

Numerical evaluation of finite element methods in convection-diffusion problems

G. Sangalli

Dipartimento di Matematica “F. Casorati”, Università di Pavia, Via Ferrata 1, 27100 Pavia, Italy
e-mail: sangalli@dimat.unipv.it

Received: December 1999 / Revised version: May 2000

Abstract. In this paper we propose an inf-sup test to identify and measure the stability of various finite element methods for the solution of multi-dimensional convection-diffusion problem.

1 Introduction

It has been recognized that many problems in fluid mechanics cannot be solved efficiently using standard finite element discretizations. In particular we refer to the convection-diffusion problem

$$\begin{cases} -\varepsilon \Delta u + \boldsymbol{\beta} \cdot \nabla u = f & \text{in } \Omega \\ u = 0 & \text{in } \partial\Omega, \end{cases} \quad (1)$$

where Ω is a convex and polygonal subset of \mathbb{R}^2 with $\text{diam}(\Omega) = 1$, $\varepsilon > 0$ is the diffusion parameter, $\boldsymbol{\beta} \in \mathbb{R}^2$ represents the constant velocity field and the source term f belongs to L^2 . This is, with respect to ε , a linear singularly perturbed boundary value problem; when the convection term (hyperbolic in nature) is dominant in this system ($\varepsilon \ll \|\boldsymbol{\beta}\|$) standard methods introduce non-physical oscillatory solutions. In recent years various techniques to obtain *robust* methods that work for all values of ε have been proposed (see, for example, [9]).

In this paper we propose an automatic procedure to measure the stability properties of a given method. Generally the numerical testing of a given algorithm is accomplished by solving model problems (i.e., choosing a particular source term) that present some typical features. The procedure we present now constitutes in some sense a deeper approach, as it tries to

identify automatically the most critical configuration. This is obtained by performing a proper eigenvalue investigation on the discrete operator that represents the numerical scheme under analysis. A similar technique has already been presented in [2], although a different theoretical justification was proposed and the simpler one-dimensional convection-diffusion equation was considered.

Similar procedures are also used in other contexts: for instance, in incompressible elasticity (see [4]), in the plate problem (see [6]), and in the case of conservation laws when a finite difference scheme is evaluated by performing a numerical Von Neumann analysis.

Here the idea is to investigate numerically a stability condition for the numerical operator that holds true for the exact differential operator. Because of the particular singularly perturbed behavior of (1), it is impossible to prove a stability condition, with uniformity with respect to ε , that involves derivatives on the whole domain Ω . So in the following analysis we consider a stability condition which holds on a proper subdomain that excludes exponential boundary layers, i.e., narrow regions near the outflow boundary where the exact solution changes very rapidly. This is inspired by the local analysis developed by Johnson et al. (see [7] and its references) and by boundary layers theory (see, for example, [8]).

The paper is divided into three sections. In the first we present the theoretical result for the analytical solution, which motivates our choice of the proposed test. In the second the test is presented in a general context. The last section is devoted to numerical examples.

2 Behavior of the exact solution

Throughout the sequel C always denotes a positive constant independent of ε and f (not necessarily the same at all occurrences), $L^2(\Omega) = H^0(\Omega)$ and $H^k(\Omega)$, with k integer, denote the usual Sobolev spaces, equipped with usual norms and seminorms; $H_0^1(\Omega)$ denotes the subset of $H^1(\Omega)$ whose elements vanish on the boundary. Finally, given a non-negative function ψ on Ω , we define the ψ -weighted norm (or seminorm)

$$\|v\|_{\psi}^2 := \int_{\Omega} v^2 \psi \, dx.$$

We consider the weak formulation of (1); this is obtained by multiplying the equation by a test function $v \in H_0^1(\Omega)$, integrating over Ω :

$$-\varepsilon \int_{\Omega} \Delta u \, v \, dx + \int_{\Omega} \boldsymbol{\beta} \cdot \nabla u \, v \, dx = \int_{\Omega} f v \, dx,$$

and integrating by parts:

$$\varepsilon \int_{\Omega} \nabla u \cdot \nabla v \, dx + \int_{\Omega} \boldsymbol{\beta} \cdot \nabla u \, v \, dx = \int_{\Omega} f v \, dx.$$

The so-called *weak solution* of (1) is the (unique) function $u \in H_0^1(\Omega)$ such that

$$\varepsilon \int_{\Omega} \nabla u \cdot \nabla v \, dx + \int_{\Omega} \boldsymbol{\beta} \cdot \nabla u \, v \, dx = \int_{\Omega} f v \, dx \quad \forall v \in H_0^1(\Omega). \quad (2)$$

Since $f \in L^2(\Omega)$, by the virtue of the elliptic regularity property $u \in H^2(\Omega)$; moreover the following well-known result holds. We recall the proof for the reader's convenience.

Proposition 2.1 *Suppose $\varepsilon \leq \|\boldsymbol{\beta}\|$ and let u be the weak solution of (1). Then*

$$\varepsilon^{3/2} \|u\|_{H^2(\Omega)} + \varepsilon^{1/2} \|u\|_{H^1(\Omega)} + \|u\|_{L^2(\Omega)} \leq C \|f\|_{L^2(\Omega)}. \quad (3)$$

Proof Define $\psi(\mathbf{x}) := e^{-\boldsymbol{\beta} \cdot \mathbf{x} / \|\boldsymbol{\beta}\|}$ and choose $v = u\psi$ in (2); this gives

$$\begin{aligned} \varepsilon \int_{\Omega} \psi \nabla u \cdot \nabla u \, dx - \varepsilon \int_{\Omega} u \psi \frac{\boldsymbol{\beta}}{\|\boldsymbol{\beta}\|} \cdot \nabla u \, dx + \int_{\Omega} u \psi \boldsymbol{\beta} \cdot \nabla u \, dx \\ = \int_{\Omega} f u \psi \, dx. \end{aligned} \quad (4)$$

Integrating by parts we obtain

$$\begin{aligned} \int_{\Omega} u \psi \boldsymbol{\beta} \cdot \nabla u \, dx &= \int_{\Omega} \boldsymbol{\beta} \cdot \nabla \left(\frac{u^2}{2} \right) \psi \, dx \\ &= \frac{\|\boldsymbol{\beta}\|}{2} \int_{\Omega} u^2 \psi \, dx; \end{aligned}$$

moreover, using the Cauchy–Schwarz inequality and the hypothesis, we get

$$\begin{aligned} -\varepsilon \int_{\Omega} u \psi \frac{\boldsymbol{\beta}}{\|\boldsymbol{\beta}\|} \cdot \nabla u \, dx &\leq 2 \left(\frac{\varepsilon}{2\|\boldsymbol{\beta}\|} \right)^{1/2} \left(\varepsilon \|\nabla u\|_{\psi}^2 \right)^{1/2} \left(\frac{\|\boldsymbol{\beta}\|}{2} \|u\|_{\psi}^2 \right)^{1/2} \\ &\leq 2^{-1/2} \left(\varepsilon \|\nabla u\|_{\psi}^2 + \frac{\|\boldsymbol{\beta}\|}{2} \|u\|_{\psi}^2 \right). \end{aligned}$$

So, using the Cauchy–Schwarz's inequality again, we get from (4)

$$\varepsilon \|\nabla u\|_{\psi}^2 + \frac{\|\boldsymbol{\beta}\|}{2} \|u\|_{\psi}^2 \leq C \|f\|_{\psi} \|u\|_{\psi},$$

and then by the equivalence between $\|\cdot\|_\psi$ and $\|\cdot\|_0$ we obtain

$$\varepsilon^{1/2} |u|_{H^1(\Omega)} + \|u\|_{L^2(\Omega)} \leq C \|f\|_{L^2(\Omega)} .$$

From (1) and the previous analysis we infer that

$$\begin{aligned} \varepsilon \|\Delta u\|_{L^2(\Omega)} &\leq C (\|f\|_{L^2(\Omega)} + \|\boldsymbol{\beta} \cdot \nabla u\|_{L^2(\Omega)}) \\ &\leq C \varepsilon^{-1/2} \|f\|_{L^2(\Omega)} ; \end{aligned}$$

by the elliptic regularity property we have

$$|u|_{H^2(\Omega)} \leq C \|\Delta u\|_{L^2(\Omega)} ,$$

whence

$$\varepsilon^{3/2} |u|_{H^2(\Omega)} \leq C \|f\|_{L^2(\Omega)} ,$$

and this concludes the proof.

For problems with exponential boundary layers, estimate (3) is sharp; in subdomains that exclude layers it is possible to obtain better control for the derivative in the streamline direction of the solution u . We denote by \mathbf{n} the outward unit normal vector defined almost everywhere on the boundary of Ω . We set

$$\begin{aligned} \Gamma_+ &:= \{\mathbf{x} \in \partial\Omega : \boldsymbol{\beta} \cdot \mathbf{n} > 0\} , \\ \Gamma_0 &:= \{\mathbf{x} \in \partial\Omega : \boldsymbol{\beta} \cdot \mathbf{n} = 0\} , \\ \Gamma_- &:= \{\mathbf{x} \in \partial\Omega : \boldsymbol{\beta} \cdot \mathbf{n} < 0\} . \end{aligned}$$

Proposition 2.2 *For every open set $\Omega' \subset \Omega$ with $\text{dist}(\Gamma_+, \Omega') > 0$ we have*

$$\lim_{\varepsilon \rightarrow 0^+} \inf_{f \in L^2(\Omega)} \frac{\|f\|_{L^2(\Omega)}}{\|\boldsymbol{\beta} \cdot \nabla u\|_{L^2(\Omega')}} = 1, \tag{5}$$

where $u = u_\varepsilon(f)$ is the solution of (2) and depends on ε and f .

Proof Define δ^+ as the solution of the hyperbolic problem

$$\begin{cases} -\boldsymbol{\beta} \cdot \nabla \delta^+ = \|\boldsymbol{\beta}\| & \text{in } \Omega \\ \delta^+ = 0 & \text{on } \Gamma_+; \end{cases}$$

hence δ^+ is a continuous and almost everywhere differentiable function and its value at $\mathbf{x} \in \Omega$ equals the distance between \mathbf{x} and Γ_+ computed along the velocity field $\boldsymbol{\beta}$. Consider moreover a monotone and smooth function ξ such that

$$\begin{cases} \xi \equiv 0 & \text{in } (-\infty, 1/2) \\ \xi \equiv 1 & \text{in } (1, \infty) . \end{cases}$$

Finally, we define a pair of functions ψ_i , with $i = 1, 2$, as

$$\psi_i(\mathbf{x}) := \xi \left(i \cdot \frac{\delta^+(\mathbf{x})}{\text{dist}(\Gamma_+, \Omega')} \right);$$

it is easy to see that the functions ψ_i (for $i = 1, 2$) are continuous, a.e. differentiable and satisfy

$$0 \leq \psi_i \leq 1 \quad \text{on } \Omega, \tag{6}$$

$$\psi_i = 0 \quad \text{on } \Gamma_+, \tag{7}$$

$$\boldsymbol{\beta} \cdot \nabla \psi_i \leq 0 \quad \text{on } \Omega; \tag{8}$$

moreover,

$$\psi_1 = 1 \quad \text{on } \Omega', \tag{9}$$

$$\psi_2 = 1 \quad \text{on } \{\psi_1 \geq 0\}. \tag{10}$$

Set $u_\beta := \boldsymbol{\beta} \cdot \nabla u$. We know from Proposition 2.1 that the weak solution u of (2) satisfies (1) in $L^2(\Omega)$; so we can multiply (1) by $u_\beta \psi_2$ and integrate over Ω :

$$\int_{\Omega} (u_\beta)^2 \psi_2 \, d\mathbf{x} = \int_{\Omega} f u_\beta \psi_2 \, d\mathbf{x} + \varepsilon \int_{\Omega} \Delta u u_\beta \psi_2 \, d\mathbf{x}. \tag{11}$$

Consider the last term in (11): integrating by parts we get

$$\begin{aligned} \varepsilon \int_{\Omega} \Delta u u_\beta \psi_2 \, d\mathbf{x} &= \varepsilon \int_{\Omega} \text{div}(\nabla u) u_\beta \psi_2 \, d\mathbf{x} \\ &= -\varepsilon \int_{\Omega} \psi_2 \nabla u \cdot \nabla u_\beta \, d\mathbf{x} \\ &\quad - \varepsilon \int_{\Omega} u_\beta \nabla u \cdot \nabla \psi_2 \, d\mathbf{x} \\ &\quad + \varepsilon \int_{\partial\Omega} \psi_2 u_\beta \nabla u \cdot \mathbf{n} \, d\sigma(\mathbf{x}) \\ &= \text{I} + \text{II} + \text{III}. \end{aligned} \tag{12}$$

Integrating by parts and using (8) we have

$$\begin{aligned} \text{I} &= -\varepsilon \int_{\Omega} \psi_2 \boldsymbol{\beta} \cdot \nabla \left(\frac{\nabla u \cdot \nabla u}{2} \right) d\mathbf{x} \\ &= \frac{\varepsilon}{2} \int_{\Omega} |\nabla u|^2 \boldsymbol{\beta} \cdot \nabla \psi_2 d\mathbf{x} - \frac{\varepsilon}{2} \int_{\partial\Omega} \psi_2 |\nabla u|^2 \boldsymbol{\beta} \cdot \mathbf{n} d\sigma(\mathbf{x}) \\ &\leq -\frac{\varepsilon}{2} \int_{\partial\Omega} \psi_2 |\nabla u|^2 \boldsymbol{\beta} \cdot \mathbf{n} d\sigma(\mathbf{x}). \end{aligned}$$

Moreover the homogeneous boundary condition on u implies that ∇u is normal to $\partial\Omega$, so we have

$$u_{\boldsymbol{\beta}} = (\nabla u \cdot \mathbf{n}) (\boldsymbol{\beta} \cdot \mathbf{n}) \quad \text{on } \partial\Omega,$$

and then

$$\text{III} = \varepsilon \int_{\partial\Omega} |\nabla u|^2 \psi_2 \boldsymbol{\beta} \cdot \mathbf{n} d\sigma(\mathbf{x}).$$

Then by virtue of (7) we have

$$\begin{aligned} \text{I} + \text{III} &\leq \frac{\varepsilon}{2} \int_{\partial\Omega} |\nabla u|^2 \psi_2 \boldsymbol{\beta} \cdot \mathbf{n} d\sigma(\mathbf{x}) \\ &\leq 0. \end{aligned}$$

In II, assuming that $\varepsilon \leq \|\boldsymbol{\beta}\|$ and applying Proposition 1, we obtain

$$\begin{aligned} \text{II} &\leq C \varepsilon \|u\|_{H^1(\Omega)}^2 \\ &\leq C \|f\|_{L^2(\Omega)}^2. \end{aligned}$$

So from (12) we get

$$\begin{aligned} \|u_{\boldsymbol{\beta}}\|_{\psi_2}^2 &\leq C \left(\|f\|_{\psi_2} \|u_{\boldsymbol{\beta}}\|_{\psi_2} + \|f\|_{L^2(\Omega)}^2 \right) \\ &\leq C \left(\|f\|_{L^2(\Omega)} \|u_{\boldsymbol{\beta}}\|_{\psi_2} + \|f\|_{L^2(\Omega)}^2 \right), \end{aligned}$$

and in conclusion

$$\|u_{\boldsymbol{\beta}}\|_{\psi_2} \leq C \|f\|_{L^2(\Omega)}. \quad (13)$$

Proceeding in the same way with ψ_1 we get

$$\|u_{\boldsymbol{\beta}}\|_{\psi_1}^2 \leq \|f\|_{\psi_1} \|u_{\boldsymbol{\beta}}\|_{\psi_1} - \varepsilon \int_{\Omega} u_{\boldsymbol{\beta}} \nabla u \cdot \nabla \psi_1 d(\mathbf{x}); \quad (14)$$

By (13) and (10) we now obtain

$$\begin{aligned} -\varepsilon \int_{\Omega} u_{\beta} \nabla u \cdot \nabla \psi_1 &\leq C \varepsilon \|u_{\beta}\|_{\psi_2} \|u\|_{H^1(\Omega)} \\ &\leq C \varepsilon^{1/2} \|f\|_{L^2(\Omega)}^2, \end{aligned}$$

and (14) reduces to

$$\|u_{\beta}\|_{\psi_1}^2 \leq \|f\|_{L^2(\Omega)} \|u_{\beta}\|_{\psi_1} + C \varepsilon^{1/2} \|f\|_{L^2(\Omega)}^2,$$

that is,

$$\begin{aligned} \|u_{\beta}\|_{\psi_1} &\leq \frac{\|f\|_{L^2(\Omega)} + \sqrt{\|f\|_{L^2(\Omega)}^2 + 4C \varepsilon^{1/2} \|f\|_{L^2(\Omega)}^2}}{2} \\ &\leq (1 + C \varepsilon^{1/2}) \|f\|_{L^2(\Omega)}. \end{aligned}$$

Taking the limit as $\varepsilon \rightarrow 0^+$ and using (9) we get

$$\lim_{\varepsilon \rightarrow 0^+} \inf_{f \in L^2(\Omega)} \frac{\|f\|_{L^2(\Omega)}}{\|\beta \cdot \nabla u\|_{L^2(\Omega')}} \geq 1.$$

Now consider a regular function v on Ω which vanishes outside Ω' ; we have

$$\lim_{\varepsilon \rightarrow 0^+} \frac{\|-\varepsilon \Delta v + \beta \cdot \nabla v\|_{L^2(\Omega)}}{\|\beta \cdot \nabla v\|_{L^2(\Omega')}} = 1.$$

Hence

$$\lim_{\varepsilon \rightarrow 0^+} \inf_{f \in L^2(\Omega)} \frac{\|f\|_{L^2(\Omega)}}{\|\beta \cdot \nabla u\|_{L^2(\Omega')}} \leq 1.$$

3 Proposed numerical test

We consider a general *conforming* finite element method. The discrete problem that corresponds to (2) is then:

$$\begin{cases} \text{find } u_h \in V_h \text{ such that} \\ \mathcal{A}_h(u_h, v_h) = \int_{\Omega} f \mathcal{P}_h v_h \, d\mathbf{x} \quad \forall v_h \in V_h, \end{cases} \quad (15)$$

where V_h is a finite dimensional subspace of $H_0^1(\Omega)$ consisting of continuous piecewise polynomial functions on a quasi-uniform and regular triangulation \mathcal{T}_h of Ω (each triangle has edges whose length is of order h); \mathcal{A}_h is a non-singular bilinear form on V_h and \mathcal{P}_h is a non-singular linear,

$L^2(\Omega)$ -valued operator defined on V_h . From the numerical point of view it is useful to introduce the so-called *mesh Peclet number* $Pe := h \|\boldsymbol{\beta}\| / \varepsilon$ and to distinguish between advection-dominated problems (when $Pe \geq 1$) and diffusion-dominated problems (when $Pe < 1$).

The standard Galerkin method corresponds to considering the same variational formulation of (2). This method is not appropriate when $Pe \geq 1$ for the following reason: the exact solution of (2) is characterized by sharp boundary layers along Γ_0 and Γ_+ but, unless the mesh density around Γ_0 and Γ_+ is considerably greater than in the interior of Ω , the piecewise polynomial approximation is highly oscillatory even in regions in which the true solution is smooth.

Suppose that we have a new numerical method and are interested in knowing whether such a method is affected by unstable behaviour, i.e., whether it propagates spurious oscillations from the layer regions into the internal region, where the exact solution is regular. In this case we can get some information by looking at the value

$$s := \inf_{f \in L^2(\Omega)} \frac{\|f\|_{L^2(\Omega)}}{\|\boldsymbol{\beta} \cdot \nabla u_h\|_{L^2(\Omega')}}. \quad (16)$$

We know from Proposition 2.2 that the optimal value of s , i.e., the value corresponding to the true solution, approaches 1 when the diffusion coefficient ε approaches zero. So what we expect from a stable method is that, by choosing Ω' properly, the quantity s becomes approximately 1 when the diffusion parameter is so small that the layer regions are thinner than the mesh size. On the contrary, in the case of the propagation of non-physical oscillation outside of layer regions, the norm $\|\boldsymbol{\beta} \cdot \nabla u_h\|_{L^2(\Omega')}$, which appears in the denominator in definition (16), grows with respect to the exact counterpart $\|\boldsymbol{\beta} \cdot \nabla u\|_{L^2(\Omega')}$; hence the more a method behaves in an unstable way the more we should get small values of s . Moreover very small values of s are the signal that the method is not robust.

For the sake of simplicity we will deal only with piecewise linear functions (i.e., $v \in V_h \Leftrightarrow v|_T$ is affine $\forall T \in \mathcal{T}_h$); a reasonable assumption is to take as Ω' the union of all triangles belonging to \mathcal{T}_h except those that are adjacent to $\Gamma_0 \cup \Gamma_+$ (see Fig. 1). The computation of s can be performed numerically by standard algebraic techniques. First observe that in (16) we can take $f \in \mathcal{P}_h(V_h)$, because only this part of the source term can affect the solution u_h of (15); hence, given a basis for V_h , for example, the usual nodal basis $\mathcal{B} := \{\phi_i\}_{i=1}^N$ where ϕ_i is related to the node P_i (i.e., $\phi_i(P_j) = \delta_{ij}$),

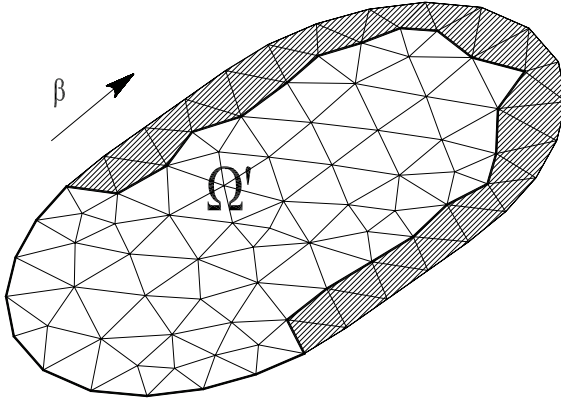


Fig. 1. Construction of Ω' given a subdivision \mathcal{T}_h of Ω

we introduce the matrices

$$A_{ij} = (\mathcal{A}_h(\phi_j, \phi_i)), \tag{17}$$

$$U_{ij} = \int_{\Omega'} \boldsymbol{\beta} \cdot \nabla \phi_i \boldsymbol{\beta} \cdot \nabla \phi_j \, d\mathbf{x}, \tag{18}$$

$$V_{ij} = \int_{\Omega} \mathcal{P}_h \phi_i \mathcal{P}_h \phi_j \, d\mathbf{x}; \tag{19}$$

note that in the assemblage of these matrices element by element, as usual in FEM codes, in the case of U we only have to consider elements in Ω' . Now we can compute s :

$$s = \inf_{f \in \mathcal{P}_h(V_h)} \sup_{v_h \in V_h} \frac{\langle f, \mathcal{P}_h v_h \rangle}{\|\mathcal{P}_h v_h\|_{L^2(\Omega)} \|\boldsymbol{\beta} \cdot \nabla u_h\|_{L^2(\Omega')}} \\ = \inf_{u_h \in V_h} \sup_{v_h \in V_h} \frac{\mathcal{A}_h(u_h, v_h)}{\|\mathcal{P}_h v_h\|_{L^2(\Omega)} \|\boldsymbol{\beta} \cdot \nabla u_h\|_{L^2(\Omega')}};$$

note that s is nothing but the value related to an inf-sup condition for the bilinear form \mathcal{A}_h , with respect to the discrete norms related to (18)–(19); then, introducing the basis representation, we have

$$s^2 = \inf_{\mathbf{x} \in \mathbb{R}^N} \sup_{\mathbf{y} \in \mathbb{R}^N} \frac{(\mathbf{y}' \mathbf{A} \mathbf{x})^2}{\mathbf{y}' \mathbf{V} \mathbf{y} \mathbf{x}' \mathbf{U} \mathbf{x}}.$$

The sup can be explicitly computed: we factorize $V = R^t R$ (note that V is positive definite) and define $\mathbf{w} = R \mathbf{y}$, then

$$\begin{aligned} \sup_{\mathbf{y} \in \mathbb{R}^N} \frac{(\mathbf{y}^t \mathbf{A} \mathbf{x})^2}{\mathbf{y}^t V \mathbf{y}} &= \sup_{\mathbf{w} \in \mathbb{R}^N} \frac{(\mathbf{w}^t R^{-t} \mathbf{A} \mathbf{x})^2}{\mathbf{w}^t \mathbf{w}} \\ &= \mathbf{x}^t A^t R^{-1} R^{-t} \mathbf{A} \mathbf{x} \\ &= \mathbf{x}^t A^t V^{-1} \mathbf{A} \mathbf{x}. \end{aligned}$$

It remains to compute

$$\inf_{\mathbf{x} \in \mathbb{R}^N} \frac{\mathbf{x}^t A^t V^{-1} \mathbf{A} \mathbf{x}}{\mathbf{x}^t U \mathbf{x}},$$

which has the same value as the minimum generalized eigenvalue λ_{\min} for the problem

$$A^t V^{-1} \mathbf{A} \mathbf{x} = \lambda U \mathbf{x}. \tag{20}$$

Finally we get $s = \lambda_{\min}^{1/2}$.

4 Testing some methods

We now give various examples in order to understand how the numerical evaluation of s (defined in (16)) allows us to distinguish between unstable, stable and monotonic methods. We therefore briefly discuss the modified finite element methods that we are going to test.

In order to solve (1), the simplest stabilizing technique for the standard Galerkin finite element method adds an *artificial diffusion* of size $O(h)$ to the diffusion coefficient ε . Thus the *artificial diffusion* (AD) method fits into the general framework (15) by setting $\mathcal{A}_h \equiv \mathcal{A}_h^{\text{AD}}$ and $\mathcal{P}_h \equiv \mathcal{P}_h^{\text{AD}}$ as

$$\left\{ \begin{aligned} \mathcal{A}_h^{\text{AD}}(w_h, v_h) &:= (\varepsilon + h \|\boldsymbol{\beta}\|) \int_{\Omega} \nabla w_h \cdot \nabla v_h \, dx \\ &\quad + \int_{\Omega} \boldsymbol{\beta} \cdot \nabla w_h \, v_h \, dx && \forall w_h, v_h \in V_h \\ \mathcal{P}_h^{\text{AD}}(v_h) &:= v_h && \forall v_h \in V_h. \end{aligned} \right.$$

Another modification of the standard Galerkin method is the *upwind triangle* (UW) method. It is based on a particular quadrature rule for the convective term that uses just the upwind part of the streamline derivative. We set, for all $w_h \in V_h$ and $\phi_i \in \mathcal{B}$,

$$\int_{\Omega} \boldsymbol{\beta} \cdot \nabla w_h \phi_i \, dx \approx Q^{\text{UW}}(w_h, \phi_i) := \frac{1}{3} \sum_{\substack{T \in \mathcal{T}_h \\ \bar{T} \cap P_i \neq \emptyset}} \text{area}(T) \boldsymbol{\beta} \cdot \nabla w_h|_{T_i^{\text{UW}}}, \tag{21}$$

and extend by linearity the definition of Q^{uw} to the whole of $V_h \times V_h$. In (21) T_i^{uw} denotes the *upwind triangle* for the node P_i , that is (for the sake of simplicity we assume that the edges of the triangles are not aligned with β):

1. P_i is a vertex of T_i^{uw} ;
2. the vector $-\beta$ points from P_i into T_i^{uw} .

So in this case we set

$$\begin{cases} \mathcal{A}_h^{uw}(w_h, v_h) := \varepsilon \int_{\Omega} \nabla w_h \cdot \nabla v_h \, dx + Q^{uw}(w_h, v_h) & \forall w_h, v_h \in V_h \\ \mathcal{P}_h^{uw}(v_h) := v_h & \forall v_h \in V_h. \end{cases}$$

This method preserves the inverse-monotonicity property of the continuous operator and, under additional hypotheses, it is L^∞ -stable uniformly with respect to ε (see [9]).

The streamline diffusion finite element method, introduced by Hughes and Brooks, adds weighted residual terms to the standard Galerkin formulation; this can be interpreted as a modification of the test function space, so the method is known as the *streamline upwind Petrov-Galerkin* (SUPG) method. We define

$$\begin{cases} \mathcal{A}_h^{supg}(w_h, v_h) := \varepsilon \int_{\Omega} \nabla w_h \cdot \nabla v_h \, dx + \int_{\Omega} \beta \cdot \nabla w_h v_h \, dx \\ \quad + \sum_{T \in \mathcal{T}_h} \tau_T^{supg} \int_T \beta \cdot \nabla w_h \beta \cdot \nabla v_h \, dx & \forall w_h, v_h \in V_h \\ \mathcal{P}_h^{supg}(v_h) := v_h + \sum_{T \in \mathcal{T}_h} \tau_T^{supg} \beta \cdot \nabla v_h & \forall v_h \in V_h, \end{cases}$$

where, as proposed by Hughes and coworkers in [5],

$$\tau_T^{supg} := \frac{\text{diam}(T)}{2 \|\beta\|}.$$

A new approach has been recently developed. Preserving the standard Galerkin formulation in a space enriched by *bubble functions* and using a *static condensation* of the bubble part, we obtain an additional stabilizing term of streamline diffusion type depending on the bubble shape (see [1] for more details). In particular we consider *residual-free bubbles* (RFB), as

proposed by Brezzi and Russo in [3], giving rise to

$$\left\{ \begin{array}{l} \mathcal{A}_h^{\text{RFB}}(w_h, v_h) := \varepsilon \int_{\Omega} \nabla w_h \cdot \nabla v_h \, dx + \int_{\Omega} \boldsymbol{\beta} \cdot \nabla w_h v_h \, dx \\ \quad + \sum_{T \in \mathcal{T}_h} \tau_T^{\text{RFB}} \int_T \boldsymbol{\beta} \cdot \nabla w_h \boldsymbol{\beta} \cdot \nabla v_h \, dx \quad \forall w_h, v_h \in V_h \\ \mathcal{P}_h^{\text{RFB}}(v_h) := v_h + \sum_{T \in \mathcal{T}_h} \tau_T^{\text{RFB}} \boldsymbol{\beta} \cdot \nabla v_h \quad \forall v_h \in V_h, \end{array} \right.$$

and

$$\tau_T^{\text{RFB}} := \frac{h_{\boldsymbol{\beta}, T}}{3 \|\boldsymbol{\beta}\|},$$

where $h_{\boldsymbol{\beta}, T}$ is the length of the longest segment in T parallel to $\boldsymbol{\beta}$.

First we perform two different tests. In Test 1 we take $\Omega = (0, 1) \times (0, 1)$, $\boldsymbol{\beta} = [1; 1]$, $\varepsilon = 10^{-4}$ and a non-structured mesh where the mesh size is around 10^{-1} (see Fig. 2). For Test 2 the domain Ω is an oval of length 2 and width 1, $\boldsymbol{\beta} = [1; 0]$, $\varepsilon = 10^{-4}$ and still a non-structured mesh with the same characteristics as before; in this second case we have a characteristic boundary (see Fig. 3). The results for both cases are presented in Table 1.

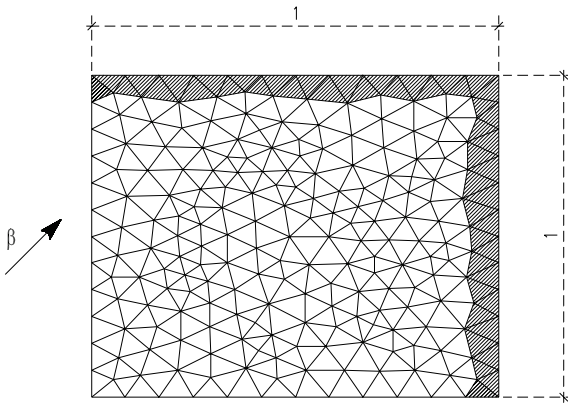


Fig. 2. Mesh in Test 1 and Test 4 (hatched elements are not considered in the assemblage of U)

In Test 3 we present the computed values of s for different mesh sizes. The unstructured meshes are refined from about 10 to about 10^3 elements and the other parameters are the same as in Test 1 (see Fig. 4). We can observe that, as we expect, the value of s for stable methods is near to 1, while, on the other hand, for Standard Galerkin this value is very small and

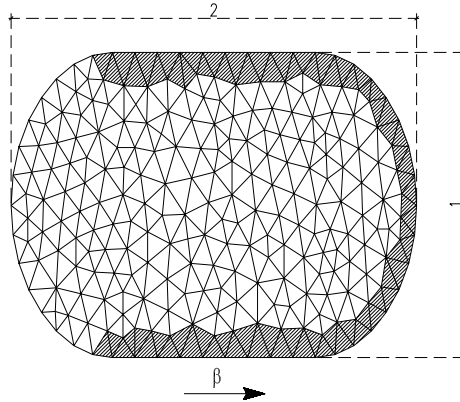


Fig. 3. Mesh in Test 2 (hatched elements are not considered in the assemblage of U)

Table 1. Values of s in Test 1 and Test 2

Method	Value of s in Test 1	Value of s in Test 2
Std.Gal.	$6.3 \cdot 10^{-3}$	$6.1 \cdot 10^{-3}$
AD	$9.3 \cdot 10^{-1}$	1.0
UW	$7.7 \cdot 10^{-1}$	$8.7 \cdot 10^{-1}$
RFB	$2.9 \cdot 10^{-1}$	$2.6 \cdot 10^{-1}$
SUPG	$2.9 \cdot 10^{-1}$	$2.6 \cdot 10^{-1}$

decreases when the mesh is made coarser. Note also that the prolongation of the diagram of standard Galerkin method reaches the optimal value 1 when $Pe = 1$ (i.e., when the element size is approximately 10^{-4} and therefore n is about 10^8). Test 4 consists of repeating Test 3 with structured meshes (the corresponding values of s are plotted in Fig. 5). In Test 5 we compute a sequence of values of s (shown in Fig. 6) corresponding to a sequence of values of ϵ from 10^{-2} to 10^{-10} (with the other conditions as in Test 1); we can confirm that, when a stable method is tested, the corresponding value of s is constant with respect to ϵ .

The following tests show how it is possible to obtain more information about the type of instability of a tested numerical method. We take into consideration the AD and RFB methods with a regular mesh and $\epsilon = 10^{-8}$, $\beta = [1; 1]$ in Test 6 and $\beta = [1; 0]$ in Test 7; computed values of s are presented in Table 2.

In Figs. 7, 8, 9 and 10 we plot the source term f such that

$$f \text{ minimizes } \frac{\|f\|_{L^2(\Omega)}}{\|\beta \cdot \nabla u_h\|_{L^2(\Omega')}};$$

this function can be computed from the corresponding eigenvector u_h which is associated to λ_{\min} in the eigenvalue problem (20). We can understand why

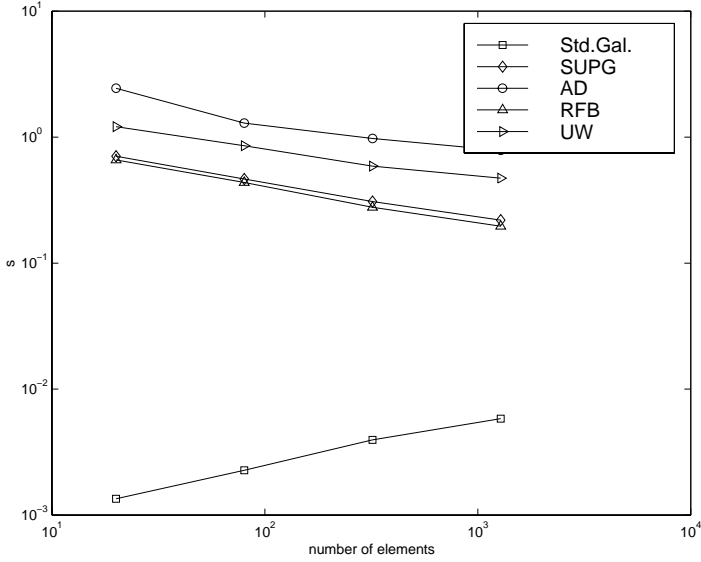


Fig. 4. Test 3; computed values of s for different mesh sizes (unstructured mesh case)

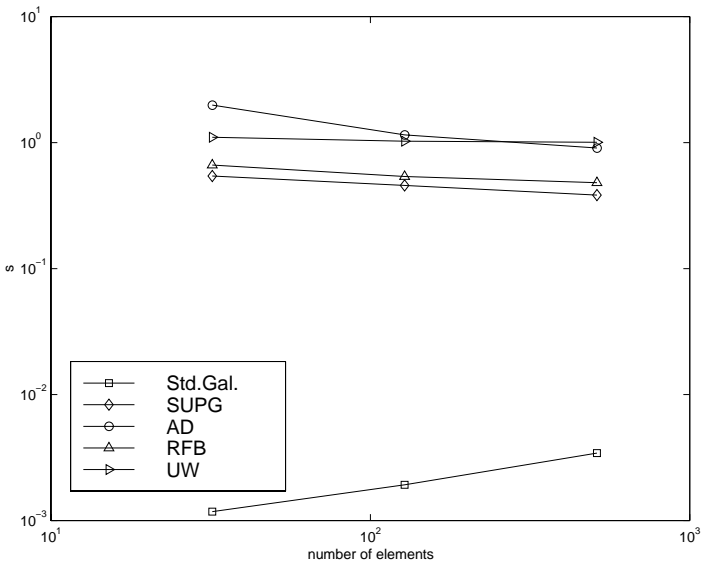


Fig. 5. Test 4; computed values of s for different mesh sizes (structured mesh case)

Table 2. Values of s in Test 6 and Test 7

Method	Value of s in Test 6	Value of s in Test 7
AD	$9.9 \cdot 10^{-1}$	1.0
RFB	$4.6 \cdot 10^{-1}$	$4.7 \cdot 10^{-1}$

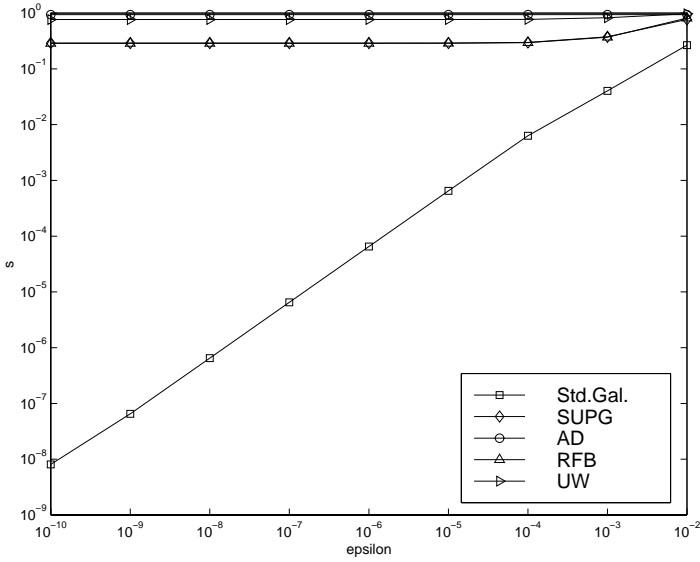


Fig. 6. Test 5; computed values of s for different values of ϵ

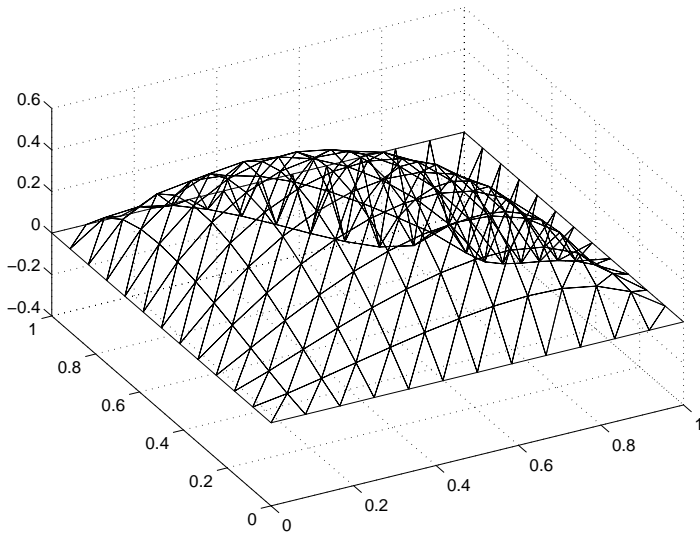


Fig. 7. Test 6, AD; plotting of f

the RFB method achieves a value of s smaller than 1: the most unstable configuration is reached when f presents characteristic shocks inside the domain, that are not stabilized (as happens, instead, with AD). As we know, this type of instability is produced by the absence of crosswind artificial diffusion, which is typical of RFB and SUPG methods. In some cases this behavior may be helpful: indeed even though RFB produces some small

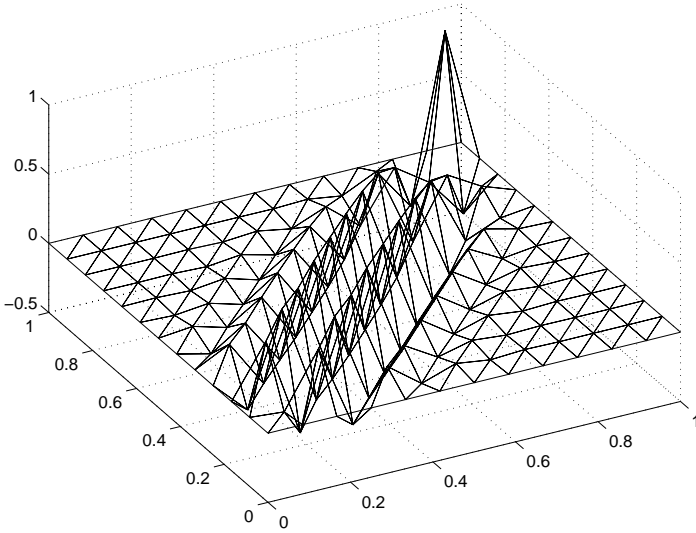


Fig. 8. Test 6, RFB; plotting of f

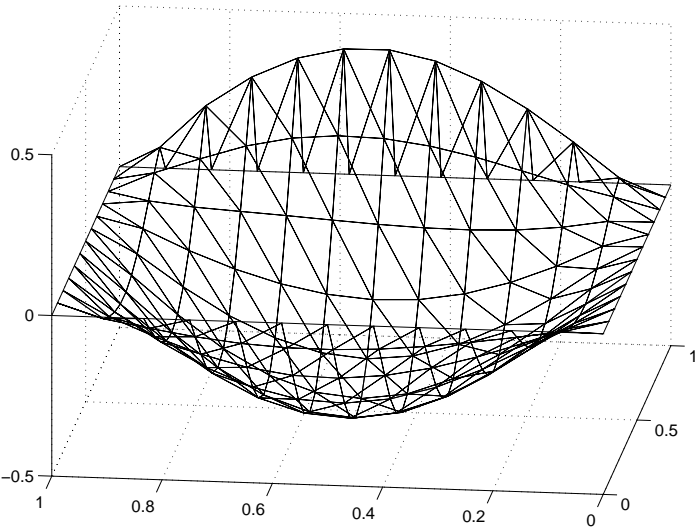


Fig. 9. Test 7, AD; plotting of f

non-physical oscillations, these are restricted to a few layers of elements and so it offers a possible indicator of the presence of shock that the AD method may not detect.

In the last test (Test 8) we note that the previous analysis enables us to distinguish among different types of instability: we compare, under the same conditions as in Test 6, RFB and a Modified Artificial Diffusion method, obtained by adding to the standard Galerkin formulation an artificial diffusion-

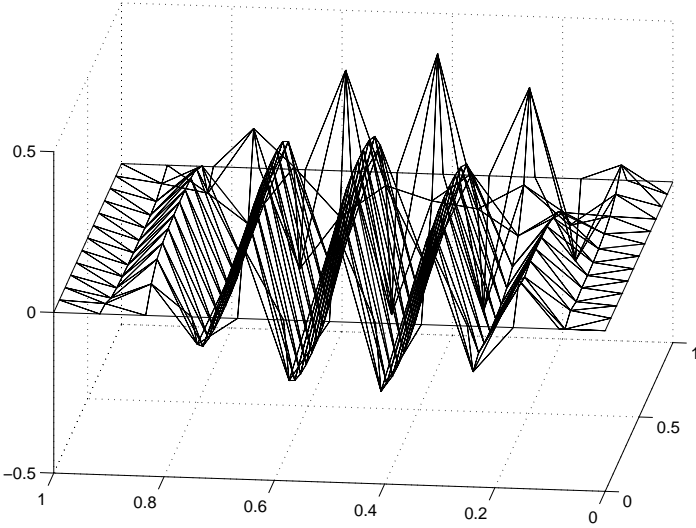


Fig. 10. Test 7, RFB; plotting of f

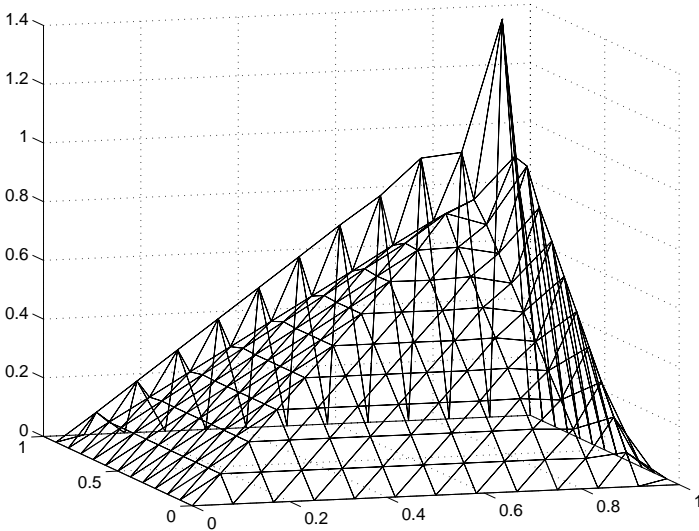


Fig. 11. Test 8, RFB; plotting of f

like term with a coefficient smaller than $h \|\beta\|$; we choose this coefficient in such a way that the two methods give approximately the same value of s , as in Table 3.

In Figs. 11 and 12 we plot the minimizing source term f . We can see that for RFB we have a wavy surface “aligned” with β : this shows that the method can allow internal shocks without smoothing them too much, as for contact discontinuity. We also observe that, on the contrary, in the case

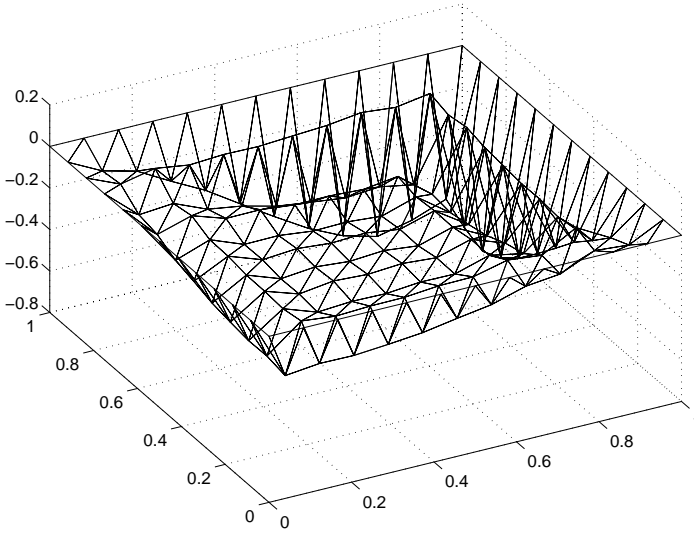


Fig. 12. Test 8, Modified AD; plotting of f

Table 3. Values of s in Test 8

Method	Value of s in Test 8
Modified AD	$4.5 \cdot 10^{-1}$
RFB	$4.6 \cdot 10^{-1}$

of Modified Artificial Diffusion the condition of maximal propagation of spurious oscillation is not related to a lack of regularity of the exact solution in the interior of Ω .

5 Conclusions

Our tests show that this automatic evaluation is significant; the analysis of s indicates the extent to which spurious oscillations are propagated in smooth regions by the tested method. Our automatic procedure distinguishes between unstable methods (standard Galerkin), stable but not monotonicity preserving methods (SUPG and RFB) and monotone methods (like UW). In case of a small amount of instability, with a deeper investigation we can also determine the nature of this instability and recognize how it is generated: for instance, in the case of the RFB method, we discover the absence of crosswind artificial diffusion. In conclusion this procedure can assist in the setting-up of good numerical methods, the analysis of a new method and the choice of the most appropriate numerical procedure for a particular situation.

References

- [1] Baiocchi, C., Brezzi, F., Franca, L.P.: Virtual bubbles and Galerkin-least-squares type methods (Ga.L.S.). *Comput. Methods Appl. Mech. Engrg.* **105**, 125–141 (1993)
- [2] Bathe, K.-J., Brezzi, F., Hendriana, D., Sangalli, G.: Inf-sup testing of upwind methods. *Internat. J. Numer. Methods Engrg.* **48**, 745–760 (2000)
- [3] Brezzi, F., Russo, A.: Choosing bubbles for advection-diffusion problems. *Math. Models Methods Appl. Sci.* **4**, 571–587 (1994)
- [4] Chapelle, D., Bathe, K.-J.: The inf-sup test. *Comput. & Structures* **47**, 537–545 (1993)
- [5] Franca, L.P., Frey, S.L., Hughes, T.J.R.: Stabilized finite element methods. I. Application to the advective-diffusive model. *Comput. Methods Appl. Mech. Engrg.* **95**, 253–276 (1992)
- [6] Iosilevich, A., Bathe, K.-J., Brezzi, F.: On evaluating the inf-sup condition for plate bending elements. *Internat. J. Numer. Methods Engrg.* **40**, 3639–3663 (1997)
- [7] Johnson, C., Schatz, A.H., Wahlbin, L.B.: Crosswind smear and pointwise errors in streamline diffusion finite element methods. *Math. Comp.* **49**, 25–38 (1987)
- [8] Shih, S.-D., Kellogg, B.: Asymptotic analysis of a singular perturbation problem. *SIAM J. Math. Anal.* **18**, 1467–1511 (1987)
- [9] Roos, H.-G., Stynes, M., Tobiska, L.: Numerical methods for singularly perturbed differential equations. *Convection-diffusion and flow problems*. Berlin: Springer 1996