Subgrid Stabilization of Galerkin Approximations of Linear Contraction Semi-Groups of Class C^0 in Hilbert Spaces

Jean–Luc Guermond

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

Received 10 February 1999; accepted 20 April 2000

This article presents a stabilized Galerkin technique for approximating linear contraction semi-groups of class C^0 in a Hilbert space. The main result of this article is that this technique yields an optimal approximation estimate in the graph norm. The key idea is two-fold. First, it consists in introducing an approximation space that is broken up into resolved scales and subgrid scales, so that the bilinear form associated with the generator of the semi-group satisfies a uniform inf-sup condition with respect to this decomposition. Second, the Galerkin approximation is slightly modified by introducing an artificial diffusion on the subgrid scales. Numerical tests show that the method applies also to nonlinear semi-groups. Numer Methods Partial Differential Eq 17: 1–25, 2001

Keywords: *finite elements; Galerkin methods; stabilization; linear hyperbolic equations; semi-groups; multiscale methods; subgrid modeling; artificial viscosity*

I. INTRODUCTION

In this article, we are concerned with the following abstract linear problem: For $f \in C^1([0,+\infty[;L)$ and $u_0 \in D(A)$,

$$\begin{cases} \text{Find } u \in C^1([0,+\infty[;L) \cap C^0([0,+\infty[;D(A)) \text{ so that} \\ u_{|t=0} = u_0, \\ \dfrac{\mathrm{d}u}{\mathrm{d}t} + Au = f, \end{cases} \tag{1.1}$$

where L is a separable Hilbert space, A is the generator of a linear contraction semi-group of class C^0, and $D(A) \subset L$ is the domain of A. The Hille–Yosida theorem guarantees that this problem is well posed. Hereafter, we shall think of A as being a first-order differential operator; for instance

Correspondence to: Jean–Luc Guermond, (e-mail: guermond@limsi.fr)
Contract grant sponsor: ASCI, Orsay; contract grant number: UPR-CNRS 9029

$\Omega =]0,1[^2$, $A = \partial_{x_1}$, $D(A) = \{v \in L^2(\Omega), \partial_{x_1} v \in L^2(\Omega),\ v_{|x_1=0} = 0\}$, $L = L^2(\Omega)$. The goal of the present article is to present a stabilized Galerkin technique for approximating (1.1). The most novel feature of the proposed technique is that it yields optimal error estimates in the graph norm of A.

Classical approximation theories based on the Galerkin technique yield approximate semi-groups that are uniformly bounded in the norm of L only (see for instance Brenner et al. [1]). It is in general difficult to obtain optimal error bounds in the graph norm, although stability in norms, which are intermediate between that of L and the graph norm, can sometimes be obtained. For instance, in the case of scalar hyperbolic equations, the discontinuous Galerkin technique yields stability in a mesh-dependent norm that is related to the graph norm (see Lesaint–Raviart [2] and Johnson-Pitkäranta [3]). Also for this class of problems, another possibility consists in using the Galerkin/Least-Square technique both in space and time as in Johnson et al. [4].

Another situation that we shall also consider in this article consists in the following problem: For $f \in C^1([0,+\infty[; L)$ and $u_0 \in X$,

$$\begin{cases} \text{Find } u \in C^1([0,+\infty[; L) \cap C^0([0,+\infty[; X) \text{ so that} \\ u_{|t=0} = u_0, \\ \dfrac{\mathrm{d}u}{\mathrm{d}t} + Au + \epsilon Du = f, \end{cases} \tag{1.2}$$

where X is a separable Hilbert space that is dense and continuously embedded in $D(A)$, $D : X \longrightarrow X'$ is linear, continuous, and

$$\exists c > 0,\ \forall v \in X, \qquad \langle Dv, v\rangle + (Av, v)_L + \|v\|_L^2 \geq c\|v\|_X^2. \tag{1.3}$$

This situation corresponds to parabolic equations (see, e.g., Lions–Magenes [5, p. 253]). In practice, one may think of $D = -\Delta : H_0^1(\Omega) \subset D(A) \longrightarrow H^{-1}(\Omega)$. When ϵ is small or if D is degenerate, the stability induced by the elliptic term ϵDu is not strong enough to guarantee the Galerkin approximation to be free from spurious numerical wiggles. The second result of the present article is that, by adopting the stabilized Galerkin technique tailored for solving (1.1), one obtains an approximate solution of (1.2) that converges in the Graph norm of A uniformly with respect to ϵ. Hence, contrary to the space-time Galerkin/Least-Square method, no stability parameter needs to be tuned as a function of ϵ.

The theory developed in this article is based on the following two principles:

(i) The first principle is that any valuable internal approximation theory of e^{tA} should provide some stability in the graph norm, because controlling the graph norm guarantees the approximate solution to be free of spurious numerical wiggles.To this end, the theory presented herein involves two approximation spaces X_H and X_h, so that the triplet (X_h, X_H, A) satisfies the following discrete inf-sup condition:

$$\exists c_a,\ \forall (H,h),\ \forall u_H \in X_H, \qquad \sup_{v_h \in X_h} \frac{(Au_H, v_h)_L}{\|v_h\|_L} \geq c_a \|Au_H\|_L. \tag{1.4}$$

It is shown in [6, 7] that, for a large class of linear PDE's of first order, it is indeed possible 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. The spaces X_H and X_h^H are referred to as the resolved scales space and the subgrid scales space, respectively. The inequality (1.4) is important, because it yields stability on the graph norm of the resolved scales of the approximate solution.

(ii) The second principle upon which the present work is based is that, for monotone operators, the graph norm of the subgrid scales of the approximate solution can be controlled by means of a small artificial diffusion mechanism; the control being provided by a simple energy argument.

The ideas of scale separation and subgrid viscosity are rooted in many works: subgrid modeling and spectral viscosity (Smagorinsky [8], Tadmor [9]), the Nonlinear Galerkin Method (Foias–Manley–Temam [10], Marion–Temam [11]), the stabilizing property of bubble functions (Arnold–Brezzi–Fortin [12], Brezzi et al. [13], Baiocchi–Brezzi–Franca [14], Crouzeix–Raviart [15]).

The material of the article is organized as follows. In Section II, the abstract functional framework is introduced; stability and quasi-optimal convergence results in the graph norm are proved. The singular perturbation problem (1.2) is considered in Section III, and the results of Section II are extended to this context. Section IV is devoted to examples and applications of the present theory.

II. APPROXIMATION OF A MODEL PROBLEM

A. Model Problem

Let L be a real separable Hilbert space and $(\cdot,\cdot)_L$ be its inner product. Let $A : D(A) \subset L \longrightarrow L$ be an unbounded linear operator. We assume that A is monotone:

$$\forall v \in D(A), \qquad (Av, v)_L \geq 0; \tag{2.1}$$

and A is maximal:

$$\forall f \in L,\ \exists v \in D(A), \qquad v + Av = f. \tag{2.2}$$

A first series of consequences of these hypotheses are as follows.

Lemma 2.1. *If $A : D(A) \subset L \longrightarrow L$ is monotone and maximal, then*

(i) $D(A)$ is dense in L.
(ii) A is closed (i.e. the graph of A is closed).
(iii) For all $\lambda > 0$, $I + \lambda A : D(A) \longrightarrow L$ is bijective and $\|(I+\lambda A)^{-1}\|_{\mathcal{L}(L,L)} \leq 1$.

Proof. See Brezis [16, p. 101] or Yosida [17, p. 246]. ■

Furthermore, the Hille–Yosida theorem states that A is the generator of a contraction semi-group of class C^0 and, conversely, generators of such semi-groups are maximal monotone operators (see Brezis [16, p. 110] or Yosida [17, p. 248]). In practice, this result implies the following.

Theorem 2.2. *For $f \in C^1([0,+\infty[; L)$ and $u_0 \in D(A)$, the problem*

$$\begin{cases} u_{t=0} = u_0, \\ \dfrac{\mathrm{d}u}{\mathrm{d}t} + Au = f, \end{cases} \tag{2.3}$$

has a unique u in $C^1([0,+\infty[; L) \cap C^0([0,+\infty[; D(A))$.

Having in mind a Galerkin approximation of the evolution equation in (2.3), we want to introduce a bilinear form $a(u,v) = (Au,v)_L$. To this end, we set $V = D(A)$ and we equip

V with the graph norm: $\|v\|_V = (\|v\|_L^2 + \|Av\|_L^2)^{1/2}$. Since the graph of A is closed, V is a Banach space. Furthermore, it is clear that equipped with the inner product $(u,v)_L + (Au, Av)_L$, V is a Hilbert space. Since $D(A) = V$ is dense in L, we are in the following classical situation $V \subset L \equiv L' \subset V'$. Furthermore, we introduce the semi-norm $|v|_V = \|Av\|_L$. From this definition, we deduce

$$\forall u \in V, \qquad \sup_{v \in L} \frac{a(u,v)}{\|v\|_L} = |u|_V. \tag{2.4}$$

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

$$\forall (u,v) \in V \times V, \qquad a_s(u,v) = \frac{1}{2}(a(u,v) + a(v,u)). \tag{2.5}$$

It is clear that for all u in V we have $a(u,u) = a_s(u,u) \geq 0$; we shall hereafter refer to this property as a and a_s being monotone bilinear forms. We shall make use of the following classical property.

Lemma 2.3. *Let E be a vector space and $x : E \times E \longrightarrow \mathbb{R}$ be a symmetric monotone bilinear form, then*

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

The problem (2.3) can be recast into the following equivalent form: For $f \in C^1([0,+\infty[;L)$ and $u_0 \in D(A)$, find u in $C^1([0,+\infty[;L) \cap C^0([0,+\infty[;D(A))$ so that

$$\begin{cases} (u(0),v) = (u_0,v), \quad \forall v \in L \\ (\dfrac{\mathrm{d}u}{\mathrm{d}t}, v)_L + a(u,v) = (f,v)_L, \quad \forall v \in L,\ \forall t \geq 0. \end{cases} \tag{2.6}$$

Remark. *Problem (2.6) is essentially a Petrov–Galerkin problem; that is, the solution space and the space of the test functions are different; i.e., $u(t)$ is expected to be in $D(A)$, whereas the test functions span L. The failure of classical discrete Galerkin techniques to approximate this problem properly is rooted in this basic fact.*

Remark. *Note that since u is in $C^1([0,+\infty[;L)$, the bound $\|Au\|_L \leq \|du/dt\|_L + \|f\|_L$ provides an estimate on $\|Au\|_L$. The key idea of the present work is that good Galerkin approximations to this problem should yield a similar estimate.*

B. Discrete Setting

We introduce two sequences of finite dimensional subspaces of V, say $(X_H)_H$ and $(X_h)_h$, where indices $0 < H \leq 1$ and $0 < h \leq 1$ denote two positive parameters tending to zero. In the applications described in §IV., we take $h \approx H/2$.

The space X_H is assumed to have the following approximation property: there are W, a dense subspace of V, a linear operator $I_H \in \mathcal{L}(W, X_H)$ and $k > 0$, $c > 0$, so that

$$\forall v \in W, \qquad \|v - I_H v\|_L + H\|v - I_H v\|_V \leq cH^{k+1}\|v\|_W. \tag{2.7}$$

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

$$\forall v_H \in X_H, \qquad \sup_{\phi_h \in X_h} \frac{a(v_H, \phi_h)}{\|\phi_h\|_L} \geq c_a |v_H|_V. \tag{2.8}$$

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:

$$\exists c > 0,\ \forall (H, h),\ \forall v_h \in X_h \qquad \|P_H v_h\|_L \leq c \|v_h\|_L. \tag{2.9}$$

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 $v_h^H = v_h - v_H$; that is to say,

$$X_h = X_H \oplus X_h^H. \tag{2.10}$$

We shall hereafter refer to X_H and X_h^H as the space of the resolved scales and the space of the subgrid scales, respectively. The reader may think of P_H as a filter.

In addition, we assume that X_h satisfies the following inverse stability property: there is $c_i > 0$ so that

$$\forall v_h \in X_h, \qquad \|v_h\|_V \leq c_i H^{-1} \|v_h\|_L. \tag{2.11}$$

Remark. *It is at this point that we implicitly assume that A is a first-order differential operator. Actually, in the case of a finite element approximation, (2.11) holds uniformly if the mesh is quasi-uniform (see, e.g., Girault–Raviart [18, p. 103]), A is a first-order differential operator, and $c_1 h \leq H \leq c_2 h$. The last constraint is equivalent to assuming that the dimension of X_H is a fraction of that of X_h. In other words, X_h is not significantly larger than X_H, but it is large enough for the discrete inf-sup inequality (2.8) to hold uniformly with respect to H and h.*

Remark. *The present theory can be extended to nonuniform meshes by proceeding as in Guermond [7].*

Remark. *The reader is referred to Guermond [6, 7] for examples of $\mathbb{P}_1$ and $\mathbb{P}_2$ finite element frameworks satisfying the hypotheses (2.7), (2.8), (2.9), and (2.11) for a large class of first-order differential operators.*

C. Discrete Problem

The subgrid scales have been introduced for the sole purpose of controlling the resolved scales by means of the inf-sup inequality (2.8). By doing so, we are left with the problem of controlling the subgrid scales. To this end, we introduce an artificial diffusion mechanism. More precisely, 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 are a semi-norm $|\cdot|_b$ and $c_B > 0$ so that

$$\begin{cases} b_h(v_h^H, v_h^H) \geq H |v_h^H|_b^2, \\ b_h(v_h^H, w_h^H) \leq c_B H |v_h^H|_b \, |w_h^H|_b, \end{cases} \tag{2.12}$$

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

$$\forall v_h^H \in X_h^H, \qquad c_{e1} |v_h^H|_V \leq |v_h^H|_b \leq c_{e2} H^{-1} \|v_h^H\|_L. \tag{2.13}$$

Example 1. *The simplest choice for* b_h *is* $b_h(v_h^H, w_h^H) = H(Av_h^H, Aw_h^H)_L$.

Example 2. *Let* X *be a dense subspace of* V*, continuously embedded in* V*. Assume that there is a semi-norm* $|\cdot|_X$ *in* X *so that* $c_{e1}|v_h^H|_V \le |v_h^H|_X \le c_{e2}H^{-1}\|v_h^H\|_L$ *for all* v_h^H *in* X_h^H*. Let* $(\cdot,\cdot)_X$ *be the inner product associated with* $|\cdot|_X$*. By assuming* $X_h \subset X$*, one can set* $b_h(v_h^H, w_h^H) = H(v_h^H, w_h^H)_X$.

Example 3. *For the scalar transport equation* $\mathrm{d}u/\mathrm{d}t + \beta\cdot\nabla u = f$ *in* Ω*, with suitable assumptions on the vector field* β*, we have* $Au = \beta\cdot\nabla u$*,* $L = \mathrm{L}^2(\Omega)$ *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. By assuming* $X_h \subset \mathrm{H}^1(\Omega)$ *and* $X_h \subset V$*, the following two definitions are possible for* b_h*:*

$$b_h(v_h^H, w_h^H) = \begin{cases} H\displaystyle\int_\Omega (\beta\cdot\nabla v_h^H)(\beta\cdot\nabla w_h^H), \\ H\displaystyle\int_\Omega (\nabla v_h^H)\cdot(\nabla w_h^H). \end{cases} \tag{2.14}$$

The second model may be helpful in practice for two reasons. First, it may help dampening cross-wind oscillations when approximating very stiff problems; second, if, in practice, β *is time-dependent (though the present theory assumes* β *to be time-independent), this model is time-independent (hence, the assembling of the matrix is done only once).*

Lemma 2.4. *There is* $c_b > 0$ *so that*

$$\forall v_h^H \in X_h^H, \qquad \sup_{w_h\in X_h} \frac{b_h(v_h^H, w_h^H)}{\|w_h\|_L} \le c_b|v_h^H|_b. \tag{2.15}$$

Proof. The stability hypothesis (2.9) on P_H together with the inverse stability property (2.13) yields

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

The desired result follows readily. ∎

For the sake of simplicity we assume that u_0 is in W, and we approximate the initial data by $I_H u_0$. The discrete problem we consider hereafter consists in finding u_h in $C^1([0,+\infty[;X_h)$ so that

$$\begin{cases} u_{h|t=0} = I_H u_0, \\ (\dfrac{\mathrm{d}u_h}{\mathrm{d}t}, v_h)_L + a(u_h, v_h) + b_h(u_h^H, v_h^H) = (f, v_h), \qquad \forall v_h \in X_h. \end{cases} \tag{2.16}$$

Theorem 2.5. *The discrete problem (2.16) has a unique solution.*

Proof. (2.16) is a system of linear ODEs. ∎

Remark. *The subgrid stabilization technique proposed here has some similarities with the spectral viscosity method proposed by Tadmor [9], Maday, Ould Kaber and Tadmor [19] to stabilize spectral methods for nonlinear conservation laws.*

D. Error Analysis

Let T be a strictly positive real number.

Lemma 2.6. *Let ϕ be in $W^{1,1}((0,T);\mathbb{R})$ (the derivative being understood in the usual distribution sense) and ψ be in $L^1((0,T);\mathbb{R})$ so that $\phi \geq 0$, $\psi \geq 0$. Assume there are a, $b \geq 0$ two real numbers so that $\phi^{(1)}(t) + \psi \leq a\phi^{1/2} + b$, then*

$$\|\phi\|_{L^\infty(0,T)} + \|\psi\|_{L^1(0,T)} \leq e\,(\frac{a^2}{4}T^2 + bT + \phi(0)).$$

Proof. By using the inequality $xy \leq \gamma x^2 + y^2/4\gamma$, which is valid for any positive constant γ, we infer

$$\frac{\mathrm{d}\phi}{\mathrm{d}t} + \psi \leq \frac{a^2 T}{4} + \frac{\phi}{T} + b,$$

which yields

$$\frac{\mathrm{d}}{\mathrm{d}t}(\phi(t)e^{-t/T}) + \psi(t)e^{-t/T} \leq e^{-t/T}(\frac{a^2}{4}T + b).$$

By integrating this inequality, we obtain for a.e. t in $(0,T)$

$$\begin{aligned}\phi(t) + \int_0^t \psi(\tau)e^{(t-\tau)/T}\,\mathrm{d}\tau &\leq \phi(0)e^{t/T} + (\frac{a^2}{4}T + b)\int_0^t e^{(t-\tau)/T}\,\mathrm{d}\tau,\\ \phi(t) + \int_0^t \psi(\tau)\,\mathrm{d}\tau &\leq \phi(0)e + (\frac{a^2}{4}T + b)T(e-1)\\ &\leq e(\frac{a^2}{4}T^2 + bT + \phi(0)).\end{aligned}$$

The proof is complete. ■

The main convergence result of this section is as follows.

Theorem 2.7. *Assume u is in $W^{2,\infty}([0,T];W)$, then the discrete solution u_h of (2.16) satisfies*

$$\|u-u_h\|_{L^\infty([0,T];L)} + \left[\int_0^T a_s(u-u_h, u-u_h)\right]^{1/2} \leq c_1(T,u)H^{k+1/2}, \tag{2.17}$$

$$\left[\frac{1}{T}\int_0^T \|u-u_h\|_V^2\right]^{1/2} \leq c_2(T,u)H^k, \tag{2.18}$$

where constants c_1 and c_2 can be bounded from above as follows:

$$c_1 \leq c\,[H + T\,[1+T]]^{1/2}\,\|u\|_{W^{2,\infty}([0,T];W)},$$

$$c_2 \leq c\,[1+T]\,\|u\|_{W^{2,\infty}([0,T];W)}.$$

Proof. To simplify the notations, let us set $\eta_h(t) = u(t) - I_H u(t)$, and $e_h(t) = I_H u(t) - u_h(t)$. Note that we have $u - u_h = \eta_h + e_h$. From the definition of $\eta_h(t)$, we deduce

$$\forall j \in \{0,1,2\}, \qquad \|\eta_h^{(j)}\|_{L^\infty([0,T];H)} + H\|\eta_h^{(j)}\|_{L^\infty([0,T];V)} \leq cH^{k+1}\|u\|_{W^{2,\infty}([0,T];W)}.$$

The equation that controls e_h is obtained by subtracting (2.16) from (2.6), where the test functions span X_h:

$$\forall v_h \in X_h, \qquad (\frac{\mathrm{d}e_h}{\mathrm{d}t}, v_h)_L + a(e_h, v_h) - b_h(u_h^H, v_h^H) = -(\frac{\mathrm{d}\eta_h}{\mathrm{d}t}, v_h)_L - a(\eta_h, v_h).$$

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

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

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

$$\forall v_h \in X_h, \quad (\frac{\mathrm{d}e_h}{\mathrm{d}t}, v_h)_L + a(e_h, v_h) + b_h(e_h^H, v_h^H) = -(\frac{\mathrm{d}\eta_h}{\mathrm{d}t}, v_h)_L - a(\eta_h, v_h). \tag{2.19}$$

Furthermore, $\mathrm{d}e_h/\mathrm{d}t = e_h^{(1)}$ is controlled by

$$\forall v_h \in X_h, \quad (e_h^{(2)}, v_h)_L + a(e_h^{(1)}, v_h) + b_h(e_h^{(1)H}, v_h^H) = -(\eta_h^{(2)}, v_h)_L - a(\eta_h^{(1)}, v_h). \tag{2.20}$$

Let us derive some bounds on the initial data. It is clear that $e_h(0) = 0$; furthermore, by using (2.19) at $t = 0$, we infer $\|e_h^{(1)}(0)\|_L \le \|\eta_h^{(1)}(0)\|_L + \|\eta_h(0)\|_V$. As a result, we have the following error estimates at $t = 0$:

$$\begin{cases} \|e_h(0)\|_L = 0, \\ \|e_h^{(1)}(0)\|_L \le cH^k \|u\|_{W^{2,\infty}([0,T];W)}. \end{cases}$$

Now, we seek a bound on $\|e_h^{(1)}\|_{L^\infty([0,T];L)}$. We take $e_h^{(1)}$ as test function in (2.20). Owing to the coercivity property of b_h, we obtain

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|e_h^{(1)}\|_L^2 + a_s(e_h^{(1)}, e_h^{(1)}) + H\|e_h^{H(1)}\|_b^2 \le (\|\eta_h^{(2)}\|_L + |\eta_h^{(1)}|_V)\|e_h^{(1)}\|_L.$$

Since a_s is monotone, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\|e_h^{(1)}\|_L^2 \le 2(\|\eta_h^{(2)}\|_L + |\eta_h^{(1)}|_V)\|e_h^{(1)}\|_L.$$

By using Lemma 2.6, we infer

$$\|e_h^{(1)}\|_{L^\infty([0,T];L)}^2 \le c(\|e_h^{(1)}(0)\|_L^2 + T^2\|\eta_h\|_{W^{2,\infty}([0,T];V)}^2).$$

As a result, we have

$$\|e_h^{(1)}\|_{L^\infty([0,T];L)} \le cH^k(1+T)\|u\|_{W^{2,\infty}([0,T];W)}.$$

To obtain a bound on e_h, we use e_h as a test function in (2.19). Since a_s is symmetric and monotone, owing to Lemma 2.3 we have

$$\begin{aligned} \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|e_h\|_L^2 + a_s(e_h, e_h) + H\|e_h^H\|_b^2 &\le \|\eta_h^{(1)}\|_L\|e_h\|_L + a(e_h, \eta_h) - 2a_s(e_h, \eta_h) \\ &\le \|\eta_h^{(1)}\|_L\|e_h\|_L + |e_h|_V\|\eta_h\|_L \\ &\quad + \gamma a_s(e_h, e_h) + c_\gamma a_s(\eta_h, \eta_h), \end{aligned}$$

where we have used the inequality $2xy \le \gamma x^2 + y^2/\gamma$, which is valid for any positive constant γ. 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 value of γ and c_γ may change on different occurrences. By choosing $\gamma = 1/2$, we obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\|e_h\|_L^2 + a_s(e_h, e_h) + 2H\|e_h^H\|_b^2 \le 2\|e_h\|_L\|\eta_h^{(1)}\|_L \\ + c(|e_h|_V\|\eta_h\|_L + \|\eta_h\|_V\|\eta_h\|_L).\end{aligned} \tag{2.21}$$

Note that the term $|e_h|_V\|\eta_h\|_L$ on the right-hand side of (2.21) is not controlled yet; it is the most critical one in this error analysis. It is a this point that the inf-sup inequality (2.8) plays its role:

$$\begin{aligned}c_a|e_H|_V &\le \sup_{\phi_h \in X_h} \frac{a(e_H, \phi_h)}{\|\phi_h\|_L} \\ &\le \sup_{\phi_h \in X_h} \frac{-(e_h^{(1)}, \phi_h) - a(e_h^H, \phi_h) - b_h(e_h^H, \phi_h^H)}{\|\phi_h\|_L} \\ &\quad + \sup_{\phi_h \in X_h} \frac{-(\eta_h^{(1)}, \phi_h) - a(\eta_h, \phi_h)}{\|\phi_h\|_L} \\ &\le \|e_h^{(1)}\|_L + |e_h^H|_V + \|\eta_h^{(1)}\|_L + |\eta_h|_V + c_b|e_h^H|_b \\ &\le c(\|e_h^{(1)}\|_L + |e_h^H|_b + \|\eta_h^{(1)}\|_L + |\eta_h|_V).\end{aligned}$$

By using the bound already obtained on $\|e_h^{(1)}\|_L$ together with the triangular inequality, we infer

$$\begin{aligned}|e_h|_V &\le |e_H|_V + |e_h^H|_V \\ &\le c(|e_h^H|_b + (1+T)H^k\|u\|_{W^{2,\infty}([0,T];W)}).\end{aligned}$$

Coming back to (2.21), we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\|e_h\|_L^2 + a_s(e_h, e_h) + 2H\|e_h^H\|_b^2 &\le 2\|e_h\|_L\|\eta_h^{(1)}\|_L + cH^{2k+1}\|u\|^2_{W^{2,\infty}([0,T];W)} \\ &\quad + c'\left[|e_h^H|_b + (1+T)H^k\|u\|_{W^{2,\infty}([0,T];W)}\right]\|\eta_h\|_L \\ &\le 2\|e_h\|_L\|\eta_h^{(1)}\|_L \\ &\quad + c(1+T)H^{2k+1}\|u\|^2_{W^{2,\infty}([0,T];W)} \\ &\quad + \gamma H|e_h^H|_b^2 + c_\gamma H^{-1}\|\eta_h\|_L^2.\end{aligned}$$

By choosing $\gamma = 1$, we obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\|e_h\|_L^2 + a_s(e_h, e_h) + H\|e_h^H\|_b^2 \le 2\|e_h\|_L\|\eta_h^{(1)}\|_L \\ + c(1+T)H^{2k+1}\|u\|^2_{W^{2,\infty}([0,T];W)}.\end{aligned}$$

Owing to lemma 2.6, we infer

$$\begin{aligned}\|e_h\|^2_{L^\infty([0,T];L)} + \int_0^T (a_s(e_h,e_h) + H\|e_h^H\|_b^2) \\ &\leq c\left[\|e_h(0)\|^2_L + T^2\|\eta_h^{(1)}\|^2_L \right. \\ &\quad \left. + T(1+T)H^{2k+1}\|u\|^2_{W^{2,\infty}([0,T];W)}\right] \\ &\leq cT\left[1+T\right]H^{2k+1}\|u\|^2_{W^{2,\infty}([0,T];W)},\end{aligned}$$

which yields

$$\|u-u_h\|_{L^\infty([0,T];L)} \leq cH^{k+1/2}\left[H + T(1+T)\right]^{1/2}\|u\|_{W^{2,\infty}([0,T];W)}.$$

Now we derive an error estimate in the graph norm. By using the bound already obtained on $|e_h|_V$, we have

$$\begin{aligned}\int_0^T |e_h|^2_V &\leq c\int_0^T \left[|e_h^H|_b^2 + (1+T)^2H^{2k}\|u\|^2_{W^{2,\infty}([0,T];W)}\right] \\ &\leq cTH^{2k}\left[(1+T)+(1+T)^2\right]\|u\|^2_{W^{2,\infty}([0,T];W)} \\ &\leq cTH^{2k}(1+T)^2\|u\|^2_{W^{2,\infty}([0,T];W)}.\end{aligned}$$

As a result,

$$\left[\frac{1}{T}\int_0^T |u-u_h|^2_V\right]^{1/2} \leq cH^k\left[1+T\right]\ \|u\|_{W^{2,\infty}([0,T];W)}.$$

The final estimate in the graph norm is obtained by combining this bound and that in the L-norm. ■

Remark. *The bound (2.18) is optimal in the graph norm.*

Remark. *The bound (2.17) is not optimal: a factor $H^{1/2}$ is missing. Actually, by proceeding as in Zhou [20], optimality can be recovered if the mesh underlying the approximation space X_h satisfies special geometrical properties.*

Remark. *The estimate (2.17) is identical to the one that could be obtained by applying the counterpart of the discontinuous Galerkin method to the present problem (see Johnson–Pitkäranta [3]).*

Remark. *Note that for large time T, $c_1 = \mathcal{O}(T)$ and $c_2 = \mathcal{O}(T)$, that is, in the most unfavorable case, the error grows linearly with respect to T.*

Remark. *When looking back at the proof of theorem 2.7, one observes that the stability hypothesis (2.8) can be slightly weakened as follows: there are $c_a > 0$ and $c_\delta > 0$, independent of (H,h) so that*

$$\forall v_h \in X_h, \qquad \sup_{\phi_h\in X_h}\frac{a(v_H,\phi_h)}{\|\phi_h\|_L} \geq c_a|v_H|_V - c_\delta[\|v_h\|_L + a_s(v_h,v_h)^{1/2} + |v_h^H|_V]. \quad (2.22)$$

III. SINGULAR PERTURBATION PROBLEM

This section is devoted to the analysis of problem (1.2). In terms of PDEs, this situation corresponds in practice to hyperbolic equations perturbed by a small elliptic term or a degenerate elliptic operator.

A. Abstract Framework

In addition to the two Hilbert spaces, L and $D(A) = V$, already defined, we introduce a new Hilbert space X that is dense and continuously embedded in $D(A) = V$ (for the sake of simplicity).

We introduce a continuous bilinear form $d \in \mathcal{L}(X \times X, \mathbb{R})$, and we assume that there is a semi-norm $|\cdot|_X$ in X so that $d(u,v) \leq c_d|u|_X|v|_X$ 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,

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

We shall now consider the following problem: For $f \in C^1([0,+\infty[;L)$ and $u_0 \in X$, find u in $C^1([0,+\infty[;L) \cap C^0([0,+\infty[;X)$ so that

$$\begin{cases} (u(0),v) = (u_0,v), \quad \forall v \in L \\ (\dfrac{\mathrm{d}u}{\mathrm{d}t}, v)_L + a(u,v) + \epsilon d(u,v) = (f,v)_L, \quad \forall v \in X,\ \forall t \geq 0, \end{cases} \tag{3.2}$$

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 $\epsilon \leq 1$. Furthermore, to ensure that problem (3.2) is well-posed, we assume the following property: $\|v\|_X \leq c(\|v\|_L + |v|_X)$.

Theorem 3.1. *Problem (3.2) has a unique solution.*

Proof. Apply Lions's theorem (*cf.* Lions–Magenes [5, p. 253]). ■

B. Discrete Setting

We introduce X_H and X_h, two finite dimensional subspaces of X that satisfy the same hypotheses as in Section II; namely, hypotheses (2.7), (2.22), (2.9), (2.11). For the sake of simplicity, we assume that there is $c > 0$ so that

$$\forall v_h \in X, \qquad |v_h|_X \leq cH^{-1}\|v_h\|_L. \tag{3.3}$$

Remark. *In practice, the hypothesis (3.3) means that X and V control derivatives of the same order. For instance, think of $\Omega =]0,1[^2$, $V = \{v \in L^2(\Omega), \partial_{x_1} v \in L^2(\Omega),\ v_{|x_1=0} = 0\}$, $L = L^2(\Omega)$, $a(u,v) = \int_\Omega v\partial_{x_1} u$, $X = H_0^1(\Omega)$, and $d(u,v) = \int_\Omega \nabla u \cdot \nabla v$.*

For the sake of simplicity, the initial data u_0 is assumed to be in W, and we approximate it by $I_H u_0$. The discrete problem that we shall consider hereafter consists in finding u_h in $C^1([0,+\infty[;X_h)$ so that

$$\begin{cases} u_{h|t=0} = I_H u_0, \\ (\dfrac{\mathrm{d}u_h}{\mathrm{d}t}, v_h)_L + a(u_h,v_h) + \epsilon d(u_h,v_h) + b_h(u_h^H, v_h^H) = (f,v_h), \quad \forall v_h \in X_h. \end{cases} \tag{3.4}$$

Theorem 3.2. *Problem (2.16) is well-posed.*

Proof. It is a system of linear ODEs. ∎

C. Error Analysis

The main convergence result of this section is as follows.

Theorem 3.3. *Assume u is in $W^{2,\infty}([0,T];W)$, then the discrete solution u_h of (3.4) satisfies*

$$\|u-u_h\|_{L^\infty([0,T];L)} + \left[\int_0^T a_s(u-u_h,u-u_h)\right]^{1/2} + \epsilon^{1/2}\|u-u_h\|_{L^2([0,T];X)} \le c_1(T,u)\left[H^{k+1/2}+\epsilon^{1/2}H^k\right], \tag{3.5}$$

$$\left[\frac{1}{T}\int_0^T \|u-u_h\|_V^2\right]^{1/2} \le c_2(T,u)H^k, \tag{3.6}$$

where constants c_1 and c_2 can be bounded from above as follows:

$$c_1 \le c\left[H+T\left[1+T\right]\right]^{1/2}\|u\|_{W^{2,\infty}([0,T];W)},$$
$$c_2 \le c\left[1+T\right]\|u\|_{W^{2,\infty}([0,T];W)}.$$

Proof. As in the proof of theorem 2.7, set $\eta_h(t)=u(t)-I_Hu(t)$, and $e_h(t)=I_Hu(t)-u_h(t)$.

The equation that controls e_h is obtained by subtracting (3.4) from (3.2), where the test functions span X_h; that is, $e_h(0)=0$ and for all v_h in X_h

$$(e_h^{(1)},v_h)_L + a(e_h,v_h) + \epsilon d(e_h,v_h) + b_h(e_h^H,v_h^H) = -(\eta_h^{(1)},v_h)_L - a(\eta_h,v_h) - \epsilon d(\eta_h,v_h), \tag{3.7}$$

where we have used $u_h^H=-e_h^H$. Likewise, the ODE that controls $e_h^{(1)}$ is

$$(e_h^{(2)},v_h)_L + a(e_h^{(1)},v_h) + \epsilon d(e_h^{(1)},v_h) + b_h(e_h^{H(1)},v_h^H) = -(\eta_h^{(2)},v_h)_L - a(\eta_h^{(1)},v_h) - \epsilon d(\eta_h^{(1)},v_h). \tag{3.8}$$

First, we derive upper bounds for the initial data. By using (3.7) at $t=0$ and owing to (3.3) we obtain $\|e_h^{(1)}(0)\|_L \le \|\eta_h^{(1)}(0)\|_L + |\eta_h^{(1)}(0)|_V + c_d\epsilon H^{-1}|\eta_h^{(1)}(0)|_X$. That is to say,

$$\begin{cases} \|e_h(0)\|_L = 0, \\ \|e_h^{(1)}(0)\|_L \le c(1+\epsilon H^{-1})H^k\|u\|_{W^{2,\infty}([0,T];W)}. \end{cases}$$

Second, we bound from above the L-norm of the time-derivative of e_h. By using $e_h^{(1)}$ as test function in (3.8), we obtain

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|e_h^{(1)}\|_L^2 + (1-\frac{\epsilon}{2})a_s(e_h^{(1)},e_h^{(1)}) + \frac{\epsilon}{2}|e_h^{(1)}|_X^2 + H\|e_h^{H(1)}\|_b^2 \le c_d\epsilon|\eta^{(1)}|_X|e_h^{(1)}|_X + (\|\eta_h^{(2)}\|_L + |\eta_h^{(1)}|_V)\|e_h^{(1)}\|_L.$$

Since $\epsilon \leq 1$ and a_s is monotone, we infer

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|e_h^{(1)}\|_L^2 + (\frac{1}{2}-\gamma)\epsilon|e_h^{(1)}|_X^2 + H\|e_h^{H(1)}\|_b^2 \leq c_\gamma\epsilon|\eta_h^{(1)}|_X^2 + (\|\eta_h^{(2)}\|_L + |\eta_h^{(1)}|_V)\|e_h^{(1)}\|_L.$$

We choose $\gamma = 1/4$ and apply Lemma 2.6

$$\begin{aligned}\|e_h^{(1)}\|^2_{L^\infty([0,T];L)} &\leq c[\|e_h^{(1)}(0)\|_L^2 + T^2(\|\eta_h^{(2)}\|_L^2 + |\eta_h^{(1)}|_V^2) + \epsilon T|\eta_h^{(1)}|_X^2)] \\ &\leq c(1+T^2+\epsilon^2H^{-2})H^{2k}\|u\|^2_{W^{2,\infty}([0,T];W)}.\end{aligned}$$

As a result, we have

$$\|e_h^{(1)}\|_{L^\infty([0,T];L)} \leq cH^k(1+T+\epsilon H^{-1})\|u\|_{W^{2,\infty}([0,T];W)}.$$

Third, we derive a bound on $\|e_h\|_L$ by using e_h as test function in (3.7). Owing to Lemma 2.3, we deduce

$$\begin{aligned}&\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|e_h\|_L^2 + \frac{1}{2}a_s(e_h,e_h) + \frac{\epsilon}{2}|e_h|_X^2 + H\|e_h^H\|_b^2 \\ &\quad\leq \|\eta_h^{(1)}\|_L\|e_h\|_L + \epsilon c_d|\eta_h|_X|e_h|_X + a(e_h,\eta_h) - 2a_s(e_h,\eta_h) \\ &\quad\leq \|\eta_h^{(1)}\|_L\|e_h\|_L + |e_h|_V\|\eta_h\|_L + \gamma\epsilon|e_h|_X^2 + c_\gamma\epsilon|\eta_h|_X^2 \\ &\qquad + \gamma a_s(e_h,e_h) + c_\gamma a_s(\eta_h,\eta_h) \\ &\quad\leq \|\eta_h^{(1)}\|_L\|e_h\|_L + |e_h|_V\|\eta_h\|_L + \gamma\epsilon|e_h|_X^2 + \gamma a_s(e_h,e_h) \\ &\qquad + c_\gamma(\epsilon|\eta_h|_X^2 + |\eta_h|_V\|\eta_h\|_L).\end{aligned}$$

By choosing $\gamma = 1/4$, we obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\|e_h\|_L^2 + \frac{1}{2}a_s(e_h,e_h) + \frac{\epsilon}{2}|e_h|_X^2 + 2H\|e_h^H\|_b^2 &\leq 2\|\eta_h^{(1)}\|_L\|e_h\|_L + 2|e_h|_V\|\eta_h\|_L \\ &\quad + c(\epsilon H^{2k} + H^{2k+1})\|u\|^2_{W^{2,\infty}([0,T];W)}.\end{aligned} \tag{3.9}$$

Now, the critical step consists in finding a bound from above on $|e_h|_V$. To this end, we shall use (2.22) and we shall investigate two possibilities: either $\epsilon \leq H$ or $\epsilon > H$.

First case: $\epsilon \leq H$. (Note that this case is the most important one in practical applications.) The weakened stability hypothesis (2.22) yields

$$
\begin{aligned}
c_a|e_H|_V &\leq \sup_{\phi_h\in X_h} \frac{a(e_H,\phi_h)}{\|\phi_h\|_L} + c_\delta[\|e_h\|_L + |e_h^H|_V + a_s(e_h,e_h)^{1/2}] \\
&\leq \sup_{\phi_h\in X_h} \frac{-(e_h^{(1)},\phi_h) - a(e_h^H,\phi_h) - \epsilon d(e_h,\phi_h) - b_h(e_h^H,\phi_h^H) - (\eta_h^{(1)},\phi_h)}{\|\phi_h\|_L} \\
&\quad + \sup_{\phi_h\in X_h} \frac{-a(\eta_h,\phi_h) - \epsilon d(\eta_h,\phi_h)}{\|\phi_h\|_L} + c_\delta[\|e_h\|_L + |e_h^H|_V + a_s(e_h,e_h)^{1/2}] \\
&\leq \|e_h^{(1)}\|_L + |e_h^H|_V + \|\eta_h^{(1)}\|_L + |\eta_h|_V + c_b|e_h^H|_b \\
&\quad + \epsilon c_d(|e_h|_X + |\eta_h|_X) \sup_{\phi_h\in X_h} \frac{|\phi_h|_X}{\|\phi_h\|_L} + c_\delta[\|e_h\|_L + |e_h^H|_V + a_s(e_h,e_h)^{1/2}] \\
&\leq c(\|e_h^{(1)}\|_L + |e_h^H|_b + \|\eta_h^{(1)}\|_L + |\eta_h|_V + \epsilon H^{-1}(|e_h|_X + |\eta_h|_X) \\
&\quad + \|e_h\|_L + a_s(e_h,e_h)^{1/2}) \\
&\leq c(\|e_h^{(1)}\|_L + |e_h^H|_b + \epsilon H^{-1}|e_h|_X + \|e_h\|_L + a_s(e_h,e_h)^{1/2} \\
&\quad + H^k\|u\|_{W^{2,\infty}([0,T];W)}).
\end{aligned}
$$

By using the bound already obtained on $\|e_h^{(1)}\|_L$ together with the triangular inequality and the hypothesis $\epsilon \leq H$, we infer

$$
\begin{aligned}
|e_h|_V &\leq |e_H|_V + |e_h^H|_V \\
&\leq c(|e_h^H|_b + \epsilon H^{-1}|e_h|_X + \|e_h\|_L + a_s(e_h,e_h)^{1/2} + (1+T)H^k\|u\|_{W^{2,\infty}([0,T];W)}).
\end{aligned}
$$

Coming back to (3.9), we have

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}&\|e_h\|_L^2 + \frac{1}{2}a_s(e_h,e_h) + \frac{\epsilon}{2}|e_h|_X^2 + 2H\|e_h^H\|_b^2 \\
&\leq 2\|e_h\|_L\|\eta_h^{(1)}\|_L + c(\epsilon H^{2k} + (1+T)H^{2k+1})\|u\|^2_{W^{2,\infty}([0,T];W)} \\
&\quad + c'\left[|e_h^H|_b + \epsilon H^{-1}|e_h|_X + \|e_h\|_L + a_s(e_h,e_h)^{1/2}\right]\|\eta_h\|_L \\
&\leq c\|e_h\|_L(\|\eta_h^{(1)}\|_L + \|\eta_h\|_L) + c'(\epsilon H^{2k} + (1+T)H^{2k+1})\|u\|^2_{W^{2,\infty}([0,T];W)} \\
&\quad + \gamma H|e_h^H|_b^2 + \gamma\epsilon|e_h|_X^2 + \gamma a_s(e_h,e_h) + c_\gamma(1 + \epsilon H^{-2} + H^{-1})\|\eta_h\|_L^2.
\end{aligned}
$$

By choosing $\gamma = 1/4$, we obtain

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\|e_h\|_L^2 + \frac{1}{4}a_s(e_h,e_h) + \frac{\epsilon}{4}|e_h|_X^2 + H\|e_h^H\|_b^2 &\leq c\|e_h\|_L(\|\eta_h^{(1)}\|_L + \|\eta_h\|_L) \\
&+ \quad c'[\epsilon H^{2k} + (1+T)H^{2k+1}]\|u\|^2_{W^{2,\infty}([0,T];W)}.
\end{aligned}
$$

Owing to Lemma 2.6, we infer

$$\begin{aligned}\|e_h\|^2_{L^\infty([0,T];L)} + \int_0^T [a_s(e_h,e_h) + H\|e_h^H\|_b^2 + \epsilon|e_h|_X^2] \\ \leq c\Big[\|e_h(0)\|_L^2 + T^2(\|\eta_h^{(1)}\|_L^2 + \|\eta_h\|_L^2) \\ +T(\epsilon H^{2k} + (1+T)H^{2k+1})\|u\|^2_{W^{2,\infty}([0,T];W)}\Big] \\ \leq cT\,[1+T]\,(H^{2k+1} + \epsilon H^{2k})\|u\|^2_{W^{2,\infty}([0,T];W)}.\end{aligned}$$

Second case: $\epsilon > H$. In this case, it is not necessary to control $|e_h|_V$ by means of the discrete inf-sup condition, since there is enough stability in the L-norm and X semi-norm to control this term. More precisely, we have

$$\begin{aligned}2|e_h|_V\|\eta_h\|_L &\leq \|e_h\|_V\|\eta_h\|_L \\ &\leq c\|e_h\|_X\|\eta_h\|_L \\ &\leq c(\|e_h\|_L + |e_h|_X)\|\eta_h\|_L \\ &\leq c\|e_h\|_L\|\eta_h\|_L + \gamma\epsilon|e_h|_X^2 + c_\gamma H^{-1}\|\eta_h\|_L^2.\end{aligned}$$

Now coming back to (3.9) and by choosing $\gamma = 1/4$, we deduce

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\|e_h\|_L^2 + \frac{1}{2}a_s(e_h,e_h) + \frac{\epsilon}{4}|e_h|_X^2 + 2H\|e_h^H\|_b^2 \\ \leq c\|e_h\|_L(\|\eta_h^{(1)}\|_L + \|\eta_h\|_L) + c(\epsilon H^{2k} + H^{2k+1})\|u\|^2_{W^{2,\infty}([0,T];W)}.\end{aligned}$$

By applying Lemma 2.6, we infer

$$\begin{aligned}\|e_h\|^2_{L^\infty([0,T];L)} + \int_0^T [a_s(e_h,e_h) + H\|e_h^H\|_b^2 + \epsilon|e_h|_X^2] \\ \leq c\Big[\|e_h(0)\|_L^2 + T^2(\|\eta_h^{(1)}\|_L^2 + \|\eta_h\|_L^2) \\ +T(\epsilon H^{2k} + H^{2k+1})\|u\|^2_{W^{2,\infty}([0,T];W)}\Big] \\ \leq cT\,[1+T]\,(H^{2k+1} + \epsilon H^{2k})\|u\|^2_{W^{2,\infty}([0,T];W)}.\end{aligned}$$

In both cases, we infer

$$\begin{aligned}\|u-u_h\|_{L^\infty([0,T];L)} + \left[\int_0^T a_s(u-u_h,u-u_h)\right]^{1/2} + \epsilon^{1/2}\|u-u_h\|_{L^2([0,T];X)} \\ \leq c(H^{k+1/2} + \epsilon^{1/2}H^k)\,[H + T(1+T)]^{1/2}\,\|u\|_{W^{2,\infty}([0,T];W)}.\end{aligned}$$

To obtain the error estimate in the graph norm (3.6), we proceed by looking alternatively at the case $\epsilon \leq H$ and the case $\epsilon > H$.

First case: $\epsilon \leq H$. By using the bound already obtained on $|e_h|_V$, we deduce

$$\begin{aligned}H|e_h|_V^2 \leq c(H|e_h^H|_b^2 + \epsilon|e_h|_X^2 + a_s(e_h,e_h) + H\|e_h\|_L^2 \\ + H(1+T)^2H^{2k}\|u\|^2_{W^{2,\infty}([0,T];W)}).\end{aligned}$$

As a result, we infer

$$H\int_0^T |e_h|_V^2 \le c\int_0^T \left[H|e_h^H|_b^2 + \epsilon|e_h|_X^2 + a_s(e_h,e_h)\right]$$
$$+ HT(\|e_h\|^2_{L^\infty([0,T];L)} + (1+T)^2H^{2k}\|u\|^2_{W^{2,\infty}([0,T];W)}),$$
$$\le [1+HT]T(1+T)H^{2k+1}\|u\|^2_{W^{2,\infty}([0,T];W)}.$$

The desired result follows readily.

Second case: $\epsilon \ge H$.

$$\int_0^T |e_h|_V^2 \le c(T\|e_h\|^2_{L^\infty([0,T];L)} + \frac{1}{\epsilon}\int_0^T \epsilon|e_h|_X^2)$$
$$\le cT(1+T)(H^{2k+1}+\epsilon H^{2k})(T+\frac{1}{\epsilon})\|u\|^2_{W^{2,\infty}([0,T];W)}$$
$$\le cT(1+T)^2H^{2k}(1+\frac{H}{\epsilon})\|u\|^2_{W^{2,\infty}([0,T];W)}$$
$$\le cT(1+T)^2H^{2k}\|u\|^2_{W^{2,\infty}([0,T];W)}.$$

This completes the proof. ∎

IV. NUMERICAL IMPLEMENTATION

A. $\mathbb{P}_1$ and $\mathbb{P}_2$ Interpolations

We describe in this section two 2D finite element settings that we use in our numerical tests. For the sake of simplicity, we assume hereafter that Ω is a polygon and $\mathcal{T}_H$ is a regular triangulation of Ω composed of affine simplexes, (T_H).

Two-Level $\mathbb{P}_1$ ***Interpolation*** Assuming that we shall deal with m-valued vector functions, we define a $\mathbb{P}_1$ resolved scale space, X_H, by

$$X_H = \{v_H \in H^1(\Omega)^m \mid v_{H|T_H} \in \mathbb{P}_1(T_H)^m,\ \forall T_H \in \mathcal{T}_H\}. \tag{4.1}$$

To build the subgrid scale space, we proceed as follows. 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. We set $h = H/2$ and denote by $\mathcal{T}_h$ the resulting new triangulation. For each macro-triangle T_H, we denote by $\mathbb{P}$ the set of functions that are continuous on T_H, piecewise $\mathbb{P}_1$ on each subtriangle of T_H, and vanish at the three vertices of T_H. Now we set

$$X_h^H = \{v_h^H \in H^1(\Omega)^m \mid v^H_{h|T_H} \in \mathbb{P}^m,\ \forall T_H \in \mathcal{T}_H\}. \tag{4.2}$$

It is clear that X_h has the following simple characterization:

$$X_h = \{v_h \in H^1(\Omega)^m \mid v_{h|T_h} \in \mathbb{P}_1(T_h)^m,\ \forall T_h \in \mathcal{T}_h\}. \tag{4.3}$$

We shall call the couple (X_H, X_h) the two-level $\mathbb{P}_1$ approximation (see Fig. 1).

Two-Level $\mathbb{P}_2$ ***Interpolation*** Now we build the $\mathbb{P}_2$ extension of the two-level $\mathbb{P}_1$ setting. First, we define the $\mathbb{P}_2$ finite element space for the resolved scales:

$$X_H = \{v_H \in H^1(\Omega)^m \mid v_{H|T_H} \in \mathbb{P}_2(T_H)^m,\ \forall T_H \in \mathcal{T}_H\}. \tag{4.4}$$

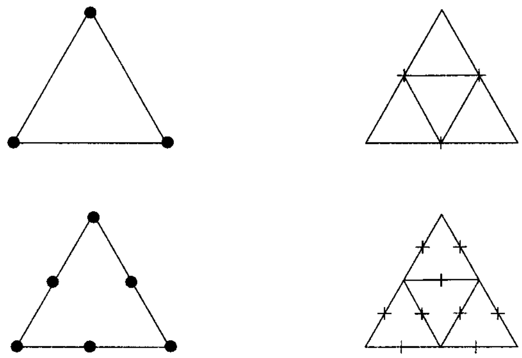

FIG. 1. Two examples of admissible finite elements: (left) resolved scale finite element; (right) subgrid scale finite element. Top: two-level $\mathbb{P}_1$ finite element; bottom: two-level $\mathbb{P}_2$ finite element.

To define the subgrid scale space, 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 subtriangles. For each triangle T_h in the new triangulation $\mathcal{T}_h$, we denote by ψ_1, ψ_2, ψ_3 the three $\mathbb{P}_2$ nodal functions associated with the middle of each edges of T_h. We define the subgrid scale space by

$$X_h^H = \{v_h^H \in H^1(\Omega)^m \mid v_{h|T_h}^H \in \text{span}(\psi_1, \psi_2, \psi_3)^m,\ \forall T_h \in \mathcal{T}_h\}. \tag{4.5}$$

X_h has the following simple characterization:

$$X_h = \{v_h \in H^1(\Omega)^m \mid v_{h|T_h} \in \mathbb{P}_2(T_h)^m,\ \forall T_h \in \mathcal{T}_h\}. \tag{4.6}$$

We shall hereafter refer to the couple (X_H, X_h) as the two-level $\mathbb{P}_2$ approximation (see also Fig. 1).

Subgrid Viscosity In all the numerical tests reported hereafter, the bilinear form associated with the subgrid viscosity is defined by

$$b_h(v_h^H, w_h^H) = c_b \sum_{T_h \in \mathcal{T}_h} \text{mes}(T_h)^{1/2} \int_{T_h} \nabla v_h^H \cdot \nabla w_h^H, \tag{4.7}$$

where c_b is a fixed parameter.

B. Convergence Tests

To illustrate the efficiency of the present method, we make convergence tests in $\Omega =]0, 1[$ on the model problem

$$\begin{cases} u_{|t=0} = \sin(2\pi x^\alpha), \\ \dfrac{\partial u}{\partial t} + \dfrac{\partial u}{\partial x} = 0, \end{cases} \tag{4.8}$$

with periodic boundary conditions. To avoid superconvergence phenomena, we use random quasi-uniform grids. The subgrid parameter c_b in (4.7) is set to 0.1; $\mathcal{O}(1)$ variations around this value do not change the conclusions we shall draw. The march in time is done by means of a fully implicit time-stepping strategy based on the second-order, 3-level, backward differentiation formula. In each case, the time-step is chosen small enough to guarantee that the error in time is much smaller than the error in space.

We use $\mathbb{P}_1$ finite elements, and we make tests with $\alpha = 0.6$ and $\alpha = 0.8$. In both cases, u_0 is in $H^1(\Omega)$. The results at time $t = 1$ are shown in Fig. 2.

It is clear that, in the two cases considered, the convergence in the L^2 norm is monotone for the stabilized 2-level $\mathbb{P}_1$ solution, whereas it is erratic for the $\mathbb{P}_1$ Galerkin solution. Note also that the Galerkin solution does not converge in the H^1 norm, whereas the stabilized one does, though slowly, because the exact solution is not in $H^2(\Omega)$.

To further illustrate the convergence problem in the Graph norm of the Galerkin solution, we show in Fig. 3 the two-level $\mathbb{P}_1$ stabilized solution and the $\mathbb{P}_1$ Galerkin solution on three different grids: $h = 1/60$, $h = 100$, and $h = 1/200$. For all grids, the Galerkin solution is plagued by oscillations that spread all over the computational domain, whereas the stabilized solution is smooth everywhere except close to the point where the space derivative of the solution is singular.

To give an idea on the efficiency of the proposed method on 2D problems, we now solve (4.8) in $\Omega =]0, 1[^2$, still with periodic boundary conditions. The mesh is composed of 3728 triangles and of 1945 $\mathbb{P}_1$ nodes; the mesh size is approximatively $h \approx 1/40$. The results at time $t = 1$ are shown in Fig. 4. As in 1D, the Galerkin solution oscillates widely throughout the domain. The 2-level $\mathbb{P}_1$ stabilized solution is smooth, and the error is localized in the region where the solution is rough.

C. Convergence Tests with Rough Data

The tests performed above show that the subgrid stabilization technique can efficiently dampen oscillations that otherwise propagate throughout the domain. However, when the solution, or one of its derivative, is rough, the stabilization technique cannot control the highly localized Gibbs phenomenon. To eliminate these residual, unwelcome oscillations, we introduce a subgrid shock

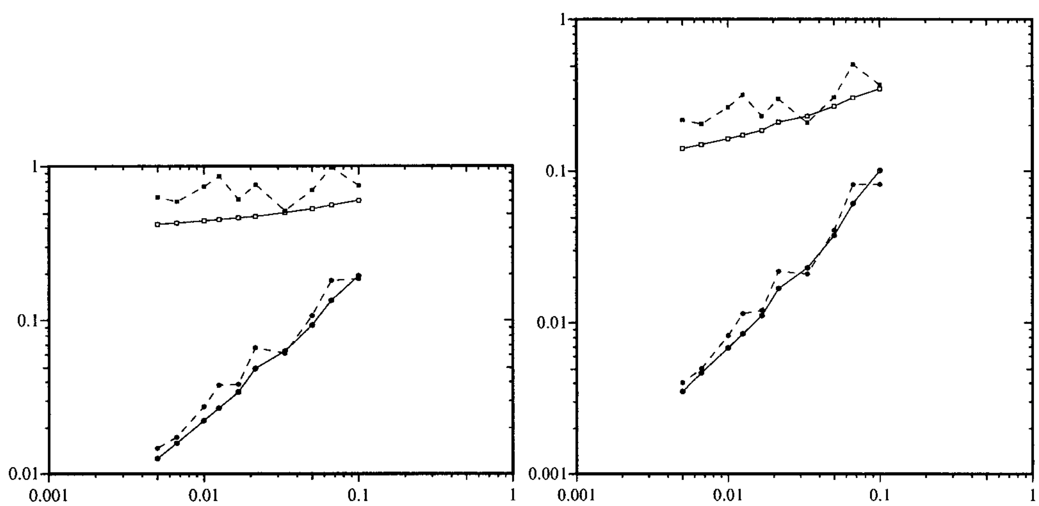

FIG. 2. Convergence tests; L^2 norm and H^1 norm of error as a function mesh size h: (solid line) the two-level $\mathbb{P}_1$ solution; (dashed line) the $\mathbb{P}_1$ Galerkin solution. Left: $u_0 = \sin(2\pi x^{0.6})$; right:$u_0 = \sin(2\pi x^{0.8})$.

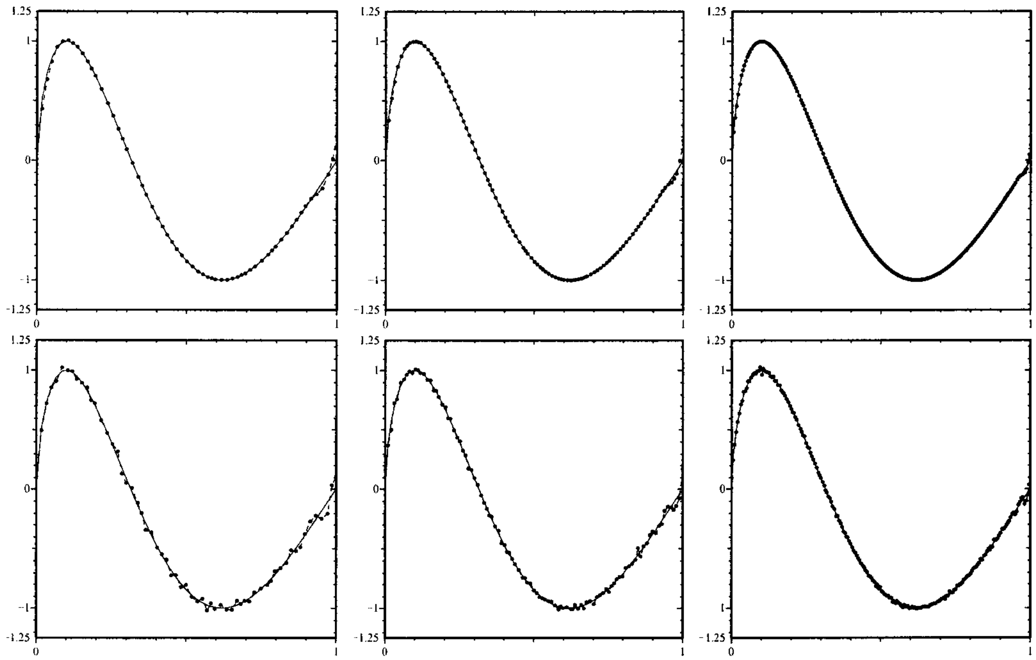

FIG. 3. Convergence tests with $u_0 = \sin(2\pi x^{0.6})$; top: 2-level $\mathbb{P}_1$ stabilized solution; bottom: $\mathbb{P}_1$ Galerkin solution; from left to right: $h = 1/60$, $h = 1/100$, $h = 1/200$.

capturing form as follows:

$$c_h(u_h^H, v_h, w_h) = c_{sc} \sum_{T_h \in \mathcal{T}_{2h}} \text{mes}(T_h)^{1/2} \frac{\|\nabla u_h^H\|_{0,T_h}}{\|\nabla u_h\|_{0,T_h}} \int_{T_h} \nabla v_h \cdot \nabla w_h, \tag{4.9}$$

where we recall that $u_h^H = (1 - P_H)u_h$ is the subgrid scale (i.e., the fluctuating part) of u_h. Basically, this term is $\mathcal{O}(H^{k+1})$ when the solution is smooth; as a result, it does not modify the convergence properties of the algorithm. The approximate problem is now: Find u_h in

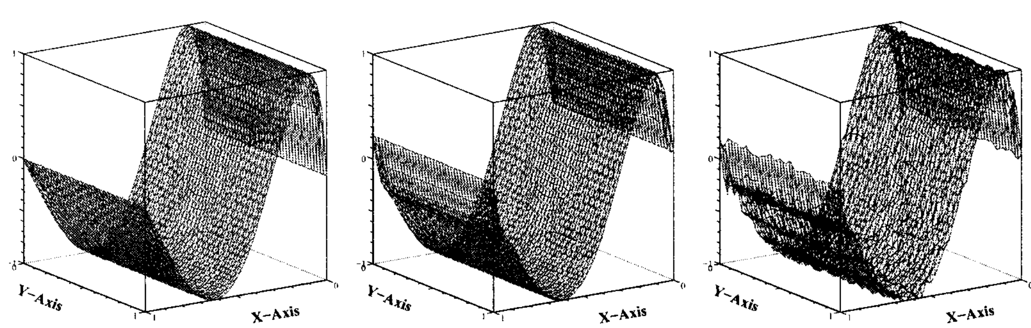

FIG. 4. Advection problem in a square, with $u_0 = \sin(2\pi x^{0.6})$; left: $\mathbb{P}_1$ interpolate of the exact solution; middle: 2-level $\mathbb{P}_1$ stabilized solution; right: $\mathbb{P}_1$ Galerkin solution.

$C^1([0,+\infty];X_h)$ so that

$$\begin{cases} u_{h|t=0} = I_H u_0, \\ (\frac{\mathrm{d}u_h}{\mathrm{d}t}, v_h)_L + a(u_h, v_h) + \epsilon d(u_h, v_h) + b_h(u_h^H, v_h^H) \\ \qquad\qquad\qquad +c_h(u_h^H, u_h, v_h) = (f, v_h), \qquad \forall v_h \in X_h. \end{cases} \tag{4.10}$$

We test the new formulation on the 1D advection problem (4.8) in $\Omega =]-1,+1[$ with initial data proposed in [21]:

$$u_0(x) = \begin{cases} e^{-300(x+0.7)^2} & \text{if } \ |x+0.7| \leq 0.25, \\ 1 & \text{if } \ |x+0.1| \leq 0.2, \\ \left(1 - \left(\dfrac{x-0.6}{0.2}\right)^2\right)^{1/2} & \text{if } \ |x-0.6| \leq 0.2, \\ 0 & \text{otherwise.} \end{cases} \tag{4.11}$$

The computation is made with 2-level $\mathbb{P}_1$ finite elements and 2-level $\mathbb{P}_2$ finite elements on three different grids, respectively composed of 50, 100, and 200 nodes. The stabilizing coefficients are $c_b = 0.05$, $c_{sc} = 0.05$. The results at time $t = 4$ are shown in Fig. 5. For both the $\mathbb{P}_1$ and $\mathbb{P}_2$ approximations, the convergence is monotone. As expected, the 2-level $\mathbb{P}_2$ approximation converges faster than the 2-level $\mathbb{P}_1$ solution.

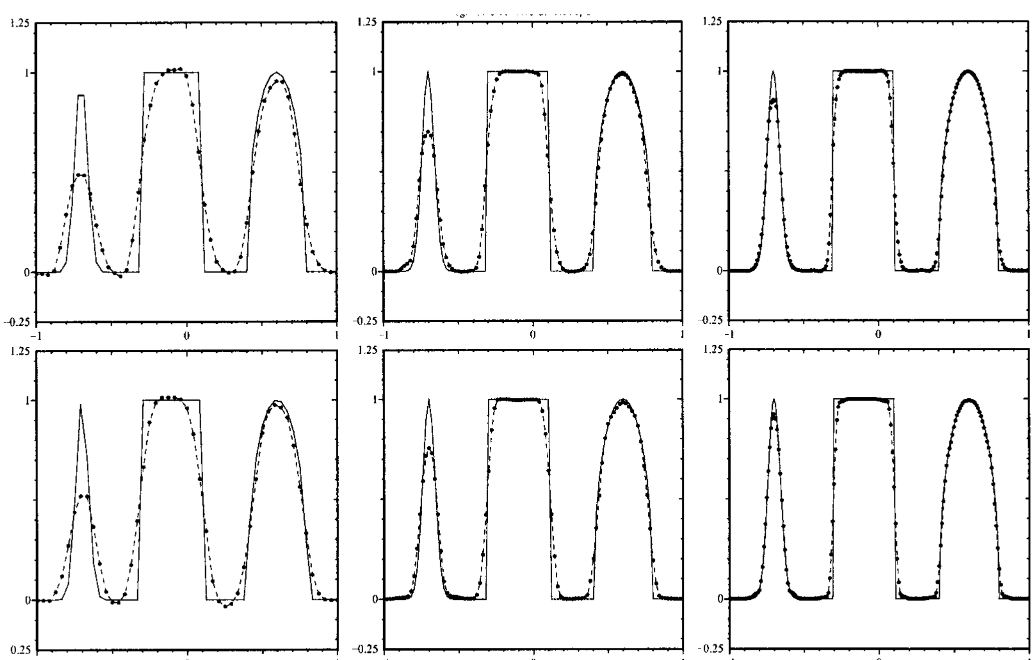

FIG. 5. Advection problem; top: $\mathbb{P}_1$ approximation; bottom: $\mathbb{P}_2$ approximation; left: 50 nodes; center: 100 elements; right: 200 nodes.

D. Degenerate Parabolic Problem

To test the capability of the proposed method to deal with degenerate parabolic problems, we consider a new class of convection-diffusion equations proposed in a series of articles by Kurganov and Rosenau (see [22] and the literature cited therein). "The novel feature of these equations is that large amplitude solutions develop spontaneous discontinuities, while small solutions remain smooth at all times."

Let us consider the following problem in $\Omega =]-3, 3[$:

$$\begin{cases} u_{|t=0} = \begin{cases} 1.2 & \text{if} \quad -3 \leq x < 0, \\ -1.2 & \text{if} \quad 0 < x \leq 3, \end{cases} \\ u(\pm 3, t) = \mp 1.2 \quad \text{for } 0 \leq t, \\ \partial_t u + \partial_x u^2 - \partial_x \left(\dfrac{\partial_x u}{\sqrt{1 + (\partial_x u)^2}} \right) = 0. \end{cases} \tag{4.12}$$

The problem is solved, up to time $t = 1.5$, by using formulation (4.10) with $\mathbb{P}_1$ finite elements on three grids: $h = 6/100$, $h = 6/200$, and $h = 6/400$. The results are shown in Fig. 6. Quite surprisingly, the Galerkin solution is not plagued by spurious oscillations, but converges to a nonentropic solution. To illustrate the insensitivity of the method to variations on the stabilizing parameters, we make two sets of computations. In the first set, we use $c_b = 0.2$, $c_{sc} = 0.2$; and in the other set, we use $c_b = 0.5$, $c_{sc} = 0.1$. The results shown in Fig. 6 demonstrate that the stabilized solution converges and does not depend too much on the choice of the stabilizing parameters.

E. The Rayleigh–Taylor Instability

We now illustrate the capability of the method to solve very stiff two-dimensional problems by testing it on a Rayleigh–Taylor flow problem.

We are interested in solving the incompressible Navier–Stokes problem with variable density in $\Omega =]-\frac{1}{2}, \frac{1}{2}[\times]-2, 2[$. We impose symmetry on $x = 0$, periodicity on $x = \pm\frac{1}{2}$, and the no-slip boundary condition on $y = \pm 2$. We consider two fluids of constant density, the ratio between the densities being 7. The heavy fluid is above the light one, and both fluids are at rest

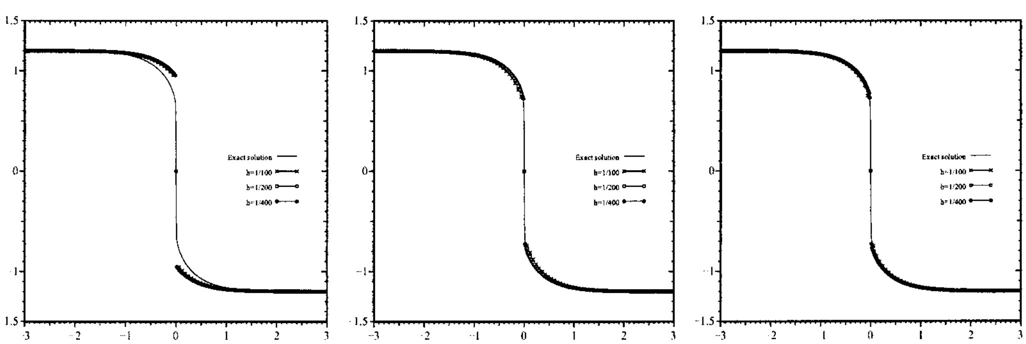

FIG. 6. Degenerate parabolic problem on three grids: $h = 6/100$, $h = 6/200$, and $h = 6/400$; left: $\mathbb{P}_1$ Galerkin solution; center: 2-level $\mathbb{P}_1$ approximation with $c_b = c_{sc} = 0.2$; right: 2-level $\mathbb{P}_1$ approximation with $c_b = 0.5$, $c_{sc} = 0.1$.

at the initial time. It is almost impossible to approximate the solution to this problem using the Galerkin technique alone. The problem is solved by means of mixed $\mathbb{P}_2/\mathbb{P}_1$ finite elements: $\mathbb{P}_2$ for velocity and density, $\mathbb{P}_1$ for pressure (see [23] for more details on this problem). We stabilize the computation on the density equation by using the 2-level $\mathbb{P}_2$, shock-capturing subgrid viscosity technique described above. To initiate the instability, the interface, initially set at $y = 0$, is slightly deformed with a sine law whose amplitude is 1% of the domain width. The Reynolds number, defined by $Re = \rho_{\min} d^{3/2} g^{1/2}/\mu$, where d is the width of the domain and g the gravity constant, is $Re = 1000$. In Fig. 7, we show the time evolution of the interface between the two fluids at times: 1, 1.5, 2, 2.5, 3, 3.5, 3.75, 4, and 4.25. It is clear that the interface remains sharp during the time evolution.

F. Shock Tube Problem

We finish this article by illustrating the capability of the proposed stabilization method to solve nonlinear conservation laws.

We treat the shock tube problem known in the literature as the Lax problem (see [21] for other details). The velocity, the pressure, the density, and the total energy are denoted by u, p, ρ, and e, respectively. By setting $\phi = (\rho, \rho u, e)$, the Euler equations read

$$\begin{cases} \phi_{t=0} = \phi_0, & \text{for } -\infty < x < +\infty \\ \partial_t \phi + \partial_x f(\phi) = 0, & \text{for } -\infty < x < +\infty, \text{ and } 0 < t, \end{cases}$$

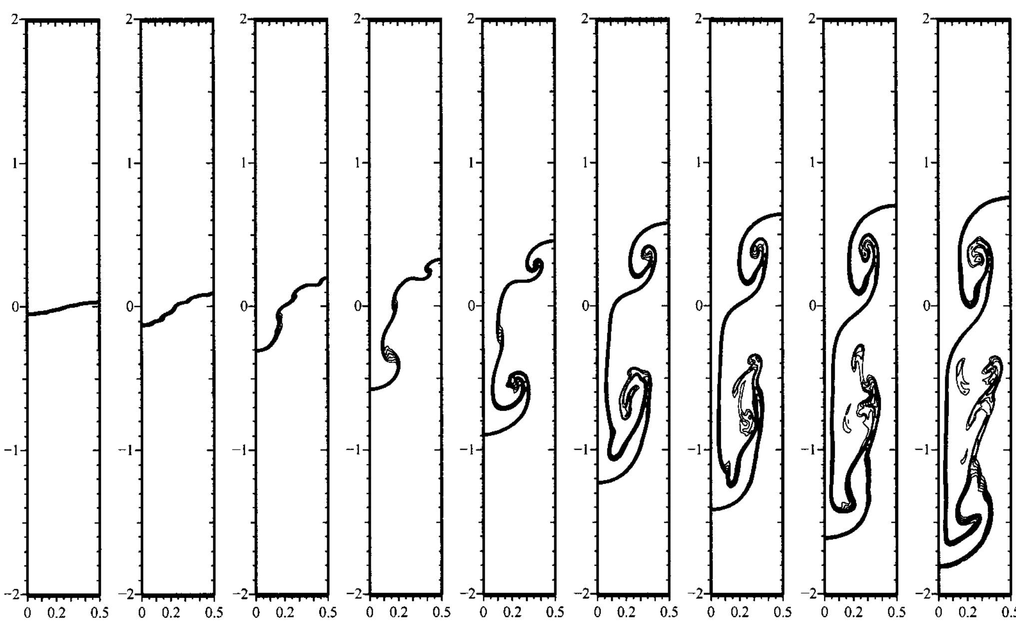

FIG. 7. Rayleigh–Taylor instability; $Re = 1000$; density ratio 7. The initial amplitude is 1% of the wavelength. The interface is shown at times: 1, 1.5, 2, 2.5, 3, 3.5, 3.75, 4, and 4.25 (density contours $2 \leq \rho \leq 4$.).

where $f(\phi) = (\rho u, \rho u^2 + p, ue + pu)$ and $p = (\gamma - 1)(e - \frac{1}{2}\rho u^2)$ with $\gamma = 1.4$. We solve this problem in $\Omega =] - 5, +5[$ with initial data

$$\begin{cases} \rho_0 = \begin{cases} 0.445 & \text{if } x < 0 \\ 0.5 & \text{if } 0 < x \end{cases} \\ u_0 = \begin{cases} 0.698 & \text{if } x < 0 \\ 0 & \text{if } 0 < x \end{cases} \\ p_0 = \begin{cases} 3.528 & \text{if } x < 0 \\ 0.571 & \text{if } 0 < x \end{cases} \end{cases}$$

The results obtained by means of the 2-level $\mathbb{P}_1$ stabilized technique on three different grids are shown in Fig. 8. The stabilizing parameters are $c_b = 0.5$, $c_{sc} = 0.3$. The approximate solution converges to the entropic solution. Similar results have been obtained with the 2-level $\mathbb{P}_2$ approximation.

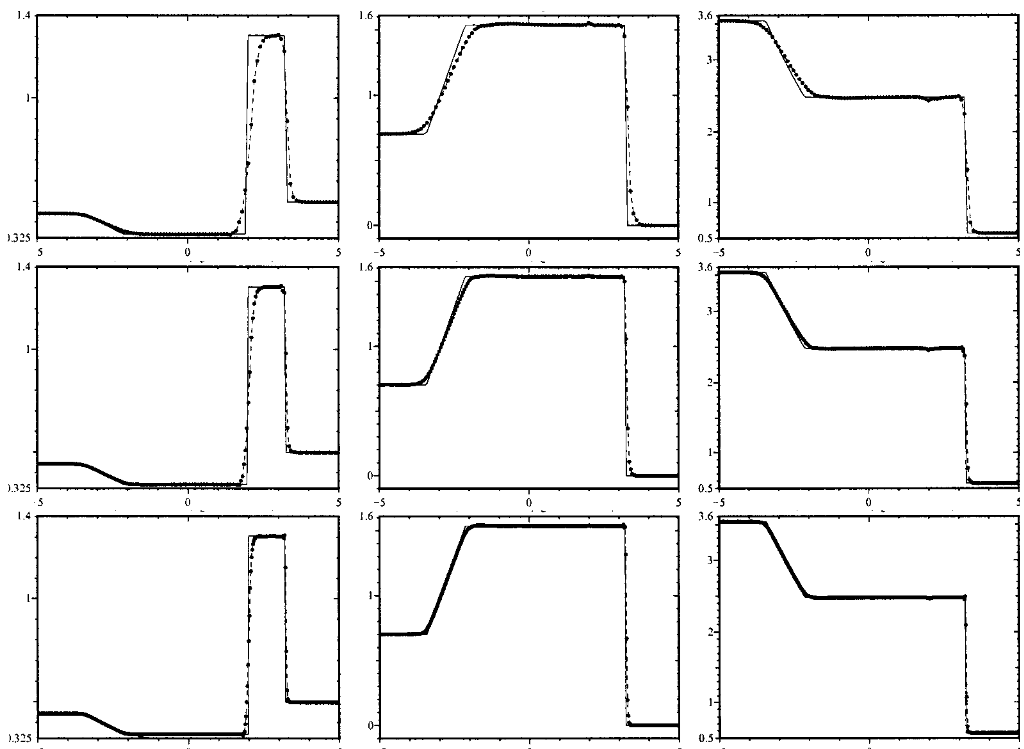

FIG. 8. Lax problem; 2-level $\mathbb{P}_1$ approximation; top: 100 nodes; middle: 200 nodes; bottom: 400 nodes; (left) density; (center) velocity; (right) pressure.

V. CONCLUSION

This article is the third part of a work initiated in [6]. The objective of this research is to propose a framework to stabilize Galerkin approximations of linear problems that do not possess a coercivity property. In [6, 7], the analysis was restricted to steady problems. The present article extends the arguments developed for steady problems to the approximation of linear contraction semi-groups of class C^0. The technique proposed is based on a hierarchical 2-level decomposition of the approximation space. The stability in the graph norm is obtained by introducing an artificial diffusion on the subgrid scales. As a result, optimal convergence in the graph norm has been proved.

The convergence proofs given in the present article assume that the grid is quasi-uniform, because uniform inverse inequalities have been used. This hypothesis can be weakened by using the local mesh size in the definition of the artificial diffusion bilinear form b_h and by proceeding as in Guermond [7].

One goal of the present research is to understand and (hopefully) to theoretically justify some dynamical subgrid viscosity models that are popular in Computational Fluid Dynamics. Consequently, we are investigating the generalization of the present technique to turbulent flows.

VI. ACKNOWLEDGMENTS

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

References

1. P. Brenner, M. Crouzeix, and V. Thomée, Single step methods for inhomogeneous linear differential equations in Banach spaces, RAIRO Anal Numer 16 (1982), 5–26.
2. P. Lesaint and P. A. Raviart, ‘‘On a finite element method for solving the neutron transport equation,’’ Mathematical aspects of finite elements in partial differential equations, C. de Boors (Editor), Academic, New York, 1974, pp. 89–123.
3. C. Johnson and J. Pitkäranta, An analysis of the discontinuous Galerkin method for a scalar hyperbolic equation, Math Comp 46 (1986), 1–26.
4. C. Johnson, U. Nävert, and J. Pitkäranta, Finite element methods for linear hyperbolic equations, Comp Meth Appl Mech Eng 45 (1984), 285–312.
5. J. L. Lions and E. Magenes, Problèmes aux limites non homogènes et applications, Dunod, Paris, 1968.
6. J.-L. Guermond, Stabilization of Galerkin approximations of transport equations by subgrid modeling, Mod Math Anal Num (M^2AN), 33 (1999), 1293–1316.
7. J.-L. Guermond, Subgrid stabilizations of Galerkin approximations of monotone operators, Zeitschrift für angewandte Mathematik und Mechanik, ZAMM 79 (1999), 29–32.
8. J. Smagorinsky, General circulation experiments with the primitive equations. I. The basic experiments, J Atmos Sci 32 (1963), 680–689.
9. E. Tadmor, Convergence of spectral methods for nonlinear conservation laws, SIAM J Numer Anal 26 (1989), 30–44.
10. C. Foias, O. P. Manley, and R. Temam, Modelization of the interaction of small and large eddies in two-dimensional turbulent flows, Math Mod Num Anal 22 (1988), 93–114.
11. M. Marion and R. Temam, Nonlinear Galerkin methods: the finite element case, Numer Math 57 (1990), 1–22.

12. D. N. Arnold, F. Brezzi, and M. Fortin, A stable finite element for the Stokes equations, Calcolo 21 (1984), 337–344.

13. F. Brezzi, M. O. Bristeau, L. Franca, M. Mallet, and G. Rogé, A relationship between stabilized finite element methods and the Galerkin method with bubble functions, Comp Meth Appl Mech Eng 96 (1992), 117–129.

14. C. Baiocchi, F. Brezzi, and L. P. Franca, Virtual bubbles and Galerkin-least-square type methods (Ga.L.S), Comp Meth Appl Mech Eng 105 (1993), 125–141.

15. M. Crouzeix and P. A. Raviart, Conforming and nonconforming finite element methods for solving the stationnary Stokes equations, RAIRO 3 (1973), 33–75.

16. H. Brezis, Analyse fonctionnelle, théorie et applications, Masson, Paris, 1983.

17. K. Yosida, Functional analysis, SCSM 123, Sixth edition, Springer–Verlag, New York, 1980.

18. V. Girault and P. A. Raviart, Finite element methods for Navier–Stokes equations, Springer Series in Comp Math 5, Springer–Verlag, New York, 1986.

19. Y. Maday, S. M. Ould Kaber, and E. Tadmor, Legendre pseudo-spectral viscosity method for nonlinear conservation laws, SIAM J Numer Anal 30 (1993), 321–342.

20. G. Zhou, How accurate is the streamline diffusion finite element method? Math Comp 66 (1997), 31–44.

21. C. W. Shu and S. Osher, Efficient implementation of essentially non-oscillatory shock-capturing schemes, II, J Comp Phys 83 (1989), 32–78.

22. J. Goodman, A. Kurganov, and P. Rosenau, Breakdown in Burgers-type equations with saturating dissipation fluxes, Nonlinearity 12 (1999), 247–268.

23. J.-L. Guermond and L. Quartapelle, A projection FEM for variable density incompressible flows, J Comp Phys, to appear.