale (i.e. the fluctuating part) of u_h. In practice we solve the following non linear problem:

$$\begin{cases} \text{Find } u_h \text{ in } X_h \text{ such that} \\ a(u_h, v_h) + b_h(u_h^H, v_h^H) + c_h(u_h^H, u_h, v_h) = (f, v_h) \qquad \forall v_h \in X_h. \end{cases}$$

Given that the non-linearity is very mild, this problem is easily solved by means of a very crude fixed point algorithm. The projection in the plane $x = 0$ of the solution to this problem is shown in Fig. 5(bottom centre). The effectiveness of the proposed shock capturing technique is clear. The boundary layer is captured in one element by using $c_{sc} = 1$.

5.5 *Example 3: the Bürgers equation*

To further compare the effects of the proposed subgrid stabilization and those of the subgrid shock capturing technique we propose solving the Bürgers problem:

$$\begin{cases} u\partial_y u - \nu\nabla^2 u = 0 \qquad \text{in } \Omega =]0, 1[\times[-1, 1[\\ u_{|x=-1} = -1, \quad u_{|x=1} = 1 \quad u_{|y=0} = u_{|y=1}. \end{cases}$$

For the viscosity we set $\nu = 10^{-4}$. We test the two-level $\mathbb{P}_1$ and $\mathbb{P}_2$ approximation techniques. For the $\mathbb{P}_1$ solution we use a mesh with $h = 1/50$, $H = 1/25$ and for the $\mathbb{P}_2$ solution we use a mesh with $h = 1/25$, $H = 2/25$, so that in the x-direction we have 101 nodes on both fine meshes.

We have plotted the projection in the plane $x = 0$ of the graph of the solution in Fig. 6. The three figures at the top are for the $\mathbb{P}_1$ approximation and those at the bottom are for the $\mathbb{P}_2$ approximation. The Galerkin solution is on the left. For both finite elements, the solution oscillates widely throughout the domain. The stabilized solution is shown in the centre. Some overshoots and undershoots are still present in the vicinity of the shock. These remaining oscillations are symptoms of Gibbs' phenomena. Note that except near the shock, all the spurious oscillations have disappeared. The results of the combination of the subgrid stabilization and the shock capturing techniques are shown on the top right and bottom ($c_b = 0{\cdot}1$ and $c_{sc} = 2$). The solution is very satisfactory considering the quite coarse mesh that is used.

6. Concluding remarks

A subgrid stabilization technique has been analysed in a quite general framework. It has been proved to yield quasi-optimal error estimates for a problem without coercivity. The effectiveness of the method has been illustrated by means of numerical examples. A shock capturing technique based on the subgrid scales of the solution has been proposed. It has been shown to be numerically efficient, though its mathematical analysis remains to be done. Hopefully, the combination of the two techniques proposed herein may contribute to the justification of Large eddy simulation models that are popular in CFD.

Although some of the ideas on which the subgrid stabilization is based stem from the framework of residual free bubbles, the connection between the present theory and the RFB theory is not clear to the author (see Baiocchi *et al.*, 1993; Brezzi *et al.*, 1992, 1997, for details on RFB). It seems, however, that there are major differences between the two approaches:

(i) The subgrid scale space X_h^H is composed of problem-independent shape functions, whereas in the RFB theory these functions are problem dependent, and 'the computation of [these functions] could be as difficult as the original problem' (Franca & Russo, 1996).
(ii) The RFB theory relies heavily on static condensation. Although for the $\mathbb{P}_1$/bubble and $\mathbb{P}_2$/bubble frameworks the subgrid scales can be eliminated by static condensation, this procedure is not feasible for the two-level $\mathbb{P}_1$ and $\mathbb{P}_2$ finite elements.
(iii) To the author's knowledge, the RFB analysis never refers to the inf-sup condition (2.7). This condition is the keystone of the present theory and seems to be new.
(iv) As the present theory only requires a to be continuous and monotone, it can be quite readily extended to approximating linear contraction semi-groups of class C^0 without relying on the discontinuous Galerkin technique as will be shown in a forthcoming paper.

Acknowledgement

The present work has been partly supported by ASCI (UPR-CNRS 9029), Orsay.

REFERENCES

ARNOLD, D. N., BREZZI, F., & FORTIN, M. 1984 A stable finite element for the Stokes equations. *Calcolo* **21**, 337–344.

AZERAD, P. & POUSIN, G. 1996 Inégalité de Poincaré courbe pour le traitement variationnel de l'équation de transport. *C. R. Acad. Sci. Paris,* Série I **322**, 721–727.

BAIOCCHI, C., BREZZI, F., & FRANCA, L. P. 1993 Virtual bubbles and Galerkin-least-square type methods (Ga.L.S). *Comput. Methods Appl. Mech. Eng.* **105**, 125–141.

BARDOS, C. 1970 Problèmes aux limites pour les équations aux dérivées partielles du premier ordre à coefficients réels; théorèmes d'approximation; application à l'équation de transport. *Ann. Scient. Éc. Norm. Sup.* 4e série **3**, 185–233.

BARLES, G. 1994 *Solutions de Viscosité des Équations de Hamilton–Jacobi. Mathématiques & Applications*. SMAI & Springer, **17**.

BREZIS, H. 1983 *Analyse Fonctionnelle, Théorie et Applications*. Paris: Masson.

BREZZI, F., BRISTEAU, M. O., FRANCA, L., MALLET, M., & ROGÉ, G. 1992 A relationship between stabilized finite element methods and the Galerkin method with bubble functions. *Comput. Methods Appl. Mech. Eng.* **96**, 117–129.

BREZZI, F., FRANCA, L., HUGHES, T. J. R., & RUSSO, A. 1997 $b = \int g$. *Comput. Methods Appl. Mech. Eng.* **145**, 329–339.

BROOKS, A. N. & HUGHES, T. J. R. 1982 Streamline upwind/Petrov–Galerkin formulations for convective dominated flows with particular emphasis on the incompressible Navier–Stokes equations. *Comput. Methods Appl. Mech. Eng.* **32**, 199–259.

CROUZEIX, M. & RAVIART, P.-A. 1973 Conforming and nonconforming finite element methods for solving the stationary Stokes equations. *R.A.I.R.O.* **3**, 33–75.

FOIAS, C., MANLEY, O. P., & TEMAM, R. 1988 Modelization of the interaction of small and large eddies in two dimensional turbulent flows. *Math. Modelling Numer. Anal.* **22**, 93–114.

FRANCA, L. P. & RUSSO, A. 1996 Deriving upwinding, mass lumping and selective reduced integration by residual-free bubbles. *Appl. Math. Lett.* **9**, 83–88.

GERMANO, M., PIOMELLI, U., MOIN, P., & CABOT, W. H. 1991 A dynamic subgrid-scale eddy viscosity model. *Phys. Fluids* A **3**, 1760–1765.

GUERMOND, J.-L. 1999a Stabilization of Galerkin approximations of transport equations by subgrid modelling. *Math. Modelling Numer. Anal.* **33**, 1293–1316.

GUERMOND, J.-L. 1999b Stabilisation par viscosité de sous-maille pour l'approximation de Galerkin des opérateurs monotones. *C. R. Acad. Sci. Paris,* Série I **328**, 617–622.

GUERMOND, J.-L. 1999c *Proc. GAMM 98 in ZAMM* **79**, 29–32.

GIRAULT, V. & RAVIART, P.-A. 1986 *Finite Element Methods for Navier–Stokes Equations. Springer Series in Computational Mathematics* **5**. Berlin: Springer.

HUGHES, T. J. R., FRANCA, L. P., & HULBERT, G. M. 1989 A new finite element formulation for computational fluid dynamics: VIII. The Galerkin/least-squares method for advection–diffusive equations. *Comput. Methods Appl. Mech. Eng.* **73**, 173–189.

JOHNSON, C., NÄVERT, U., & PITKÄRANTA, J. 1984 Finite element methods for linear hyperbolic equations. *Comput. Methods Appl. Mech. Eng.* **45**, 285–312.

MADAY, Y., OULD KABER, S. M., & TADMOR, E. 1993 Legendre pseudospectral viscosity method for nonlinear conservation laws. *SIAM J. Numer. Anal.* **30**, 321–342.

MARION, M. & TEMAM, R. 1990 Nonlinear Galerkin methods: the finite element case. *Numer. Math.* **57**, 1–22.

SMAGORINSKY, J. 1963 General circulation experiments with the primitive equations. I. The basic experiments. *J. Atmos. Sci.* **32**, 680–689.

ZHOU, G. 1997 How accurate is the streamline diffusion finite element method? *Math. Comput.* **66**, 31–44.