MATHEMATICS OF COMPUTATION
VOLUME 60, NUMBER 201
JANUARY 1993, PAGES 167–188

ADAPTIVE STREAMLINE DIFFUSION FINITE ELEMENT METHODS FOR STATIONARY CONVECTION-DIFFUSION PROBLEMS

KENNETH ERIKSSON AND CLAES JOHNSON

ABSTRACT. Adaptive finite element methods for stationary convection- diffusion problems are designed and analyzed. The underlying discretization scheme is the Shock-capturing Streamline Diffusion method. The adaptive algorithms proposed are based on a posteriori error estimates for this method leading to *reliable* methods in the sense that the desired error control is guaranteed. A priori error estimates are used to show that the algorithms are *efficient* in a certain sense.

0. INTRODUCTION

The Streamline Diffusion method (SD-method for short) is a general finite element method for hyperbolic problems developed during the 1980s with applications in particular to convection-diffusion and compressible and incompressible flow problems (see [5–14, 18–19]). The SD-method is a generalization of the Standard Galerkin method obtained by two modifications. First, the test functions are modified by adding a multiple of a linearized form of the hyperbolic operator involved, which gives a weighted least squares control of the *residual* of the finite element solution (where the residual, roughly speaking, is the deviation from equality when the computed finite element solution is inserted into the given differential equation), and secondly, artificial viscosity of a particular form is added with the viscosity coefficient depending on the local mesh size and the absolute value of the local residual of the finite element solution. We refer to the first and second modification as *streamline diffusion* and *shock-capturing artificial viscosity*, respectively. The SD-method combines good stability with high accuracy, so that, e.g., shocks are resolved within few mesh points without under- or overshoots, and the precision in regions of smoothness of the exact solution is high. Extensive theoretical results are available, ranging from local error estimates for scalar linear convection problems to global error estimates for the incompressible Euler or Navier-Stokes equations with no lower bound on the viscosity. Further, convergence results for scalar conservation laws in several space dimensions and entropy consistency results for systems of conservation laws like the compressible Euler equations have been derived. Numerical results for a wide range of problems, including compressible and incompressible flow, are also available.

Received by the editor September 23, 1991 and, in revised form, December 31, 1991.
1991 *Mathematics Subject Classification.* Primary 65N15, 65N30.
This research was supported by the Swedish National Board for Technical Development (STUF).

In recent years we have developed adaptive finite element methods for elliptic and parabolic problems based on a posteriori error estimates (see [1–3]). In each of these cases we consider the following problem (P): Given a problem with exact solution u and a norm $\|\cdot\|$, design an efficient adaptive algorithm (A) for constructing a finite element mesh T such that

$$\|u - U\| \leq \text{TOL}, \tag{0.1}$$

where U is the finite element solution on the mesh T and $\text{TOL} > 0$ is the error tolerance. Clearly, our problem (P) has two ingredients: First, we want the adaptive algorithm (A) to be *reliable* in the sense that the error control (0.1) is guaranteed by the construction. Secondly, we want (A) to be *efficient* in the sense that, ideally, the constructed mesh T is nowhere overly refined, as compared to an *optimal* mesh $\tilde{T}$ which is a mesh with minimal degrees of freedom such that $\|u - \tilde{u}\| \leq \text{TOL}$, where $\tilde{u}$ is a standard nodal interpolant on $\tilde{T}$ of u. In [1–3] we have demonstrated theoretically and in numerical experiments that problem (P) may be solved (with varying degree of precision concerning the efficiency) in the case of linear model problems of elliptic and parabolic type. For such problems the adaptive algorithms are based on a posteriori error estimates of the form

$$\|u - U\| \leq \mathscr{E}(U, h, \text{data}), \tag{0.2}$$

where the error bound $\mathscr{E}$ depends on the computed solution U, the mesh size parameter h (which is here a function of space (and time)) and the data of the problem. The typical form of the a posteriori error bound (0.2) is as follows for an elliptic problem with piecewise linear basis functions and with $\|\cdot\|$, e.g., the L_2-norm:

$$\|u - U\| \leq C\|h^2 R(U)\|, \tag{0.3}$$

where $R(U)$ is the residual of U properly evaluated. The adaptive algorithm based on (0.2) seeks to construct a mesh T with mesh size h and corresponding discrete solution U such that $\mathscr{E}(U, h, \text{data}) \leq \text{TOL}$, usually by constructing a sequence of meshes T_j of mesh size h_j with corresponding solutions U_j, where T_{j+1} is constructed from U_j by equidistribution of the element contributions to $\mathscr{E}(U_j, h_{j+1}, \text{data})$ such that $\mathscr{E}(U_j, h_{j+1}, \text{data}) \lessapprox \text{TOL}$. Clearly, an adaptive algorithm based on an a posteriori error estimate of the form (0.2) will be *reliable* in the sense that if $\mathscr{E}(U, h, \text{data}) \leq \text{TOL}$, then (0.1) will be satisfied. To prove efficiency we have used a priori error estimates to bound $\mathscr{E}(U, h, \text{data})$ by quantities measuring relevant interpolation errors. Note that in order to demonstrate the efficiency of the adaptive algorithm, both the a posteriori error estimate (0.2) and the a priori error estimates used to bound $\mathscr{E}(U, h, \text{data})$ need to be (reasonably) sharp.

The main purpose of the present work is to extend our results on adaptive finite element methods for linear elliptic and parabolic problems to the SD-method for linear convection-diffusion problems in model form. In this case our adaptive algorithms will not be fully efficient in the above sense, compared to interpolation, but may be viewed as being *reasonably* efficient in the sense that in typical cases the meshes generated by the algorithms may be only mildly over-refined. As far as we know, the present work is the first to show that reliable and reasonably efficient adaptive error control based on a posteriori error estimates

is indeed possible also for hyperbolic problems. Previous adaptive techniques (see, e.g., [16, 17]) for hyperbolic problems such as convection-diffusion problems have been based either on ad hoc criteria suggested by interpolation error estimates, refining the mesh locally according to the size of the gradient or a difference quotient of the computed solution, or on simple a posteriori error estimates of the form

$$\|u - U\| \leq C\|R(U)\|, \tag{0.4}$$

where $\|\cdot\|$ is the L_2-norm and $R(U)$, again, the residual of the finite element solution U. In the first case the *reliability*, in particular, can be questioned since the relation between the mesh refinement criteria and the actual error is unclear, and in the second case the *efficiency* may be very low since in the interesting cases of a nonsmooth exact solution, $\|R(U)\|$ may tend to infinity as h tends to zero because $R(U)$ may be large (typically of order $O(h^{-1/2})$ or $O(h^{-1})$) in regions of nonsmoothness. Comparing (0.4) and (0.3), one notes the presence of the factor h^2 in (0.3), which reflects the ellipticity of the underlying problem in that case.

The a posteriori error estimate for the SD-method to be presented in this paper may be formulated roughly as follows:

$$\|u - U\| \leq C\|\min(1, R(U))\|, \tag{0.5}$$

where now the right-hand side may tend to zero (at close to optimal rate) as h tends to zero, leading to *reliable* and *efficient* adaptive methods. In the proofs of the a posteriori error estimates for the SD-method and the reliability and efficiency of the associated adaptive methods, we use in an essential way the special features of the SD-method, both the streamline diffusion modification and the shock-capturing modification. Both modifications were originally designed from stability and accuracy considerations without having adaptivity in mind, but this paper shows that the SD-method in fact has the basic features required to make reliable and efficient adaptive error control possible (which is not the case, e.g., for the Standard Galerkin method for convection-diffusion problems). In particular, we note that the residual plays a fundamental role in both the design of the SD-method and in the a posteriori error estimates.

The proofs of the a posteriori error estimates follow the same general pattern in both the elliptic and parabolic cases, and also in the present hyperbolic case: An error representation formula involving the computed discrete solution and the exact solution of an associated dual problem is established, and the error is estimated in terms of the residual of the finite element solution and the local discretization parameter h, using the orthogonality present in the discrete equations and elliptic regularity of the dual problem. Note that the improved estimate (0.5), as compared to the standard estimate (0.4), results from using the elliptic regularization built into the SD-method through the shock-capturing artificial viscosity (and not simply by cutoff and localization, cf. below), which gives a new way of viewing the advantages (and necessity) of elliptic regularization through artificial viscosity in hyperbolic problems.

Our long-term goal is to design adaptive algorithms with some degree of reliability and efficiency for complex hyperbolic problems such as the Navier-Stokes equations for compressible or incompressible flow. Formally, we may extend our techniques for proving a posteriori error estimates also to these

complex problems, by linearization and introducing again a certain continuous dual problem. The main technical problem is now to quantitatively estimate the stability of the dual solution, which in a certain sense reflects the stability of a linearized version of the given equations. In general (e.g., for systems in several dimensions), it seems impossible to establish the required stability estimates including certain solution-dependent constants by theoretical analysis, but it may still be possible to obtain the desired estimates by solving the dual problem numerically. Extensions of the results of this note to time-dependent linear convection-diffusion problems are given in [4]. Further extensions to adaptivity, including mesh orientation and stretching, will be presented in future work.

For numerical experiments based on the adaptive methods presented in this note we refer to [13].

The remaining part of this paper is organized as follows: In §1 we introduce the two stationary convection-diffusion type model problems to be considered and derive stability estimates for their solutions in terms of data. In §§2 and 3 we introduce the SD-method for the approximate solution of these problems and derive the a posteriori error estimates to be used in the final §4, where we formulate corresponding adaptive algorithms and discuss their reliability and efficiency.

1. Two model problems

We shall consider, in parallel, the two model problems

$$u_x - \operatorname{div}(\varepsilon \nabla u) = f \quad \text{in } \Omega, \tag{1.1a}$$

$$u = 0 \quad \text{on } \Gamma_- \cup \Gamma_0, \tag{1.1b}$$

$$\frac{\partial u}{\partial n} = 0 \quad \text{on } \Gamma_+, \tag{1.1c}$$

and

$$u_x - \operatorname{div}(\varepsilon \nabla u) = f \quad \text{in } \Omega, \tag{1.2a}$$

$$u = 0 \quad \text{on } \Gamma, \tag{1.2b}$$

where Ω is a bounded, convex polygonal domain in R^2 with boundary $\Gamma = \Gamma_- \cup \Gamma_0 \cup \Gamma_+$, $u_x = \partial u/\partial x$, $\partial u/\partial n = \nabla u \cdot n$, $n = (n_1, n_2)$ is the exterior unit normal to Γ, and f and $\varepsilon > 0$ are given data. The pieces Γ_-, Γ_0 and Γ_+ denote the parts of Γ where the x-component n_1 of n is negative, zero, and positive, respectively.

We shall mainly be concerned with the case when ε is small, in which case (1.1a) and (1.2a) models a stationary, convection-dominated, convection-diffusion type process with flow velocity (1, 0). Note that we are seeking estimates valid for arbitrary $\varepsilon > 0$ with, in particular, all constants appearing being independent of ε. For convenience, we shall assume that $\varepsilon \leq \frac{1}{2}$ in Ω, which, if not already satisfied, can be achieved by the change of variables $x' = sx$, $y' = sy$ with $s = 1/(2\bar{\varepsilon})$ and $\bar{\varepsilon} = \max_{\bar{\Omega}} \varepsilon$.

It is well known that the solutions of problems (1.1) and (1.2) may have singular layers of width $O(\sqrt{\varepsilon})$ along characteristics $\{(x, y) : y = y_0\}$ of the corresponding reduced equation with $\varepsilon = 0$. This will be the case, e.g., if f has a jump discontinuity along such a characteristic or, in a more general class

of problems, if there is a jump discontinuity in the boundary data along the 'inflow' part Γ_- of Γ. Moreover, the solution of problem (1.2), with Dirichlet boundary conditions all around the domain, will in general also have an 'outflow' singular layer of width $O(\varepsilon)$ along Γ_+.

Stability estimates. Below we shall use the following basic stability estimate for the problem (1.1), where the control of u_x and $\varepsilon D^2 u$ is of particular interest:

Lemma 1.1. *Assume there is a constant c such that*

$$-c \le \varepsilon_x \le c \min(1, \varepsilon) \quad \text{in } \Omega, \tag{1.3a}$$

$$|\varepsilon_y| \le c \min(1, \varepsilon^{1/2}) \quad \text{in } \Omega, \tag{1.3b}$$

and let u be the solution of (1.1). Then there is a constant $C = C(c, \Omega)$ such that

$$\|\varepsilon^{1/2}\nabla u\| + \|u\| + \|u_x\| + \|\varepsilon D^2 u\| + \left(\int_{\Gamma_+} u^2 n_1 d\Gamma\right)^{1/2} + \left(\int_{\Gamma} \varepsilon |\nabla u|^2 |n_1| d\Gamma\right)^{1/2} \le C\|f\|, \tag{1.4}$$

where $\|v\| = (\int_\Omega v^2 d\Omega)^{1/2}$ and $D^2 u = (u_{xx}^2 + 2u_{xy}^2 + u_{yy}^2)^{1/2}$.

Proof. We first multiply (1.1a) by u and integrate over Ω to obtain, after integration by parts using (1.1b, c),

$$\frac{1}{2}\int_{\Gamma_+} u^2 n_1 d\Gamma + \|\varepsilon^{1/2}\nabla u\|^2 = (f, u) \le \|f\|\,\|u\|, \tag{1.5}$$

where $(\cdot, \cdot)$ denotes the $L_2(\Omega)$ inner product.

We then multiply (1.1a) by u_x and integrate to obtain

$$\|u_x\|^2 + (\varepsilon \nabla u, \nabla u_x) - \int_{\Gamma_-} \varepsilon \frac{\partial u}{\partial n} u_x d\Gamma = (f, u_x).$$

On Γ_-, where u is constant ($= 0$), we have that $u_x n_2 = u_y n_1$ and, consequently,

$$\frac{\partial u}{\partial n} u_x = (u_x n_1 + u_y n_2) u_x = |\nabla u|^2 n_1.$$

After integration by parts in the term $(\varepsilon \nabla u, \nabla u_x) = \frac{1}{2}\int_\Omega \varepsilon \frac{d}{dx}|\nabla u|^2 d\Omega$ we thus have

$$\|u_x\|^2 + \frac{1}{2}\int_\Gamma \varepsilon |\nabla u|^2 |n_1| d\Gamma = (f, u_x) + \frac{1}{2}(\varepsilon_x \nabla u, \nabla u).$$

We now use our assumption (1.3a) together with (1.5) and the fact that $\|u\| \le C\|u_x\|$ (since $u = 0$ on part of the boundary) to conclude that

$$\|u_x\|^2 + \frac{1}{2}\int_\Gamma \varepsilon |\nabla u|^2 |n_1| d\Gamma \le \|f\|(\|u_x\| + C\|u\|) \le C\|f\|\,\|u_x\|. \tag{1.6}$$

The desired estimate for $\|\varepsilon^{1/2}\nabla u\|$, $\|u\|$, $\|u_x\|$, and the boundary integrals in (1.4) now follows at once.

In order to estimate the term $\|\varepsilon D^2 u\|$ in (1.4) we first note that after integration by parts twice we have

$$\int_\Omega \varepsilon^2 u_{xy} u_{xy} d\Omega = \int_\Gamma \varepsilon^2 (u_{xy} n_2 - u_{yy} n_1) u_x d\Gamma + 2 \int_\Omega \varepsilon(\varepsilon_x u_{yy} - \varepsilon_y u_{xy}) u_x d\Omega + \int_\Omega \varepsilon^2 u_{yy} u_{xx} d\Omega = \mathrm{I} + \mathrm{II} + \mathrm{III}.$$

Here, by our assumptions (1.3ab), $|\mathrm{II}| \le C\|\varepsilon D^2 u\| \, \|u_x\|$. We shall now estimate the boundary integral I. On Γ_0, the integral vanishes, since $u_x = 0$. On Γ_+, we have that $u_x = -u_y n_2 / n_1$ and consequently $(u_{xy} n_2 - u_{yy} n_1) u_x = \frac{1}{2} \frac{n_2}{n_1} \frac{\partial}{\partial \tau} u_y^2$, where $\frac{\partial}{\partial \tau} = -n_2 \frac{\partial}{\partial x} + n_1 \frac{\partial}{\partial y}$ denotes the tangential derivative along Γ. For each (straight) line segment $\Gamma_+^{(i)}$ of Γ_+, we thus get

$$\int_{\Gamma_+^{(i)}} \varepsilon^2 (u_{xy} n_2 - u_{yy} n_1) u_x d\Gamma = \frac{1}{2} \frac{n_2}{n_1} \int_{\Gamma_+^{(i)}} \varepsilon^2 \frac{\partial}{\partial \tau} u_y^2 d\Gamma = -\frac{n_2}{n_1} \int_{\Gamma_+^{(i)}} \varepsilon \frac{\partial \varepsilon}{\partial \tau} u_y^2 d\Gamma.$$

Here we have used the fact that ∇u vanishes at the endpoints of $\Gamma_+^{(i)}$, or $n_2 = 0$ on $\Gamma_+^{(i)}$. Similarly, on Γ_- we have $u_x = u_y n_1 / n_2$ and, consequently, for each line segment $\Gamma_-^{(i)}$ of Γ_-,

$$\int_{\Gamma_-^{(i)}} \varepsilon^2 (u_{xy} n_2 - u_{yy} n_1) u_x d\Gamma = -\frac{1}{2} \frac{n_1}{n_2} \int_{\Gamma_-^{(i)}} \varepsilon^2 \frac{\partial}{\partial \tau} u_y^2 d\Gamma = \frac{n_1}{n_2} \int_{\Gamma_-^{(i)}} \varepsilon \frac{\partial \varepsilon}{\partial \tau} u_y^2 d\Gamma.$$

In the special case $n_2 = 0$ the integrand vanishes. Using our assumption (1.3ab), we now conclude that $|\mathrm{I}| \le C \int_\Gamma \varepsilon |\nabla u|^2 |n_1| d\Gamma$. Putting things together, we find that

$$\begin{aligned} \|\varepsilon D^2 u\|^2 &= \int_\Omega \varepsilon^2 (u_{xx}^2 + 2u_{xy}^2 + u_{yy}^2) dx dy \\ &= \int_\Omega \varepsilon^2 (u_{xx}^2 + 2u_{xx} u_{yy} + u_{yy}^2) dx dy + 2\mathrm{I} + 2\mathrm{II} \\ &\le \|\varepsilon \Delta u\|^2 + C \int_{\Gamma_+} \varepsilon |\nabla u|^2 n_1 d\Gamma + \frac{1}{2} \|\varepsilon D^2 u\|^2 + C \|u_x\|^2. \end{aligned}$$

Now using the differential equation (1.1a) and the assumptions (1.3a, b) together with our previous estimates (1.5) and (1.6), we conclude that

$$\|\varepsilon D^2 u\| \le C \left(\|f\| + \|u_x\| + \|\nabla \varepsilon \cdot \nabla u\| + \left(\int_{\Gamma_+} \varepsilon |\nabla u|^2 n_1 d\Gamma \right)^{1/2} \right) \le C \|f\|.$$

This completes the proof of Lemma 1.1. □

For problem (1.2) with Dirichlet data also along Γ_+, we have the following counterpart of Lemma 1.1:

Lemma 1.2. *Assume* (1.3), *let* u *be the solution of* (1.2), *and let* φ *be a 'cutoff' function such that* $0 \le \varphi \le 1$ *in* Ω, $\varphi = 0$ *on* Γ_+ *and*

(1.7a) $$0 \le -\varphi_x \le \max(1, \varphi/\varepsilon) \quad \text{in } \Omega,$$

(1.7b) $$|\varphi_y| \le \max(1, \varphi/\sqrt{\varepsilon}) \quad \text{in } \Omega,$$

(1.7c) $$\varphi \varepsilon_x \le c\varepsilon \quad \text{in } \Omega.$$

Then there is a constant C depending only on Ω and c such that

$$\|u\| + \|\nabla u\|_\varepsilon + \|u_x\|_\varphi + \left(\int_\Gamma \varepsilon\varphi|\nabla u|^2|n_1|d\Gamma\right)^{1/2} \le C\|f\|, \tag{1.8}$$

where $\|v\|_\varphi = (\int_\Omega \varphi v^2 d\Omega)^{1/2}$.

Proof. Multiplication of (1.2) by u now gives $\|\nabla u\|_\varepsilon^2 = (f, u) \le \|f\|\,\|u\|$.

Given a positive lower bound c for ε, we would have $\|u\| \le C\|\nabla u\|_\varepsilon$, and it would follow that

$$\|u\| + \|\nabla u\|_\varepsilon \le C\|f\|, \tag{1.9}$$

with $C = C(c)$. By considering the corresponding equation for the transformed dependent variable $v = e^{-x}u$ (with $0 < \varepsilon \le \frac{1}{2}$ as above) it is easy to see that, in fact, (1.9) holds with C independent of ε.

We now multiply (1.2a) by φu_x and integrate to obtain

$$\|u_x\|_\varphi^2 + (\varepsilon\nabla u, \varphi\nabla u_x) + \overbrace{(\varepsilon\nabla u, \nabla\varphi u_x)}^{\text{I}} \overbrace{- \int_{\Gamma_-} \varepsilon\varphi \frac{\partial u}{\partial n} u_x\, d\Gamma}^{\text{II}}$$

$$= (f, \varphi u_x) \le C\|f\|_\varphi^2 + \frac{1}{8}\|u_x\|_\varphi^2.$$

Here,

$$(\varepsilon\nabla u, \varphi\nabla u_x) = \overbrace{\frac{1}{2}\int_{\Gamma_-} \varphi\varepsilon|\nabla u|^2 n_1\, d\Gamma}^{\text{III}} + \overbrace{\frac{1}{2}\|\nabla u\|_{\varepsilon(-\varphi_x)}^2}^{\text{IV}} \overbrace{- \frac{1}{2}(\varepsilon_x\varphi\nabla u, \nabla u)}^{\text{V}}.$$

Consequently, using (1.7a,b), we obtain

$$\text{I} + \text{IV} = -\frac{1}{2}\|u_x\|_{\varepsilon(-\varphi_x)}^2 + (\varepsilon u_y, \varphi_y u_x) + \frac{1}{2}\|u_y\|_{\varepsilon(-\varphi_x)}^2 \ge -\frac{5}{8}\|u_x\|_\varphi^2 - C\|\nabla u\|_\varepsilon^2.$$

Finally, as in the proof of Lemma 1.1 we deduce that

$$\text{II} + \text{III} = \int_{\Gamma_-} \varepsilon\varphi\left(\frac{1}{2}|\nabla u|^2 n_1 - \frac{\partial u}{\partial n}u_x\right) d\Gamma = \frac{1}{2}\int_{\Gamma_-} \varepsilon\varphi|\nabla u|^2|n_1|\,d\Gamma \ge 0.$$

Together, our estimates now show that

$$\|u_x\|_\varphi + \left(\int_\Gamma \varepsilon\varphi|\nabla u|^2|n_1|d\Gamma\right)^{1/2} \le C(\|f\| + \|\nabla u\|_\varepsilon) \tag{1.10}$$

if we also take into account the boundary condition on φ. The desired estimate (1.8) now follows from (1.9) and (1.10). □

2. The streamline diffusion method

We now formulate the SD-method for the discretization of (1.1) and (1.2). Let $T = \{K\}$ be a partition of Ω into 'edge-to-edge' triangular elements K such that

$$ch_K^2 \le \int_K d\Omega, \quad \forall K \in T, \tag{2.1}$$

where h_K is the diameter of K and c is a positive constant. Depending on the problem under consideration, (1.1) or (1.2), we define V by

$$V = \{v \in C(\Omega) : v|_K \text{ is linear in } (x, y), \forall K \in T, v|_{\Gamma_- \cup \Gamma_0} = 0\}, \tag{2.2a}$$

or

$$V = \{v \in C(\Omega) : v|_K \text{ is linear in } (x, y), \forall K \in T, v|_{\Gamma} = 0\}, \tag{2.2b}$$

and seek $U \in V$ such that

$$B(U, v) = L(v), \quad \forall v \in V, \tag{2.3}$$

where

$$B(w, v) = (w_x, v + \delta v_x) + (\hat{\varepsilon}\nabla w, \nabla v) - (\operatorname{div}(\hat{\varepsilon}\nabla w), \delta v_x)_T,$$

$$L(v) = (f, v + \delta v_x), \quad (w, v)_T = \sum_{K \in T} \int_K wv \, d\Omega,$$

$$\delta = c_1 \max(0, h - \hat{\varepsilon}), \tag{2.4}$$

$$\hat{\varepsilon}(U, h) = \max(\varepsilon, c_2 h^2 |f - U_x|), \tag{2.5}$$

h being the mesh function defined by $h|_K = h_K$, and c_1 and c_2 positive constants. (Concerning the definition of $\hat{\varepsilon}$, cf. Remark 1.2 below.)

Note that in general, since $\hat{\varepsilon}$ depends on U (unless $\hat{\varepsilon} = \varepsilon$), the discrete problem (2.3) is nonlinear, even though the continuous problems (1.1) and (1.2) are linear. In practice, when iterative methods are used to solve the discrete equations, the additional complication due to the nonlinearity introduced by $\hat{\varepsilon}$ is small (cf. below).

For technical reasons we shall assume that the modified diffusion coefficient $\hat{\varepsilon}$ does not vary too abruptly from one element to another. In particular, we shall assume that for some constant C independent of K,

$$\max_K \hat{\varepsilon} \leq C \min_{N(K)} \hat{\varepsilon}, \tag{2.6}$$

where $N(K)$ denotes the neighborhood of K consisting of the elements K' sharing at least one node with K. Note that by smoothing of $\hat{\varepsilon}$ by local averaging we may guarantee that (2.6) holds. Below we shall make further regularity assumptions on $\hat{\varepsilon}$, which may require additional smoothing. For simplicity we shall assume that $\hat{\varepsilon}$ already as defined by (2.5) has the desired regularity properties and leave the analysis of the general case to future work. Note that smoothing of $\hat{\varepsilon}$ also improves the convergence properties of iterative methods when solving the nonlinear discrete problem (2.3).

3. A POSTERIORI ERROR ESTIMATES

In this section we shall derive a posteriori error estimates for the discretization method (2.3) for the problems (1.1) and (1.2). It is then natural to split the error $u - U$ into two parts, $\rho = u - \hat{u}$ and $\theta = \hat{u} - U$, where $\hat{u}$ is the solution of the problem obtained by replacing ε by $\hat{\varepsilon}$ in (1.1) and (1.2), respectively, so that $\hat{u}$ satisfies

$$\hat{u}_x - \operatorname{div}(\hat{\varepsilon}\nabla\hat{u}) = f \text{ in } \Omega, \quad \hat{u} = 0 \text{ on } \Gamma_- \cup \Gamma_0, \quad \frac{\partial \hat{u}}{\partial n} = 0 \text{ on } \Gamma_+, \tag{3.1}$$

or

$$\hat{u}_x - \operatorname{div}(\hat{\varepsilon}\nabla\hat{u}) = f \text{ in } \Omega, \quad \hat{u} = 0 \text{ on } \Gamma. \tag{3.2}$$

Note that $\rho = u - \hat{u}$ is the difference between the solutions of two continuous problems with different diffusion coefficients, and that $\theta = \hat{u} - U$ is the error in an SD-approximation of (3.1) or (3.2) in a case when the shock-capturing viscosity is equal to the given viscosity, and thus the discrete problem may be viewed as being linear. We remark that we may view $\hat{u}$ as a regularization of the exact solution u such that the current mesh is sufficiently fine to resolve all details of $\hat{u}$, whereas if $\hat{u} \neq u$ (i.e., if $\hat{\varepsilon} > \varepsilon$), then some details of u may be left unresolved. Since the mesh fits with the regularity of $\hat{u}$ we can prove almost optimal a posteriori estimates for $\theta = \hat{u} - U$, using the elliptic regularization built in through the artificial viscosity $\hat{\varepsilon}$, whereas for $\rho = u - \hat{u}$ we will obtain somewhat less precise results. In the next section, we shall formulate different adaptive methods based on controlling each of the error bounds for ρ and θ on, say, the tolerance level TOL/2. Alternatively, we shall force ρ to be zero by refining until $\hat{\varepsilon} = \varepsilon$, in which case $\hat{u} = u$. Clearly, in the second case there is no need to estimate ρ and it suffices to estimate θ.

Below we shall denote by $D_h^2 U$ the piecewise constant function defined by

$$D_h^2 U|_K = \left(\frac{1}{2} \sum_{\tau \in \partial K} (|[\nabla U]_\tau| / h_\tau)^2 \right)^{1/2}, \qquad K \in T, \tag{3.3}$$

where h_τ is the length of edge τ of K and $[\,]_\tau$ denotes the jump across τ. Note that $D_h^2 U$ may be viewed as a discrete counterpart of $D^2 u$. Below we shall also be using the notation

$$\min{}^*(1, s) = \begin{cases} 1 & \text{in } K \text{ if } \partial K \cap \Gamma_- \neq \varnothing, \\ \min(1, s) & \text{otherwise.} \end{cases}$$

We now first consider the case of Neumann data along the outflow boundary. For the θ-part of the error we then have the following estimate:

Lemma 3.1. *Assume (cf. Remark 3.1 below) that* $|\hat{\varepsilon}_x| \leq c \min(1, \hat{\varepsilon})$ *and* $|\hat{\varepsilon}_y| \leq c \min(1, \hat{\varepsilon}^{1/2})$ *in* Ω. *Let* $\hat{u}$ *be the solution of* (3.1) *and let* $U \in V$ *with* V *defined as in* (2.2a) *be the corresponding discrete solution determined by* (2.3). *Then there is a constant* C *such that*

$$\|\hat{u} - U\| \leq \mathscr{E}_\theta(U, h, f), \tag{3.4}$$

where

$$\mathscr{E}_\theta(U, h, f) = C \left(\| \min{}^*(1, h^2 \hat{\varepsilon}^{-1}) R(U) \| + \max_{\Gamma_- \cup \Gamma_+} \hat{\varepsilon}^{1/2} \|f\| \right),$$

$$R(U) = r(U) + \hat{\varepsilon} D_h^2 U \quad \text{and} \quad r(U) = |f - U_x + \nabla\hat{\varepsilon} \cdot \nabla U|.$$

Proof. With $\theta = \hat{u} - U$, let z be the solution of the associated continuous 'dual' problem

$$-z_x - \operatorname{div}(\hat{\varepsilon}\nabla z) = \theta \text{ in } \Omega, \quad z = 0 \text{ on } \Gamma_+ \cup \Gamma_0, \quad \frac{\partial z}{\partial n} = 0 \text{ on } \Gamma_-. \tag{3.5}$$

Then

$$\|\theta\|^2 = (\theta, -z_x - \operatorname{div}(\hat{\varepsilon}\nabla z)) = (\theta_x, z) + (\hat{\varepsilon}\nabla\theta, \nabla z) - \int_{\Gamma_+} \hat{\varepsilon}\theta\frac{\partial z}{\partial n}d\Gamma$$

$$= (\hat{u}_x, z) + (\hat{\varepsilon}\nabla\hat{u}, \nabla z) - (U_x, z) - (\hat{\varepsilon}\nabla U, \nabla z) - \int_{\Gamma_+} \hat{\varepsilon}\theta\frac{\partial z}{\partial n}d\Gamma$$

$$= (f, z) + \int_{\Gamma_-} \hat{\varepsilon}\frac{\partial\hat{u}}{\partial n}z d\Gamma - (U_x, z) - (\hat{\varepsilon}\nabla U, \nabla z) - \int_{\Gamma_+} \hat{\varepsilon}\theta\frac{\partial z}{\partial n}d\Gamma.$$

Using (2.3), we get for any interpolant $\tilde{z} \in V$ of z,

$$\|\theta\|^2 = (f - U_x, z - \tilde{z} - \delta\tilde{z}_x) - (\hat{\varepsilon}\nabla U, \nabla(z - \tilde{z})) - (\nabla\hat{\varepsilon}\cdot\nabla U, \delta\tilde{z}_x)$$
$$+ \int_{\Gamma_-} \hat{\varepsilon}\frac{\partial\hat{u}}{\partial n}z d\Gamma - \int_{\Gamma_+} \hat{\varepsilon}\theta\frac{\partial z}{\partial n}d\Gamma$$

$$= (f - U_x + \nabla\hat{\varepsilon}\cdot\nabla U, z - \tilde{z} - \delta\tilde{z}_x) + \sum_K \sum_{\tau\in\partial K} \int_\tau \hat{\varepsilon}\frac{\partial U}{\partial n_K}(z - \tilde{z})d\tau$$
$$+ \int_{\Gamma_-} \hat{\varepsilon}\frac{\partial\hat{u}}{\partial n}z d\Gamma - \int_{\Gamma_+} \hat{\varepsilon}\theta\frac{\partial z}{\partial n}d\Gamma = \mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV},$$

where n_K denotes the exterior unit normal to ∂K, $K \in T$.

Let us first estimate the term I. We note that on one hand,

$$|z - \tilde{z} - \delta\tilde{z}_x| \le |z| + |\tilde{z}| + \delta|\tilde{z}_x|,$$

and on the other hand,

$$|z - \tilde{z} - \delta\tilde{z}_x| \le Ch^2\hat{\varepsilon}^{-1}(h^{-2}\hat{\varepsilon}|z - \tilde{z}| + h^{-1}\hat{\varepsilon}|\tilde{z}_x - z_x| + |z_x|).$$

Here we have used the fact that $\delta \le C\min(h, h^2\hat{\varepsilon}^{-1})$, since $\delta = 0$ whenever $\hat{\varepsilon} \ge h$. We recall that the interpolant $\tilde{z}$ of z may be defined so that

$$\|\tilde{z}\|_K \le C\|z\|_{N(K)}, \tag{3.6a}$$

$$\|z - \tilde{z}\|_K \le Ch_K^i\|D^i z\|_{N(K)}, \qquad i = 1, 2, \tag{3.6b}$$

$$\|\nabla(z - \tilde{z})\|_K \le Ch_K\|D^2 z\|_{N(K)}, \tag{3.6c}$$

$$\|\tilde{z}_x\|_K \le Ch_K^{-1}\|\tilde{z}\|_K, \tag{3.6d}$$

where $D^1 = \nabla$ and $N(K)$ is defined as above. Note that in order to be able to estimate the interpolant in terms of the function values as in (3.6a), $\tilde{z}$ is defined from local averages of z around each nodal point. This explains why the norms on the right-hand side in the estimates (3.6a–c) have to be taken over a small neighborhood of K and not just over K. Note also that (2.1) gives an upper bound (depending on c) on the number of elements in $N(K)$ for any K. The last inequality (3.6d) is an 'inverse' estimate based on the fact that $\tilde{z}$ is a polynomial on each K.

From the above properties of $\tilde{z}$ and our assumptions on $\hat{\varepsilon}$ which, in particular, imply (2.6), and the obvious counterpart of the regularity estimate (1.4) for the dual problem (3.5), we now get

$$|I| \le C\|\min^*(1, h^2\hat{\varepsilon}^{-1})r(U)\|(\|z\| + \|\tilde{z}\| + \|\delta\tilde{z}_x\| + \|h^{-2}\hat{\varepsilon}(z - \tilde{z})\|$$
$$+ \|h^{-1}\hat{\varepsilon}(z_x - \tilde{z}_x)\| + \|z_x\|)$$
$$\le C\|\min^*(1, h^2\hat{\varepsilon}^{-1})r(U)\|(\|z\| + \|\hat{\varepsilon}D^2 z\| + \|z_x\|)$$
$$\le C\|\min^*(1, h^2\hat{\varepsilon}^{-1})r(U)\|\,\|\theta\|.$$

Here we have taken into account the fact that the estimates (3.6b, c) are not valid on the elements along Γ_-, since $\tilde{z}$, as a function in V, has to vanish there, and this is not the case with z in general.

Further, as in the proof of Lemma 4.3 of [3], we have

$$\begin{aligned}|\mathrm{II}| &\le \left|\sum_\tau \int_\tau \hat{\varepsilon}\left[\frac{\partial U}{\partial n_\tau}\right]_\tau (z-\tilde{z})d\tau\right| \\ &\le E(U)\left(\sum_\tau \int_\tau \max^*(1, \hat{\varepsilon}h_\tau^{-2})h_\tau(z-\tilde{z})^2 d\tau\right)^{1/2} \\ &\le CE(U)\|\max^*(1, \hat{\varepsilon}h^{-2})(|z-\tilde{z}| + h|\nabla z - \nabla\tilde{z}|)\| \\ &\le CE(U)(\|z\| + \|\hat{\varepsilon}D^2 z\|) \le CE(U)\|\theta\|,\end{aligned}$$

where $\max^*(1, s)$ is the obvious counterpart of $\min^*(1, s)$ and

$$E(U) = \left(\sum_\tau \int_\tau \min^*(1, h_\tau^2\hat{\varepsilon}^{-1})h_\tau^{-1}\left(\hat{\varepsilon}\left[\frac{\partial U}{\partial n_\tau}\right]_\tau\right)^2 d\tau\right)^{1/2}.$$

Here the sum is taken over all edges which are not part of $\Gamma_0 \cup \Gamma_+$ (where $z - \tilde{z} = 0$), and for the edges along Γ_- we set $[\frac{\partial U}{\partial n_\tau}] = \frac{\partial U}{\partial n}$. As in [3, Remark 2.3], the sum $E(U)$ may be estimated in terms of $D_h^2 U$ defined by (3.3), and we have that

$$E(U) \le C\|\min^*(1, h^2\hat{\varepsilon}^{-1})\hat{\varepsilon}D_h^2 U\|.$$

It now remains to estimate the boundary integrals III and IV. From Lemma 1.1 we have

$$\begin{aligned}|\mathrm{III}| &\le C\max_{\Gamma_-}\hat{\varepsilon}^{1/2}\left(\int_{\Gamma_-}\hat{\varepsilon}\left(\frac{\partial\hat{u}}{\partial n}\right)^2|n_1|d\Gamma\right)^{1/2}\left(\int_{\Gamma_-} z^2|n_1|d\Gamma\right)^{1/2} \\ &\le C\max_{\Gamma_-}\hat{\varepsilon}^{1/2}\|f\|\,\|\theta\|,\end{aligned}$$

where C may depend on Ω, for instance, on the upper bound for $-n_1$ on Γ_-.

In order to estimate IV, we first note that by the triangle inequality,

$$\int_{\Gamma_+}\theta^2 d\Gamma \le C\int_{\Gamma_+}(\hat{u}^2 + U^2)n_1 d\Gamma,$$

where again $C = C(\Omega)$. From Lemma 1.1 we know that $\int_{\Gamma_+}\hat{u}^2 n_1 d\Gamma \le C\|f\|^2$. By putting $v = U$ in (2.3) we find similarly, using our assumptions on $\hat{\varepsilon}$, that

$$\begin{aligned}\frac{1}{2}\int_{\Gamma_+} U^2 n_1 d\Gamma + \|\nabla U\|_{\hat{\varepsilon}}^2 + \|U_x\|_\delta^2 &\le (f, U + \delta U_x) + (\nabla\hat{\varepsilon}\cdot\nabla U, \delta U_x) \\ &\le \|f\|(\|U\| + \|\delta U_x\|) + c\max_{\bar{\Omega}}\delta^{1/2}\|\nabla U\|_{\hat{\varepsilon}}\|U_x\|_\delta.\end{aligned}$$

Since δ is small, we may here use a standard kick-back argument to conclude that

$$\int_{\Gamma_+} U^2 n_1 d\Gamma \le C(\|f\|(\|U\| + \|\delta U_x\|)) \le C\|f\|\,\|U\|,$$

where we have also used a counterpart of (3.6d) for U. Since $U = \hat{u} - \theta$, and $\hat{u}$ may be estimated in terms of f using Lemma 1.1, we have that $\|U\| \leq C(\|f\| + \|\theta\|)$. By Lemma 1.1, we further have that

$$\int_{\Gamma_+} \hat{\varepsilon} \left(\frac{\partial z}{\partial n}\right)^2 d\Gamma \leq C\|\theta\|^2.$$

We may thus conclude that

$$|\mathrm{IV}| \leq C \max_{\Gamma_+} \hat{\varepsilon}^{1/2}(\|f\|^2 + \|f\|\,\|\theta\|)^{1/2}\|\theta\| \leq \frac{1}{2}\|\theta\|^2 + C \max_{\Gamma_+} \hat{\varepsilon}\|f\|^2,$$

where $C = C(\Omega)$.

Putting all these estimates together, we find that

$$\|\hat{u} - U\| \leq C \Big(\|\min^*(1, h^2\hat{\varepsilon}^{-1})r(U)\| + \|\min^*(1, h^2\hat{\varepsilon}^{-1})\hat{\varepsilon}D_h^2 U\| + \max_{\Gamma_- \cup \Gamma_+} \hat{\varepsilon}^{1/2}\|f\| \Big).$$

Clearly, by estimating the two terms I and II simultaneously, we can derive the somewhat more precise estimate (3.4). This completes the proof of Lemma 3.1. □

We now turn to the case of problem (1.2) with Dirichlet outflow boundary data. The counterpart of the estimate (3.4) for the θ-part of the error then reads:

Lemma 3.2. *Assume* $\operatorname{dist}(\Gamma_-, \Gamma_+) > 0$; *let* $\hat{u}$ *be the solution of* (3.2) *and let* $U \in V$ *be the corresponding discrete solution determined by* (2.3) *with* V *as in* (2.2b). *Then, under the assumptions on* $\hat{\varepsilon}$ *of Lemma* 3.1 (*cf. Remark* 3.1 *below*), *there is a constant* C *such that*

$$\|\hat{u} - U\| \leq \mathscr{E}_\theta(U, h, f), \tag{3.7}$$

where $\mathscr{E}_\theta(U, h, f) = C(\|\min^*(1, h^2\hat{\varepsilon}^{-1})R(U)\| + \max_{\Gamma_-} \hat{\varepsilon}^{1/2}\|f\|)$.

Proof. The proof is essentially the same as that of Lemma 3.1. The only differences are that now the boundary integral IV vanishes, so that the 'max' is taken over Γ_- only, and that in order to be able to estimate $\partial\hat{u}/\partial n$ on Γ_- we now need to assume that Γ_- and Γ_+ are separated so as to be able to find a cutoff function φ as in Lemma 1.2 with $\varphi = 1$, say, on Γ_-. □

Remark 3.1. The assumptions in Lemmas 3.1 and 3.2 on $\hat{\varepsilon}$ may be considerably relaxed. For instance, it suffices to assume, in addition to (1.3b), that the inequality $\hat{\varepsilon}_x \leq c\hat{\varepsilon}$ (which we cannot expect to hold near an outflow singular layer) holds in a neighborhood of Γ_-, since this condition is used *only* to bound $(\partial u/\partial n)|_{\Gamma_-}$, and that $-c\min(1, \hat{\varepsilon}) \leq \hat{\varepsilon}_x \leq C$ in Ω. Note that in a typical application with an outflow boundary layer (of width $O(h)$) we expect to have $\hat{\varepsilon} = O(h^3)$ outside the layer and $\hat{\varepsilon} = O(h)$ in the layer, which is consistent with the stated requirements on $\hat{\varepsilon}$.

Remark 3.2. Note that in a singular layer where $|f - U_x|$ is large, we expect to have $|f - U_x| \approx R(U)$, and consequently, with $\hat{\varepsilon} = ch^2|f - U_x|$,

$$\min(1, h^2\hat{\varepsilon}^{-1})R(U) = \min(R(U), R(U)/(c|f - U_x|)) \leq C\min(R(U), 1).$$

The term involving $R(U)$ in the error bound $\mathscr{E}_\theta$ may therefore be small also in the presence of characteristic and outflow singular layers (of width $O(\hat{\varepsilon}^{1/2})$ and $O(\hat{\varepsilon})$, respectively) in which $R(U)$ is big, cf. (0.4), (0.5). Note that it is natural to use instead of (2.5) the following slightly different (implicit) definition of $\hat{\varepsilon}$: $\hat{\varepsilon} = \max(\varepsilon, c_2 h^2 R(U))$, involving the full residual $R(U)$ and not just the part $|f - U_x|$ as in (2.5). With this choice of $\hat{\varepsilon}$, we clearly have $\min(1, h^2\hat{\varepsilon}^{-1})R(U) \le C\min(R(U), 1)$.

We shall now proceed by deriving error bounds also for the $\rho = u - \hat{u}$ part of the error. We shall then first consider the case of Neumann outflow boundary data as in problem (1.1) and the problem of characteristic singular layers. In the following two lemmas we first derive a preliminary estimate for ρ in terms of $\hat{u}$, $\hat{\varepsilon}$ and data, and then complete this via an estimate for $\nabla\theta$ to a full a posteriori estimate for ρ. The proofs of these two lemmas can be found in the Supplement section.

Lemma 3.3. *Assume $f \in H_0^1(\Omega)$ and (for simplicity) that ε is constant in Ω. Let u and $\hat{u}$ be the solutions of* (1.1) *and* (3.1), *respectively. Then there is a constant C such that*

$$\|u - \hat{u}\| \le \widehat{\mathscr{E}}_\rho(\hat{u}, \hat{\varepsilon}, \text{data}), \tag{3.8}$$

where

$$\widehat{\mathscr{E}}_\rho(\hat{u}, \hat{\varepsilon}, \text{data}) = C(\|\sigma\nabla\hat{u}\|_{\hat{\varepsilon}} + \|\varphi\nabla f\|_{\hat{\varepsilon}}),$$

$\varphi = \hat{\varepsilon}^{-2}(\hat{\varepsilon} - \varepsilon)^2$ *and* $\sigma = \max(3|\hat{\varepsilon}^{-3/2}\varepsilon\nabla\hat{\varepsilon}| + 1 + (\varphi\hat{\varepsilon})_x, 0)$.

Lemma 3.4. *Let $\hat{u}$ be the solution of* (3.1), *and let U be determined by* (2.3) *with V as in* (2.2a). *Then, under the assumption* (2.6) *there is a constant C such that for $\theta = \hat{u} - U$*

$$\|\nabla\theta\|_{\hat{\varepsilon}} \le C\|\min^*(1, h\hat{\varepsilon}^{-1/2})R(U)\|. \tag{3.9}$$

We may now put the estimates of Lemmas 3.1, 3.3 and 3.4 together to obtain a full a posteriori estimate for $u - U$ in the case of Neumann outflow boundary data:

Theorem 3.1. *Assume $f \in H_0^1(\Omega)$ and that ε is constant in Ω, and let u and U be the solutions of* (1.1) *and* (2.3), *respectively, with V defined as in* (2.2a). *Then, under the assumptions of Lemma* 3.1 *on $\hat{\varepsilon}$, there are constants C such that*

$$\|u - U\| \le \mathscr{E}_\theta(U, h, f) + \mathscr{E}_\rho(U, h, f), \tag{3.10}$$

where

$$\mathscr{E}_\rho(U, h, f) = C(\|\min^*(1, h\hat{\varepsilon}^{-1/2})R(U)\| + \|\nabla U\|_{\hat{\varepsilon}} + \|\varphi\nabla f\|_{\hat{\varepsilon}}),$$

$\varphi = \hat{\varepsilon}^{-2}(\hat{\varepsilon} - \varepsilon)^2$, *and $\mathscr{E}_\theta$ is defined as in Lemma* 3.1.

Proof. Under the assumptions of Lemma 3.1, the function σ defined in Lemma 3.3 is bounded uniformly in Ω, so that by the triangle inequality,

$$\|\sigma\nabla\hat{u}\|_{\hat{\varepsilon}} \le C(\|\nabla\theta\|_{\hat{\varepsilon}} + \|\nabla U\|_{\hat{\varepsilon}}),$$

and consequently, by Lemmas 3.3 and 3.4,

$$\begin{aligned} \|u - \hat{u}\| &\le \mathscr{E}_\rho(U, h, f) \\ &= C(\|\min^*(1, h\hat{\varepsilon}^{-1/2})R(U)\| + \|\nabla U\|_{\hat{\varepsilon}} + \|\varphi\nabla f\|_{\hat{\varepsilon}}). \end{aligned} \tag{3.11}$$

The a posteriori error bound for the $\theta = \hat{u} - U$ part of the error is given directly by Lemma 3.1. This completes the proof. □

It is possible to extend Theorem 3.1 to the case $f \notin H_0^1(\Omega)$ by simply replacing f by an appropriate approximation $\tilde{f} \in H_0^1(\Omega)$ and using the L_2 stability in problem (1.1) to get the result:

Theorem 3.2. *Let u be the solution of* (1.1) *and let $U \in V$ be determined by* (2.3), *with V as in* (2.2a) *and f replaced by any $\tilde{f} \in H_0^1(\Omega)$. Then there is a constant C such that*

$$\|u - U\| \leq \mathscr{E}_\theta(U, h, \tilde{f}) + \tilde{\mathscr{E}}_\rho(U, h, f, \tilde{f}), \tag{3.12}$$

where

$$\tilde{\mathscr{E}}_\rho(U, h, f, \tilde{f}) = C(\|\min^*(1, h\hat{\varepsilon}^{-1/2})R(U)\| + \|\varphi\nabla\tilde{f}\|_{\hat{\varepsilon}} + \|f - \tilde{f}\| + \|\nabla U\|_{\hat{\varepsilon}}),$$

$$R(U) = |\tilde{f} - U_x + \nabla\hat{\varepsilon}\cdot\nabla U| + \hat{\varepsilon}D_h^2 U \text{ and } \hat{\varepsilon} = \max(\varepsilon, ch^2|\tilde{f} - U_x|).$$

Proof. Let $\tilde{u}$ be the solution of (1.1) with f replaced by $\tilde{f}$. Then by Lemma 1.1,

$$\|u - \tilde{u}\| \leq C\|f - \tilde{f}\|,$$

and from Theorem 3.1 we get the desired estimate for $\tilde{u} - U$. □

We now turn our attention to problem (1.2) with Dirichlet data along Γ_+ and the additional problem of an outflow singular layer. We note at once that in this case the estimate of Lemma 3.3 is not sharp, since we expect to have $|\nabla\hat{u}| = O(\hat{\varepsilon}^{-1})$ in an outflow layer and thus in this case $\|\sigma\nabla\hat{u}\|_{\hat{\varepsilon}} = O(\hat{\varepsilon}^{-1/2})$. One possible way of deriving analogues of Theorems 3.1 and 3.2 for problem (1.2) would then be to replace the L_2-norm estimate of Lemma 3.3 by an estimate in a weighted L_2 norm, using $O(h)$ *cutoff* at the outflow boundary, together with a separate estimate for the boundary layer error using maximum-norm error control of $\hat{u}$ and U. Although probably feasible, such a procedure has the disadvantage of not being 'automatic', requiring, in particular, the specification of an appropriate (problem-dependent) cutoff procedure and a special treatment of the outflow layer. We shall therefore consider another possibility of replacing Lemma 3.3 by a sharper estimate in terms of $\hat{u}$, where, however, the step replacing $\hat{u}$ by U leading to an adequate full a posteriori error estimate is technically more complex and therefore will be omitted in the present paper. We shall thus derive a sharp estimate for ρ in terms of $\hat{u}$, $\hat{\varepsilon}$ and data, as before, and base an adaptive method for ρ directly on this estimate by simply replacing the unknown argument $\hat{u}$ by the computed solution U, leaving the problem of deriving a full a posteriori error estimate for ρ for future work. We note that since the finite element mesh for U fits with the regularity of $\hat{u}$, it is natural to expect that replacing $\hat{u}$ by U in this way is possible, whereas a direct replacement of u by U in an a priori error estimate involving u may be more difficult to justify.

Alternatively, as we shall see below, it is possible to derive an adaptive algorithm for full error control simply by adding $\hat{\varepsilon} = \varepsilon$ as an additional control, in which case, of course, $\rho = 0$.

The sharper estimate replacing Lemma 3.3 reads as follows:

Lemma 3.5. *Assume* $\operatorname{dist}(\Gamma_-, \Gamma_+) > 0$ *and that (for simplicity)* ε *is constant,* $|\hat{\varepsilon}_y| \le c\hat{\varepsilon}^{1/2}$, *and* $|\hat{\varepsilon}_x| \le c\hat{\varepsilon}$ *in* Ω (*cf. Remark* 3.1), *and let* u *and* $\hat{u}$ *be the solutions of* (1.2) *and* (3.2), *respectively. Then there is a constant* C *such that*

$$\|u - \hat{u}\| \le \hat{\mathscr{E}}_\rho(\hat{u}, \hat{\varepsilon}, \text{data}), \tag{3.13}$$

where

$$\hat{\mathscr{E}}_\rho(\hat{u}, \hat{\varepsilon}, \text{data}) = C\left(\|(\hat{\varepsilon} - \varepsilon)\hat{u}_x\| + \|d_+((\hat{\varepsilon} - \varepsilon)\hat{u}_y)_y\| + \max_{\Gamma_-} \hat{\varepsilon}^{1/2}\|f\|\right),$$

and d_+ *is the distance to the outflow boundary* Γ_+ *in the direction* $(1, 0)$.

Proof. As in the proof of Lemma 3.3 (see the Supplement section), we have that

$$\rho_x - \operatorname{div}(\varepsilon\nabla\rho) = \phi \quad \text{in } \Omega,$$

where $\phi = -\operatorname{div}((\hat{\varepsilon} - \varepsilon)\nabla\hat{u})$. Let z be the solution of

$$-z_x - \operatorname{div}(\varepsilon\nabla z) = \rho \text{ in } \Omega, \quad z = 0 \text{ on } \Gamma_+ \cup \Gamma_0, \quad \frac{\partial z}{\partial n} = 0 \text{ on } \Gamma_-. \tag{3.14}$$

We then find that

$$\begin{aligned}\|\rho\|^2 &= (\rho, -z_x - \operatorname{div}(\varepsilon\nabla z)) = (\rho_x - \operatorname{div}(\varepsilon\nabla\rho), z) + \mathrm{I} \\ &= (-\operatorname{div}((\hat{\varepsilon} - \varepsilon)\nabla\hat{u}), z) + \mathrm{I} = ((\hat{\varepsilon} - \varepsilon)\hat{u}_x, z_x) \\ &\quad - (((\hat{\varepsilon} - \varepsilon)\hat{u}_y)_y, z) + \mathrm{I} + \mathrm{II},\end{aligned}$$

where

$$\mathrm{I} = \int_{\Gamma_-} \varepsilon \frac{\partial\rho}{\partial n} z\, d\Gamma, \qquad \mathrm{II} = -\int_{\Gamma_-} (\hat{\varepsilon} - \varepsilon)\hat{u}_x z n_1\, d\Gamma.$$

Since z vanishes along Γ_+, we have that

$$z(x, y) = -\int_{x_+}^{x} z_x(s, y)ds,$$

where (x_+, y) is a point on Γ_+. From the boundedness in L_2 of the Hilbert transform $F(x) = \frac{1}{x}\int_0^x f(s)ds$, we easily get that $\|d_+^{-1}z\| \le 2\|z_x\|$. We now need to estimate I and II. Using Lemma 1.2, we have that

$$\begin{aligned}|\mathrm{I} + \mathrm{II}| &\le C \max_{\Gamma_-} \hat{\varepsilon}^{1/2} \left(\int_{\Gamma_-} (\varepsilon|\nabla u|^2 + \hat{\varepsilon}|\nabla\hat{u}|^2)|n_1|d\Gamma\right)^{1/2} \left(\int_{\Gamma_-} z^2|n_1|d\Gamma\right)^{1/2} \\ &\le C \max_{\Gamma_-} \hat{\varepsilon}^{1/2}\|f\| \left(\int_{\Gamma_-} z^2|n_1|d\Gamma\right)^{1/2},\end{aligned}$$

where $C = C(\Omega)$. Putting things together, we now obtain

$$\begin{aligned}\|\rho\|^2 \le C &\left(\|(\hat{\varepsilon} - \varepsilon)\hat{u}_x\| + \|d_+((\hat{\varepsilon} - \varepsilon)\hat{u}_y)_y\| + \max_{\Gamma_-} \hat{\varepsilon}^{1/2}\|f\|\right) \\ &\times \left(\|z_x\| + \left(\int_{\Gamma_-} z^2|n_1|d\Gamma\right)^{1/2}\right).\end{aligned}$$

If we now use the counterpart of (1.4) for the solution z of the dual problem (3.14), we obtain the desired estimate at once. This completes the proof. □

Let us now put the estimates of Lemmas 3.2 and 3.5 together to obtain the following result:

Theorem 3.3. *Under the assumptions of Lemmas* 3.2 *and* 3.5 *there is a constant* C *such that*

$$\|u - U\| \leq \mathscr{E}_\theta(U, h, f) + \hat{\mathscr{E}}_\rho(\hat{u}, \hat{\varepsilon}, \text{data}). \tag{3.15}$$

Note that this is not a full a posteriori estimate, since $\hat{u}$ is unknown. As indicated above, we shall design an adaptive algorithm for the SD-method for (1.2) based simply on replacing $\hat{u}$ by U in (3.15). To prove that this is possible, leading to a reliable algorithm, would require a counterpart of Lemma 3.4 in certain weighted norms, the technicalities of which we leave to future investigations. Thus, in this paper we do not prove that the adaptive algorithm based on (3.15) to be presented below is fully reliable.

Remark 3.3. As indicated in the introduction, it is easy to prove a posteriori error estimates of the form

$$\|u - U\| \leq C\|R(U)\|, \tag{3.16}$$

for various Galerkin methods for convection-diffusion problems, including the SD-method and also standard Galerkin methods, by using only the L_2-stability of the continuous (dual) problem (i.e., $\|u\| \leq C\|f\|$ for problem (1.1) or (1.2)). However, an a posteriori error estimate of the form (3.16) cannot be used as a basis for an adaptive algorithm for controlling $\|u - U\|$ in general, since in the presence of characteristic or outflow layers we will have that $\|R(U)\| = O(h^{-\alpha})$ with $\alpha = 1/8$ or $\alpha = 1/2$ unless $h = O(\varepsilon^{2/3})$ or $h = O(\varepsilon)$, respectively (cf. the discussion in §4.1 below). In particular, (3.16) is useless in the initial stages of an adaptive procedure when the mesh is not yet properly refined.

4. Adaptive procedures

We shall now design adaptive algorithms for the SD-method for the problems (1.1) and (1.2) based on the error estimates of the previous section, seeking procedures for which we can demonstrate *reliability* and *efficiency* as discussed in the introduction. In each case our basic computational goal is to solve the following (optimization) problem (O): With TOL a given error tolerance, find, using the SD-method, an approximate solution U of problem (1.1) or (1.2) on a mesh T such that

$$\|u - U\| \leq \text{TOL}, \tag{4.1}$$

at minimal computational 'cost' (here measured simply in terms of the total number of nodes of the mesh T). We recall that a mesh $\tilde{T}$ with minimal number of nodes such that $\|u - \tilde{u}\| \leq \text{TOL}$, where $\tilde{u}$ is an interpolant of u on $\tilde{T}$, is referred to as an *optimal* mesh. When discussing the efficiency of our adaptive procedures, we shall compare the constructed mesh T with the optimal mesh $\tilde{T}$.

4.1. **Neumann outflow boundary data.** We first consider the case of problem (1.1) without the singular outflow layer complication. Clearly, the adaptive method suggested by the a posteriori error estimates (3.10) and (3.12) is to seek a mesh T with (nearly) minimal number of nodes such that

$$\mathscr{E}_\theta(U, h, f) + \mathscr{E}_\rho(U, h, f) \leq \text{TOL} \tag{4.2}$$

or

$$\mathscr{E}_\theta(U, h, \tilde{f}) + \tilde{\mathscr{E}}_\rho(U, h, f, \tilde{f}) \le \mathrm{TOL}. \tag{4.3}$$

For the purpose of our discussion below we label this adaptive method (or strategy) (M_1). An *algorithm* designed for the search of the mesh T with minimal number of nodes satisfying (4.2) and the associated U could be designed roughly as follows:

1°. Start with a (coarse) quasi-uniform mesh T_0.

2°. Given a mesh T_j with mesh size h_j, compute the corresponding approximate solution $U_j \in V_j$ determined by (2.3), where V_j is the space of piecewise linear functions defined by (2.2a) based on T_j.

3°. If (4.2) or (4.3) holds with $U = U_j$ and $h = h_j$, then stop and accept U_j and T_j as an (approximate) solution of (O). Otherwise, construct a new mesh T_{j+1} with corresponding mesh size h_{j+1} with (approximately) as few nodes as possible such that

$$\mathscr{E}_\theta(U_j, h_{j+1}, f) + \mathscr{E}_\rho(U_j, h_{j+1}, f) \le \mathrm{TOL}, \tag{4.4}$$

and then go back to 2°.

In practice, to construct a mesh T_{j+1} with (approximately) as few degrees of freedom as possible, we seek an equidistributed mesh T_{j+1} in the sense that all element contributions in the integrals in the L_2-norms $\|\cdot\|$ in $\mathscr{E}_\theta$ and $\mathscr{E}_\rho$ are approximately equal (see [3] for details).

As we shall see below, the a posteriori error estimate of Theorem 3.2 does not appear to be quite sharp, and as a consequence the adaptive method (M_1) will possibly not be fully efficient. As an alternative we may therefore consider the following method based on the 'quasi' a posteriori error estimate obtained by combining the estimates of Lemmas 3.1 and 3.3. This gives us the following method (M_2): Seek a mesh T with (nearly) minimal number of nodes such that for the corresponding U,

$$\mathscr{E}_\theta(U, h, f) + \hat{\mathscr{E}}_\rho(U, \hat{\varepsilon}, \mathrm{data}) \le \mathrm{TOL}. \tag{4.5}$$

Note that this method is based on heuristically replacing the unknown argument $\hat{u}$ in $\hat{\mathscr{E}}_\rho$ by the known computed solution U.

As a third possibility, changing the computational goal somewhat, we consider the following method (M_3): Seek a mesh T with (nearly) minimal number of nodes such that for the corresponding U,

$$\mathscr{E}_\theta(U, h, f) \le \mathrm{TOL} \quad \text{and} \quad \hat{\varepsilon} = \hat{\varepsilon}(U, h) = \varepsilon. \tag{4.6}$$

We note that once $\hat{\varepsilon} = \varepsilon$, then $\rho = 0$, and thus (4.2) will be guaranteed if only $\mathscr{E}_\theta(U, h, f) \le \mathrm{TOL}$. As we will indicate below, it appears that the requirement $\hat{\varepsilon} = \varepsilon$ in (4.6) will force the refinement to continue until all details of the flow have been resolved to their true scale, in particular, the mesh size will be smaller than $O(\sqrt{\varepsilon})$ in a characteristic layer. This is not necessarily the case with the methods (M_1) and (M_2) where, depending on the tolerance chosen and the given diffusion coefficient ε, characteristic layers may be left unresolved.

Remark. In the implementation of the adaptive algorithm 1° – 3° above one faces, in particular, the problem of assigning appropriate values to the constants C appearing in the definitions of $\mathscr{E}_\theta$ and $\mathscr{E}_\rho$. Clearly, for efficiency reasons one

would like to choose these constants as small as possible. Consider for instance the constant C in the definition of $\mathscr{E}_\theta$ in Lemma 3.1. Tracing the origin of this constant, one realizes, first of all, that one should actually have different weighting constants associated with the different terms in $\mathscr{E}_\theta$ and write

$$\mathscr{E}_\theta(U, h, f) = \|\min^*(1, h^2\hat{\varepsilon}^{-1})(C_1 r(U) + C_2\hat{\varepsilon}D_h^2 U)\| + C_3 \max_{\Gamma_-\cup\Gamma_+} \hat{\varepsilon}^{1/2}\|f\|.$$

Secondly, the constants C_1 and C_2 associated with the two most important terms basically are of the form $C_s C_i$, where C_i is an interpolation error constant which can be determined rather easily (cf. the discussion in [3]) and C_s is the stability constant in Lemma 1.1. In the case of the simple model problems under consideration here, the stability constant C_s could easily be evaluated theoretically simply by following the proof of Lemma 1.1. For more complicated problems, however, it seems more realistic to seek an appropriate computational replacement for such a procedure and estimate C_s from a numerical solution of the dual problem (3.5). The latter problem is the subject of ongoing research. Note here that in Lemma 1.1, too, one should actually use individual stability constants for the individual terms ($\|\varepsilon^{1/2}\nabla u\|$, $\|u\|$, $\|u_x\|$, $\|\varepsilon D^2 u\|$, ...) in the estimate.

In the discussions below, we shall always assume that the problem of finding suitable values for the constants C has been appropriately solved.

Reliability and efficiency. Let us now discuss the *reliability* and *efficiency* of the adaptive methods (M_1), (M_2), and (M_3). The methods (M_1) and (M_3), of course, will be *reliable* in the sense that *if* the corresponding algorithm reaches its stopping criterion, by finding a mesh T and the corresponding U such that (4.2), (4.3), or (4.6) holds, then we know from the corresponding a posteriori estimates that (4.2) will be guaranteed. For the method (M_2), on the other hand, our analysis does not guarantee full reliability in the above sense, since the stopping criterion (4.3) is based on replacing $\hat{u}$ by U in Lemma 3.3, which has not been justified. Nevertheless, replacing $\hat{u}$ by U seems to be a reasonable thing to do, since $\hat{u}$ is sufficiently regular and the finite element mesh for U fits with the regularity of $\hat{u}$ so as to admit error estimates for $\hat{u} - U$, e.g., as in Lemma 3.1 (cf. also the discussion preceding Lemma 3.5).

Concerning the *efficiency*, we would like to know that a mesh generated by the adaptive algorithm corresponding to the method (M_1), (M_2), or (M_3) under consideration is (reasonably) close to an *optimal* mesh and not excessively overrefined for the computational goal (4.1). As the weakest possible demand on efficiency, we would like to know that the method is *operational* in the sense that (4.2) may be realized by refining the mesh.

For the purpose of a brief discussion of these matters in a simple model situation, we consider the case of an interior singular layer due to a jump discontinuity in f across a characteristic $y = \text{const}$. With our interest focused on the case when ε is very small, we shall assume for simplicity that $\varepsilon \leq Ch^3$ everywhere in Ω, and further that ε is constant. As we shall see below, we then expect to have that $\hat{\varepsilon} = O(h^\alpha)$ with $\alpha \geq 1$ everywhere in Ω, so that, in particular, we may control the terms depending directly on the data f in (3.4), (3.7), (3.10) etc. We may thus concentrate on the terms involving the computed solution U and its residual $R(U)$. In order to be able to estimate these terms, we shall first derive a preliminary estimate for $f - U_x$ in the different parts of

the computational domain. To this end, we shall use the fact that

$$f - U_x = \hat{u}_x - U_x - \operatorname{div}(\hat{\varepsilon}\nabla\hat{u}) \tag{4.7}$$

and estimate separately $\hat{u}_x - U_x$ and $\operatorname{div}(\hat{\varepsilon}\nabla\hat{u})$. We recall (see [8]) that one can derive a priori error estimates for the SD-method of the form

$$\|\hat{u} - U\|_\psi \le C(\|h^{3/2}D^2\hat{u}\|_\psi + \|h^2 f\|) \tag{4.8a}$$

and

$$\|\hat{u}_x - U_x\|_\psi \le C(\|hD^2\hat{u}\|_\psi + \|h^2 f\|), \tag{4.8b}$$

where ψ is a cutoff function subject to certain conditions similar to those in (1.7) with ε replaced by h. Strictly speaking, existing proofs of these estimates require additional assumptions, such as a constant $\hat{\varepsilon}$ and certain quasi-uniformity of the mesh, but it seems reasonable to believe that the results of [8] should be extendable to the case of a variable $\hat{\varepsilon}$ subject to the conditions of Lemma 1.1 and for a fairly general class of locally refined meshes such as in [3]. The estimates (4.8a, b) indicate that in the parts of Ω where $\hat{u}$ is smooth we should have $\hat{u} - U = O(h^{3/2})$ and $\hat{u}_x - U_x = O(h)$. In order to estimate the $\operatorname{div}(\hat{\varepsilon}\nabla\hat{u})$-term in (4.7), we first note that $\operatorname{div}(\hat{\varepsilon}\nabla\hat{u}) = \nabla\hat{\varepsilon}\cdot\nabla\hat{u} + \hat{\varepsilon}\Delta\hat{u}$. Here, $\hat{\varepsilon}$ is of order $O(h^2|f - U_x|))$, and under reasonable mesh assumptions as in [3] it follows that $\nabla\hat{\varepsilon} = O(h|f - U_x|)$. We may therefore conclude that, in regions of smoothness of $\hat{u}$, we have $|f - U_x| = O(h)$ and consequently $\hat{\varepsilon} = O(h^3)$, $\hat{\varepsilon}^{1/2}\nabla U = O(h^{3/2})$, $R(U) = O(h)$ and $\min(1, h^2\hat{\varepsilon}^{-1})R(U) = R(U) = O(h)$.

We now note that for all three methods, (M_1), (M_2) and (M_3), owing to the presence of the term $\|\min(1, h^2\hat{\varepsilon}^{-1})R(U)\|$ in $\mathscr{E}_\theta$, the a posteriori error estimates do not seem to be fully sharp in the smooth parts of the domain. In particular, it seems as if we have lost a factor of $h^{1/2}$ compared to (4.8a), according to which $\hat{u} - U = O(h^{3/2})$ in the parts of the domain where $\hat{u}$ is smooth (and a factor h as compared to the interpolation error of order $O(h^2)$ in smooth regions). This indicates that the method will overrefine slightly in smooth regions. However, as we shall soon see, in the case of a characteristic singular layer, the majority of the elements will be located in the characteristic layer, so that the total number of elements is not much affected by a moderate overrefinement in the smooth region.

In the characteristic layer we expect to have $f - U_x = O(h^{-1/2})$ and $\hat{\varepsilon} = O(h^{3/2})$. This is based on the following heuristic argument. We recall the a priori error estimate (4.8b). Since this estimate is more or less local, we expect to have a corresponding pointwise estimate under reasonable assumptions. Since $D^2\hat{u} = O(\hat{\varepsilon}^{-1})$ and $|\nabla\hat{\varepsilon}| = O(\varepsilon^{-1/2})$ in the layer, and since we know that $\hat{\varepsilon} = O(h)$ just by applying an inverse estimate, we are led to believe, neglecting the f-term in (4.8b) and under appropriate assumptions on $\nabla\hat{\varepsilon}$, that

$$|f - U_x| \le |\hat{u}_x - U_x| + |\operatorname{div}(\hat{\varepsilon}\nabla\hat{u})| \le C(h\hat{\varepsilon}^{-1} + 1) \le C(h^{-1}|f - U_x|^{-1} + 1).$$

From this we conclude that $|f - U_x| = O(h^{-1/2})$, $\hat{\varepsilon} = O(h^{3/2})$ and $R(U) = O(h^{-1/2})$ in the layer. Further, we expect the width of the numerical singular layer to be $O(\max(\hat{\varepsilon}^{1/2}, h^{3/4})) = O(h^{3/4})$ (cf. [10]), and $\nabla\hat{\varepsilon}\cdot\nabla U$ and $D_h^2 U$ should both be of order $O(1)$. From this we conclude that $\hat{\varepsilon}^{1/2}\nabla U = O(1)$, $\min(1, h^2\hat{\varepsilon}^{-1})R(U) = h^2\hat{\varepsilon}^{-1}R(U) = O(1)$ and $\min(1, h\hat{\varepsilon}^{-1/2})R(U) =$

$h\hat{\varepsilon}^{-1/2}R(U) = O(h^{-1/4})$ in the layer. If we square all these quantities, multiply by the width of the layer and then take the square root, we find that the contribution to $\mathscr{E}_\theta$, $\mathscr{E}_\rho$, and $\widehat{\mathscr{E}}_\rho$ coming from the layer should be of order $O(h^{3/8})$, $O(h^{1/8})$, and $O(h^{3/8})$, respectively. (Note that by defining an $\tilde{f}$ such that $|\nabla\tilde{f}| = O(h^{-1/4})$ in the layer and with $\tilde{f} = f$ in the smooth region, and using (3.12) rather than (3.11), we may include also the contributions from the data terms in these estimates.) We see that we have lost a factor of $h^{1/4}$ in the $\mathscr{E}_\rho$-and $\widetilde{\mathscr{E}}_\rho$-terms, since $O(h^{1/2})$ is optimal according to interpolation theory. The reason is, of course, that we have not been able to derive an error bound of second order in h for ρ, since we have been using (3.9). The estimates for the $\mathscr{E}_\theta$-and $\widehat{\mathscr{E}}_\rho$-terms, on the other hand, appear to be close to the optimal interpolation error $O(h^{1/2})$.

4.2. **Dirichlet outflow boundary data.** Let us now turn our attention to the case of problem (1.2) and the additional complication due to the presence of an outflow singular layer. Now we base our adaptive method on the estimate of Theorem 3.1 and consider the following method (M_4): Seek a mesh T with (nearly) minimal number of nodes such that for the corresponding U

$$\mathscr{E}_\theta(U, h, f) + \widehat{\mathscr{E}}_\rho(U, \hat{\varepsilon}, \text{data}) \leq \text{TOL}, \tag{4.9}$$

where

$$\widehat{\mathscr{E}}_\rho(U, \hat{\varepsilon}, \text{data}) = C\left(\|(\hat{\varepsilon} - \varepsilon)U_x\| + \|d_+(\hat{\varepsilon} - \varepsilon)D_y^h U_y\| + \max_{\Gamma_-} \hat{\varepsilon}^{1/2}\|f\|\right)$$

and (with notation as in the definition of D_h^2)

$$D_y^h v|_K = \frac{1}{3}\sum_{K' \in N(K)} \frac{(v(P') - v(P))(P' - P)_2}{|P' - P|^2}.$$

Clearly, this method is based on replacing $\hat{u}$ by U in the $\widehat{\mathscr{E}}_\rho$-term in the 'quasi' a posteriori error estimate (3.15).

Alternatively, we consider here also (M_3), which applies to the outflow singular layer problem as well, with $\mathscr{E}_\theta$ defined as in Lemma 3.2.

Reliability and efficiency. Method (M_3), of course, is reliable as before, whereas the method (M_4) based partly on a heuristic step is not fully reliable (cf. the discussion preceding Lemma 3.5). Concerning the *efficiency*, let us first consider method (M_4). Again, the term $\min(1, h^2\hat{\varepsilon}^{-1})R(U)$ will be of order $O(h)$ in the parts of the domain where $\hat{u}$ is smooth. In a characteristic layer we will have $\min(1, h^2\hat{\varepsilon}^{-1})R(U) = h^2\hat{\varepsilon}^{-1}R(U) = O(1)$. The width of the characteristic layer is expected to be $O(h^{3/4})$, which would give an error contribution to $\mathscr{E}_\theta$ of order $O(h^{3/8})$, which is close to optimal. In the outflow singular layer we expect to have $|f - U_x| = O(h^{-1})$, simply by applying an inverse estimate, and consequently $\hat{\varepsilon} = O(h)$ and $R(U) = O(h^{-1})$, so that

$$\min(1, h^2\hat{\varepsilon}^{-1})R(U) = O(1).$$

The width of the outflow singular layer is known to be $O(h)$, so that the contribution to the $\mathscr{E}_\theta$-term will be $O(h^{1/2})$, which is optimal.

In smooth parts, the first two terms in $\widehat{\mathscr{E}}_\rho$ seem to be of order $O(h^2)$ and the f-term of order $O(h^{3/2})$, which is optimal compared to (4.8a). In a characteristic layer, the second term in $\widehat{\mathscr{E}}_\rho$ determines the order and appears to be $O(h^{3/8})$, since $\hat{u}_{yy} = O(\hat{\varepsilon}^{-1})$ and the width of the layer is $O(h^{3/4})$. In the outflow layer, finally, $\hat{u}_x = O(\hat{\varepsilon}^{-1})$ and $\hat{u}_{yy} = O(\hat{\varepsilon}^{-2})$ or better, so that $d_+((\hat{\varepsilon} - \varepsilon)\hat{u}_y)_y = O(1)$ and consequently $\widehat{\mathscr{E}}_\rho = O(h^{1/2})$, which again is optimal.

Concerning the efficiency of method (M_3) we have already seen that $\mathscr{E}_\theta$ appears to be a sharp bound for the θ-part of the error. We now would like to analyze the additional effect of the control $\hat{\varepsilon} = \varepsilon$. We shall give an argument which indicates that the additional control $\hat{\varepsilon}(U, h) = \varepsilon$ in (4.6) will lead to resolution of both outflow layers and characteristic layers. To see this, note that in an outflow layer, in order to have $\hat{\varepsilon} = \varepsilon$, we must have that

$$\varepsilon = \hat{\varepsilon} \geq ch^2|f - U_x| \geq ch,$$

since $|f - U_x| \geq ch^{-1}$ in the layer, so that the outflow layer of width $O(\hat{\varepsilon})$ will be resolved. In a characteristic layer we expect to have $h \leq c\varepsilon^{2/3}$, since

$$\hat{\varepsilon} \geq ch^2|f - U_x| \simeq ch^3 D^2\hat{u} \simeq ch^3\hat{\varepsilon}^{-1},$$

which states that

$$\hat{\varepsilon} \geq ch^{3/2} \quad \text{or} \quad h^{3/4} \leq c\varepsilon^{1/2},$$

if $\hat{\varepsilon} = \varepsilon$, which again indicates resolution, since the width of the characteristic layer is $O(\varepsilon^{1/2})$ and the width of the corresponding numerical layer (according to recent but yet preliminary results, see also [10]) is $O(h^{3/4})$. Note that the adaptive method (M_4) may or may not, depending on the tolerance TOL, lead to resolution of characteristic and outflow singular layers. Roughly speaking, if the tolerance is set greater than $O(\varepsilon^{1/4})$, then neither an outflow singular layer nor characteristic singular layers will be resolved. If the tolerance is between $O(\varepsilon^{1/2})$ and $O(\varepsilon^{1/4})$, then characteristic layers will be resolved, but not an outflow singular layer. Finally, if TOL is of order $O(\varepsilon^{1/2})$, then all layers will be resolved. Thus, imposing $\hat{\varepsilon} = \varepsilon$ corresponding to resolution of all layers, may lead to an overrefinement. On the other hand, resolution of the layers may be a computational goal as well, in addition to (4.1). Observe that the condition $\hat{\varepsilon} = \varepsilon$ in (4.6) corresponds to continuing the refinement until no shock-capturing artificial viscosity is effectively added. Even in this case, the shock-capturing artificial viscosity plays an important role during the adaptive refinement process when the mesh is not fine enough to resolve all details of the flow, but "disappears" on the final mesh where $\hat{\varepsilon} = \varepsilon$ and all details are resolved. Note that in a natural adaptive process we start with a coarse mesh and refine locally (e.g., until $\hat{\varepsilon} = \varepsilon$, i.e., $h \leq \varepsilon$ in outflow layers) instead of starting with an extremely fine quasi-uniform mesh (satisfying $h \leq \varepsilon$ everywhere) and then unrefine locally, which clearly would be an inefficient procedure.

BIBLIOGRAPHY

1. K. Eriksson and C. Johnson, *An adaptive finite element method for linear elliptic problems*, Math. Comp. **50** (1988), 361–383.
2. ———, *Adaptive finite element methods for parabolic problems* I: *A linear model problem*, SIAM J. Numer. Anal. **28** (1991), 43–77.

3. ——, *Adaptive finite element methods for parabolic problems* II: *A priori error estimates in* $L_\infty(L_2)$ *and* $L_\infty(L_\infty)$, Department of Mathematics, Chalmers University of Technology, Göteborg, 1992.
4. ——, *Adaptive streamline diffusion finite element methods for time dependent convection diffusion problems*, (to appear).
5. P. Hansbo, *Adaptivity and streamline diffusion procedures in the finite element method*, Thesis, Department of Structural Mechanics, Chalmers University of Technology, Göteborg, 1989.
6. P. Hansbo and A. Szepessy, *A velocity-pressure streamline diffusion finite element method for the incompressible Navier-Stokes equations*, Comput. Methods Appl. Mech. Engrg. **84** (1990), 175–192.
7. T. Hughes et al., *A new finite element method formulation for computational fluid dynamics* I–IV, Comput. Methods Appl. Mech. Engrg. **54** (1986), 223–234, 341–355; **58** (1986), 305–328, 329–336.
8. C. Johnson, U. Nävert, and J. Pitkäranta, *Finite element methods for linear hyperbolic equations*, Comput. Methods Appl. Mech. Engrg. **45** (1984), 285–312.
9. C. Johnson and J. Saranen, *Streamline diffusion methods for the incompressible Euler and Navier-Stokes equations*, Math. Comp. **47** (1986), 1–18.
10. C. Johnson, A. Schatz, and L. Wahlbin, *Crosswind smear and pointwise errors in streamline diffusion finite element methods*, Math. Comp. **49** (1987), 25–38.
11. C. Johnson and A. Szepessy, *On the convergence of a finite element method for a nonlinear hyperbolic conservation law*, Math. Comp. **49** (1987), 427–444.
12. C. Johnson, A. Szepessy, and P. Hansbo, *On the convergence of shock-capturing streamline diffusion finite element methods for hyperbolic conservation laws*, Math. Comp. **54** (1990), 107–129.
13. C. Johnson, *Adaptive finite element methods for diffusion and convection problems*, Comput. Methods Appl. Mech. Engrg. **82** (1990), 301–322.
14. C. Johnson and P. Hansbo, *Adaptive streamline diffusion finite element methods for compressible flow*, Comput. Methods Appl. Mech. Engrg. **87** (1991), 267–280.
15. R. Löhner, K. Morgan, and M. Vahdati, *FEM-FCT: combining unstructured grids with high resolution*, Comm. Appl. Numer. Methods **4** (1988), 717–729.
16. R. Löhner, K. Morgan, and O. Zienkiewicz, *Adaptive grid refinement for the compressible Euler equations*, Accuracy Estimates and Adaptive Refinements in Finite Element Computations (I. Babuška et al., eds.), Wiley, New York, 1986, pp. 281–297.
17. T. Strouboulis and T. Oden, *A posteriori estimation of the error in finite-element approximations of convection dominated problems*, Finite Element Analysis in Fluids (T.J. Chung and G.R. Karr, eds.), Univ. of Alabama in Huntsville Press, 1989, pp. 125–136.
18. A. Szepessy, *Convergence of the streamline diffusion finite element method for conservation laws*, Thesis, Mathematics Department, Chalmers University of Technology, Göteborg, 1989.
19. ——, *Convergence of a shock-capturing streamline diffusion finite element method for a scalar conservation law in two space dimensions*, Math. Comp. **53** (1989), 527–545.

DEPARTMENT OF MATHEMATICS, CHALMERS UNIVERSITY OF TECHNOLOGY, S-412 96 GÖTEBORG, SWEDEN
E-mail address: kenneth@math.chalmers.se
E-mail address: claes@math.chalmers.se

MATHEMATICS OF COMPUTATION
VOLUME 60, NUMBER 201
JANUARY 1993, PAGES S1–S2

Supplement to

ADAPTIVE STREAMLINE DIFFUSION FINITE ELEMENT METHODS FOR STATIONARY CONVECTION-DIFFUSION PROBLEMS

KENNETH ERIKSSON AND CLAES JOHNSON

Proof of Lemma 3.3. We find that $\rho = u - \hat{u}$ solves the following problem:

$$\begin{aligned} \rho_x - \operatorname{div}(\varepsilon \nabla \rho) &= \phi \quad \text{in } \Omega, \\ \rho &= 0 \quad \text{on } \Gamma_- \cup \Gamma_0, \\ \frac{\partial \rho}{\partial n} &= 0 \quad \text{on } \Gamma_+, \end{aligned}$$

where $\phi = -\operatorname{div}((\hat{\varepsilon} - \varepsilon)\nabla \hat{u})$.

From Lemma 1.1 and the fact that $\|\rho\| \le C\|\rho_x\|$ we get that

$$\|\rho\| \le C\|\operatorname{div}((\hat{\varepsilon} - \varepsilon)\nabla \hat{u})\|. \tag{A}$$

Since ε is constant, we have that

$$\begin{aligned} \operatorname{div}((\hat{\varepsilon} - \varepsilon)\nabla \hat{u}) &= \nabla \hat{\varepsilon} \cdot \nabla \hat{u} + (\hat{\varepsilon} - \varepsilon)\Delta \hat{u} \\ &= \nabla \hat{\varepsilon} \cdot \nabla \hat{u} + \hat{\varepsilon}^{-1}(\hat{\varepsilon} - \varepsilon)(\hat{u}_x - f - \nabla \hat{\varepsilon} \cdot \nabla \hat{u}) \\ &= \hat{\varepsilon}^{-1}\varepsilon \nabla \hat{\varepsilon} \cdot \nabla \hat{u} + \hat{\varepsilon}^{-1}(\hat{\varepsilon} - \varepsilon)(\hat{u}_x - f). \end{aligned} \tag{B}$$

Further, with $\varphi = \hat{\varepsilon}^{-2}(\hat{\varepsilon} - \varepsilon)^2$ we have that

$$\begin{aligned} \|\hat{u}_x - f\|_\varphi^2 &= (\varphi(\hat{u}_x - f), \operatorname{div}(\hat{\varepsilon}\nabla \hat{u})) \\ &= \int_\Gamma \varphi(\hat{u}_x - f)\hat{\varepsilon}\frac{\partial \hat{u}}{\partial n}\, d\Gamma - (\nabla \varphi(\hat{u}_x - f), \hat{\varepsilon}\nabla \hat{u}) \\ &\quad - (\varphi \nabla \hat{u}_x, \hat{\varepsilon}\nabla \hat{u}) + (\varphi \nabla f, \hat{\varepsilon}\nabla \hat{u}) \\ &= I + II + III + IV. \end{aligned}$$

Here

$$III = -\frac{1}{2}\int_\Gamma \varphi\hat{\varepsilon}|\nabla \hat{u}|^2 n_1\, d\Gamma + \frac{1}{2}((\varphi\hat{\varepsilon})_x \nabla \hat{u}, \nabla \hat{u}) = V + VI,$$

and since f vanishes on Γ, $\partial \hat{u}/\partial n$ on Γ_+ and $\hat{u}$ on Γ_-, we get as in (1.8)

$$\begin{aligned} I + V &= \int_\Gamma \varphi\hat{\varepsilon}(\hat{u}_x \frac{\partial \hat{u}}{\partial n} - \frac{1}{2}|\nabla \hat{u}|^2 n_1)d\Gamma \\ &= -\frac{1}{2}\int_\Gamma \varphi\hat{\varepsilon}|\nabla \hat{u}|^2 |n_1|\, d\Gamma \le 0. \end{aligned}$$

0025-5718/93 $1.00 + $.25 per page

Using obvious estimates for the terms II and IV, and the fact that $\varphi^{-1/2}\nabla\varphi = 2\varepsilon\hat{\varepsilon}^{-2}\nabla\hat{\varepsilon}$, we obtain

$$\|\hat{u}_x - f\|_\varphi^2 \le \frac{1}{2}\|\hat{u}_x - f\|_\varphi^2 + 2\|\varepsilon\hat{\varepsilon}^{-1}\nabla\hat{\varepsilon}\cdot\nabla\hat{u}\|^2$$
$$+ \frac{1}{2}\|\varphi\hat{\varepsilon}^{1/2}\nabla f\|^2 + \frac{1}{2}\|\hat{\varepsilon}^{1/2}\nabla\hat{u}\|^2 + \frac{1}{2}((\varphi\hat{\varepsilon})_x\nabla\hat{u},\nabla\hat{u}).$$

From the identity (B) we may thus conclude, with σ as above, that

$$\|\mathrm{div}((\hat{\varepsilon}-\varepsilon)\nabla\hat{u})\|^2 \le 2(\|\hat{\varepsilon}^{-1}\varepsilon\nabla\hat{\varepsilon}\cdot\nabla\hat{u}\|^2 + \|\hat{u}_x - f\|_\varphi^2)$$
$$\le 2(\|\sigma\nabla\hat{u}\|_{\hat{\varepsilon}}^2 + \|\varphi\nabla f\|_{\hat{\varepsilon}}^2).$$

Together with (A) this proves the desired error bound for ρ and completes the proof of Lemma 3.3. □

Proof of Lemma 3.4. We have that

$$(\theta_x, v + \delta v_x) + (\hat{\varepsilon}\nabla\theta,\nabla v) - (\mathrm{div}(\hat{\varepsilon}\nabla\theta),\delta v_x)_T = 0, \quad \forall v \in V.$$

Now let z be the solution of

$$\text{(C)}\qquad \begin{aligned} -z_x - \mathrm{div}(\hat{\varepsilon}\nabla z) &= -\mathrm{div}(\hat{\varepsilon}\nabla\theta) && \text{in } \Omega,\\ z &= 0 && \text{on } \Gamma_- \cup \Gamma_0,\\ \hat{\varepsilon}\frac{\partial z}{\partial n} + n_1 z &= \hat{\varepsilon}\frac{\partial\theta}{\partial n} && \text{on } \Gamma_+. \end{aligned}$$

We find that

$$\|\nabla\theta\|_{\hat{\varepsilon}}^2 = (\theta, -z_x - \mathrm{div}(\hat{\varepsilon}\nabla z)) + \int_{\Gamma_+}\hat{\varepsilon}\theta\frac{\partial\theta}{\partial n}d\Gamma = (\theta_x, z) + (\hat{\varepsilon}\nabla\theta,\nabla z)$$
$$= (\theta_x, z - \tilde{z} - \delta\tilde{z}_x) + (\hat{\varepsilon}\nabla\theta,\nabla(z-\tilde{z})) + (\mathrm{div}(\hat{\varepsilon}\nabla\theta),\delta\tilde{z}_x)_T$$
$$= (f - U_x + \nabla\hat{\varepsilon}\cdot\nabla U, z - \tilde{z} - \delta\tilde{z}_x) + \sum_\tau\int_\tau \hat{\varepsilon}[\frac{\partial U}{\partial n_\tau}](z-\tilde{z})d\Gamma$$
$$= I + II.$$

With $r = |f - U_x + \nabla\hat{\varepsilon}\cdot\nabla U|$, we have that

$$|I| \le (r, |z-\tilde{z}| + \delta|\tilde{z}_x|) \le C\|\min_*(1, h\hat{\varepsilon}^{-1/2})r\|\,(\|z\| + \|\nabla z\|_{\hat{\varepsilon}}).$$

Similarly

$$|II| \le C\|\min_*(1, h\hat{\varepsilon}^{-1/2})\hat{\varepsilon}D_h^2 U\|\,(\|z\| + \|\nabla z\|_{\hat{\varepsilon}}).$$

Multiplying the equation in (C) by z and integrating over Ω we obtain, using the given boundary conditions,

$$\frac{1}{2}\int_{\Gamma_+} z^2 n_1 + \|\nabla z\|_{\hat{\varepsilon}}^2 = (\hat{\varepsilon}\nabla z,\nabla\theta) \le \|\nabla z\|_{\hat{\varepsilon}}\,\|\nabla\theta\|_{\hat{\varepsilon}},$$

from which follows at once that $\|\nabla z\|_{\hat{\varepsilon}} \le \|\nabla\theta\|_{\hat{\varepsilon}}$. If we repeat the arguments in the first part of the proof of Lemma 1.2 we easily derive the same control for z, or $\|z\| \le C\|\nabla\theta\|_{\hat{\varepsilon}}$. Together our estimates now show that

$$\|\nabla\theta\|_{\hat{\varepsilon}} \le C(\|\min_*(1, h\hat{\varepsilon}^{-1/2})r\| + \|\min_*(1, h\hat{\varepsilon}^{-1/2})\hat{\varepsilon}D_h^2 U\|).$$

Clearly, by estimating the terms I and II together, we can derive the more precise estimate (3.9). This completes the proof of Lemma 3.4. □