SIAM J. Numer. Anal.
Vol. 48, No. 3, pp. 953–979

STABILIZATION OF GALERKIN FINITE ELEMENT APPROXIMATIONS TO TRANSIENT CONVECTION-DIFFUSION PROBLEMS*

JAVIER DE FRUTOS†, BOSCO GARCÍA-ARCHILLA‡, AND JULIA NOVO§

Abstract. A postprocessing technique to improve Galerkin finite element approximations to linear evolutionary convection-reaction-diffusion equations is considered. A steady convection-reaction-diffusion problem with data based on the computed standard Galerkin approximation is solved at any fixed time. The postprocessing approximation is obtained using the SUPG method over the same Galerkin finite element space. Error bounds for the method are obtained in the convection-dominated regime. The numerical experiments we present show a substantial reduction of spurious oscillations achieved by means of this procedure.

Key words. stabilized methods, finite element methods, transient convection-diffusion problems

AMS subject classifications. 65M60, 65M20, 65M15, 35M25

DOI. 10.1137/090763378

1. Introduction.

A new technique to improve the accuracy of the spatial discretization of evolutionary convection-reaction-diffusion problems is studied. In this procedure a steady convection-reaction-diffusion problem, with data based on the computed standard Galerkin approximation, is solved at any time level where the output is desired. More precisely, we consider the problem

$$\begin{aligned} &u_t - \epsilon\Delta u + b\cdot\nabla u + cu = f \quad \text{in} \quad \Omega,\\ &u = 0 \quad \text{in} \quad \partial\Omega, \\ &u(0,x) = u_0(x) \quad \text{in} \quad \Omega, \end{aligned} \tag{1.1}$$

where Ω is a bounded open domain with smooth boundary in $\mathbb{R}^n$, $n = 1, 2, 3$, b and c are given functions, $\epsilon \geq 0$ is a constant diffusion coefficient, and u_0 a given initial data. Typically, in some applications, the size of the diffusion is much smaller than the size of the convective term and solutions develop sharp layers. In this case, it is well known that standard finite element methods perform poorly and develop nonphysical spurious oscillations, especially when using low order piecewise polynomials. Stabilization techniques (see, e.g., [23], [24], [25] and the references therein) are widely used in steady problems to suppress oscillations so that physically reasonable approximations can be obtained.

The method we study is as follows. Let us denote by $u_h(t)$, $t \in (0, T]$, the semidiscrete Galerkin finite element approximation to (1.1). Assume that the output is wanted at the final time T. Then we postprocess $u_h(T)$ to get a new approximation $\tilde u_h$ by solving numerically the steady problem of finding $v \in H_0^1(\Omega)$ such that $L(v) =$

*Received by the editors June 30, 2009; accepted for publication (in revised form) April 20, 2010; published electronically June 29, 2010. The research of the first and third authors was supported by the Spanish MEC under grant MTM2007-60528.
http://www.siam.org/journals/sinum/48-3/76337.html

†Departamento de Matemática Aplicada, Universidad de Valladolid, 47011 Valladolid, Spain (frutos@mac.uva.es).

‡Departamento de Matemática Aplicada II, Universidad de Sevilla, 41002 Sevilla, Spain (bosco@esi.us.es). This author's research was supported by the Spanish MEC under grant MTM2006-0847.

§Departamento de Matemáticas, Universidad Autónoma de Madrid, Instituto de Ciencias Matemáticas CSIC-UAM-UC3M-UCM, 28049 Madrid, Spain (julia.novo@uam.es).

$f - u_{h,t}$, where $L(v) = -\epsilon\Delta v + b \cdot \nabla v + cv$. This steady problem is solved using the stabilized streamline-upwind Petrov–Galerkin (SUPG) method [14], [4]. In this paper we show that the new approximation $\tilde{u}_h$ is essentially free of oscillations. This is most remarkable since the oscillation-free postprocessed approximation $\tilde{u}_h$ is computed at a fixed time T based on the highly oscillatory $u_{h,t}(T)$ (which have been obtained using the plain Galerkin method along the full time interval $(0, T]$).

This technique was first introduced in [11] for nonlinear convection-diffusion problems of evolution. It was proved in [11] that the new technique possesses a rate of convergence one unit higher up to a logarithmic term than that of the Galerkin method when ϵ is away from zero. In [12] the fully discrete case was also addressed. In this paper we carry out the error analysis of the spatial postprocessed semidiscretizations of (1.1) in the convection-dominated regime, that is, when ϵ tends to zero. We carry out the analysis when the SUPG method is used in the steady problem of the postprocessing step, although error bounds for other stabilized methods as the DWG (Douglas–Wang/Galerkin) or the GALS (Galerkin/least-squares) can be obtained in a similar way.

The new technique is an alternative for the discretization of time-dependent problems with stabilized methods. These have been studied in [24] (see also [2], [3], [5], [6] [8], [9], [15], [16], [17], [18], [19], [20], [21]). Thus, in the present paper, we examine the possibility of computing with a standard method (without stabilization techniques) until a selected time and then use a stabilized method in a single steady problem. This procedure simplifies the task of stabilizing in evolutionary convection-diffusion problems and has the advantage of being extensible (see [11]) to more involved nonlinear problems.

In this paper we obtain error bounds in the norm associated to the SUPG method that do not deteriorate when ϵ tends to zero. The rate of convergence we prove is suboptimal in the sense that it differs from the rate of convergence of the SUPG approximation to a steady problem by a factor of order $h^{1/2}$. However, we remark that, to our knowledge, error bounds for the SUPG method in the evolutionary case are scarce or simply not available. We refer to [21], where error bounds for the GALS method are obtained in the fully discrete case using the θ-method assuming that the stabilization parameter is $O(\Delta t)$, Δt being the time step; see also [5], where the limit case $\epsilon = 0$ is studied.

In practice, both the Galerkin u_h and postprocessed approximation $\tilde{u}_h$ cannot be computed exactly and, instead, approximations $U_h^n \approx u_h(t_n)$ and $\widetilde{U}_h^n \approx \tilde{u}(t_n)$ on time levels $t_0 < t_1 < \cdots < t_N = T$ are computed. In this paper we get error bounds for the error in the fully discrete postprocessed approximation $\widetilde{U}_h^n$. We prove that the temporal error of the fully discrete postprocessed approximation can be bounded in terms of the temporal error of the Galerkin approximation. The results are valid for any convergent time-stepping procedure. We include a refined analysis when the time integrator chosen is the popular midpoint rule.

We also obtain improved error bounds in the one-dimensional case using linear finite elements. For this case, we analyze the postprocessed method in both the coercive and noncoercive cases and obtain a quasi-optimal error bound in the $H^1(0, x_{N-1})$-norm (i.e., in the H^1-norm but excluding the last interval). The analysis is carried out for nonuniform meshes and for variable coefficients in the convective and reactive term (i.e., general functions b and c in (1.1)). Some numerical experiments are provided to show the reduction of the spurious oscillations that is achieved when using this postprocessing technique. The extension of the error analysis when ϵ tends to zero to the nonlinear problems considered in [11] will be the subject of future research.

The outline of the paper is as follows. In section 2 we introduce the notation and state some preliminaries. In section 3 we carry out the error analysis. First, we consider the general case and analyze both the semidiscrete in space and the fully discrete cases. Then the one-dimensional case for linear finite elements is considered. In order to simplify the paper the variable coefficient case is left for the appendix. Finally, in section 4 we present some numerical experiments.

2. Preliminaries and notation.

We will assume that b and c are sufficiently smooth functions of x and define

$$\mu(x) = \left(c - \frac{1}{2}\mathrm{div} b\right)(x), \quad \mu_0 = \inf_{x\in\Omega} \mu(x).$$

We assume that Ω is a domain with smooth $\mathcal{C}^r$ boundary. Let $\mathcal{T}_h = (K_i^h, \phi_i^h)_{i\in J_h}$, $h > 0$, be a family of shape-regular and quasi-uniform partitions of suitable domains Ω_h, where h is the maximum diameter of the elements $K_i^h \in \mathcal{T}_h$, and ϕ_i^h are the mappings of the reference simplex K_0 onto K_i^h. We denote

$$V_{h,r}(\Omega_h) = \left\{ v_h \in (C^0(\overline{\Omega_h}))^n \mid v_h \circ \phi_i^h \in (\mathbb{P}_{r-1}(K_0)) \ \forall K_i^h \in \mathcal{T}_h, v_h = 0 \text{ in } \partial\Omega_h \right\},$$

where $\mathbb{P}_{r-1}(K_0)$ is the space of polynomials of (total) degree less than or equal to $r-1$ over K_0. We notice that when only linear elements are used, we may also assume that Ω is a convex polygonal or polyhedral domain and $\Omega_h = \Omega$.

Let us denote by $A_h : V_{h,r} \to V_{h,r}$ the positive self-adjoint operator defined by

$$(A_h v_h, w_h) = (\nabla v_h, \nabla w_h) \quad \forall v_h, w_h \in V_{h,r},$$

where $(\cdot,\cdot)$ denotes the standard inner product in $L^2(\Omega)$. The standard $L^2(\Omega)$ orthogonal projection onto $V_{h,r}$ will be denoted by P_h. We will denote by $\pi_h u \in V_{h,r}$ the elliptic projection defined by

$$(\nabla \pi_h u, \nabla \varphi_h) = (\nabla u, \nabla \varphi_h) \quad \forall \varphi_h \in V_{h,r}.$$

Assuming that the meshes are quasi-uniform, the following inverse inequality holds for each $v_h \in V_{h,r}$ (see, e.g., [7, Theorem 3.2.6]):

$$\|v_h\|_{W^{m,q}(K)} \le C h^{l-m-d(\frac{1}{q'}-\frac{1}{q})} \|v_h\|_{W^{l,q'}(K)}, \tag{2.1}$$

where $0 \le l \le m \le 1$, $1 \le q' \le q \le \infty$, and K is an element in the partition $\mathcal{T}_h$.

The following bound holds for any $u \in H_0^1(\Omega) \cap H^r(\Omega)$

$$\|u - \pi_h u\|_0 + h\|u - \pi_h u\|_1 \le C h^r \|u\|_r. \tag{2.2}$$

Here and in the rest of the paper $\|\cdot\|_s$ denotes the norm of the Sobolev space $H^s(\Omega)$ or, if $s = 0$, the norm of $L^2(\Omega)$. We remark that for (2.2) to hold, either $\Omega = \Omega_h$ or $\delta(h) = \max\{\mathrm{dist}(x, \partial\Omega) \mid x \in \partial\Omega_h\} = O(h^{2(r-1)})$. Following an idea by Wahlbin [26] (see also [1]), we may assume that $\Omega_h \subset \Omega$, and functions in $V_{h,r}(\Omega_h)$ are extended by 0 to Ω, so that $V_{h,r}(\Omega_h) \subset H_0^1(\Omega)^n$.

We denote by $u_h : (0, T] \to V_{h,r}$ the spatial semidiscrete Galerkin approximation to (1.1) satisfying $u_h(0,\cdot) = u_h^0 \in V_{h,r}$ and

$$(u_{h,t}, \varphi_h) + \epsilon(\nabla u_h, \nabla \varphi_h) + (b \cdot \nabla u_h, \varphi_h) + (c u_h, \varphi_h) = (f, \varphi_h) \quad \forall \varphi_h \in V_{h,r}. \tag{2.3}$$

It is well known that the Galerkin approximation develops spurious oscillations in the advection-dominated regime. In this paper, we consider a procedure that is able to stabilize the Galerkin approximation at any time T. Let us fix any positive time T; we define the postprocessed approximation $\tilde u_h = \tilde u_h(T) \in V_{h,r}$ as the solution of the following stationary convection-reaction-diffusion problem:

$$\begin{aligned}\epsilon(\nabla \tilde u_h, \nabla\varphi_h) + (b\cdot\nabla\tilde u_h, \varphi_h) + (c\tilde u_h, \varphi_h) &= (f - u_{h,t}, \varphi_h)\\ &\quad + (f - u_{h,t} + \epsilon\Delta\tilde u_h - b\cdot\nabla\tilde u_h - c\tilde u_h, b\cdot\nabla\varphi_h)_h \quad \forall\varphi_h \in V_{h,r},\end{aligned}\tag{2.4}$$

where all time-dependent functions are evaluated at the fixed time $t = T$. Here and in what follows $(\cdot,\cdot)_h$ denotes the broken inner product

$$(f,g)_h = \sum_{K\in\mathcal{T}_h} \delta_K (f,g)_K,\tag{2.5}$$

δ_K being the stabilization parameter and $(\cdot,\cdot)_K$ the standard inner product in $L^2(K)$.

We denote by $\|\cdot\|_h$ its associated norm. Let us observe that we solve a stationary convection-diffusion problem with data based on the already computed Galerkin approximation using the SUPG method introduced by Hughes and Brooks [14], [4]. The new approximation belongs to the same finite element space as that of the Galerkin approximation.

We will denote by $a_{\mathrm{S}}(\cdot,\cdot)$ the bilinear form associated to the SUPG method, defined by

$$\begin{aligned}a_{\mathrm{S}}(v_h, w_h) &= \epsilon(\nabla v_h, \nabla w_h) + (b\cdot\nabla v_h, w_h) + (cv_h, w_h)\\ &\quad + (-\epsilon\Delta v_h + b\cdot\nabla v_h + cv_h, b\cdot\nabla w_h)_h \quad \forall v_h, w_h \in V_{h,r}.\end{aligned}\tag{2.6}$$

3. Error analysis.

3.1. General case.

3.1. General case. In this section we obtain error bounds for the method in which the constants do not deteriorate when ϵ tends to zero. In what follows, for simplicity we will assume that

$$u_h(0) = \pi_h u_0.$$

We notice, however, that as long as

$$\|u_h(0) - u(0)\|_0 + h\,\|u_h(0) - u(0)\|_1 \le Ch^r\,\|u(0)\|_r\,,$$

the results that follow below are still valid.

LEMMA 1. *Let u be the solution of* (1.1) *and let $\pi_h u$ be its elliptic projection. Let u_h be the Galerkin approximation* (2.3)*. Then, for $T > 0$, there exists a constant $C > 0$, independent of ϵ, such that the following bound holds:*

$$\max_{0\le t\le T} \|(u_h - \pi_h u)_t\|_0 \le Ch^{r-1}\big(\|u(0)\|_r + \max_{0\le t\le T}(\|u_t\|_r + \|u_{tt}\|_{r-1})\big).\tag{3.1}$$

Proof. Let us denote by $e_h = u_h - \pi_h u$. Let us call

$$g_1 = (\pi_h u_t - u_t) + c(\pi_h u - u), \quad g_2 = (b\cdot\nabla(\pi_h u - u)).$$

It is easy to see that for all $\varphi_h \in V_{h,r}$

$$(e_{h,t}, \varphi_h) + \epsilon(\nabla e_h, \nabla\varphi_h) + (b\cdot\nabla e_h, \varphi_h) + (ce_h, \varphi_h) = (-g_1 - g_2, \varphi_h),$$

which we can write as

$$e_{h,t} = -G_h e_h - P_h(g_1 + g_2), \tag{3.2}$$

where $G_h : V_{h,r} \to V_{h,r}$ is the operator defined by

$$G_h = \epsilon A_h + P_h(b \cdot \nabla + cI), \tag{3.3}$$

I being the identity operator. Since the solution of $y_{h,t} = -G_h y_h$ can be written as $y_h(t) = \exp(-tG_h)y_h(0)$, standard energy arguments show that

$$\|\exp(-tG_h)\|_0 \le \mathrm{e}^{-\mu_0 t}.$$

Taking derivatives with respect to t in (3.2) we have

$$e_{h,tt} = -G_h e_{h,t} - P_h\left(g_{1,t} + g_{2,t}\right),$$

and, thus,

$$e_{h,t}(t) = \exp(-tG_h)e_{h,t}(0) + \int_0^t \exp(-(t-s)G_h)P_h\left(g_{1,t} + g_{2,t}\right)\,\mathrm{d}s.$$

Since by applying (2.2) we have $\|(g_{1,t} + g_{2,t})\|_0 \le Ch^{r-1}(\|u_t\|_r + \|u_{tt}\|_{r-1})$, the proof will be finished if we show that $\|e_{h,t}(0)\|_0$ can be bounded by the right-hand side of (3.1). In view of (3.2) and applying (2.2) we have that

$$\|e_{h,t}(0)\|_0 \le \|G_h e_h(0)\|_0 + Ch^{r-1}(\|u(0)\|_r + \|u_t(0)\|_{r-1}),$$

and since we are assuming that $e_h(0) = 0$, the result follows. □

The following lemma states the coerciveness of the bilinear form associated to the SUPG method. The result is standard and can be found, for example, in [25, Lemma 10.3].

LEMMA 2. *Let us assume that* $\mu_0 > 0$ *and, further,*

$$\delta_K \|c\|_{K,\infty}^2 \le \frac{\mu_0}{2}, \quad \epsilon\delta_K \le \frac{h_K^2}{2C^2}, \tag{3.4}$$

where C *is the constant in the inverse inequality* (2.1)*. Then the bilinear form* $a_{\mathrm{S}}(\cdot,\cdot)$ *associated to the SUPG method satisfies*

$$a_{\mathrm{S}}(u_h, u_h) \ge \frac{1}{2}\Bigg(\epsilon\|\nabla u_h\|_0^2 + \sum_{K\in\mathcal{T}_h} \delta_K \|b\cdot\nabla u_h\|_{0,K}^2 + \|\mu^{1/2} u_h\|_0^2\Bigg). \tag{3.5}$$

For linear elements the second condition in (3.4) *can be omitted.*

In what follows we will assume $\delta_K = \delta_0 h_K / \|b\|_{\infty,K}$ whenever $Pe_K = \|b\|_{\infty,K} h_K / (2\epsilon) > 1$ and $\delta_K = \delta_1 h_K^2/\epsilon$ for $Pe_K \le 1$, where δ_0 and δ_1 are user-chosen positive constants. No precise general formula for an optimal value of the stabilization parameter δ_K is known; see [25, Remark 10.4].

Let us denote by $w_h \in V_{h,r}$ the SUPG approximation to the steady problem

$$-\epsilon\Delta v + b \cdot \nabla v + cv = g, \tag{3.6}$$

subject to homogeneous Dirichlet boundary conditions. Then w_h satisfies

$$a_S(w_h, \varphi_h) = (g, \varphi_h) + (g, b\cdot\nabla\varphi_h)_h \quad \forall \varphi_h \in V_{h,r}.$$

Under the conditions of Lemma 2 it is easy to obtain the following (well-known) error bound for the SUPG approximation (see, for example, [25, Theorem 10.5], [22]):

$$\left(\epsilon\|\nabla(v-w_h)\|_0^2+\sum_{K\in\mathcal{T}_h}\delta_k\|b\cdot\nabla(v-w_h)\|_{0,K}^2+\|\mu^{1/2}(v-w_h)\|_0^2\right)^{1/2} \le C(\epsilon^{1/2}+h^{1/2})h^{r-1}\|v\|_r. \tag{3.7}$$

We note that the solution u of the evolutionary problem (1.1) is also the solution of (3.6) for the particular case in which $g=f-u_t$.

THEOREM 1. *Let us fix $T>0$, let u be the solution of* (1.1), *and let $\tilde u_h\in V_{h,r}$ be the postprocessed approximation* (2.4). *Assume that $\mu_0>0$ and condition* (3.4) *is satisfied. Then there exists a constant $C>0$ that does not depend on ϵ such that the following bound holds:*

$$\begin{aligned}&\left(\epsilon\|\nabla(u-\tilde u_h)\|_0^2+\sum_{K\in\mathcal{T}_h}\delta_k\|b\cdot\nabla(u-\tilde u_h)\|_{0,K}^2+\|\mu^{1/2}(u-\tilde u_h)\|_0^2\right)^{1/2}\\ &\quad\le Ch^{r-1}(\|u\|_r+\|u_t\|_r+\|u_{tt}\|_{r-1})+C(\epsilon^{1/2}+h^{1/2})h^{r-1}\|u\|_r.\end{aligned} \tag{3.8}$$

Proof. Let us denote by w_h the SUPG approximation to the steady problem (3.6) with $g=f-u_t$. It is easy to see that this approximation satisfies

$$\begin{aligned}\epsilon(\nabla w_h,\nabla\varphi_h)+(b\cdot\nabla w_h,\varphi_h)+(cw_h,\varphi_h)&=(f-u_t,\varphi_h)\\ &\quad+(f-u_t+\epsilon\Delta w_h-b\cdot\nabla w_h-cw_h,b\cdot\nabla\varphi_h)_h\quad\forall\varphi_h\in V_{h,r}.\end{aligned} \tag{3.9}$$

Let us decompose $u-\tilde u_h=(u-w_h)+(w_h-\tilde u_h)$. To bound the first term we apply (3.7). Let us now get a bound for the second term. Let us denote

$$\tilde e_h=\tilde u_h-w_h.$$

Subtracting (3.9) from (2.4) we get

$$\begin{aligned}&\epsilon(\nabla\tilde e_h,\nabla\varphi_h)+(b\cdot\nabla\tilde e_h,\varphi_h)+(c\tilde e_h,\varphi_h)+(-\epsilon\Delta\tilde e_h+b\cdot\nabla\tilde e_h+c\tilde e_h,b\cdot\nabla\varphi_h)_h\\ &\qquad=(u_t-u_{h,t},\varphi_h)+\sum_{K\in\mathcal{T}_h}\delta_K(u_t-u_{h,t},b\cdot\nabla\varphi_h)_K.\end{aligned}$$

Taking $\varphi_h=\tilde e_h$ and applying (3.5) we get

$$\begin{aligned}\epsilon\|\nabla\tilde e_h\|_0^2+\|b\cdot\nabla\tilde e_h\|_h^2+\|\mu^{1/2}\tilde e_h\|_0^2&\le 2(u_t-u_{h,t},\tilde e_h)+2(u_t-u_{h,t},b\cdot\nabla\tilde e_h)_h\\ &\le\frac{2}{\mu_0}\|u_t-u_{h,t}\|_0^2+\frac12\|\mu^{1/2}\tilde e_h\|_0^2\\ &\quad+2\|u_t-u_{h,t}\|_h^2+\frac12\|b\cdot\nabla\tilde e_h\|_h^2.\end{aligned}$$

Now notice that for both the convection-dominated or the diffusion-dominated regime we have $\|u_t-u_{h,t}\|_h^2\le Ch\|u_t-u_{h,t}\|_0^2$. Thus, by writing

$$u_t-u_{h,t}=(u_t-\pi_hu_t)+(\pi_hu_t-u_{h,t})$$

and applying (2.2) and Lemma 1 the proof is concluded. □

Remark 1. The case $\mu_0 = 0$, which includes the interesting case $c = 0$ and $\nabla \cdot b = 0$, can be treated in the following way. The change of variables $v = u\mathrm{e}^{-\alpha t}$, with $\alpha > 0$ arbitrary, transforms the equation into one of the same type satisfying $\mu_0 = \alpha > 0$. We can apply the numerical method to this equation and then transform back to the original variables. The result of Theorem 1 is still valid with an $\mathcal{O}(1)$ multiplicative constant in the error bound if, for example, we choose $\alpha = 1/T$. A similar comment applies if $\mu_0 < 0$. Alternatively, one can argue as in [25, Remark 7.3].

Remark 2. Let us observe that while in the SUPG method for the evolutionary problem (1.1) the stabilization terms are computed along all the time integration, in our method we carry out the time integration using the standard Galerkin method and compute the stabilization terms only at a fixed time. This procedure, besides being easier to implement, can be easily extended to more complicated nonlinear problems (see [11]). On the contrary, the extension of the SUPG method to nonlinear evolutionary problems is not trivial (see, for example, the discussion in [13]).

If we compare the error bound (3.8) that we have obtained for the new method with the error bound (3.5) for the SUPG approximation to a steady problem, we observe a difference of half an order in the rate of convergence. However, to the best of our knowledge, optimal error bounds for the semidiscrete SUPG method in the evolutionary case are not available, even in the linear case. In view of the results obtained in section 3.3 for the one-dimensional case using linear elements where we prove that the error in the energy norm in the last interval is only $O(h^{1/2})$ (i.e., the same rate of convergence that provides the bound (3.8) for linear elements ($r = 2$) in the L^2-norm of the streamline derivative) we think that the error bound (3.8) cannot be improved in general.

3.2. Fully discrete case. Throughout this section we will assume that $\mu_0 > 0$. The general case can be treated as explained in Remark 1. We consider the case in which approximations $U_h^n \approx u_h(t_n)$ on time levels $0 = t_0 < t_1 < \cdots < t_N = T$ are computed by means of any convergent time integrator. Given an approximation $d_t^* U_h^n$ to $u_{h,t}(t_n)$ the fully discrete postprocessed approximation $\widetilde{U}_h^n \in V_{h,r}$ is obtained as the solution of the following problem:

$$a_S(\widetilde{U}_h^n, \varphi_h) = (f - \mathrm{d}_t^* U_h^n, \varphi_h) + (f - \mathrm{d}_t^* U_h^n, b \cdot \nabla \varphi_h)_h \quad \forall \varphi_h \in V_{h,r}. \tag{3.10}$$

For the approximation $d_t^* U_h^n$ to the derivative $u_{h,t}(t_n)$ we propose

$$d_t^* U_h^n = -\epsilon A_h U_h^n - P_h \left(b \cdot \nabla U_h^n + c U_h^n\right) + P_h f(t_n). \tag{3.11}$$

This is the approximation used in the numerical experiments of section 4 and the same that has been considered in [10], [12]. To estimate the error $u(t_n) - \widetilde{U}_h^n$ we write

$$u(t_n) - \widetilde{U}_h^n = (u(t_n) - \tilde{u}_h(t_n)) + \tilde{e}_n,$$

where $\tilde{e}_n = \tilde{u}_h(t_n) - \widetilde{U}_h^n$. The first term on the right-hand side above is the error in the spatial semidiscrete postprocessed approximation that has been bounded in Theorem 1. Next, we analyze the time discretization error $\tilde{e}_n$. We estimate the size of $\tilde{e}_n$ in terms of $e_n = u_h(t_n) - U_h^n$, the temporal error of the fully discrete Galerkin approximation. We carry out the error analysis for any convergent time-stepping procedure satisfying

$$\lim_{k\to 0} \max_{0\le n\le N} \|e_n\|_0 = 0, \quad \lim_{k\to 0} \max_{0\le n\le N} \|e_n\|_1 = 0,$$

where $k = \max\{k_{n-1} = t_n - t_{n-1} \mid 1 \le n \le N\}$.

THEOREM 2. *Let us fix $t_n > 0$, and let $\tilde{e}_n = \tilde{u}(t_n) - \tilde{U}_h^n$ be the temporal error of the fully discrete postprocessed approximation. Then there exists a constant $C > 0$ that does not depend on ϵ such that the following bound holds:*

$$\begin{aligned}(3.12)\qquad &\left(\epsilon\|\nabla\tilde{e}_n\|_0^2 + \sum_{K\in\mathcal{T}_h}\delta_k\|b\cdot\nabla\tilde{e}_n\|_{0,K}^2 + \|\mu^{1/2}\tilde{e}_n\|_0^2\right)^{1/2}\\ &\qquad\le C\left(\epsilon h^{-1}\|e_n\|_1 + \|b\|_\infty\|e_n\|_1 + \|c\|_\infty\|e_n\|_0\right),\end{aligned}$$

where $e_n = u_h(t_n) - U_h^n$ is the temporal error of the Galerkin approximation.

Proof. The proof is similar to the proof of Theorem 1. Subtracting (3.10) from (2.4) we get

$$a_S(\tilde{e}_n,\varphi_h) = (d_t^* U_h^n - u_{h,t},\varphi_h) + (d_t^* U_h^n - u_{h,t}, b\cdot\nabla\varphi_h)_h \quad \forall\varphi_h\in V_{h,r}.$$

Taking $\varphi_h = \tilde{e}_h$ and applying (3.4) we get

$$\begin{aligned}\epsilon\|\nabla\tilde{e}_n\|_0^2 + \|b\cdot\nabla\tilde{e}_n\|_h^2 + \|\mu^{1/2}\tilde{e}_n\|_0^2 &\le 2(d_t^* U_h^n - u_{h,t}(t_n),\tilde{e}_n)\\ &\quad +2(d_t^* U_h^n - u_{h,t}(t_n), b\cdot\nabla\tilde{e}_n)_h\\ &\le \frac{2}{\mu_0}\|d_t^* U_h^n - u_{h,t}(t_n)\|_0^2 + \frac{1}{2}\|\mu^{1/2}\tilde{e}_n\|_0^2\\ &\quad + 2\|d_t^* U_h^n - u_{h,t}(t_n)\|_h^2 + \frac{1}{2}\|b\cdot\nabla\tilde{e}_n\|_h^2.\end{aligned}$$

Now, since we have $\|\cdot\|_h \le Ch\|\cdot\|_0$ it is sufficient to get a bound for $\|d_t^* U_h^n - u_{h,t}(t_n)\|_0$ to conclude the proof. Taking into account that

$$u_{h,t} = -\epsilon A_h u_h(t_n) - P_h\left(b\cdot\nabla u_h(t_n) + cu_h(t_n)\right) + P_h f(t_n)$$

and using (3.11) we get

$$(3.13)\qquad \|d_t^* U_h^n - u_{h,t}(t_n)\|_0 \le \|\epsilon A_h e_n + P_h\left(b\cdot\nabla e_n + ce_n\right)\|_0.$$

Applying the inverse inequality (2.1) we finally arrive at

$$(3.14)\qquad \|d_t^* U_h^n - u_{h,t}(t_n)\|_0 \le C\epsilon h^{-1}\|e_n\|_1 + \|b\|_\infty\|e_n\|_1 + \|c\|_\infty\|e_n\|_0. \qquad \square$$

In view of bound (3.12) we deduce that the temporal error of the fully discrete postprocessed approximation can be bounded in terms of the temporal error of the Galerkin approximation. Analogous results were obtained in [10] for an earlier postprocessed technique applied to the nonlinear evolutionary Navier–Stokes equations, and in [12], where the same postprocessing technique was applied to nonlinear evolutionary convection-diffusion equations in the diffusive regime. In [10] and [12] we prove that the temporal errors of the Galerkin finite element approximations satisfy

$$\|e_n\|_0 + \|A_h e_n\|_0 \le Ck^{l_0}, \quad 1\le n\le N,$$

where k is the time step, $l_0 = 1$ for the backward Euler method, and $l_0 = 2$ for the two-step backward differentiation formula. Let us point out that, as was already observed in [10] and [12], better bounds are obtained in the diffusion-dominated regime if $\epsilon\|A_h e_n\|_0$ is used in (3.12) instead of $\epsilon h^{-1}\|e_n\|_1$.

Let us now consider the case in which the fully discrete approximations U_h^n are obtained by means of the trapezoidal rule. Then

$$\frac{U_h^{n+1}-U_h^n}{k_n}+G_h\left(\frac{U_h^{n+1}+U_h^n}{2}\right)=\frac{f(t_{n+1})+f(t_n)}{2},\quad n=0,\dots,N-1,$$

where, as in (3.3), $G_h=\epsilon A_h+P_h(b\cdot\nabla+cI)$. We have that

$$\|v_h\|_G^2=(G_hv_h,v_h)\geq\epsilon\|v_h\|_1^2+\mu_0\|v_h\|_0^2. \tag{3.15}$$

In the following theorem we bound the temporal error of the fully discrete postprocessed approximation for this particular case. Let us remark, though, that the result can be similarly proved for other A-stable time integrators.

THEOREM 3. *Let us fix $t_n>0$, let us assume that we integrate in time with the trapezoidal rule, and let $\tilde{e}_n=\tilde{u}(t_n)-\tilde{U}_h^n$ be the temporal error of the fully discrete postprocessed approximation. Then there exists a constant $C>0$ that does not depend on ϵ such that the following bound holds:*

$$\begin{aligned}&\left(\epsilon\|\nabla\tilde{e}_n\|_0^2+\sum_{K\in\mathcal{T}_h}\delta_k\|b\cdot\nabla\tilde{e}_n\|_{0,K}^2+\|\mu^{1/2}\tilde{e}_n\|_0^2\right)^{1/2}\\&\qquad\leq Ck^2\left(\int_0^{t_n}\left(\left\|\frac{\mathrm{d}^3}{\mathrm{d}t^3}f(t)\right\|_0^2+\left\|\frac{\mathrm{d}^4}{\mathrm{d}t^4}u_h(t)\right\|_0^2\right)\,\mathrm{d}t\right)^{1/2}.\end{aligned}\tag{3.16}$$

Proof. Reasoning as in the proof of Theorem 2 but using (3.13) instead of (3.14) we arrive at

$$\left(\epsilon\|\nabla\tilde{e}_n\|_0^2+\sum_{K\in\mathcal{T}_h}\delta_k\|b\cdot\nabla\tilde{e}_n\|_{0,K}^2+\|\mu^{1/2}\tilde{e}_n\|_0^2\right)^{1/2}\leq C\|G_he_n\|_0.$$

In the rest of this proof we get a bound for $\|G_he_n\|_0$. For simplicity we will assume that

$$e_0=0. \tag{3.17}$$

We first observe that for $n=0,1,\dots,N-1$

$$\frac{1}{k_n}\left(e_{n+1}-e_n,\varphi_h\right)+\big(G_he_{n+1/2},\varphi_h\big)=(\tau_n,\varphi_h)\quad\forall\varphi_h\in V_{h,r}, \tag{3.18}$$

where here and in the rest of the proof

$$e_{n+1/2}=\frac{e_{n+1}+e_n}{2},\quad n=0,1,\dots,N-1,$$

and τ_n is the truncation error, that is,

$$\tau_n=\frac{u_h(t_{n+1})-u_h(t_n)}{k_n}+G_h\left(\frac{u_h(t_{n+1})+u_h(t_n)}{2}\right)-\frac{f(t_{n+1})+f(t_n)}{2}.$$

A simple calculation shows that

$$\begin{aligned}\tau_n=&\frac{1}{k_n}\int_{t_n}^{t_n+1}u_{h,t}(t)\,\mathrm{d}t-\left(\frac{u_{h,t}(t_{n+1})+u_{h,t}(t_n)}{2}\right)\\=&-\frac{1}{k_n}\int_{t_n}^{t_{n+1}}\frac{(t_{n+1}-t)(t-t_n)}{2}\frac{\mathrm{d}^3}{\mathrm{d}t^3}u_h(t)\,\mathrm{d}t.\end{aligned}\tag{3.19}$$

Let us now take $\varphi_h = G_h^* G_h e_{n+1/2}$ in (3.18) to get

$$\frac{1}{2k_n}(\|G_h e_{n+1}\|_0^2 - \|G_h e_n\|_0^2) + (G_h e_{n+1/2}, G_h^* G_h e_{n+1/2}) = (G_h \tau_n, G_h e_{n+1/2}). \tag{3.20}$$

Taking into account that $(G_h e_{n+1/2}, G_h^* G_h e_{n+1/2}) = (G_h G_h e_{n+1/2}, G_h e_{n+1/2})$ and using (3.15) we have

$$(G_h e_{n+1/2}, G_h^* G_h e_{n+1/2}) = \big\|G_h e_{n+1/2}\big\|_G^2 \ge \mu_0 \big\|G_h e_{n+1/2}\big\|_0^2.$$

Thus, from (3.20) it follows that

$$\frac{1}{k_n}(\|G_h e_{n+1}\|_0^2 - \|G_h e_n\|_0^2) + \mu_0 \big\|G_h e_{n+1/2}\big\|_0^2 \le \frac{1}{\mu_0}\,\|G_h \tau_n\|_0^2\,.$$

Multiplying by k_n in the inequality above, summing from $j = 0$ up to $j = n - 1$, and recalling (3.17), we have

$$\|G_h e_n\|_0^2 + \mu_0 \sum_{j=0}^{n-1} k_j \big\|G_h e_{j+1/2}\big\|_0^2 \le \frac{1}{\mu_0} \sum_{j=0}^{n-1} k_j\,\|G_h \tau_j\|_0^2\,.$$

To conclude, there remains only to bound the right-hand side above. Applying Hölder's inequality in the expression of τ_n in (3.19) we obtain

$$\|G_h \tau_n\|_0 \le \frac{k_n^{3/2}}{2\sqrt{30}} \left(\int_{t_n}^{t_{n+1}} \left\| \frac{\mathrm{d}^3}{\mathrm{d}t^3} G_h u_h(t) \right\|_0^2 \,\mathrm{d}t \right)^{1/2}.$$

Finally, since $u_{h,t} + G_h u_h = P_h f$ and, consequently, $\frac{\mathrm{d}^4}{\mathrm{d}t^4} u_h + G_h \frac{\mathrm{d}^3}{\mathrm{d}t^3} u_h = P_h \frac{\mathrm{d}^3}{\mathrm{d}t^3} f$, we finally arrive at

$$\|G_h e_n\|_0 \le C k^2 \left(\int_0^{t_n} \left(\left\| \frac{\mathrm{d}^3}{\mathrm{d}t^3} f(t) \right\|_0^2 + \left\| \frac{\mathrm{d}^4}{\mathrm{d}t^4} u_h(t) \right\|_0^2 \right) \mathrm{d}t \right)^{1/2}. \qquad \square$$

Remark 3. Observe that the time derivatives of the Galerkin approximation u_h (see (3.16)) are (up to $O(h)$ terms) of the same size as those of u, since the arguments used in the proof of Lemma 1 can be iterated with further time derivatives. It must be pointed out, however, that unless some compatibility conditions are satisfied at $t = 0$ further time derivatives of u blow up when $t \to 0$. In this case, and at the price of a much more cumbersome analysis, bounds similar to (3.16) can be proved if negative powers of t are allowed to appear on the right-hand side (see, e.g., [10]).

3.3. One-dimensional case. For problems in one spatial dimension, we now show that the results in Theorem 1 can be improved if linear elements are used. For this purpose, we consider the problem

$$\begin{aligned} &u_t - \epsilon u_{xx} + b u_x = f, \quad 0 < x < 1, \quad t > 0, \\ &u(0,t) = u(1,t) = 0, \quad u(x,0) = u_0, \end{aligned} \tag{3.21}$$

where, in the present section and for simplicity, we consider b, a positive constant. We denote by u_h the Galerkin approximation based on linear finite elements. We consider partitions $0 = x_0 < x_1 < \cdots < x_N = 1$ of $[0,1]$, and we will denote

$$h_j = x_j - x_{j-1}, \quad j = 1, \ldots, N, \quad h = \max_{1\le j\le N} h_j, \quad h_0 = \min_{1\le j\le N} h_j.$$

We will assume that the meshes are quasi-uniform so that for certain $\lambda > 1$,

$$\frac{h}{h_0} < \lambda, \quad h > 0.$$

The following sets will appear in several estimates below:

$$I_n = (x_0, x_1) \cup (x_n, x_{n+1}), \quad n = 1, \ldots, N-1. \tag{3.22}$$

As before, we denote by V_h the finite element space and by $\tilde{u}_h$ the postprocessed approximation based on linear finite elements that satisfies, for all $\varphi_h \in V_h$,

$$\begin{aligned} \epsilon(\tilde{u}_{h,x}, \varphi_{h,x}) + (b\tilde{u}_{h,x}, \varphi_h) + (b\tilde{u}_{h,x}, b\varphi_{h,x})_h = (f - u_{h,t}, \varphi_h) \\ + (f - u_{h,t}, b\varphi_{h,x})_h. \end{aligned} \tag{3.23}$$

In what follows, for simplicity, we will assume that we are in the convection-dominated regime, so that

$$\frac{bh_0}{2\epsilon} > 1, \tag{3.24}$$

and, accordingly, we set the stabilization parameters δ_K in (2.5) to be

$$\delta_K = \delta_j = h_j/(2b) \quad \text{for all elements } K = [x_{j-1}, x_j], \quad j = 1, \ldots, N. \tag{3.25}$$

With this choice of stabilization parameters, we have the relation

$$\frac{h_0}{2b} \|v_h\|_0^2 < \|v_h\|_h^2 \le \frac{h}{2b} \|v_h\|_0^2 \quad \forall v_h \in V_h. \tag{3.26}$$

In the rest of the paper we denote by φ_j, $j = 1, \ldots, N-1$, the nodal basis functions of the linear finite element space,

$$\varphi_j(x) = \begin{cases} (x - x_{j-1})/h_j, & x \in [x_{j-1}, x_j], \\ (x_j - x)/h_{j+1}, & x \in [x_j, x_{j+1}], \\ 0, & x \notin [x_{j-1}, x_{j+1}]. \end{cases}$$

Given $v_h \in V_h$, for the sake of brevity we will write $v_j = v_h(x_j)$, so that

$$v_h(x) = \sum_{j=1}^{N-1} v_j \varphi_j(x),$$

and we also denote

$$Dv_j = \frac{v_j - v_{j-1}}{h_j}, \quad j = 1, \ldots, N,$$

where we assume $v_0 = v_N = 0$. Observe that $v_{h,x} = Dv_j$ in (x_{j-1}, x_j).

We now state and prove two auxiliary results.

LEMMA 3. *Assume that* (3.24) *holds and for* $v_h \in V_h$ *and* $j = 1, \ldots, N-1$, *let*

$$s_j = \epsilon(v_{h,x}, \varphi_{j,x}) + (bv_{h,x}, \varphi_j) + (bv_{h,x}, b\varphi_{j,x})_h$$

and $S_j = s_1 + \cdots + s_j$. *Then the following bounds hold:*

$$\|v_{h,x}\|_{0,(x_0,x_{N-1})} \le \frac{2}{bh_0^{1/2}}\Big(\sqrt{s_1^2+\cdots+s_{N-1}^2} + \epsilon\,|Dv_N|\Big), \tag{3.27}$$

$$b\,|Dv_N| \le \frac{1}{h_N}\sqrt{s_1^2+\cdots+s_{N-1}^2} + \frac{1}{h_N}\,|S_{N-1}|, \tag{3.28}$$

$$b\,|v_n| \le |S_n| + 2\frac{\epsilon}{h_0^{1/2}}\,\|v_{h,x}\|_{0,(x_0,x_{n+1})} \tag{3.29}$$

for $n = 1,\dots,N-1$.

Proof. A simple calculation shows that

$$s_j = (\epsilon + bh_j)Dv_j - \epsilon Dv_{j+1}, \quad j = 1,\dots,N-1. \tag{3.30}$$

Since

$$\|v_{h,x}\|_{L^2(x_{l-1},x_m)} = \left(\sum_{j=l}^{m} h_j\,|Dv_j|^2\right)^{1/2} \tag{3.31}$$

from (3.30) it follows that

$$\left(\frac{\epsilon}{h^{1/2}} + bh_0^{1/2}\right)\|v_{h,x}\|_{L^2(x_0,x_{N-1})} \le \frac{\epsilon}{h_0^{1/2}}\,\|v_{h,x}\|_{L^2(x_1,x_{N-1})} + \epsilon\,|De_N|\,. \\ + \left(\sum_{j=1}^{N-1}|s_j|^2\right)^{1/2},$$

and since, according to (3.24), $\epsilon/h_0^{1/2} \le bh_0^{1/2}/2$, the bound (3.27) follows.

Since $v_0 = 0$, summation in (3.30) from $j = 1$ to $j = n$ gives

$$bv_n = \epsilon(Dv_{n+1} - Dv_1) + S_n, \quad n = 1,\dots,N-1, \tag{3.32}$$

and since $v_N = 0$, we can write $Dv_N = -v_{N-1}/h_N$, so that thanks to (3.32) we have

$$bDv_N = -\frac{\epsilon}{h_N}(Dv_N - Dv_1) - \frac{1}{h}S_{N-1},$$

so that

$$\left(b + \frac{\epsilon}{h_N}\right)|Dv_N| \le \frac{\epsilon}{h_N}\,|Dv_1| + \frac{1}{h_N}\,|S_{N-1}|\,. \tag{3.33}$$

Noticing that

$$|Dv_j| \le h_0^{-1/2}\,\|v_{h,x}\|_{L^2(x_0,x_j)}\,, \quad j = 1,\dots,N, \tag{3.34}$$

so that, in particular, we have $|Dv_1| \le h_0^{-1/2}\,\|v_{h,x}\|_{L^2(x_0,x_{N-1})}$, from (3.27) and (3.33) it follows that

$$\left(b + \frac{\epsilon}{h_N}\right)|Dv_N| \le \frac{2\epsilon^2}{bh_Nh_0}\,|Dv_N| + \frac{2\epsilon}{bh_0h_N}\sqrt{s_1^2+\cdots+s_{N-1}^2} + \frac{1}{h_N}\,|S_{N-1}|\,.$$

Recalling that (3.24) holds, we have

$$\frac{\epsilon}{h_N} - \frac{2\epsilon^2}{b h_N h_0} = \frac{\epsilon}{h_N}\left(1 - \frac{2\epsilon}{b h_0}\right) > 0,$$

and the bound (3.28) follows.

Finally, (3.29) follows from (3.32) and (3.34). □

LEMMA 4. *For δ_K as specified in* (3.25)*, the following bounds hold for $v, w \in L^2(0,1)$ and $n = 1, \ldots, N-1$:*

$$\left(\sum_{j=1}^{n} |(v,\varphi_j)|^2 + |(w, b\varphi_{j,x})_h|^2\right)^{1/2} \le \frac{2\sqrt{6}}{3}\Big(\|v\|_{L^2(x_0,x_{n+1})} + \|w\|_{L^2(x_0,x_{n+1})}\Big) h^{1/2}, \tag{3.35}$$

$$\left|\sum_{j=1}^{n} (v,\varphi_j) + (w, b\varphi_{j,x})_h\right| \le \sqrt{2}\big(\|v\|_{L^2(x_0,x_{n+1})} + \|w\|_{L^2(I_n)}\, h^{1/2}\big), \tag{3.36}$$

$$\left|\sum_{j=1}^{n} (v,\varphi_j) + (w, b\varphi_{j,x})_h\right| \le \sqrt{2}\Big(\frac{\lambda^{1/2}\,\|V\|_{L^2(I_n)}}{h} + \|w\|_{L^2(I_n)}\Big) h^{1/2}, \tag{3.37}$$

where I_n are the sets defined in (3.22)*, and $V_x = v$.*

Proof. Since the support of φ_j is $[x_{j-1}, x_{j+1}]$ and $\|\varphi_j\|_0^2 = (h_{j-1} + h_j)/3$, by direct application of Hölder's inequality we have

$$|(v,\varphi_j)| \le \frac{\sqrt{6}}{3} h^{1/2}\, \|v\|_{L^2(x_{j-1},x_{j+1})}\,, \tag{3.38}$$

and since $\delta_j = h_j/(2b)$, for any $w \in L^2(0,1)$,

$$(w, b\varphi_{j,x})_h = \frac{1}{2}\int_{x_{j-1}}^{x_j} w(x)\,\mathrm{d}x - \frac{1}{2}\int_{x_j}^{x_{j+1}} w(x)\,\mathrm{d}x, \tag{3.39}$$

so that

$$|(w, b\varphi_{j,x})_h| \le \frac{\sqrt{2}}{2} h^{1/2}\, \|w\|_{L^2(x_{j-1},x_{j+1})}\,. \tag{3.40}$$

Hence, the bound (3.35) follows easily. To prove (3.37) we first notice that

$$\sum_{j=1}^{n} (v,\varphi_j) + (w, b\varphi_{j,x})_h = (V_x, \varphi_1 + \cdots + \varphi_n) + (w, b(\varphi_1 + \cdots + \varphi_n)_x)_h,$$

and integrating by parts the first term on the right-hand side above,

$$\sum_{j=1}^{n} (v,\varphi_j) + (w, b(\varphi_j)_x)_h = -(V, (\varphi_1 + \cdots + \varphi_n)_x) + (w, b(\varphi_1 + \cdots + \varphi_n)_x)_h.$$

Now observe that $\varphi_{j-1} + \varphi_j = 1$ in (x_{j-1}, x_j), so that $(\varphi_1 + \cdots + \varphi_n)_x$ is null in $[x_1, x_n]$, and takes value $1/h_1$ in (x_0, x_1) and $-1/h_{n+1}$ in (x_n, x_{n+1}), respectively. Thus, the bound (3.37) follows. Finally, (3.36) is obtained by the same argument as (3.37) omitting the integration by parts performed above. □

Next, we get an error bound for the postprocessed approximation (3.23) that improves the bound obtained for the general case in Theorem 1. We consider the steady problem

$$-\epsilon u_{xx} + b u_x = g, \quad u(0) = u(1) = 0, \tag{3.41}$$

with $g = f - u_t$, and its SUPG approximation w_h defined as the solution of

$$\epsilon(w_{h,x}, \varphi_{h,x}) + (b w_{h,x}, \varphi_h) + (b w_{h,x}, b\varphi_{h,x})_h = (g, \varphi_h) + (g, b\varphi_{h,x})_h \tag{3.42}$$

for all $\varphi_h \in V_h$. Subtracting (3.23) from (3.42), for the difference $\tilde{e}_h = w_h - \tilde{u}_h$ we obtain

$$\epsilon(\tilde{e}_{h,x}, \varphi_{h,x}) + (b\tilde{e}_{h,x}, \varphi_h) + (b\tilde{e}_{h,x}, b\varphi_{h,x})_h = (u_t - u_{h,t}, \varphi_h) + (u_t - u_{h,t}, b\varphi_{h,x})_h.$$

Let us first observe that we can apply (2.2) and Lemma 1 to get the bound

$$\|u_t - u_{h,t}\|_0 \le Ch\Big(\|u(0)\|_2 + \max_{0\le t\le T}(\|u_t(t)\|_2 + \|u_{tt}(t)\|_1)\Big). \tag{3.43}$$

To simplify the notation let us denote

$$\mathcal{K} = \|u(0)\|_2 + \max_{0\le t\le T}(\|u_t(t)\|_2 + \|u_{tt}(t)\|_1). \tag{3.44}$$

Also, we denote

$$\tilde{r}_j = (u_t - u_{h,t}, \varphi_j) + (u_t - u_{h,t}, b\varphi_{j,x})_h$$

and

$$\tilde{R}_j = \tilde{r}_1 + \cdots + \tilde{r}_j$$

and apply Lemma 4 to estimate them. More precisely, applying (3.35), (3.36), and (3.43), we have that there exists a constant $C > 0$ such that

$$\sqrt{|\tilde{r}_1|^2 + \cdots + |\tilde{r}_n|^2} \le C\mathcal{K}h^{3/2}, \tag{3.45}$$

$$\big|\tilde{R}_n\big| \le C\mathcal{K}h. \tag{3.46}$$

Notice that since the antiderivative of $u_t - u_{h,t}$ does not enjoy a decay rate better than $O(h)$; the bound (3.37) applied to $v = w = u_t - u_{h,t}$ allows only for an $O(h^{1/2})$ rate of decay, which is worse than that in (3.46).

THEOREM 4. *Let u be the solution of* (3.21) *and let* (3.25) *hold. Then there exist a positive constant C that does not depend on ϵ such that the postprocessed approximation $\tilde{u}_h$ solution of* (3.23) *satisfies the following bounds:*

$$\|u - \tilde{u}_h\|_{1,[0,x_{N-1}]} \le C\left(h + \frac{\epsilon}{h^{1/2}}\right)(\|u\|_2 + \mathcal{K}), \tag{3.47}$$

$$\|u - \tilde{u}_h\|_{1,[x_{N-1},x_N]} \le Ch^{1/2}(\|u\|_2 + \mathcal{K}), \tag{3.48}$$

where $\mathcal{K}$ is the constant in (3.44).

Proof. Let us decompose the error $u - \tilde{u}_h = (u - w_h) + (w_h - \tilde{u}_h)$. To bound the first term we apply

$$\|w_h - u\|_1 \le C(h + \epsilon)\|u\|_2. \tag{3.49}$$

To get (3.49) we first observe that although we now have $\mu_0 = 0$, Lemma 2 can still be applied, without requiring condition (3.4) to be satisfied and changing (3.5) by

$$a_{\rm S}(u_h, u_h) \geq \left(\epsilon\|\nabla u_h\|_0^2 + \sum_{K\in\mathcal{T}_h} \delta_K \|b\cdot\nabla u_h\|_{0,K}^2\right).$$

Then, reasoning as in the proof of (3.7), one gets (3.49); see [22].

Next we concentrate on the bound for the second term, for which we will apply Lemma 3 together with (3.45) and (3.46) above. Indeed, observe that for $v_h = \tilde e_h$ in Lemma 3, we have $s_j = \tilde r_j$ and $S_j = \tilde R_j$. Thus, applying (3.27), (3.45) and recalling that $h \leq \lambda h_0$, we have

$$\|\tilde e_{h,x}\|_{L^2(x_0,x_{N-1})} \leq \frac{C}{b}\lambda^{1/2}\mathcal{K}h + \frac{\epsilon}{bh_0^{1/2}}\left|D\tilde e_N\right| . \tag{3.50}$$

Also, applying (3.28) and in view of (3.45) and (3.46) we have

$$|D\tilde e_N| \leq \frac{C}{b}\lambda\mathcal{K}(h^{1/2}+1). \tag{3.51}$$

Then the bound (3.47) follows from (3.50) and (3.51). The bound (3.48) follows from (3.51) and the fact that $\|\tilde e_{h,x}\|_{L^2(x_{N-1},x_N)} = h_N^{1/2}\left|D\tilde e_N\right|$. □

Remark 4. Let us observe that in view of (3.47), whenever $\epsilon < h$, the bound of the $H^1(0,x_{N-1})$-norm of the error improves the bound (3.8) of Theorem 1. Let us also observe that, in the proof of Lemma 3, for $v_h = \tilde e_h$ and, consequently, $s_j = \tilde r_j$, (3.30) becomes $\tilde r_j = (\epsilon + bh_j)D\tilde e_j - \epsilon D\tilde e_{j+1}$ for $j = 1,\ldots,N-1$, so that in the limit of the convection-dominated regime ($\epsilon = 0$) we get

$$\tilde e_j - \tilde e_{j-1} = \frac{1}{b}(u_t - u_{h,t}, \varphi_j) + (u_t - u_{h,t}, \varphi_{j,x})_h, \quad j = 1,\ldots,N-1,$$

from which

$$|\tilde e_j - \tilde e_{j-1}| \leq \frac{2}{b}\|u_t - u_{h,t}\|_{\infty,[x_{j-1},x_{j+1}]}h, \quad j = 1,\ldots,N-1. \tag{3.52}$$

Since the SUPG approximation w_h is nonoscillatory, the size of $|\tilde e_j - \tilde e_{j-1}|$ gives a measure of the size of the oscillations in the postprocessed approximation. These, as the bound (3.52) shows, can be expected to be considerably smaller than those of the Galerkin approximation, since they are of the size of the local $L^\infty[x_{j-1},x_{j+1}]$-norm of $u_t - u_{h,t}$ times the size of the Galerkin mesh h.

Not only is the result of Theorem 4 valid for the advection-diffusion equation (3.21), but it can also be extended to equations with a reactive term. Let us consider the equation

$$\begin{aligned} &u_t - \epsilon u_{xx} + bu_x + cu = f, \quad 0 < x < 1, \quad t > 0,\\ &u(0,t) = u(1,t) = 0, \quad u(x,0) = u_0, \end{aligned} \tag{3.53}$$

with b and c positive constants. We denote as before by u_h the Galerkin finite element approximation and by $\tilde u_h$ the postprocessed approximation that satisfies

$$\begin{aligned} \epsilon(\tilde u_{h,x}, \varphi_{h,x}) + (b\tilde u_{h,x} + c\tilde u_h, \varphi_h) + (b\tilde u_{h,x} + c\tilde u_h, b\varphi_{h,x})_h &\\ = (f - u_{h,t}, \varphi_h) + (f - u_{h,t}, b\varphi_{h,x}) \quad \forall\varphi_h \in V_h.& \end{aligned} \tag{3.54}$$

In the next theorem we obtain the same error bound of Theorem 4 for this approximation. Since we are in the coercive case, Lemma 2 can be applied. The first condition on (3.4) must hold, whereas the second one is not required since we are dealing with linear elements. A simple calculation shows that this happens if $h \le b/c$.

THEOREM 5. *Let u be the solution of* (3.53) *and $\tilde u_h$ the postprocessed approximation* (3.54)*, and set δ_K as in* (3.25)*. Then there exists a constant C that does not depend on ϵ such that the following bounds hold for $h \le b/c$:*

$$\|u-\tilde u_h\|_{1,[0,x_{N-1}]} \le C\left(h+\frac{\epsilon}{h^{1/2}}\right)(\|u\|_2+\mathcal{K}), \tag{3.55}$$

$$\|u-\tilde u_h\|_{1,[x_{N-1},x_N]} \le Ch^{1/2}(\|u\|_2+\mathcal{K}), \tag{3.56}$$

where $\mathcal{K}$ is the constant in (3.44).

Proof. Let us denote by w_h the SUPG approximation to the steady convection-reaction-diffusion equation

$$-\epsilon u_{xx}+bu_x+cu=g, \quad u(0)=u(1)=0,$$

for $g=f-u_t$. Then the function w_h solves

$$\begin{aligned}\epsilon(w_{h,x},\varphi_{h,x})+(bw_{h,x}+cw_h,\varphi_h)+(bw_{h,x}+cw_h,b\varphi_{h,x})_h&\\ =(g,\varphi_h)+(g,b\varphi_{h,x})_h.&\end{aligned} \tag{3.57}$$

Applying (3.7) we obtain

$$\|u-w_h\|_0 \le Ch^{3/2}\|u\|_2, \quad \|u-w_h\|_1 \le Ch\|u\|_2.$$

To bound the error in the postprocessed approximation we decompose as usual $u-\tilde u_h=(u-w_h)+(w_h-\tilde u_h)$. Let us obtain a bound for the second term. We denote by $\tilde e_h=\tilde u_h-w_h$. Taking into account the equivalence (3.26) and applying Theorem 1, we obtain

$$\|\tilde e_h\|_0 \le C\mathcal{K}h, \quad \|\tilde e_{h,x}\|_0 \le C\mathcal{K}h^{1/2}. \tag{3.58}$$

Let us first observe that from (3.58) we get (3.56) and as a consequence $|De_N| \le C$. To prove (3.55) we will apply Lemmas 3 and 4. Notice that subtracting (3.57) from (3.54) we get

$$\begin{aligned}\epsilon(\tilde e_{h,x},\varphi_{h,x})+(b\tilde e_{h,x}+c\tilde e_h,\varphi_h)+(b\tilde e_{h,x}+c\tilde e_h,b\varphi_{h,x})_h&\\ =(u_t-u_{h,t},\varphi_h)+(u_t-u_{h,t},b\varphi_{h,x})_h.&\end{aligned}$$

Thus, taking $v_h=\tilde e_h$ in Lemma 3, we have that $s_j=\tilde r_j$ and $S_j=\tilde r_1+\cdots+\tilde r_j$, where

$$\tilde r_j=((u_t-u_{h,t})+c\tilde e_h,\varphi_j)+((u_t-u_{h,t})+c\tilde e_h,b\varphi_{j,x})_h, \quad j=1,\ldots,N-1. \tag{3.59}$$

Observe that we have an estimate of $\|\tilde e_h\|_0$ in (3.58). Also, as a consequence of (2.2) and Lemma 1, we have $\|u_t-u_{h,t}\|_0 \le C\mathcal{K}h$. Thus, applying Lemma 4 (estimate (3.35)) to the residuals in (3.59), we have that, as in Theorem 4, (3.45) holds. Then, applying (3.27), we reach (3.55). □

4. Numerical experiments. Next, we show a simple numerical experiment that illustrates the behavior of the postprocessed approximation. We consider (3.21) with forcing term $f = 1$, convection coefficient $b = 1$, and initial condition $u_0(x) = 0$. We compute the Galerkin approximation based on linear finite elements over a uniform partition of $[0, 1]$ of size $h = 1/N$ at time $T = 0.6$. For the time integration we use the implicit Euler method with fixed time step $k = 0.001$. To get the postprocessed approximation $\tilde{u}_h$ we solve (3.23). On the left of Figure 4.1 we have represented the Galerkin (solid line) and postprocessed (dashed line) approximations for $\epsilon = 1e - 4$ and $N = 80$. We can observe that the postprocessing step annihilates the spurious oscillations of the Galerkin approximation. On the right of Figure 4.1 we have represented the Galerkin time derivative $u_{h,t}$. The bound (3.52) indicates that the difference between two values in the postprocessed approximation is bounded in terms of the local L^∞-norm of the error in the Galerkin time derivative $u_{h,t}$. Indeed, if we look at a zoom of Figure 4.1 (see Figure 4.2) we can observe that the postprocessing step does not annihilate completely the Galerkin oscillations, although it considerably

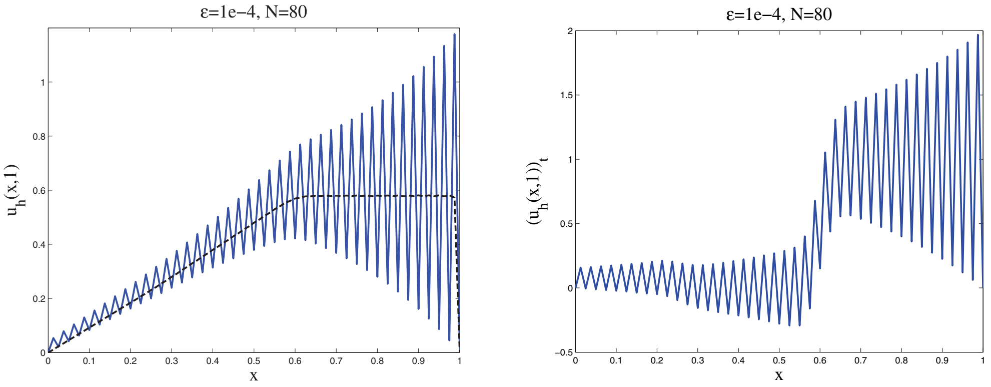

FIG. 4.1. *On the left: Galerkin and postprocessed approximations. On the right: Galerkin time derivative.*

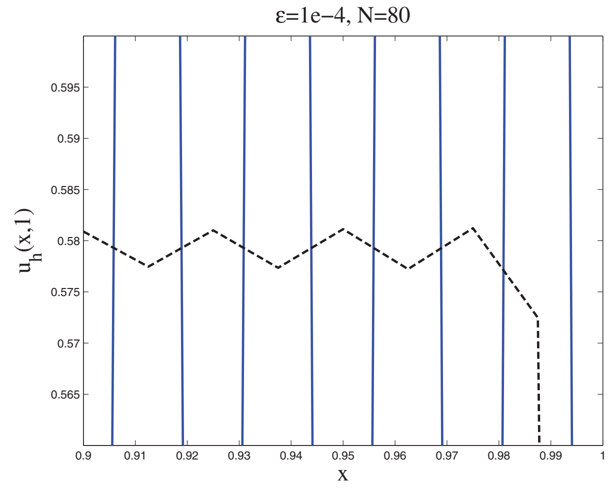

FIG. 4.2. *Zoom of Figure* 4.1 *(left).*

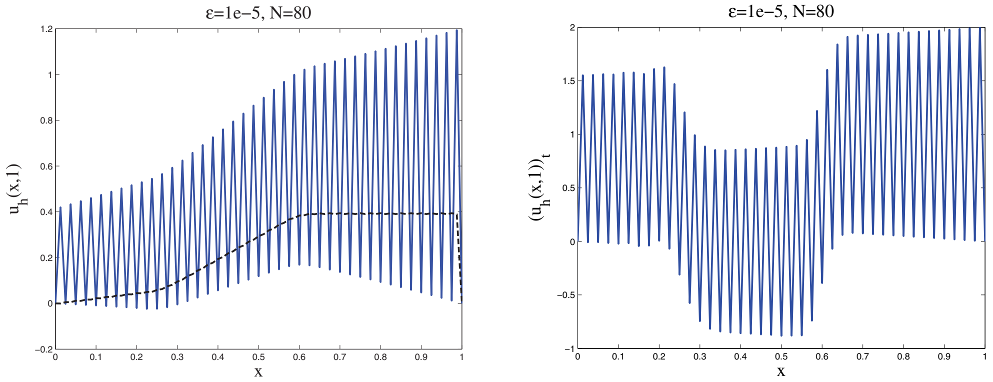

FIG. 4.3. *On the left: Galerkin and postprocessed approximations. On the right: Galerkin time derivative.*

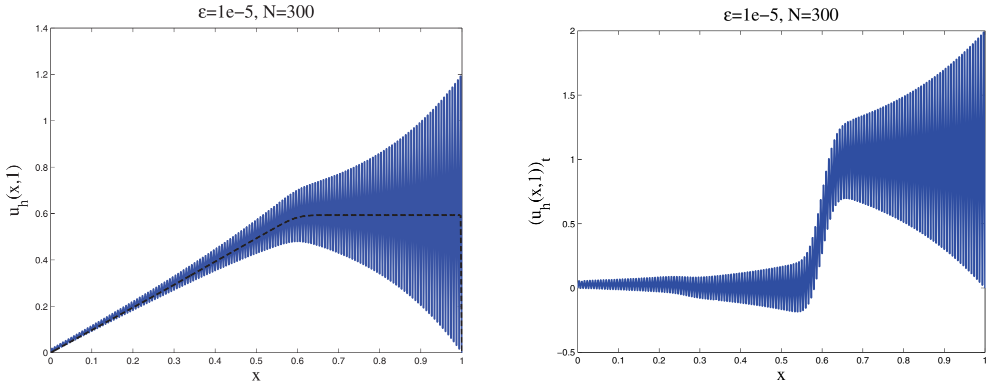

FIG. 4.4. *On the left: Galerkin and postprocessed approximations. On the right: Galerkin time derivative.*

diminishes them. To check the good behavior of our method when ϵ tends to zero we have computed the Galerkin and postprocessed approximations with $N = 80$ for $\epsilon = 1e - 5$. In Figure 4.3 we have plotted the approximations on the left and the Galerkin time derivative on the right. We can observe in the figure that although the postprocessed approximation annihilates again the Galerkin oscillations, it is not accurate enough since the Galerkin approximation used in the postprocessing step (3.23) is completely inaccurate; observe the picture of the Galerkin time derivative on the right of Figure 4.3. This lack of accuracy can be solved by computing the Galerkin approximation over a refined mesh. In Figure 4.4 we show the results obtained using a partition of $[0, 1]$ into 300 subintervals. Now, even though the Galerkin approximation is still completely contaminated, the postprocessing step is able to produce an accurate and oscillation-free approximation.

The lack of accuracy of $\tilde{u}_h$ when the mesh is not fine enough, which is observed in Figure 4.3, can be a drawback of the method in practice. However, we can still take advantage of the nonoscillatory character of the postprocessed approximation. The main idea is to use the difference $\eta_h = \tilde{u}_h - u_h$ between the postprocessed oscillation-free approximation $\tilde{u}_h$ and the polluted Galerkin approximation u_h as an a posteriori indicator, at each element of the mesh, of the oscillations presented in the

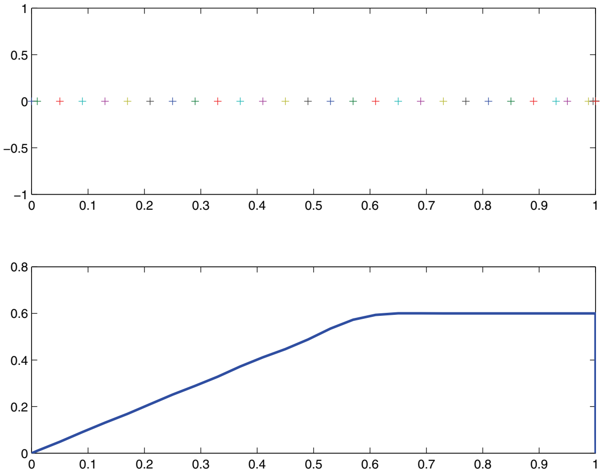

FIG. 4.5. *Approximation obtained using the adaptive algorithm for $\epsilon = 1e-6$.*

Galerkin approximation. Using this indicator at each time step, we can detect the oscillations developed in the Galerkin approximation and consequently locally refine the mesh before these oscillations become excessively large and globally pollute the approximation. The adaptive procedure we now present provides a wider application of the postprocessing technique since it produces accurate and oscillation-free Galerkin approximations with a very small number of degrees of freedom.

- Choose an initial subdivision of the interval $[0,1]$.
- Compute the Galerkin approximation at the first time step.
- Compute the postprocessed approximation.
- Compute the error indicator as the difference between the postprocessed and the Galerkin approximations: $\eta_h = \tilde{u}_h - u_h$.
- For every element $I_j = [x_{j-1}, x_j]$ if the difference $|\eta_h(x_j) - \eta_h(x_{j-1})|$ is greater than a given tolerance TOL_1, halve the interval I_j. If the difference is less than a given tolerance $\mathrm{TOL}_2 < \mathrm{TOL}_1$, suppress the point x_{j-1} whenever the new interval does not exceed a maximum prescribed size. Interpolate the approximation and use it as initial condition for the next time step.
- Continue with the procedure until the final time T.

We now show a numerical experiment to illustrate the behavior of our algorithm. We consider the same experiment as before with a smaller value of ϵ, more precisely $\epsilon = 1e-6$, for which, in view of Figure 4.3, we cannot expect an accurate postprocessed approximation. The initial mesh for this experiment has 100 nodes and the maximum h is set to 0.04. The parameters TOL_1 and TOL_2 were set to $\mathrm{TOL}_1 = 0.01$ and $\mathrm{TOL}_2 = \mathrm{TOL}_1/100$. In Figure 4.5 we show the approximation obtained at the final time $T = 0.6$ (bottom) and the final mesh (top). Our algorithm ends with only 45 nodes from which 19 lie on the interval $[0.9, 1]$ and, as we can observe in the figure, produces an excellent approximation in which the boundary layer is perfectly solved. The extension of this adaptive procedure to more than one spatial dimension as well as to nonlinear problems will be the subject of future research.

Next, we show a numerical experiment in a two-dimensional problem. Let us consider the equation

$$u_t - \epsilon \Delta u + u_x + u_y = f$$

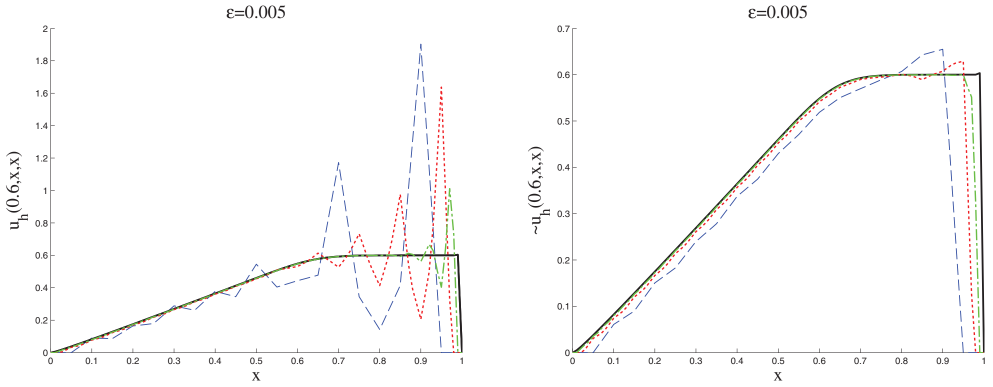

FIG. 4.6. *Sections with the plane $y = x$. Exact solution shown as solid line. On the left, Galerkin approximations. On the right, postprocessed approximations.*

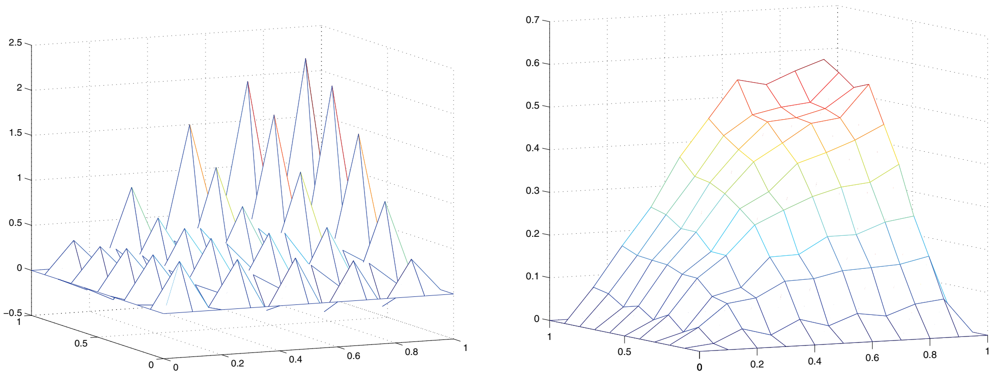

FIG. 4.7. *On the left, Galerkin approximation for $h = 1/10$. On the right, postprocessed approximation.* $\epsilon = 0.001$.

in the domain $\Omega = [-1,1] \times [-1,1]$ subject to homogeneous Dirichlet boundary conditions. We take $f = 1$ and as initial condition $u_0(x,y) = 0$. Let us consider regular triangulations of Ω induced by the set of nodes $(i/N, j/N)$, $0 \le i, j \le N$, where $N = 1/h$ is an integer. We use linear finite elements. The final time chosen is $T = 0.6$. All the experiments are carried out using MATLAB. For the time integration we use the midpoint rule with fixed time step. To compare the methods a reference approximation was computed with the Galerkin method on a very fine mesh and with sufficiently small time steps. In Figure 4.6 we have represented the sections of the Galerkin and postprocessed approximations along the plane $y = x$ for $\epsilon = 0.005$. The exact solution is plotted using a solid line. On the left, we plot the Galerkin approximations for $h = 1/10, 1/20$, and $1/40$ using dashed, dotted, and dash-dotted lines, respectively. The same lines are used on the right for the postprocessed approximations. It can be observed, in agreement with the experiments shown before in one spatial dimension, that the postprocessing step considerably reduces the spurious oscillations for all the values of h in the figure. For the last value $h = 1/40$, the postprocessed approximation does not oscillate at all and matches very precisely the section of the exact solution.

To observe the behavior of our method when ϵ decreases, we have represented in Figures 4.7 and 4.8 the Galerkin (left) and postprocessed (right) approximations

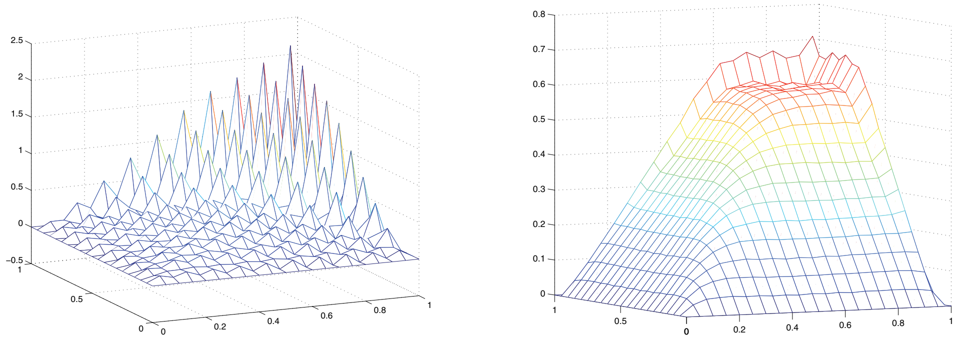

Fig. 4.8. *On the left, Galerkin approximation for* $h = 1/20$. *On the right, postprocessed approximation.* $\epsilon = 0.001$.

for $\epsilon = 0.001$. In Figure 4.7 we represent the approximations for $h = 1/10$ and in Figure 4.8 for $h = 1/20$. We can observe that although the Galerkin approximation is completely contaminated by spurious oscillations all over the whole domain Ω, the postprocessed method provides quite accurate approximations with a coarse mesh of only $N = 10$ or $N = 20$ nodes for each variable. The postprocessed approximation with $N = 10$ still has some small oscillations away from the boundary layers, while in the case $N = 20$ the small oscillations remain only in the neighborhood of the boundary layer. Of course, the oscillations can be completely annihilated by increasing the number of degrees of freedom. However, the aim of these figures is to show that from a “completely wrong” Galerkin approximation this postprocessing procedure can recover enough information to compute quite accurate approximations.

5. Appendix: Variable coefficients.

We analyze the one-dimensional case allowing for b to depend on x. We denote

$$b_M = \max_{0\le x\le 1} b(x), \quad b_0 = \min_{0\le x\le 1} b(x),$$

and we will assume that $b_0 > 0$.

We will follow the results on the constant coefficient case, commenting on the differences. We start by noticing that the postprocessed approximation satisfies (3.23), and the SUPG approximation w_h to the solution u of (3.41) satisfies (3.42). We now set

$$\delta_K = \delta_j = h_j/(2\underline{b}_j) \quad \text{for all elements } K = [x_{j-1}, x_j], \quad j = 1, \dots, N, \tag{5.1}$$

where

$$\underline{b}_j = \frac{1}{2}\left(\int_{x_{j-1}}^{x_j} b^2(x)\,\mathrm{d}x\right) \Big/ \left(\int_{x_{j-1}}^{x_j} b(x)\varphi_{j-1}(x)\,\mathrm{d}x\right), \quad j = 1, \dots, N. \tag{5.2}$$

With this choice of the parameter δ_K the following relation holds:

$$(bv_{h,x}, \varphi_j) + (bv_{h,x}, b\varphi_{j,x})_h = b_j(v_h(x_j) - v_h(x_{j-1})), \quad j = 1, \dots, N-1, \tag{5.3}$$

where

$$b_j = \frac{1}{h_j}\int_{x_{j-1}}^{x_j} b(x)\,\mathrm{d}x.$$

Remark 5. The choice of δ_K here is not restrictive. With other choices, like, for example, $\delta_K = \delta_j = h_j/(2\,\|b\|_{L^\infty(x_{j-1},x_j)})$, the right-hand side of (5.3) should be replaced by

$$\tilde{b}_j(v_h(x_j) - v_h(x_{j-1})) - \varepsilon_{j+1}(v_h(x_{j+1}) - v_h(x_j)),$$

with $|b_j - \tilde{b}_j| = O(h_j^q)$ and $|\varepsilon_{j+1}| = O(h_{j+1}^q)$ for some $q \geq 1$. This makes the analysis much more cumbersome and lengthy, but the results are essentially those we state below.

Observe also that by the mean value theorem, $b_j = b(\xi_j)$ for some $\xi_j \in (x_{j-1}, x_j)$, so that

$$b_j \geq b_0, \quad j = 1, \dots, N.$$

Also, Hölder's inequality and the mean value theorem shows that $b_0^2/b_M < \underline{b}_j < b_M^2/b_0$. Consequently, for $j = 1, \dots, N$,

$$\frac{b_0}{2b_M^2}h_0 \leq \frac{b_0}{2b_M^2}h_j \leq \delta_j \leq \frac{b_M}{2b_0^2}h_j \leq \frac{b_M}{2b_0^2}h,$$

so that the following relations follow:

$$\frac{b_0}{2b_M^2}h_0\,\|v_h\|_0^2 < \|v_h\|_h^2 \leq \frac{b_M}{2b_0^2}h\,\|v_h\|_0^2 \quad \forall v_h \in V_h.$$

As in previous sections, we restrict ourselves to the convection-dominated regime, so that

$$\frac{b_0 h_0}{2\epsilon} > 1, \tag{5.4}$$

but we will need to further assume

$$\frac{b_0 h_0}{2\epsilon} > 2\exp\biggl(\frac{1}{b_0}\int_0^1 |b'(x)|\ \mathrm{d}x\biggr). \tag{5.5}$$

Indeed, the bound that will be needed is

$$\frac{b_0 h_0}{2\epsilon} > 2\exp\biggl(\frac{1}{b_0}\sum_{j=1}^{N-1} |b_{j+1} - b_j|\biggr), \tag{5.6}$$

but since, as commented above, $b_j = b(\xi_j)$ for some $\xi_j \in (x_{j-1}, x_j)$, we have

$$|b_{j+1} - b_j| = \left|\int_{\xi_j}^{\xi_{j+1}} b'(x)\,\mathrm{d}x\right| \leq \int_{\xi_j}^{\xi_{j+1}} |b'(x)|\ \mathrm{d}x,$$

and then (5.5) implies (5.6).

We now state and prove the version of Lemma 3 for the variable coefficient case. In its proof, the following discrete Gronwall lemma will be needed, which can be easily proved by an induction argument.

LEMMA 5. *Let* $(y_n)_{n=1}^\infty$ *and* $(\gamma_n)_{n=1}^\infty$ *be sequences of nonnegative numbers and* $\sigma_0 > 0$ *such that*

$$y_n \leq \gamma_0 + \sum_{j=1}^{n-1} \gamma_j y_j, \quad n = 1, 2, \dots.$$

Then

$$y_n \le \gamma_0 \exp(\gamma_1 + \cdots + \gamma_{n-1}), \quad n = 1, 2, \ldots.$$

LEMMA 6. *Assume that* (5.4) *and* (5.6) *hold, and for* $v_h \in V_h$ *and* $j = 1, \ldots, N-1$, *let*

$$s_j = \epsilon(v_{h,x}, \varphi_{j,x}) + (bv_{h,x}, \varphi_j) + (bv_{h,x}, b\varphi_{j,x})_h$$

and $S_j = s_1 + \cdots + s_j$. *Then the following bounds hold:*

$$\|v_{h,x}\|_{0,(x_0,x_{N-1})} \le \frac{2}{b_0 h_0^{1/2}}\Big(\sqrt{s_1^2 + \cdots + s_{N-1}^2} + \epsilon\,|Dv_N|\Big), \tag{5.7}$$

$$b_0\,|Dv_N| \le \frac{1}{h_N}\sqrt{s_1^2 + \cdots + s_{N-1}^2} + \frac{\mathrm{e}^{\sigma_{N-1}}}{h_N}\max_{1\le j\le N-1}|S_j|\,, \tag{5.8}$$

$$b_n\,|v_n| \le \mathrm{e}^{\sigma_{n-1}}\max_{1\le j\le n}|S_j| + 2\frac{\epsilon \mathrm{e}^{\sigma_{n-1}}}{h_0^{1/2}}\,\|(v_h)_x\|_{0,(x_0,x_{n+1})} \tag{5.9}$$

for $n = 1, \ldots, N-1$, *where*

$$\sigma_n = \frac{1}{b_0}\sum_{j=1}^{n-1}|b_{j+1} - b_j|\,, \quad n = 1, \ldots, N-1.$$

Proof. In view of (5.3) we have

$$s_j = (\epsilon + b_j h_j)Dv_j - \epsilon Dv_{j+1}, \quad j = 1, \ldots, N-1. \tag{5.10}$$

Then, recalling (3.31), from (5.10) it follows that

$$\Big(\frac{\epsilon}{h^{1/2}} + b_0 h_0^{1/2}\Big)\|v_{h,x}\|_{L^2(x_0,x_{N-1})} \le \frac{\epsilon}{h_0^{1/2}}\|v_{h,x}\|_{L^2(x_1,x_{N-1})} + \epsilon\,|De_N| + \Big(\sum_{j=1}^{N-1}|s_j|^2\Big)^{1/2},$$

and since according to (5.4), $\epsilon/h_0^{1/2} \le b_0 h_0^{1/2}/2$, the bound (5.7) follows.

We will now prove (5.8). Summation in (5.10) from $j = 1$ to $j = n$ gives

$$\sum_{j=1}^{n} b_j(v_j - v_{j-1}) = \epsilon(Dv_{n+1} - Dv_1) + S_n, \quad n = 1, \ldots, N-1. \tag{5.11}$$

Summation by parts allows us to write (recall that $v_0 = 0$)

$$\sum_{j=1}^{n} b_j(v_j - v_{j-1}) = b_n v_n - \sum_{j=1}^{n-1}(b_{j+1} - b_j)v_j.$$

Thus, we can rewrite (5.11) as

$$b_n v_n = \sum_{j=1}^{n-1}(b_{j+1} - b_j)v_j + \epsilon(Dv_{n+1} - Dv_1) + S_n, \quad n = 1, \ldots, N-1. \tag{5.12}$$

Applying Lemma 5 we have

$$|v_n| \le \frac{1}{b_n}\exp\biggl(\frac{1}{b_0}\sum_{j=1}^{n-1}|b_{j+1}-b_j|\biggr)\max_{1\le j\le n}\bigl|\epsilon(Dv_{j+1}-Dv_1)+S_j\bigr|. \tag{5.13}$$

On the other hand, since $v_N=0$ we have $Dv_N=-v_{N-1}/h_N$, so that setting $n=N-1$ in (5.12) and multiplying by $-h_N^{-1}$ we have

$$b_{N-1}Dv_N = -\frac{\epsilon}{h_N}(Dv_N-Dv_1)-\frac{1}{h_N}S_{N-1}-\frac{1}{h_N}\sum_{j=1}^{N-2}(b_{j+1}-b_j)v_j,$$

that is,

$$\Bigl(b_{N-1}+\frac{\epsilon}{h_N}\Bigr)Dv_N = \frac{\epsilon}{h_N}Dv_1-\frac{1}{h_N}S_{N-1}-\frac{1}{h_N}\sum_{j=1}^{N-2}(b_{j+1}-b_j)v_j. \tag{5.14}$$

In order to finish the proof of (5.8) we need to express the first and third terms on the right-hand side above in terms of the previous bounds. We start with the third term. Observe that

$$|b_{j+1}-b_j|\,|v_j| \le \frac{|b_{j+1}-b_j|}{b_0}|\,|b_jv_j| = (\sigma_{j+1}-\sigma_j)|\,|b_jv_j|\,.$$

Then, in view of (5.13), we have

$$\biggl|\sum_{j=1}^{N-2}(b_{j+1}-b_j)v_j\biggr| \le \biggl(\sum_{j=1}^{N-2}(\sigma_{j+1}-\sigma_j)\mathrm{e}^{\sigma_j}\biggr)\max_{1\le j\le N-2}\bigl|\epsilon(Dv_{j+1}-Dv_1)+S_j\bigr|.$$

Notice that the sum above is a lower Riemann sum of the exponential function and, thus, smaller than the corresponding integral. Hence,

$$\biggl|\sum_{j=1}^{N-2}(b_{j+1}-b_j)v_j\biggr| \le (\mathrm{e}^{\sigma_{N-1}}-1)\max_{1\le j\le N-2}\bigl|\epsilon(Dv_{j+1}-Dv_1)+S_j\bigr|.$$

Recalling (3.34) we can write

$$\max_{1\le j\le N-2}\bigl|\epsilon(Dv_{j+1}-Dv_1)+S_j\bigr| \le \frac{2\epsilon}{h_0^{1/2}}\,\|v_{h,x}\|_{L^2(x_0,x_{N-1})}+\max_{1\le j\le N-2}\bigl|S_j\bigr| \tag{5.15}$$

and, thus,

$$\biggl|\sum_{j=1}^{N-2}(b_{j+1}-b_j)v_j\biggr| \le (\mathrm{e}^{\sigma_{N-1}}-1)\left(\frac{2\epsilon}{h_0^{1/2}}\,\|v_{h,x}\|_{L^2(x_0,x_{N-1})}+\max_{1\le j\le N-2}\bigl|S_j\bigr|\right).$$

Then, going back to (5.14), we get

$$\Bigl(b_{N-1}+\frac{\epsilon}{h_N}\Bigr)|Dv_N| \le \frac{2\epsilon\mathrm{e}^{\sigma_{N-1}}}{h_Nh_0^{1/2}}\|v_{h,x}\|_{L^2(x_0,x_{N-1})}+\frac{\mathrm{e}^{\sigma_{N-1}}}{h_N}\max_{1\le j\le N-1}|S_j|\,,$$

and applying (5.7) we obtain

$$\Big(b_{N-1}+\frac{\epsilon}{h_N}\Big)\,|Dv_N| \le \frac{4\epsilon \mathrm{e}^{\sigma_{N-1}}}{h_N b_0 h_0}\Big(\sqrt{s_1^2+\cdots+s_{N-1}^2}+\epsilon\,|Dv_N|\Big)\frac{\mathrm{e}^{\sigma_{N-1}}}{h_N}\max_{1\le j\le N-1}|S_j|\,,$$

so that, rearranging terms, we have

$$\Big(b_{N-1}+\frac{\epsilon}{h_N}\Big(1-4\frac{\epsilon \mathrm{e}^{\sigma_{N-1}}}{b_0 h_0}\Big)\Big)\,|Dv_N| \le \frac{4\epsilon \mathrm{e}^{\sigma_{N-1}}}{h_N b_0 h_0}\sqrt{s_1^2+\cdots+s_{N-1}^2}\,\frac{\mathrm{e}^{\sigma_{N-1}}}{h_N}\max_{1\le j\le N-1}|S_j|\,.$$

Since we are assuming that (5.6) holds, the left-hand side above can be bounded below by $b_{N-1}\,|Dv_N|$, and the coefficient multiplying the square root on the right-hand side can be bounded by $1/h_N$. Hence, (5.8) follows.

Finally, (5.9) follows from (5.13) and (5.15). □

LEMMA 7. *For δ_K as specified in* (5.1), (5.2)*, the bounds* (3.35–3.37) *hold for v, $w \in L^2(0,1)$ and $n = 1,\dots,N-1$.*

Proof. We follow the proof of Lemma 4. Observe that (3.35) follows from (3.38) and (3.40). In the present case (3.38) obviously holds, so that we are left to show that (3.40) also holds in the present case. To do this we notice that instead of (3.39) we now have

$$(w, b(\varphi_j)_x)_h = \frac{1}{2\underline{b}_j}\int_{x_{j-1}}^{x_j} b(x)w(x)\,\mathrm{d}x - \frac{1}{2\underline{b}_{j+1}}\int_{x_j}^{x_{j+1}} b(x)w(x)\,\mathrm{d}x.$$

However, in view of the expression of $\underline{b}_j$ in (5.2), applying Hölder's inequality we have

$$\frac{1}{\underline{b}_j} \le \frac{\|\varphi_{j-1}\|_{L^2(x_{j-1},x_j)}}{\|b\|_{L^2(x_{j-1},x_j)}} = \frac{h_j^{1/2}}{\sqrt{3}\,\|b\|_{L^2(x_{j-1},x_j)}},$$

so that applying Hölder's inequality again it follows that

$$\frac{1}{2\underline{b}_j}\left|\int_{x_{j-1}}^{x_j} b(x)w(x)\,\mathrm{d}x\right| \le \frac{\sqrt{3}}{6}h_j^{1/2}\,\|w\|_{L^2(x_{j-1},x_j)}\,,\quad j=1,\dots,N, \tag{5.16}$$

and thus $|(w,b(\varphi_j)_x)_h| \le (\sqrt{6}/6)h^{1/2}\,\|w\|_{L^2(x_{j-1},x_{j+1})}$, which implies (3.40). Thus, (3.35) follows. Also, thanks to (5.16), the same arguments used to prove (3.36) and (3.37) in Lemma 4 also apply in the present case. □

LEMMA 8. *Let u be the solution of* (3.41) *and let* (5.4) *and* (5.6) *hold. There exists a positive constant C independent of ϵ such that the SUPG approximation w_h satisfies the error estimate* (3.49).

Proof. The proof can be obtained by reasoning as in the proof of Theorem 4, using Lemmas 6 and 7; see [22]. □

For the postprocessed approximation, we state Theorem 6 below, which is the variable coefficient version of Theorem 4. It can be proved by following the arguments in the proof of Theorem 4, provided that references to Lemmas 3, 4 and (3.49) are replaced by references to Lemmas 6, 7, and 8, respectively.

THEOREM 6. *Let u be the solution of* (3.21) *and let* (5.4) *and* (5.6) *hold. Then there exists a positive constant C that does not depend on ϵ such that the postprocessed approximation $\tilde u_h$ solution of* (3.23) *satisfies estimates* (3.47) *and* (3.48).

Finally we consider the equivalent of Theorem 5 with variable coefficients, that is, we consider the problem

$$\begin{aligned} u_t - \epsilon u_{xx} + bu_x + cu = f, &\quad 0<x<1, \quad t>0;\\ u(0,t)=u(1,t)=0 &\quad u(x,0)=u_0, \end{aligned} \tag{5.17}$$

with $b = b(x)$ and $c = c(x)$ positive functions of x. We assume that $\mu_0 > 0$, and we denote $c_M = \max_{0\le x\le 1} c(x)$. Similarly to the constant coefficient case, the first condition in (3.4) (recall that the second one is not needed in the case of linear elements) can be shown to hold if $h \le (\mu_0 b_0^2)/(b_M c_M^2)$. Then, following the arguments in the proof of Theorem 5 (with references to Lemmas 6 and 7 instead of Lemmas 3 and 4, respectively), the following result is obtained.

THEOREM 7. *Let u be the solution of* (5.17) *and $\tilde u_h$ the postprocessed approximation* (3.54)*; set δ_K as in* (5.1), (5.2). *Assume that $\mu_0 > 0$ and that* (5.4), (5.6) *hold. Then there exists a constant C that does not depend on ϵ such that the bounds* (3.55) *and* (3.56) *hold for $h \le (\mu_0 b_0^2)/(b_M c_M^2)$.*

REFERENCES

[1] B. Ayuso and B. García-Archilla, *Regularity constants of the Stokes problem. Application to finite-element methods on curved domains*, Math. Models Methods Appl. Sci., 15 (2005), pp. 437–470.

[2] M. I. Asensio, B. Ayuso, and G. Sangalli, *Coupling stabilized finite element methods with finite difference time integration for advection-diffusion-reaction problems*, Comput. Methods Appl. Mech. Engrg., 196 (2007), pp. 3475–3491.

[3] P. B. Bochev, M. D. Gunzburger, and J. N. Shadid, *Stability of the SUPG finite element method for transient advection-diffusion problems*, Comput. Methods Appl. Mech. Engrg., 193 (2004), pp. 2301–2323.

[4] A. N. Brooks and T. J. R. Hughes, *Streamline upwind/Petrov-Galerkin formulations for convection dominated flows with particular emphasis on the incomprehensible Navier-Stokes equations*, Comput. Methods Appl. Mech. Engrg., 32 (1982), pp. 199–259.

[5] E. Burman, *Consistent SUPG method for transient transport problems: Stability and convergence*, Comput. Methods Appl. Mech. Engrg., 199 (2010), pp. 1114–1123.

[6] E. Burman and M. A. Fernández, *Finite element methods with symmetric stabilization for the transient convection-diffusion-reaction equation*, Comput. Methods Appl. Mech. Engrg., 198 (2009), pp. 2508–2519.

[7] P. G. Ciarlet, *The Finite Element Method for Elliptic Problems*, North-Holland, Amsterdam, 1978.

[8] R. Codina, *Comparison of some finite element methods for linear systems of convection-diffusion-reaction equations*, Comput. Methods Appl. Mech. Engrg., 156 (1998), pp. 185–210.

[9] R. Codina and J. Blasco, *Analysis of a stabilized finite element approximation of the transient convection-diffusion-reaction equation using orthogonal subscales*, Comput. Visual Sci., 4 (2002), pp. 167–174.

[10] J. de Frutos, B. García-Archilla, and J. Novo, *Postprocessing finite-element methods for the Navier–Stokes equations: The fully discrete case*, SIAM J. Numer. Anal., 47 (2008), pp. 596–621.

[11] J. de Frutos, B. García-Archilla, and J. Novo, *Accurate approximations to time-dependent nonlinear convection-diffusion problems*, IMA J. Numer. Anal., to appear.

[12] J. de Frutos, B. García-Archilla, and J. Novo, *Nonlinear convection-diffusion problems: fully discrete approximations and a posteriori error estimates*, IMA J. Numer. Anal., to appear.

[13] J. de Frutos and J. Novo, *Bubble stabilization of linear finite element methods for nonlinear evolutionary convection-diffusion equations*, Comput. Methods Appl. Mech. Engrg., 197 (2008), pp. 3988–3999.

[14] T. J. R. Hughes and A. N. Brooks, *A multidimensional upwind scheme with no crosswind diffusion*, in Finite Element Methods for Convection Dominated Flows, T. J. R. Hughes, ed., ASME, New York, 1979, pp. 19–35.

[15] V. John and E. Schemeyer, *Finite element methods for time-dependent convection-diffusion-reaction equations with small diffusion*, Comput. Methods Appl. Mech. Engrg., 198 (2008), pp. 475–494.

[16] G. Hauke and M. H. Doweidar, *Fourier analysis of semidiscrete and space-time stabilized methods for the advective-reactive-diffusive equation: I. SUPG*, Comput. Methods Appl. Mech. Engrg., 194 (2005), pp. 45–81.

[17] G. Hauke and M. H. Doweidar, *Fourier analysis of semidiscrete and space-time stabilized*

methods for the advective-reactive-diffusive equation: II. *SGS*, Comput. Methods Appl. Mech. Engrg., 194 (2005), pp. 691–725.

[18] A. Huerta and J. Donea, *Time-accurate solution of stabilized convection-diffusion-reaction equations:* I, *Time and space discretization*, Comm. Numer. Methods Engrg., 18 (2002), pp. 565–573.

[19] A. Huerta, B. Roig, and J. Donea, *Time-accurate solution of stabilized convection-diffusion-reaction equations:* II, *Accuracy analysis and examples*, Comm. Numer. Methods Engrg., 18 (2002), pp. 575–584.

[20] T. J. R. Hughes, L. P. Franca, and M. Mallet, *A new finite element formulation for computational fluid dynamics:* VI. *Convergence analysis of the generalized SUPG formulation for linear time-dependent multidimensional advective-diffusive systems*, Comput. Methods Appl. Mech. Engrg., 63 (1987), pp. 97–112.

[21] G. Lube and D. Weiss, *Stabilized finite element methods for singularly perturbed parabolic problems*, Appl. Numer. Math., 17 (1995), pp. 431–459.

[22] C. Johnson, U. Navert, and U. Pitkaranka, *Finite element methods for linear hyperbolic problems*, Comput. Methods Appl. Mech. Engrg., 45 (1984), pp. 285–312.

[23] A. Quarteroni and A. Valli, *Numerical Approximation of Partial Differential Equations*, Springer Ser. Comput. Math. 23, Springer-Verlag, Berlin, 1997.

[24] H. G. Roos, M. Stynes, and L. Tobiska, *Numerical Methods for Singularly Perturbed Differential Equations-Convection-Diffusion and Flow Problems*, Springer Ser. Comput. Math. 24, Springer-Verlag, Berlin, 1996.

[25] M. Stynes, *Steady-state convection-diffusion problems*, Acta Numer., 14 (2005), pp. 445–508.

[26] L. B. Wahlbin, *Maximum norm error estimates in the finite element method with isoparametric quadratic elements and numerical integration*, RAIRO Anal. Numér., 12 (1978), pp. 173–202.