COMBINING STABILIZED FINITE ELEMENT METHODS

by

Frederic G. C. Valentin

Instituto de Matemática - UFRJ, Ilha do Fundão, 21945 Rio de Janeiro RJ - Brasil

Leopoldo P. Franca

Department of Mathematics, University of Colorado at Denver, P.O. Box 173364, Campus Box 170, Denver, CO 80217-3364

Abstract

We combine the Galerkin-least-squares (GLS) and the Galerkin-gradient-least-squares (GGLS) methods to simulate scalar linear second order partial differential equations which include second, first and zero order terms. Assuming a strictly positive coefficient for the second order term, the resulting method is proven convergent for a wide range of the coefficients multiplying each term. The stability parameters are designed to take into account the presence of both modifications to the Galerkin method. Numerical results atest the good stability characteristics of the method.

March 1995

1 Introduction

We study the simulation of a scalar linear second order partial differential equation with second, first and zero order terms. Such an equation models heat transfer in flows, approximates the transport of the turbulence energy and dissipation rate of turbulence energy for the $k - \epsilon$ model of turbulence, etc.

Employing the standard Galerkin method with linear elements for this equation, yields spurious oscillations in singular regimes, i.e, when the the second-order term coefficient is much smaller than either or both the coefficients for the zeroth and first order terms.

For the case when the zero term is absent (the advective-diffusive model), Hughes and Brooks [1],[7] introduced the SUPG (Streamline Upwind Petrov Galerkin) method. The method consists in adding a mesh dependent term premultiplied by a stability parameter, a term which is a function of the Euler-Lagrange equation in each element, so that additional numerical stability is attained. The method was originally analyzed by Johnson and Navert [9] and with an improved design of the stability parameter that deals with high order approximations by Franca et al. [4]. The Galerkin least-squares method (GLS) [8] crystallized this idea by using a mesh-dependent least-squares term. This GLS method is amenable to applications in other partial differential equations such as e.g., the ones governing structural problems (see [5] and references therein).

When the first order term is absent and the singular behavior emanates from a large zero order term compared to the second order term, Franca and Dutra do Carmo [3] have introduced the Galerkin-gradient-least-squares method (GGLS) which uses the gradient of the Euler-Lagrange equation in least-squares form as the perturbation to the Galerkin method. This method yields the additional stability to deal with the spurious oscillations present in the Galerkin method using linear elements.

To deal with singular behavior in the presence of both zero and first order terms, we combine the GLS and GGLS methods, as presented in Section 2. Next, an error analysis is carried out in Section 3, and some numerical experiments are reported in Section 4.

2 The method

Let $\mathbf{\Omega} \subset \mathbf{R}^N$, $N = 2, 3$, denote an open bounded domain with a polygonal or polyhedral boundary $\mathbf{\Gamma}$. The model consists in finding a scalar field $u(\mathbf{x})$, $\mathbf{x} \in \mathbf{\Omega}$ such that

$$\sigma u + \mathbf{a} . \nabla u - \alpha \Delta u = f \qquad \text{in} \quad \mathbf{\Omega} \tag{1}$$

$$u = 0 \qquad \text{on} \quad \mathbf{\Gamma} \tag{2}$$

where $\mathbf{a}(\mathbf{x})$ is the given flow velocity with $\nabla . \mathbf{a} = 0$ in $\mathbf{\Omega}$, $\sigma(\mathbf{x}) \geq 0$ is the source coefficient, $\alpha(\mathbf{x}) > 0$ is the diffusivity and $f(\mathbf{x})$ is a prescribed source function. The method that follows is also applicable to more general boundary conditions than the homogeneous case discussed herein.

Consider a partition C_h of $\overline{\mathbf{\Omega}}$ into elements consisting of convex quadrilaterals (hexahedra in $\mathbf{R}^3$) performed in the usual way (i.e., no overlapping is allowed between any two elements of the partition; the union of all elements domains K reproduces $\mathbf{\Omega}$, etc.). Let $Q_m(K)$ denote the space of polynomials of degree less or equal to m in each coordinate direction defined on an element $K \in C_h$.

The scalar field $u(\mathbf{x})$ is approximated with the following standard conforming subspace

$$W_h = \left\{ v \in H_0^1(\mathbf{\Omega}) \ | \ v_{|_K} \in Q_m(K), \, K \in C_h \right\} \tag{3}$$

where $H_0^1(\mathbf{\Omega})$ is the Sobolev space of functions with square-integrable value and derivatives in $\mathbf{\Omega}$ with zero value on the boundary $\mathbf{\Gamma}$.

The finite element method we wish to consider is: Find $u_h \in W_h$ such that

$$B(u_h, v) = F(v) \quad , \ \forall v \in W_h \tag{4}$$

with

$$\begin{aligned} B(u,v) = {} & (\sigma u ,v) + (\mathbf{a}.\nabla u,v) + (\alpha \nabla u, \nabla v) \\ & + \sum_{K \in C_h} (\sigma u + \mathbf{a}.\nabla u - \alpha \Delta u , \tau (\sigma v + \mathbf{a}.\nabla v - \alpha \Delta v))_K \\ & + \sum_{K \in C_h} \left(\sigma \nabla u + (\nabla \nabla u)^T \mathbf{a} - \alpha \nabla \Delta u, \gamma \left(\sigma \nabla v + (\nabla \nabla v)^T \mathbf{a} - \alpha \nabla \Delta v\right)\right)_K \end{aligned} \tag{5}$$

and

$$\begin{aligned} F(v) = {} & (f, v) + \sum_{K \in C_h} (f, \tau (\sigma v + \mathbf{a}.\nabla v - \alpha \Delta v))_K \\ & + \sum_{K \in C_h} \left(\nabla f, \gamma \left(\sigma \nabla v + (\nabla \nabla v)^T \mathbf{a} - \alpha \nabla \Delta v\right)\right)_K \end{aligned} \tag{6}$$

where $(.,.)$ denotes the L_2-inner product in $\mathbf{\Omega}$, $L_2(\mathbf{\Omega})$ is the space of square-integrable functions in $\mathbf{\Omega}$, $(.,.)_K$ is the L_2-inner product in element domain K. The stability parameters τ and γ are defined from error analysis considerations as follows:

$$\tau\left(\mathbf{x}, Pe_K\left(\mathbf{x}\right)\right) = \frac{h_K}{2\left(\left|\mathbf{a}\right|_p + \sigma\, h_K\right)}\xi\left(Pe_K\right) \tag{7}$$

$$\gamma\,\left(\mathbf{x}, Pe_K\left(\mathbf{x}\right)\right) = \frac{h_K^3}{4\left(\left|\mathbf{a}\right|_p + \sigma\, h_K\right)}\xi\left(Pe_K\right) \tag{8}$$

$$Pe_K\left(\mathbf{x}\right) = \frac{m_k\left(\left|\mathbf{a}\right|_p + \sigma\, h_K\right) h_K}{2\alpha} \tag{9}$$

$$\xi\left(Pe_K\right) = \begin{cases} Pe_K & , 0 \leq Pe_K \leq 1 \\ 1 & , Pe_K \geq 1 \end{cases} \tag{10}$$

$$\left|\mathbf{a}\right|_p = \begin{cases} \left(\sum_{i=1}^{N} \mid a_i \mid^p\right)^{\frac{1}{p}} & , 1 \leq p < \infty \\ \max\limits_{i=1,N} \mid a_i \mid & , p = \infty \end{cases} \tag{11}$$

$$m_k = \min\left\{\frac{1}{3}, 2\tilde{C}_k\right\} \tag{12}$$

$$\tilde{C}_k \sum_{K \in \mathbf{C}_h} h_K^2 \parallel \Delta v \parallel_{0,K}^2 \leq \parallel \nabla v \parallel_0^2 \quad , \quad \forall v \in W_h \tag{13}$$

REMARKS:

1. For the case when $\sigma = 0$, the method above differs from the Galerkin-least-squares method as tested in [4],[8], because the additional contribution of the Galerkin-gradient-least-squares method does not go to zero. The numerical results presented in section 4 indicate that the present generalization has the same good features as the methods proposed before.

2. A similar remark can be done when $\mathbf{a} = \mathbf{0}$. The method will be different to the GGLS method in [3], but again good numerical results are attainable even with the addition of the GLS contribution.

3. The key feature in the stability parameters design is their relative size. Note that for $Pe_K \geq 1$ when either the zero or the first order terms is dominant, we have $\tau = O(\,h\,)$ for the GLS term and $\gamma = O(\,h^3\,)$ for the GGLS term.

4. Combining the GLS and GGLS methods have been recently proposed independently by Harari and Hughes [6]. Their contribution has different design considerations for the stability parameters and different signs on the perturbation terms.

5. To deal with first and zero order terms, Franca and Farhat [5] have recently introduced a method with a single additional term. The method is prompted by a relationship with the Galerkin method enriched with bubble functions. The advantage of that method is that the second modification due to GGLS is unnecessary.

3 Error analysis

In this section we consider a global convergence analysis. First note that from the stability parameters design, either

$$\frac{\xi(Pe_K(\mathbf{x}))}{Pe_K(\mathbf{x})} = 1 \ , \qquad 0 \le Pe_K \le 1$$

or

$$\frac{\xi(Pe_K(\mathbf{x}))}{Pe_K(\mathbf{x})} = \frac{1}{Pe_K(\mathbf{x})} \le 1 \ , \qquad Pe_K \ge 1$$

then

$$\frac{\xi(Pe_K(\mathbf{x}))}{Pe_K(\mathbf{x})} \le 1 \tag{14}$$

Also, either

$$m_k = \frac{1}{3} \Rightarrow m_k \le 2\tilde{C}_k$$

or

$$m_k = 2\tilde{C}_k$$

therefore

$$\frac{m_k}{4\tilde{C}_k} \le \frac{1}{2} \tag{15}$$

The following stability result is immediate:

Lemma 1 (Stability) *Assume* $\mathbf{a}(\mathbf{x})$ *and* $\alpha(\mathbf{x})$ *to satisfy*

i) $\nabla.\mathbf{a}(\mathbf{x}) = 0 \qquad , \forall \mathbf{x} \in \mathbf{\Omega}$

ii) $\alpha(\mathbf{x}) = \alpha = const > 0 \quad , \forall \mathbf{x} \in \mathbf{\Omega}$

Then

$$\begin{aligned} B(v,v) \;\; \geq \;\; & \frac{1}{6}\Big(\sigma \,\|\, v\|_0^2 + \alpha \,\|\, \nabla v\|_0^2 + \left\| \,\tau^{\frac{1}{2}} \sigma v \right\|_0^2 + \left\| \,\tau^{\frac{1}{2}} \mathbf{a}.\nabla v \right\|_0^2 \\ & + \sum_{K \in C_h} \left\| \,\gamma^{\frac{1}{2}} \left(\sigma \nabla v \, + \, (\nabla \nabla v)^T \,\mathbf{a} \, - \, \alpha \, \nabla \Delta v \right) \right\|_{0,K}^2 \Big) \,, \forall v \in W_h \end{aligned}$$

Proof. Integrating by parts and using (i) and (2) it follows that

$$(\mathbf{a}.\nabla v, v) = 0 \qquad , \forall v \in W_h$$

By (14) we get for each $K \in C_h$:

$$\begin{aligned} \tau\left(\mathbf{x}, Pe_K(\mathbf{x})\right) \;\; &= \;\; \frac{h_k}{2\left(\left|\mathbf{a}\right|_p + \sigma\, h_K\right)} \xi(Pe_K(\mathbf{x})) \\ &= \;\; \frac{m_k\, h_K^2}{4\alpha} \frac{\xi(Pe_K(\mathbf{x}))}{Pe_K(\mathbf{x})} \\ &\leq \;\; \frac{m_k\, h_K^2}{4\alpha} \end{aligned} \tag{16}$$

and

$$\begin{aligned} \gamma\left(\mathbf{x}, Pe_K(\mathbf{x})\right) \;\; &= \;\; \frac{h_k^3}{4\left(\left|\mathbf{a}\right|_p + \sigma\, h_K\right)} \xi(Pe_K(\mathbf{x})) \\ &= \;\; \frac{m_k\, h_K^4}{8\alpha} \frac{\xi(Pe_K(\mathbf{x}))}{Pe_K(\mathbf{x})} \\ &\leq \;\; \frac{m_k\, h_K^4}{8\alpha} \end{aligned} \tag{17}$$

$$
\begin{aligned}
\sum_{K\in C_h} \left\| \tau^{\frac{1}{2}}\alpha\,\Delta v \right\|^2_{0,K} &\leq \frac{m_k\alpha}{4} \sum_{K\in C_h} h_K^2 \,\|\,\Delta v\|^2_{0,K} \ \text{(by } (ii) \text{ and (16))} \\
&\leq \frac{m_k}{4\tilde{C}_k}\alpha\,\|\nabla v\|_0^2 \qquad \text{(by (13))} \\
&\leq \frac{\alpha}{2}\,\|\nabla v\|_0^2 \qquad \text{(by (15))}
\end{aligned}
\tag{18}
$$

Therefore by (5) it follows that

$$
\begin{aligned}
B(v,v) &= \sigma\,\|\,v\|_0^2 + 0 + \alpha\,\|\,\nabla v\|_0^2 + \left\|\,\tau^{\frac{1}{2}}\sigma\,v\right\|_0^2 + \left\|\,\tau^{\frac{1}{2}}\mathbf{a}.\nabla v\right\|_0^2 - 2\sum_{K\in C_h}(\sigma v, \tau\,\alpha\,\Delta v)_K \\
&\quad -2\sum_{K\in C_h}(\mathbf{a}.\nabla v, \tau\,\alpha\,\Delta v)_K + \sum_{K\in C_h}\left\|\,\tau^{\frac{1}{2}}\alpha\,\Delta v\right\|^2_{0,K} \\
&\quad + \sum_{K\in C_h}\gamma\,\left\|\,\sigma\nabla v \,+\, (\nabla\nabla v)^T\,\mathbf{a} \,-\, \alpha\,\nabla\Delta v\right\|^2_{0,K} \\
&\geq \sigma\,\|\,v\|_0^2 + \alpha\,\|\,\nabla v\|_0^2 + \left(1-\frac{1}{\lambda_1}\right)\left\|\,\tau^{\frac{1}{2}}\sigma\,v\right\|_0^2 + \left(1-\frac{1}{\lambda_2}\right)\left\|\,\tau^{\frac{1}{2}}\mathbf{a}.\nabla v\right\|_0^2 \\
&\quad + (1-\lambda_1-\lambda_2)\sum_{K\in C_h}\left\|\,\tau^{\frac{1}{2}}\alpha\,\Delta v\right\|^2_{0,K} + \sum_{K\in C_h}\gamma\,\left\|\,\sigma\nabla v \,+\, (\nabla\nabla v)^T\,\mathbf{a} \,-\, \alpha\,\nabla\Delta v\right\|^2_{0,K} \\
&\geq \sigma\,\|\,v\|_0^2 + \left(\frac{3}{2} - \frac{\lambda_1+\lambda_2}{2}\right)\alpha\,\|\,\nabla v\|_0^2 + \left(1-\frac{1}{\lambda_1}\right)\left\|\,\tau^{\frac{1}{2}}\sigma v\right\|_0^2 + \left(1-\frac{1}{\lambda_2}\right)\left\|\,\tau^{\frac{1}{2}}\mathbf{a}.\nabla v\right\|_0^2 \\
&\quad + \sum_{K\in C_h}\gamma\,\left\|\,\sigma\nabla v \,+\, (\nabla\nabla v)^T\,\mathbf{a} \,-\, \alpha\,\nabla\Delta v\right\|^2_{0,K} \qquad \text{(by (18))} \\
&\geq \frac{1}{6}\Bigg(\sigma\,\|\,v\|_0^2 + \alpha\,\|\nabla v\|_0^2 + \left\|\,\tau^{\frac{1}{2}}\sigma\,v\right\|_0^2 + \left\|\,\tau^{\frac{1}{2}}\mathbf{a}.\nabla v\right\|_0^2 \\
&\quad + \sum_{K\in C_h}\gamma\,\left\|\,\sigma\nabla v \,+\, (\nabla\nabla v)^T\,\mathbf{a} \,-\, \alpha\,\nabla\Delta v\right\|^2_{0,K}\Bigg)
\end{aligned}
$$

where we selected $\lambda_1 = \lambda_2 = \dfrac{4}{3}$. □

Before starting the next Lemma recall that by standard aproximation theory [2], there exists an interpolant $\tilde{u}_h|_K \in Q_k(K)$ to $u \in H^{k+1}(K)$ such that

$$
\|\,u - \tilde{u}_h|_K\,\|_{m,K} \leq \tilde{C} h_K^{k+1-m}\,|\,u\,|_{k+1,K}\ ,\ 0 \leq m \leq k+1
$$

The following can now be established.

Lemma 2 (Interpolation Estimate) *Assume that the solution to (1)-(2) satisfies $u \in H^{k+1}(\mathbf{\Omega}) \cap H_0^1(\mathbf{\Omega})$ and that the assumptions of Lemma 1 hold. Denoting by $\eta = \tilde{u}_h - u$ the interpolation error for each $K \in C_h$:*

1. *if* $Pe_K(\mathbf{x}) \geq 1 \;\; , \forall \mathbf{x} \in K$ *then*

$$\left\| \tau^{-\frac{1}{2}} \eta \right\|_{0,K}^2 + \sigma \, \| \eta \|_{0,K}^2 + \alpha \, \| \nabla \eta \|_{0,K}^2 + \left\| \tau^{\frac{1}{2}} \, \sigma \, \eta \right\|_{0,K}^2 + \left\| \tau^{\frac{1}{2}} \, \mathbf{a} . \nabla \eta \right\|_{0,K}^2 + \left\| \tau^{\frac{1}{2}} \alpha \, \Delta \eta \right\|_{0,K}^2$$
$$+ \left\| \gamma^{\frac{1}{2}} \left(\sigma \nabla \eta + (\nabla \nabla \eta)^T \mathbf{a} - \alpha \, \nabla \Delta \eta \right) \right\|_{0,K}^2 \leq C \, \left(\sup | \, \mathbf{a} \, |_p + \sigma \, h_K \right) h_K^{2k+1} \, |u|_{k+1,K}^2$$

2. *if* $0 \leq Pe_K(\mathbf{x}) \leq 1 \;\; , \forall \mathbf{x} \in K$ *then*

$$\left\| \tau^{-\frac{1}{2}} \eta \right\|_{0,K}^2 + \sigma \, \| \eta \|_{0,K}^2 + \alpha \, \| \nabla \eta \|_{0,K}^2 + \left\| \tau^{\frac{1}{2}} \, \sigma \, \eta \right\|_{0,K}^2 + \left\| \tau^{\frac{1}{2}} \, \mathbf{a} . \nabla \eta \right\|_{0,K}^2 + \left\| \tau^{\frac{1}{2}} \alpha \, \Delta \eta \right\|_{0,K}^2$$
$$+ \left\| \gamma^{\frac{1}{2}} \left(\sigma \nabla \eta + (\nabla \nabla \eta)^T \mathbf{a} - \alpha \, \nabla \Delta \eta \right) \right\|_{0,K}^2 \leq C \, \alpha \, h_K^{2k} \, |u|_{k+1,K}^2$$

Therefore

$$\left\| \tau^{-\frac{1}{2}} \eta \right\|_0^2 + \sigma \, \| \eta \|_0^2 + \alpha \, \| \nabla \eta \|_0^2 + \left\| \tau^{\frac{1}{2}} \, \sigma \, \eta \right\|_0^2 + \left\| \tau^{\frac{1}{2}} \, \mathbf{a} . \nabla \eta \right\|_0^2 + \sum_{K \in C_h} \left\| \tau^{\frac{1}{2}} \alpha \, \Delta \eta \right\|_{0,K}^2$$
$$+ \sum_{K \in C_h} \gamma \, \left\| \sigma \nabla \eta + (\nabla \nabla \eta)^T \mathbf{a} - \alpha \, \nabla \Delta \eta \right\|_{0,K}^2$$
$$\leq C \sum_{K \in C_h} h_K^{2k} \, | \, u \, |_{k+1,K}^2 \, \left[H(Pe_K - 1) \, h_K \left(\sup | \, \mathbf{a} \, |_p + \sigma \, h_K \right) + H(1 - Pe_K) \, \alpha \right]$$

where

$$H(x - y) = \begin{cases} 0, & x < y \\ 1, & x \geq y \end{cases} \tag{19}$$

Proof. First note that

$$\begin{aligned}
\left\| \tau^{\frac{1}{2}}\mathbf{a}.\nabla\eta \right\|_{0,K}^{2} &= \left\| \left(\frac{h_K}{2(|\,\mathbf{a}\,|_p + \sigma\, h_K)}\xi \right)^{\frac{1}{2}} \mathbf{a}.\nabla\eta \right\|_{0,K}^{2} \\
&\leq \left\| \left(\frac{h_K}{2|\,\mathbf{a}\,|_p}\xi \right)^{\frac{1}{2}} \mathbf{a}.\nabla\eta \right\|_{0,K}^{2} \\
&\leq \frac{h_k}{2} \left\| \left(\frac{\xi}{|\,\mathbf{a}\,|_p} \right)^{\frac{1}{2}} |\,\mathbf{a}\,|_2 |\,\nabla\eta\,|_2 \right\|_{0,K}^{2} \qquad \text{(by Cauchy-Schwarz)} \\
&\leq C_1 \frac{h_k}{2} \left\| (\xi |\,\mathbf{a}\,|_p)^{\frac{1}{2}} \,|\,\nabla\eta\,|_2 \right\|_{0,K}^{2} \quad \text{(by equivalence of norms on } \mathbf{R}^N\text{)}
\end{aligned}$$

and

$$\begin{aligned}
\left\| \gamma^{\frac{1}{2}} (\nabla\nabla\eta)^T \mathbf{a} \right\|_{0,K}^{2} &= \left\| \left(\frac{h_K^3}{4\,(|\mathbf{a}|_p + \sigma\, h_K)}\xi \right)^{\frac{1}{2}} (\nabla\nabla\eta)^T \mathbf{a} \right\|_{0,K}^{2} \\
&\leq \left\| \left(\frac{h_K^3}{4|\mathbf{a}|_p}\xi \right)^{\frac{1}{2}} (\nabla\nabla\eta)^T \mathbf{a} \right\|_{0,K}^{2} \\
&\leq \left\| \left(\frac{h_K^3}{4|\mathbf{a}|_p}\xi \right)^{\frac{1}{2}} \left|(\nabla\nabla\eta)^T\right|_2 |\mathbf{a}|_2 \right\|_{0,K}^{2} \\
&\leq C_2 \left\| \left(\frac{h_K^3}{4|\mathbf{a}|_p}\xi \right)^{\frac{1}{2}} \left|(\nabla\nabla\eta)^T\right|_2 |\mathbf{a}|_p \right\|_{0,K}^{2} \quad \text{(by equivalence of norms on } \mathbf{R}^N\text{)} \\
&\leq \frac{C_2\, h_K^3}{4} \left\| \left(|\mathbf{a}|_p\, \xi\right)^{\frac{1}{2}} \left|(\nabla\nabla\eta)^T\right|_2 \right\|_{0,K}^{2}
\end{aligned}$$

We now divide the proof in two parts:

a) Let $Pe_K(\mathbf{x}) \geq 1$, $\forall \mathbf{x} \in K$. Then

$$\left\| \tau^{-\frac{1}{2}} \eta \right\|_{0,K}^2 + \sigma \| \eta \|_{0,K}^2 + \alpha \| \nabla \eta \|_{0,K}^2 + \left\| \tau^{\frac{1}{2}} \sigma \eta \right\|_{0,K}^2 + \left\| \tau^{\frac{1}{2}} \mathbf{a}.\nabla \eta \right\|_{0,K}^2 + \left\| \tau^{\frac{1}{2}} \alpha \Delta \eta \right\|_{0,K}^2$$

$$+ \left\| \gamma^{\frac{1}{2}} \left(\sigma \nabla \eta + (\nabla\nabla\eta)^T \mathbf{a} - \alpha \nabla \Delta \eta \right) \right\|_{0,K}^2$$

$$\leq \frac{2}{h_K} \left\| \left(|\mathbf{a}|_p + \sigma h_K \right)^{\frac{1}{2}} \eta \right\|_{0,K}^2 + \sigma \| \eta \|_{0,K}^2 + \frac{m_k \left(\sup |\mathbf{a}|_p + \sigma h_K \right) h_K}{2 Pe_K} \| \nabla \eta \|_{0,K}^2$$

$$+ \left\| \left(\frac{h_K}{2 \left(|\mathbf{a}|_p + \sigma h_k \right)} \right)^{\frac{1}{2}} \sigma \eta \right\|_{0,K}^2 + \left\| \left(\frac{C_1 h_K}{2} |\mathbf{a}|_p \right)^{\frac{1}{2}} \nabla \eta \right\|_{0,K}^2 + \frac{m_k^2 h_K^3 \left(\sup |\mathbf{a}|_p + \sigma h_K \right)}{8} \|\Delta \eta\|_{0,K}^2$$

$$+4 \left\| \left(\frac{h_K^3}{4 \left(|\mathbf{a}|_p + \sigma h_K \right)} \sigma^2 \right)^{\frac{1}{2}} \nabla \eta \right\|_{0,K}^2 + 4 \frac{C_2 h_K^3}{4} \left\| \left(|\mathbf{a}|_p \right)^{\frac{1}{2}} \left| (\nabla\nabla\eta)^T \right|_2 \right\|_{0,K}^2$$

$$+2 \frac{h_K^5 m_k^2 (\sup |\mathbf{a}|_p + \sigma h_K)}{16 Pe_K} \|\nabla \Delta \eta\|_{0,K}^2$$

$$\leq \frac{2 \left(\sup |\mathbf{a}|_p + \sigma h_K \right)}{h_K} \| \eta \|_{0,K}^2 + \sigma \| \eta \|_{0,K}^2 + \tilde{C}_k (\sup |\mathbf{a}|_p + \sigma h_K) h_K \| \nabla \eta \|_{0,K}^2 + \frac{\sigma}{2} \| \eta \|_{0,K}^2$$

$$+ \frac{C_1 h_K}{2} \sup |\mathbf{a}|_p \| \nabla \eta \|_{0,K}^2 + \tilde{C}_k^2 h_K^3 (\sup |\mathbf{a}|_p + \sigma h_K) \|\Delta \eta\|_{0,K}^2 + 4 \frac{h_K^2 \sigma}{4} \|\nabla \eta\|_{0,K}^2 +$$

$$+4 \frac{C_2 h_K^3}{4} \sup |\mathbf{a}|_p \left\| (\nabla\nabla\eta)^T \right\|_{0,K}^2 + 2 \frac{\tilde{C}_k^2 h_K^5 (\sup |\mathbf{a}|_p + \sigma h_K)}{4} \|\nabla \Delta \eta\|_{0,K}^2$$

$$\leq C \left(\sup |\mathbf{a}|_p + \sigma h_K \right) h_K^{2k+1} |u|_{k+1,K}^2$$

b) Let $0 \le Pe_K(\mathbf{x}) \le 1$, $\forall \mathbf{x} \in K$. Then

$$\left\| \tau^{-\frac{1}{2}} \eta \right\|_{0,K}^2 + \sigma \| \eta \|_{0,K}^2 + \alpha \| \nabla \eta \|_{0,K}^2 + \left\| \tau^{\frac{1}{2}} \sigma \eta \right\|_{0,K}^2 + \left\| \tau^{\frac{1}{2}} \mathbf{a}.\nabla \eta \right\|_{0,K}^2 + \left\| \tau^{\frac{1}{2}} \alpha \Delta \eta \right\|_{0,K}^2$$

$$+ \left\| \gamma^{\frac{1}{2}} \left(\sigma \nabla \eta + (\nabla \nabla \eta)^T \mathbf{a} - \alpha \nabla \Delta \eta \right) \right\|_{0,K}^2$$

$$\le \frac{2}{h_K} \left\| \left(\frac{|\mathbf{a}|_p + \sigma h_K}{\xi (Pe_K)} \right)^{\frac{1}{2}} \eta \right\|_{0,K}^2 + \frac{2\, \alpha\, Pe_K}{m_k\, h_k^2} \| \eta \|_{0,K}^2 + \alpha \| \nabla \eta \|_{0,K}^2 + \frac{m_k h_K^2}{4\alpha} \frac{4\alpha^2 Pe_K^2}{m_k^2 h_K^4} \| \eta \|_{0,K}^2$$

$$+ C_1 \frac{h_K\, Pe_K}{2} \left\| \left(|\mathbf{a}|_p \right)^{\frac{1}{2}} . \left| \nabla \eta \right|_2 \right\|_{0,K}^2 + \frac{h_K^2\, m_k\, \alpha}{4} \| \Delta \eta \|_{0,K}^2 + 4 \frac{m_k h_K^4}{8\alpha} \frac{4\alpha^2 Pe_K^2}{m_k^2 h_K^4} \| \nabla \eta \|_{0,K}^2$$

$$+ 4 \frac{C_2 h_K^3 Pe_K}{4} \left\| \left(|\mathbf{a}|_p \right)^{\frac{1}{2}} \left| (\nabla \nabla \eta)^T \right|_2 \right\|_{0,K}^2 + 2 \frac{h_K^4\, m_k\, \alpha}{8} \| \nabla \Delta \eta \|_{0,K}^2$$

$$\le \frac{4\alpha}{m_k\, h_K^2} \| \eta \|_{0,K}^2 + \frac{2\, \alpha}{m_k\, h_k^2} \| \eta \|_{0,K}^2 + \alpha \| \nabla \eta \|_{0,K}^2 + \frac{\alpha}{m_k\, h_K^2} \| \eta \|_{0,K}^2 + \frac{C_1 \alpha}{m_k} \| \nabla \eta \|_{0,K}^2$$

$$+ \frac{\tilde{C}_k\, h_K^2\, \alpha}{2} \| \Delta \eta \|_{0,K}^2 + \frac{\alpha}{2\, m_k} \| \nabla \eta \|_{0,K}^2 + \frac{C_2 \alpha h_K^2}{2\, m_k} \left\| (\nabla \nabla \eta)^T \right\|_{0,K}^2 + \frac{\tilde{C}_k\, h_K^4\, \alpha}{4} \| \nabla \Delta \eta \|_{0,K}^2$$

$$\le C\, \alpha\, h_K^{2k} | u |_{k+1,K}^2$$

and therefore the Lemma follows. □

The following (global) convergence estimate can now be established.

Theorem 1 *Under the same hypotheses as in Lemma* 1 *and 2, the solution* u_h *of the method given by (4)-(6) converges to* u*, solution of (1)-(2) as follows:*

$$\sigma \| u_h - u \|_0^2 + \alpha \| \nabla (u_h - u)\|_0^2 + \left\| \tau^{\frac{1}{2}}\sigma \, (u_h - u)\right\|_0^2 + \left\| \tau^{\frac{1}{2}}\mathbf{a}.\nabla (u_h - u)\right\|_0^2$$
$$+ \sum_{K\in C_h} \gamma \, \left\|\sigma\nabla (u_h - u) + (\nabla\nabla (u_h - u))^T \mathbf{a} - \alpha \nabla\Delta (u_h - u)\right\|_{0,K}^2$$
$$\leq C \sum_{K\in C_h} h_K^{2k} | u|_{k+1,K}^2 \left[H (Pe_K - 1)\, h_k (\sup |\mathbf{a}|_p + \sigma h_K) + H (1 - Pe_K)\, \alpha\right]$$

where $H(\cdot)$ *is defined in (19).*

Proof. Let $e_h = u_h - \tilde{u}_h$ and $e = e_h + \eta$. The convergence proof goes as follows:

$$\frac{1}{6} \left(\sigma \| e_h \|_0^2 + \alpha \|\nabla e_h\|_0^2 + \left\| \tau^{\frac{1}{2}}\sigma \, e_h\right\|_0^2 + \left\| \tau^{\frac{1}{2}}\mathbf{a}.\nabla e_h\right\|_0^2 + \sum_{K\in C_h} \gamma \, \left\|\sigma\nabla e_h + (\nabla\nabla e_h)^T \mathbf{a} - \alpha \nabla\Delta e_h\right\|_{0,K}^2 \right)$$

$$\begin{aligned}
&\leq B(e_h, e_h) && \text{(by Lemma 1)}\\
&= B(e-\eta, e_h) &&\\
&= -B(\eta, e_h) && \text{(by consistency)}\\
&\leq |B(\eta, e_h)| &&\\
&= \Big|(\sigma\,\eta, e_h) + (\mathbf{a}.\nabla\eta, e_h) + (\alpha\nabla\eta, \nabla e_h) &&\\
&\quad + \sum_{K\in C_h} (\sigma\,\eta + \mathbf{a}.\nabla\eta - \alpha\,\Delta\eta, \tau\,(\sigma\, e_h + \mathbf{a}.\nabla e_h - \alpha\,\Delta e_h))_K &&\\
&\quad + \sum_{K\in C_h} \left(\sigma\nabla\eta + (\nabla\nabla\eta)^T \mathbf{a} - \alpha\nabla\Delta\eta, \gamma\left(\sigma\nabla e_h + (\nabla\nabla e_h)^T \mathbf{a} - \alpha\nabla\Delta\, e_h\right)\right)_K \Big| &&
\end{aligned}$$

$$
\begin{aligned}
\leq\ & |(\sigma\,\eta, e_h)| + |(\mathbf{a}.\nabla\eta, e_h)| + |(\alpha\nabla\eta, \nabla e_h)| \\
& + \left| \sum_{K\in C_h} (\sigma\eta + \mathbf{a}.\nabla\eta - \alpha\,\Delta\eta, \tau\,(\sigma e_h + \mathbf{a}.\nabla e_h - \alpha\,\Delta e_h))_K \right| \\
& + \left| \sum_{K\in C_h} \left(\sigma\nabla\eta + (\nabla\nabla\eta)^T\,\mathbf{a} - \alpha\nabla\Delta\,\eta, \gamma\left(\sigma\nabla e_h + (\nabla\nabla e_h)^T\,\mathbf{a} - \alpha\nabla\Delta e_h\right)\right)_K \right| \\
\leq\ & 3\sigma\,\|\,\eta\,\|_0^2 + |(\eta, \mathbf{a}.\nabla e_h)| + 6\alpha\,\|\,\nabla\eta\,\|_0^2 + \frac{\sigma}{12}\,\|\,e_h\,\|_0^2 + \frac{\alpha}{24}\,\|\,\nabla e_h\,\|_0^2 \\
& + \sum_{K\in C_h} \left(\left\|\,\tau^{\frac{1}{2}}\sigma\,\eta\,\right\|_{0,K} + \left\|\,\tau^{\frac{1}{2}}\mathbf{a}.\nabla\eta\,\right\|_{0,K} + \left\|\,\tau^{\frac{1}{2}}\alpha\,\Delta\eta\,\right\|_{0,K} \right) \\
& . \left(\left\|\,\tau^{\frac{1}{2}}\sigma\,e_h\,\right\|_{0,K} + \left\|\tau^{\frac{1}{2}}\mathbf{a}.\nabla e_h\right\|_{0,K} + \left\|\,\tau^{\frac{1}{2}}\alpha\,\Delta e_h\,\right\|_{0,K} \right) \\
& + \sum_{K\in C_h} 3\gamma\,\left\|\sigma\,\nabla\eta\ +\ (\nabla\nabla\eta)^T\,\mathbf{a}\ -\alpha\,\nabla\Delta\eta\,\right\|_{0,K}^2 \\
& + \sum_{K\in C_h} \frac{\gamma}{12}\left\|\sigma\,\nabla e_h\ +\ (\nabla\nabla e_h)^T\,\mathbf{a}\ -\alpha\,\nabla\Delta e_h\,\right\|_{0,K}^2 \\
\leq\ & 3\sigma\,\|\,\eta\,\|_0^2 + 6\alpha\,\|\nabla\,\eta\,\|_0^2 + 6\left\|\,\tau^{-\frac{1}{2}}\eta\,\right\|_0^2 \\
& +36\left(\left\|\,\tau^{\frac{1}{2}}\mathbf{a}.\nabla\eta\,\right\|_0^2 + \sum_{K\in C_h} \left\|\,\tau^{\frac{1}{2}}\alpha\,\Delta\eta\,\right\|_{0,K}^2 + \left\|\,\tau^{\frac{1}{2}}\sigma\,\eta\,\right\|_0^2 \right) \\
& + \sum_{K\in C_h} 3\gamma\,\left\|\sigma\,\nabla\eta\ +\ (\nabla\nabla\eta)^T\,\mathbf{a}\ -\alpha\,\nabla\Delta\eta\,\right\|_{0,K}^2 + \frac{\sigma}{12}\,\|\,e_h\|_0^2 + \frac{\alpha}{24}\,\|\,\nabla e_h\|_0^2 \\
& + \frac{1}{12}\left\|\,\tau^{\frac{1}{2}}\mathbf{a}.\nabla\,e_h\right\|_0^2 + \frac{1}{12}\left\|\,\tau^{\frac{1}{2}}\sigma\,e_h\,\right\|_0^2 + \frac{\alpha}{24}\sum_{K\in C_h} h_K^2\,\tilde{C}_k\ \|\Delta e_h\|_{0,K}^2 \\
& + \sum_{K\in C_h} \frac{\gamma}{12}\left\|\sigma\,\nabla e_h\ +\ (\nabla\nabla e_h)^T\,\mathbf{a}\ -\alpha\,\nabla\Delta e_h\,\right\|_{0,K}^2
\end{aligned}
$$

Using the inverse estimate (13) and Lemma 2 we get

$$
\begin{aligned}
\frac{1}{12}\ & \left(\sigma\,\|\,e_h\,\|_0^2 + \alpha\,\|\nabla e_h\|_0^2 + \left\|\,\tau^{\frac{1}{2}}\sigma\,e_h\right\|_0^2 + \left\|\,\tau^{\frac{1}{2}}\mathbf{a}.\nabla e_h\right\|_0^2 \right. \\
& \left. + \sum_{K\in C_h} \gamma\,\left\|\sigma\nabla e_h\ \ +\ (\nabla\nabla e_h)^T\,\mathbf{a}\ -\ \alpha\,\nabla\Delta e_h\right\|_{0,K}^2 \right) \\
\leq\ & C\sum_{K\in C_h} h_K^{2k}\,|\,u\,|_{k+1,K}^2\ [H(Pe_K - 1)\,h_K +\ (\sup|\,\mathbf{a}\,|_p + \sigma\,h_K)H(1 - Pe_K)\,\alpha]
\end{aligned} \tag{20}
$$

Since by Lemma 2 we also have

$$
\begin{aligned}
&\frac{1}{12} \quad \left(\sigma \,\| \,\eta\|_0^2 + \alpha \,\|\, \nabla\eta\|_0^2 + \left\| \,\tau^{\frac{1}{2}}\sigma\, \eta\right\|_0^2 + \left\| \,\tau^{\frac{1}{2}}\mathbf{a}.\nabla\eta\right\|_0^2 + \right. \\
&\qquad \left. \sum_{K\in C_h} \gamma \,\left\|\sigma\nabla\eta + \,(\nabla\nabla\eta)^T\,\mathbf{a} \,-\, \alpha\,\nabla\Delta\eta\right\|_{0,K}^2 \right) \\
&\leq \quad C \sum_{K\in C_h} h_K^{2k}\,|\,u\,|_{k+1,K}^2\,\left[H(Pe_K - 1)\,h_K \;+\; (\sup|\,\mathbf{a}\,|_p + \sigma\,h_K)H(1 - Pe_K)\,\alpha\right]
\end{aligned}
\tag{21}
$$

then the result follows by redefining constants in (20) and (21) and using the triangle inequality. □

4 Numerical results

We employ the test problems presented in [1], [3], [4] to illustrate the applicability of the method for different values of the coefficients. The complex numerical problem is for low values of the diffusivity α and various examples illustrate the suitability of the method in these regimes.

4.1 High zero order term

Let us consider $\mathbf{a} = \mathbf{0}$, $\sigma = 1$, $\alpha = 10^{-6}$ and $f = 1$ on a unit square

$$
\boldsymbol{\Omega} = \left\{(x,y) \in R^2 \;|\; 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1\right\}
$$

with boundary conditions

$$
u\,(0,y) = u\,(x,1) = 1 \quad , \quad 0 \leq x \leq 1,\; 0 \leq y \leq 1
$$

$$
u\,(1,y) = u\,(x,0) = 0 \quad , \quad 0 \leq x \leq 1,\; 0 \leq y \leq 1
$$

A uniform mesh with 20×20 Q1 (bilinear) elements is used, and the numerical solutions with the Galerkin and the present methods are given in Figures 1 and 2, respectively. The spurious oscillations present near the boundary layer using the Galerkin method are eliminated when employing the present method.

4.2 Advection in a rotating flow field

Consider $\sigma = f = 0$ and $\alpha = 10^{-6}$. The problem is defined on a unit square of coordinates $-0.5 \leq x, y \leq +0.5$ where the flow velocity components are given by

$$a_1 = -y\,, \quad a_2 = x$$

subject to $u = 0$ on the external boundary and with $u = \frac{1}{2}(\cos(4\pi y + \pi) + 1)$ at $x = 0$, $-0.5 \leq y \leq 0$.

A uniform mesh with 30×30 Q1 elements is used, and the Euclidean norm is adopted in computing $|\mathbf{a}|_2$.

In Figures 3 and 4 we display the solutions obtained with the Galerkin and the present methods, respectively. Since the exact solution is smooth, both methods perform well with the Galerkin method with small amplitude oscillations reduced when using the stabilized method.

4.3 Advection skew to the mesh

In this test we employ a constant unitary flow field, $|\mathbf{a}|_2 = 1$, with $\alpha = 10^{-6}$ and $\sigma = 10^{-2}$. We set a discontinuity in the data at the inflow boundary that when propagated into the domain creates an internal layer of increasing width due to the σ-term. See Figure 5 for the problem statement. A uniform mesh of 20×20 Q1 elements are used and $p = 2$.

The solution using the Galerkin method is highly oscillatory and, consequently, not shown here. The solutions for an inflow at an angle $\theta = 45^o$ is shown in Figure 6 and, with $\theta = 60^o$, in Figure 7.

References

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

[2] P. G. Ciarlet, The finite element method for elliptic problems (North-Holland, Amsterdam, 1978).

[3] L.P.Franca e E.G.Dutra do Carmo, The Galerkin gradient least-squares method, Computer Methods in Applied Mechanics and Engineering, Vol. 74 (1989), pp. 41-54.

[4] L .P. Franca, S. L. Frey e T. J. R. Hughes, Stabilized finite element methods: I. Application to the advective-diffusive model, Computer Methods in Applied Mechanics and Engineering, Vol. 95 (1992), pp. 253-276.

[5] L .P. Franca and C. Farhat, Bubble functions prompt unusual stabilized finite element methods, Computer Methods in Applied Mechanics and Engineering, to appear.

[6] I. Harari and T.J.R. Hughes, Stabilized finite element methods for steady advection-diffusion with production, Computer Methods in Applied Mechanics and Engineering, Vol. 115 (1994), pp. 165-191.

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

[8] T. J. R. Hughes, L. P. Franca and G. M. Hulbert, A new finite element formulation for computational fluid dynamics: VIII. The Galerkin/least-squares method for advective-diffusive equations, Comput. Methods Appl. Mech. Engrg. 73 (1989), pp. 173-189.

[9] C. Johnson and U. Nävert, An analysis of some finite element methods for advection diffusion problems, in: O. Axelson, L.S. Frank and A. Van Der Sluis, eds., Analytical and Numerical Approaches to Asymptotic Problems in Analysis (North-Holland, Amsterdam, 1981), pp.99-116.

Fig.1: High zero order term: Galerkin method solution.

Fig.2: High zero order term: stabilized method solution.

Fig.3: Advection in a rotating flow field: Galerkin method solution.

Fig.4: Advection in a rotating flow field: stabilized method solution.

Fig.5: Advection skew to the mesh: problem statement.

Fig.6: Advection skew to the mesh: stabilized method solution with $\theta = 45^o$.

Fig.7: Advection skew to the mesh: stabilized method solution with $\theta = 60^o$.