

## A MONOTONE FINITE ELEMENT SCHEME FOR CONVECTION-DIFFUSION EQUATIONS

JINCHAO XU AND LUDMIL ZIKATANOV

ABSTRACT. A simple technique is given in this paper for the construction and analysis of a class of finite element discretizations for convection-diffusion problems in any spatial dimension by properly averaging the PDE coefficients on element edges. The resulting finite element stiffness matrix is an  $M$ -matrix under some mild assumption for the underlying (generally unstructured) finite element grids. As a consequence the proposed edge-averaged finite element scheme is particularly interesting for the discretization of convection dominated problems. This scheme admits a simple variational formulation, it is easy to analyze, and it is also suitable for problems with a relatively smooth flux variable. Some simple numerical examples are given to demonstrate its effectiveness for convection dominated problems.

### 1. INTRODUCTION

Convection-diffusion equations, especially the convection dominated ones, are known to have many important applications. Standard finite element and/or finite difference methods are in general not suitable for these problems, in the sense that the numerical solution often contains spurious oscillations if the mesh size is not small enough. Many special techniques have been developed, including upwinding finite difference and/or finite volume methods (see [3], and [4]), finite volume methods (see [13]), streamline diffusion finite element methods [17], the Petrov-Galerkin method (see [16]), and (the hybrid) streamline-upwinding-Petrov-Galerkin (SUPG) method (see [11] and [16]). For a detailed description of numerical techniques and analytical tools in investigating convection-diffusion equations we refer to the monographs [23] and [24].

Many convection-diffusion problems satisfy a maximum principle on the continuous level. In view of numerical stability (i.e., no spurious oscillations), it is desirable that the resulting discrete equation also satisfy a maximum principle that is similar to the continuous case. Such a scheme that satisfies a maximum principle is often

---

Received by the editor May 6, 1996 and, in revised form, December 16, 1997.

1991 *Mathematics Subject Classification*. Primary 65N30, 65N15.

*Key words and phrases*. Convection dominated problems, finite element method, monotone schemes, up-winding, Scharfetter-Gummel discretization, error bounds.

The first author's work was partially supported by NSF DMS94-03915-1 and NSF DMS-9706949 through Penn State, and by NSF ASC-92-01266 and ONR-N00014-92-J-1890 through UCLA.

The second author's work was partially supported by the Bulgarian Ministry of Education and Science Grant I-504/95, by NSF Grant Int-95-06184 and ONR-N00014-92-J-1890 through UCLA, and also by the Center for Computational Mathematics and Applications of Pennsylvania State University.

known as a *monotone scheme*. A well-known sufficient condition for a scheme to be monotone is that the corresponding stiffness matrix is an  $M$ -matrix. Among the several aforementioned schemes, upwinding schemes are often monotone.

A linear monotone scheme usually has only first order accuracy. This is a rather undesirable drawback, and it certainly limits its usefulness in practical computations. Nevertheless, linear monotone schemes are still significant in many ways. Our primary interest in this type of schemes is hopefully to use this scheme as a tool to design efficient iterative and preconditioning techniques for solving other more sophisticated schemes (such as streamline diffusion methods and nonlinear monotone schemes). A linear system with an  $M$ -matrix from convection dominated problems can be efficiently solved, for example, by Gauss-Seidel method, and the convergence of Gauss-Seidel iteration can be dramatically speeded up with a proper ordering of the unknowns (cf. Hackbusch and Probst [7], Bey and Wittum [8], Wang and Xu [29], Xu [30]).

The existing monotone schemes are mostly derived by either a finite difference or a finite volume approach. One inconvenience of these approaches is that it is often not clear how to analyze theoretically the schemes derived in this way. Thus we were motivated to look for monotone schemes that fall into the standard finite element variational framework, and its theoretical analysis is more straightforward. This paper is to report our finding in this effort. The new scheme that we shall describe here has several interesting features. It is a finite element scheme with a standard variational formulation (but with a modified bilinear form) by means of the usual piecewise linear functions for both the trial and test spaces; its derivation is completely different from the other known approaches, and it does not (explicitly) use the standard upwinding techniques (such as checking the flow directions); it can be applied to very general unstructured grid in any spatial dimension; and its theoretical analysis is more transparent.

Our scheme was partially motivated by the work of Markowich and Zlamal [19] and Brezzi, Marini and Pietra [9]. In particular, a Scharfetter–Gummel type (see [25]) finite element scheme is derived in [19] for symmetric positive definite equations in two space dimensions (also with application to symmetrizable convection-diffusion equations). For the special cases considered in [19] our scheme pretty much coincides with that in [19], but our derivation is much simpler and can be applied in more general situations. For other relevant work, let us mention Mock [22], Brezzi, Marini and Pietra [10], Marini and Pietra [18], Miller, Wang and Wu [21], Miller and Wang [20], and also Babuška and Osborn [2].

In all the papers quoted here (with only one exception, [20]) the monotonicity property depends on the assumption that the triangulation is not obtuse (or weakly acute type, as it is called sometimes). A possible alternative might be quadrilateral meshes in two dimensions, where some obtuse angles can be allowed (cf. [33]) at the cost of adding other restrictive geometrical conditions. But in practice, the construction of a non-obtuse triangulation is not a simple task (see [6] for the relevant algorithmic difficulties). The monotonicity of the scheme in this paper depends on a much weaker and more practical assumption which, in two dimensions, means that the triangulation needs to be assumed to be Delaunay.

The rest of the paper is organized as follows: In Section 2 we discuss the properties of finite element discretization for the Poisson equation, which we consider as basis for the derivation of our edge-averaged finite element (EAFE) scheme. In Section 3 we derive the edge-averaged scheme for simplified convection-diffusion

equation—namely the coefficients are assumed to be continuous, and we consider only the Dirichlet problem. In this section we also give the geometrical conditions when the resulting matrix is an  $M$ -matrix. Bounds on the stiffness matrix entries are also obtained in subsection 3.3, in order to give a way of implementing this scheme. The derivation of the EAFE scheme for more general case of piecewise smooth coefficients is presented in Section 5. In Section 4, we also discuss the practically important case when the diffusion coefficient approaches zero. In Section 6 we obtain a natural convergence result, which is stated in Theorem 6.3.

## 2. PRELIMINARIES

In this section, we shall introduce some notation and describe some basic properties of finite element triangulations and finite element spaces. In particular, we shall discuss some special properties of the finite element discretization for the simple Poisson equation which, as we shall see later, will be the basis of the derivation of the EAFE scheme for convection-diffusion problems.

Let  $\Omega \subset \mathbb{R}^n$  ( $n \geq 1$ ) be a bounded Lipschitz domain. Given  $p \in [1, \infty]$  and an integer  $m \geq 0$ , we use the usual notation  $W^{m,p}(\Omega)$  to denote the Sobolev space of  $L^p$  functions whose derivatives up to order  $m$  also belong to  $L^p$ , with the standard semi-norm and norm denoted by  $|\cdot|_{m,p,\Omega}$  and  $\|\cdot\|_{m,p,\Omega}$  respectively. When  $p = 2$ ,  $H^m(\Omega) \equiv W^{m,p}(\Omega)$  with  $|\cdot|_{m,\Omega} = |\cdot|_{m,2,\Omega}$  and  $\|\cdot\|_{m,\Omega} = \|\cdot\|_{m,2,\Omega}$ .

Let  $\mathcal{T}_h$  be a family of simplicial finite element triangulations of  $\Omega$  that are shape regular and satisfy the usual conditions (see [12]). For simplicity of exposition, we assume that the triangulation covers  $\Omega$  exactly. Associated with each  $T_h$ , let  $V_h \subset H_0^1(\Omega)$  be the piecewise linear finite element space. As usual the space  $H_0^1(\Omega)$  is defined as the space of  $u \in H^1(\Omega)$  such that  $u = 0$  on  $\partial\Omega$ .

Given  $T \in \mathcal{T}_h$ , we introduce the following notation (see Figure 2.1):

- $q_j$  ( $1 \leq j \leq n + 1$ ): the vertices of  $T$ ;
- $E_{ij}$  or simply  $E$ : the edge connecting two vertices  $q_i$  and  $q_j$ ;
- $F_j$ : the  $(n - 1)$ -dimensional simplex opposite to the vertex  $q_j$ ;
- $\theta_{ij}^T$  or  $\theta_E^T$ : the angle between the faces  $F_i$  and  $F_j$ ;
- $\kappa_E^T$ :  $F_i \cap F_j$ , the  $(n - 2)$ -dimensional simplex opposite to the edge  $E$ ;
- $\delta_E\phi = \phi(q_i) - \phi(q_j)$ , for any continuous function  $\phi$  on  $E = E_{ij}$ ;
- $\tau_E = \delta_E x = q_i - q_j$ , a directional vector of  $E$ .

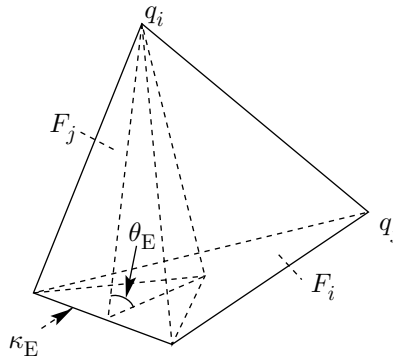


FIGURE 2.1

We shall denote the nodes in  $\mathcal{T}_h$  by  $x_j$ ,  $j = 1, \dots, N_h$ . This is a “global” notation for all the vertices on the grid. Thus, we shall use  $q_j (= q_j^T)$ ,  $j = 1, \dots, n + 1$ , in a fixed element  $T \in \mathcal{T}_h$  and  $x_j$  ( $j = 1, \dots, N_h$ ) for all the nodes. The edges  $E_{ij}$  will denote either the edge  $(q_i, q_j)$  in an element  $T$ , or the edge  $(x_i, x_j)$  living somewhere on the grid. This slight abuse of notation should not be a source of confusion.

We denote nodal basis functions in  $V_h$  by  $\varphi_i$ ,  $i = 1, \dots, N_h$ , which are continuous in  $\Omega$ , linear in each  $T$  and

$$\varphi_i(x_i) = 1, \quad \varphi_i(x_j) = 0, \quad j \neq i.$$

As we have already pointed out, we first consider the simplest and important case of the Poisson equation:

$$\begin{aligned} -\Delta u &= f, & x \in \Omega, \\ u &= 0, & x \in \partial\Omega. \end{aligned}$$

Given  $T \in \mathcal{T}_h$ , let  $(a_{ij}^T)$  be the element stiffness matrix on  $T$ . Then, for  $u_h, v_h \in V_h$ , we have

$$(2.1) \quad \int_T \nabla u_h \cdot \nabla v_h dx = \sum_{i,j} a_{ij}^T u_h(q_i) v_h(q_j).$$

Since  $a_{ii}^T = -\sum_{j \neq i} a_{ij}^T$ , we can easily obtain the simple but important identity

$$(2.2) \quad \int_T \nabla u_h \cdot \nabla v_h dx = -\sum_{i < j} a_{ij}^T (u_h(q_i) - u_h(q_j))(v_h(q_i) - v_h(q_j)), \quad u_h, v_h \in V_h.$$

Using (2.2), we can rewrite the bilinear form in the following way:

$$(2.3) \quad \int_{\Omega} \nabla u_h \cdot \nabla v_h dx = \sum_{T \in \mathcal{T}_h} \sum_{E \subset T} \omega_E^T \delta_E u_h \delta_E v_h,$$

where  $\omega_E^T = -a_{ij}^T$  with  $E$  connecting the vertices  $q_i$  and  $q_j$ . For the weights  $\omega_E^T$  the following simple identity holds:

$$(2.4) \quad \omega_E^T = \frac{1}{n(n-1)} |\kappa_E^T| \cot \theta_E^T,$$

where  $\theta_E^T$  is the angle between the faces not containing edge  $E$  (see Figure 2.1), and their intersection forms  $\kappa_E^T$  (the  $(n-2)$ -dimensional simplex opposite to the edge  $E$ ). The identity (2.4) can be found, for example, in [28] for  $n = 2$  and in [5] for  $n = 3$ . Because of its importance in our presentation we shall include a proof for any space dimension  $n$  in the Appendix.

Let  $A = (\nabla \varphi_i, \nabla \varphi_j)$  be the stiffness matrix for the Poisson equation. We are interested in conditions for  $A$  to be an  $M$ -matrix. We recall that  $A$  is an  $M$ -matrix if it is irreducible (i.e., the graph corresponding to  $A$  is connected) and

$$\begin{aligned} A_{jj} &> 0 \quad \forall j; & A_{ij} &\leq 0 \quad \forall i, j : i \neq j; \\ A_{jj} &\geq \sum_{i=1: i \neq j}^{N_h} |A_{ij}| \quad \forall j; & A_{jj} &> \sum_{i=1: i \neq j}^{N_h} |A_{ij}| \quad \text{for at least one } j. \end{aligned}$$

**Lemma 2.1.** *The stiffness matrix for the Poisson equation is an  $M$ -matrix if and only if for any fixed edge  $E$  the following inequality holds:*

$$(2.5) \quad \omega_E \equiv \frac{1}{n(n-1)} \sum_{T \supset E} |\kappa_E^T| \cot \theta_E^T \geq 0,$$

where  $\sum_{T \supset E}$  means summation over all simplexes  $T$  containing  $E$ .

For  $n = 2$ , the condition (2.5) means that the sum of the angles opposite to any edge is less than or equal to  $\pi$ , i.e., if  $T_1 \cap T_2 = \{E\}$  then  $\theta_E^{T_1} + \theta_E^{T_2} \leq \pi$ . This condition implies that the triangulation is a so-called Delaunay triangulation. It follows therefore that in  $\mathbb{R}^2$  the stiffness matrix for the Poisson equation is an  $M$ -matrix if (and only if, with some possible rare exceptions near the boundary) the triangulation is a Delaunay triangulation.

In the literature it seems to be better known that the stiffness matrix for the Poisson equation is an  $M$ -matrix if the triangulation is not obtuse, i.e., if all the interior angles in each triangle are less than or equal to  $\frac{\pi}{2}$  (below we refer to this type of triangulations as non-obtuse triangulations). Of course, a non-obtuse triangulation is a very special Delaunay triangulation. But Delaunay triangulations are certainly more general and more practical (see [5]).

### 3. THE EAFE SCHEME AND ITS BASIC PROPERTIES

In this section we give a derivation of the edge-averaged finite element (EAFE) scheme and then discuss some of its basic properties.

**3.1. Model problem.** To present the main idea more clearly, we shall first derive the discrete scheme for a simplified model problem with simplified assumptions. We shall discuss the more general case later (Section 5). Specifically, we consider

$$(3.1) \quad \begin{cases} \mathcal{L}u \equiv -\nabla \cdot (\alpha(x)\nabla u + \beta(x)u) = f(x), & x \in \Omega, \\ u = 0, & x \in \partial\Omega. \end{cases}$$

We assume that  $\alpha \in C^0(\bar{\Omega})$  with  $0 < \alpha_{min} \leq \alpha(x) \leq \alpha_{max}$  for every  $x \in \Omega$ ,  $\beta \in (C^0(\bar{\Omega}))^2$ , and  $f \in L^2(\Omega)$ .

The weak formulation of the problem (3.1) is: Find  $u \in H_0^1(\Omega)$  such that

$$(3.2) \quad a(u, v) = f(v), \quad \text{for every } v \in H_0^1(\Omega),$$

where

$$(3.3) \quad a(u, v) = \int_{\Omega} (\alpha(x)\nabla u + \beta(x)u) \cdot \nabla v dx, \quad f(v) = \int_{\Omega} f(x)v dx.$$

It can be shown (see [14]) that (3.2) is uniquely solvable and there exists a constant  $c_0 > 0$  such that for every  $v \in H_0^1(\Omega)$

$$(3.4) \quad \sup_{\phi \in H_0^1(\Omega)} \frac{a(\phi, v)}{\|\phi\|_{1,\Omega}} \geq c_0 \|v\|_{1,\Omega}; \quad \sup_{\phi \in H_0^1(\Omega)} \frac{a(v, \phi)}{\|\phi\|_{1,\Omega}} \geq c_0 \|v\|_{1,\Omega}.$$

Another important property of  $\mathcal{L}$  is that its inverse is nonnegative. More precisely (see [14]),

$$(3.5) \quad \text{If } (\mathcal{L}u)(x) \geq 0 \text{ for all } x \in \Omega \text{ then } u(x) \geq 0 \text{ for all } x \in \Omega.$$

The above condition will be referred to as the *monotonicity property*, and it holds regardless of the size of  $|\beta(x)|/\alpha(x)$ . What is interesting for applications is the convection dominated case, namely  $|\beta(x)|/\alpha(x) \gg 1, \forall x \in \Omega$ . Our goal is to construct a scheme that has a monotonicity property analogous to (3.5), namely, if  $V_h \subset H_0^1(\Omega)$  is a finite element space and  $\mathcal{L}_h$  is the corresponding discretization for  $\mathcal{L}$ , then

$$(3.6) \quad (\mathcal{L}_h^{-1} f_h)(x) \geq 0 \text{ for all } x \in \Omega, \text{ if } f_h^{(i)} = f(\varphi_i) \geq 0 \text{ for all } i = 1, \dots, N_h.$$

A finite element scheme satisfying the above condition will be called a *monotone* finite element scheme in this paper. It is known that if the stiffness matrix corresponding to  $\mathcal{L}_h$  is an  $M$ -matrix, then (3.6) holds.

**3.2. Derivation of the scheme.** Given any edge  $E$ , we introduce a function  $\psi_E$  defined locally on  $E$  (up to an arbitrary constant) by the relation

$$(3.7) \quad \frac{\partial \psi_E}{\partial \tau_E} = \frac{1}{|\tau_E|} \alpha^{-1} (\boldsymbol{\beta} \cdot \tau_E).$$

Here and also in the proof of the next lemma, with an abuse of notation  $\partial/\partial \tau_E$  denotes the tangential derivative along  $E$ . As a basis for our derivation we shall use the following result.

**Lemma 3.1.** *Let  $u \in H_0^1(\Omega) \cap C^0(\bar{\Omega})$ . Then*

$$(3.8) \quad \delta_E(e^{\psi_E} u) = \frac{1}{|\tau_E|} \int_E \alpha^{-1} e^{\psi_E} (J(u) \cdot \tau_E) ds,$$

where  $J(u) = \alpha \nabla u + \boldsymbol{\beta} u$ .

*Proof.* After multiplying both sides of  $J(u) = \alpha \nabla u + \boldsymbol{\beta} u$  by  $\alpha^{-1}$ , and taking the Euclidean inner product with the directional vector  $\tau_E$ , we obtain

$$(\nabla u \cdot \tau_E) + \alpha^{-1} (\boldsymbol{\beta} \cdot \tau_E) u = \alpha^{-1} (J(u) \cdot \tau_E).$$

Now using the definition of  $\psi_E$  in (3.7) we get

$$(3.9) \quad e^{-\psi_E} \frac{\partial (e^{\psi_E} u)}{\partial \tau_E} = \frac{1}{|\tau_E|} \alpha^{-1} (J(u) \cdot \tau_E).$$

The equality (3.8) follows from (3.9) after integration over edge  $E$ . □

Let  $\tilde{\alpha}_E(\boldsymbol{\beta})$  be the harmonic average of  $\alpha e^{-\psi_E}$  over  $E$ , defined as follows:

$$(3.10) \quad \tilde{\alpha}_E(\boldsymbol{\beta}) = \left[ \frac{1}{|\tau_E|} \int_E \alpha^{-1} e^{\psi_E} ds \right]^{-1}.$$

First we approximate  $J(u)$  over each simplex  $T$  by a constant vector  $J_T(u)$ . Then from (3.8) we have that

$$(3.11) \quad J_T(u) \cdot \tau_E \approx \tilde{\alpha}_E(\boldsymbol{\beta}) \delta_E(e^{\psi_E} u).$$

By (2.3) and (2.4), for any  $v_h \in V_h$  we get

$$(3.12) \quad \int_T J_T(u) \cdot \nabla v_h dx = \sum_E \omega_E^T (J_T(u) \cdot \tau_E) \delta_E v_h \approx \sum_{ECT} \omega_E^T \tilde{\alpha}_E(\boldsymbol{\beta}) \delta_E(e^{\psi_E} u) \delta_E v_h.$$

Thus the approximating bilinear form can be defined as

$$(3.13) \quad a_h(u_h, v_h) = \sum_{T \in \mathcal{T}_h} \left\{ \sum_{ECT} \omega_E^T \tilde{\alpha}_E(\boldsymbol{\beta}) \delta_E(e^{\psi_E} u_h) \delta_E v_h \right\}.$$

Apparently, (3.13) can be rewritten as follows:

$$(3.14) \quad a_h(u_h, v_h) = \sum_{E \in \mathcal{T}_h} \omega_E \tilde{\alpha}_E(\boldsymbol{\beta}) \delta_E(e^{\psi_E} u_h) \delta_E v_h,$$

where  $\omega_E$  is given by (2.5). Our finite element discretization is: Find  $u_h \in V_h$  such that

$$(3.15) \quad a_h(u_h, v_h) = f(v_h) \quad \text{for any } v_h \in V_h.$$

From the above derivation we can easily obtain the identity

$$(3.16) \quad a_h(u_I, v_h) = \sum_{T \in \mathcal{T}_h} \left\{ \sum_{E \subset T} \omega_E^T \left[ \frac{\tilde{\alpha}_E(\beta)}{|\tau_E|} \int_E \frac{e^{\psi_E}}{\alpha} J(u) \cdot \tau_E ds \right] \delta_E v_h \right\}$$

for all  $u \in H_0^1(\Omega) \cap C^0(\bar{\Omega})$ , where  $u_I \in V_h$  is the nodal value interpolant of  $u$ . This identity will be useful in error analysis.

*Remark 3.1.* We have pointed out that  $\psi_E$  is defined up to an arbitrary constant on  $E$  (by (3.7)), but this has no effect on the definition of the bilinear form, since (3.13) is invariant if we take  $\psi_E + \psi_E^0$  in place of  $\psi_E$  for any constant  $\psi_E^0$  on each edge.

We shall prove that the EAFE discretization is monotone.

**Lemma 3.2.** *The stiffness matrix corresponding to the bilinear form (3.13) is an M-matrix for any continuous functions  $\alpha > 0$  and  $\beta$  if and only if the stiffness matrix for the Poisson equation is an M-matrix, namely if and only if the condition (2.5) holds.*

*Proof.* Given  $j \in \{1, \dots, N_h\}$ , consider the corresponding node  $x_j$ . Obviously, if  $x_i$  is a neighbor of  $x_j$ ,

$$(3.17) \quad A_{ij} = \sum_{E \ni x_j} \omega_E \tilde{\alpha}_E(\beta) \delta_E (e^{\psi_E} \varphi_j) \delta_E \varphi_i = -\omega_E \tilde{\alpha}_E(\beta) (e^{\psi_{j,E}}) \leq 0.$$

Here  $E \ni x_j$  means all the edges having  $x_j$  as an endpoint, and  $\psi_{j,E} = \psi_E(x_j)$ .

Now, if  $x_j$  has no neighboring node on the boundary, then the  $j$ -th column sum of  $A$  is zero:

$$\sum_i A_{ij} = \sum_{E \ni x_j} \omega_E \tilde{\alpha}_E(\beta) \delta_E (e^{\psi_E} \varphi_j) \delta_E \sum_i \varphi_i = \sum_{E \ni x_j} \omega_E \tilde{\alpha}_E(\beta) \delta_E (e^{\psi_E} \varphi_j) \delta_E 1 = 0,$$

which means that  $A_{jj} = \sum_{i \neq j} |A_{ij}|$ . And if  $x_j$  has a neighboring node on the boundary, it is easy to see that  $\sum_i A_{ij} > 0$ , or  $A_{jj} > \sum_{i \neq j} |A_{ij}|$ . This completes the proof.  $\square$

*Remark 3.2.* In some applications such as semiconductor device simulation, the following equation is of special interest:

$$(3.18) \quad -\nabla \cdot (\nabla u + \nabla \psi u) = f, \quad x \in \Omega.$$

This can be viewed as a special case of our model problem (3.1) with  $\alpha = 1$  and  $\beta = \nabla \psi$ . In this case, the function  $\psi_E$  defined by (3.7) can be chosen independent of  $E$ :

$$\psi_E = \psi \quad \forall E.$$

A very special feature of this equation is that it is symmetrizable, since it can obviously be written as

$$(3.19) \quad \nabla \cdot (e^{-\psi} \nabla (e^{\psi} u)) = f, \quad x \in \Omega.$$

This equation has been studied by many authors, see for example [19], [9], [10], [18], [21]. Technically speaking, the symmetrizability this equation plays an important role in the aforementioned works. The one that is most closely related to our work is the paper by Markowich and Zlamal [19]. In fact, in this special case, our finite element scheme coincides with their scheme (which is only for symmetrizable equations in two dimensions). We note that our derivation and analysis are quite different and also much simpler. As another example of related work, let us briefly mention a hybrid finite element scheme in [9] (again only for (3.18) in two dimensions). This scheme amounts to the use of a harmonic average of coefficients over each element, and the corresponding stiffness matrix is an  $M$ -matrix provided that each triangle is non-obtuse.

**3.3. Implementation issue.** In this section we shall discuss the bounds for the stiffness matrix entries. In particular we shall show that the off-diagonal elements might have exponential decay, but they have slower growth (like  $h|\beta|$ ). This property is important for actual implementations. By (3.7) we have

$$(3.20) \quad \psi_E - \psi_{j,E} = \int_0^s \frac{(\beta \cdot \tau_E)}{\alpha} (t(x_i - x_j) + x_j) dt$$

with  $\tau_E = x_i - x_j$ . Hence

$$(3.21) \quad \alpha^{-1} e^{(\beta \cdot \tau_E)_{min} \int_0^s \frac{1}{\alpha} dt} \leq \alpha^{-1} e^{\psi_E - \psi_{j,E}} \leq \alpha^{-1} e^{(\beta \cdot \tau_E)_{max} \int_0^s \frac{1}{\alpha} dt},$$

where  $(\beta \cdot \tau_E)_{min} = \min_{x \in E} \{\beta(x) \cdot \tau_E\}$ ,  $(\beta \cdot \tau_E)_{max} = \max_{x \in E} \{\beta(x) \cdot \tau_E\}$ . We integrate over  $E$  in (3.21) and use the fact that for a given constant  $b$

$$b\alpha^{-1} \exp\left(b \int_0^s \frac{1}{\alpha} dt\right) = \frac{d}{ds} \exp\left(b \int_0^s \frac{1}{\alpha} dt\right).$$

A simple application of the fundamental theorem of calculus then yields

$$(3.22) \quad \tilde{\alpha}_E B\left(\frac{(\beta \cdot \tau_E)_{min}}{\tilde{\alpha}_E}\right) \geq \tilde{\alpha}_E(\beta) e^{\psi_{j,E}} \geq \tilde{\alpha}_E B\left(\frac{(\beta \cdot \tau_E)_{max}}{\tilde{\alpha}_E}\right),$$

where  $\tilde{\alpha}_E = \tilde{\alpha}_E(0)$  is the harmonic average of  $\alpha(x)$  on the edge and  $B(s)$  is the Bernoulli function, defined as follows:

$$B(s) = \begin{cases} \frac{s}{e^s - 1}, & s \neq 0, \\ 1, & s = 0. \end{cases}$$

By the mean-value theorem, there exists a  $t_E$  such that

$$(3.23) \quad (\beta \cdot \tau_E)_{min} \leq t_E \leq (\beta \cdot \tau_E)_{max}, \quad \tilde{\alpha}_E(\beta) e^{\psi_{j,E}} = \tilde{\alpha}_E B(t_E/\tilde{\alpha}_E).$$

Note that  $-h\|\beta\|_{0,\infty,\Omega} \leq t_E \leq h\|\beta\|_{0,\infty,\Omega}$  for all edges  $E$ .

Let us first assume that  $\beta$  is a constant. For  $i = 1, \dots, N_h$  in accordance with (3.23), the resulting system of linear equations for the nodal values of the discrete solution  $u_h$  has the form

$$(3.24) \quad \sum_{E=(x_i, x_j)} \omega_E \tilde{\alpha}_E \left[ B\left(\frac{-\beta \cdot \tau_E}{\tilde{\alpha}_E}\right) u(x_i) - B\left(\frac{\beta \cdot \tau_E}{\tilde{\alpha}_E}\right) u(x_j) \right] = G_i,$$

where  $G_i = \sum_{T \supset x_i} \int_T f \varphi_i dx$  and  $\tau_E = x_i - x_j$ . The summation is over all  $x_j \neq x_i$ , such that  $(x_i, x_j)$  is an edge.

Through this example one can easily see the advantages of the proposed scheme. In the case when  $\alpha$  is rapidly varying the entries of the matrix are smooth quantities. For example, if  $s \rightarrow +\infty$ , then  $B(s)$  approaches zero exponentially, and  $B(-s)$  behaves like  $s$ .

Related schemes based on finite differences are widely used in semiconductor device modeling. The one-dimensional derivation is due to Scharfetter and Gummel [25] (see also [27] and the references given therein). It is shown by special examples in [15] that for current continuity equations in two dimensions, the schemes based on the constant flux approximations appear to be the only ones which work successfully and give non-oscillatory solutions.

Regarding the case when  $\beta$  is a more general continuous function, we would like to point out that the moderate behavior of the stiffness matrix entries is preserved when  $\beta(x) \neq 0, x \in \Omega$ , although, if  $\beta$  has a stagnation point, then a diagonal entry in the stiffness matrix might be very small compared to the off-diagonal entries in the same row. This phenomenon occurs if  $\beta = 0$  near  $x_j$ , and all the scalar products  $(\beta \cdot \tau_E)$  are negative and  $|\beta|$  is “large” with respect to  $\alpha$ . We will give a simple example. Let  $j$  be fixed, and let  $\alpha = \varepsilon > 0$  and  $(\beta \cdot \tau_E) = -1 \forall E \ni x_j$ . Then

$$A_{jj} = \varepsilon |\varphi_j|_{1,\Omega}^2 B\left(\frac{1}{\varepsilon}\right).$$

More comments concerning similar behavior of the matrix entries in the hybrid and mixed finite element methods can be found in [9], [10]. To the authors’ knowledge, this is an issue in any monotone, linear discrete scheme for convection dominated problems.

#### 4. LIMITING CASE FOR VANISHING DIFFUSION COEFFICIENT

In this section we shall briefly discuss the limiting case when the diffusion coefficient approaches zero. The resulting scheme is a special upwinding scheme. The following simple lemma is a useful tool in investigating the limiting case.

**Lemma 4.1.** *Let  $\eta \in C^1([0, 1])$  and  $\eta(0) = 0$ . Then*

$$(4.1) \quad \lim_{\varepsilon \searrow 0} \frac{1}{\varepsilon} \int_0^1 e^{\eta(s)/\varepsilon} ds = \begin{cases} \infty, & \text{if } \eta(s) > 0, \quad 0 < s \leq 1, \\ \frac{-1}{\eta'(0)} & \text{if } \eta(s) < 0, \quad 0 < s \leq 1 \text{ and } \eta'(0) < 0. \end{cases}$$

*Proof.* Since the first identity is trivial, we shall only prove the second one. Let  $\xi(s) = 1 - \eta'(s)/\eta'(0)$ . We observe that  $\eta(s) \leq -c_0 s$  for  $s \in [0, 1]$  with some constant  $c_0 > 0$ . It follows that

$$\begin{aligned} \left| \frac{1}{\varepsilon} \int_0^1 e^{\eta(s)/\varepsilon} \xi(s) ds \right| &\leq \max_{0 \leq s \leq \sqrt{\varepsilon}} |\xi(s)| \frac{1}{\varepsilon} \int_0^{\sqrt{\varepsilon}} e^{-c_0 s/\varepsilon} ds \\ &\quad + \max_{\sqrt{\varepsilon} \leq s \leq 1} |\xi(s)| \frac{1}{\varepsilon} \int_{\sqrt{\varepsilon}}^1 e^{-c_0 s/\varepsilon} ds. \end{aligned}$$

A straightforward integration then gives

$$\begin{aligned} \left| \frac{1}{\varepsilon} \int_0^1 e^{\eta(s)/\varepsilon} \xi(s) ds \right| &\leq \max_{0 \leq s \leq \sqrt{\varepsilon}} |\xi(s)| \frac{1}{c_0} (1 - e^{-c_0/\sqrt{\varepsilon}}) \\ &\quad + \max_{0 \leq s \leq 1} |\xi(s)| \frac{1}{c_0} (e^{-c_0/\sqrt{\varepsilon}} - e^{-c_0/\varepsilon}). \end{aligned}$$

This proves that

$$\lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \int_0^1 e^{\eta(s)/\epsilon} \xi(s) ds = 0,$$

which leads to the desired result.  $\square$

Let  $\alpha(x) \equiv \epsilon > 0$ . By (3.17) and (3.20) the  $(i, j)$  entry of the stiffness matrix is given by

$$A_{ij}^\epsilon = \left[ \frac{1}{\epsilon} \int_0^1 e^{\eta(s)/\epsilon} ds \right]^{-1},$$

where  $\eta(s) = \int_0^s (\boldsymbol{\beta} \cdot \boldsymbol{\tau}_E)(tx_j + (1-t)x_i) dt$ . For simplicity let us assume that  $(\boldsymbol{\beta} \cdot \boldsymbol{\tau}_E)$  does not change sign on  $E$ . From an application of Lemma 4.1 we get

$$(4.2) \quad A_{ij}^0 = \lim_{\epsilon \searrow 0} A_{ij}^\epsilon = \begin{cases} -(\boldsymbol{\beta} \cdot \boldsymbol{\tau}_E)(x_i), & (\boldsymbol{\beta} \cdot \boldsymbol{\tau}_E) < 0 \text{ on } E, \\ 0, & (\boldsymbol{\beta} \cdot \boldsymbol{\tau}_E) > 0 \text{ on } E. \end{cases}$$

Let us denote  $\omega_{ij} = \omega_E$ ,  $N(i) = \{j : (x_i, x_j) \text{ is an edge}\}$ ,

$$x_{ij}^\beta = \begin{cases} x_i, & (x_j - x_i) \cdot \boldsymbol{\beta} < 0, \\ x_j, & (x_j - x_i) \cdot \boldsymbol{\beta} > 0. \end{cases}$$

Then the  $i$ -th equation in the resulting scheme can be written as

$$\sum_{j \in N(i)} \omega_{ij} (x_j - x_i) \cdot \boldsymbol{\beta}(x_{ij}^\beta) u_h(x_{ij}^\beta) = G_i,$$

where  $G_i = \int_\Omega f \varphi_i dx$  and the summation takes only the edges  $E = (x_i, x_j)$ .

### 5. THE EAFE SCHEME FOR MORE GENERAL EQUATIONS

In the rest of the paper, we shall study the following more general model problem:

$$(5.1) \quad \begin{cases} \mathcal{L}u \equiv -\nabla \cdot (\alpha(x)\nabla u + \boldsymbol{\beta}(x)u) + \gamma(x)u = f(x), & x \in \Omega, \\ u = 0, & x \in \bar{\Gamma}_D, \\ \alpha \frac{\partial u}{\partial \boldsymbol{\nu}} + (\boldsymbol{\beta} \cdot \boldsymbol{\nu})u = 0, & x \in \Gamma_N. \end{cases}$$

We assume that  $\alpha$ ,  $\boldsymbol{\beta}$  and  $\gamma$  are piecewise smooth functions on  $\bar{\Omega}$  and  $\alpha(x) \geq \alpha_0 > 0$ ,  $\gamma(x) \geq 0$ . We introduce the space of functions vanishing on  $\Gamma_D$ :  $H_D^1(\Omega) = \{v \in H^1(\Omega) : v(x) = 0, x \in \Gamma_D\}$ . Then the variational formulation of the above problem is: Find  $u \in H_D^1(\Omega)$  such that

$$(5.2) \quad a(u, v) = f(v) \quad \text{for every } v \in H_D^1(\Omega),$$

where

$$(5.3) \quad a(u, v) = \int_\Omega (\alpha \nabla u + \boldsymbol{\beta} u) \cdot \nabla v dx + \int_\Omega \gamma u v dx, \quad f(v) = \int_\Omega f v dx.$$

This problem is well posed and has a unique solution (see [14]).

Given  $T \in \mathcal{T}_h$  and an edge  $E \in T$ , we define a function  $\psi_E^T$  by

$$(5.4) \quad \frac{\partial \psi_E^T}{\partial \boldsymbol{\tau}_E} = \frac{1}{|\boldsymbol{\tau}_E|} \alpha^{-1} (\boldsymbol{\beta} \cdot \boldsymbol{\tau}_E).$$

We note that the superscript “ $T$ ” in  $\psi_E^T$  indicates that  $\psi_E^T$  may be different on different elements because of possible discontinuity in  $\alpha$  and  $\boldsymbol{\beta}$ . We also note that

the trace of  $\alpha$  and  $\beta$  on  $E$  from  $T$  is well defined. Let  $\tilde{\alpha}_E^T(\beta)$  be the corresponding harmonic average of  $\alpha e^{-\psi_E^T}$ ,

$$(5.5) \quad \tilde{\alpha}_E^T(\beta) = \left[ \frac{1}{|\tau_E|} \int_E \alpha^{-1} e^{\psi_E^T} ds \right]^{-1}.$$

Let us also assume that if the boundary condition changes its type at some node in  $\partial\Omega$ , this node is a vertex of a triangle  $T \in \mathcal{T}_h$ . We set  $V_h$  to be the usual piecewise linear finite element space:  $V_h \subset H_D^1(\Omega)$ . With an argument completely analogous to that in Section 3 we can obtain the discrete bilinear form (approximation to the one in (5.3)) as follows:

$$(5.6) \quad a_h(u_h, v_h) = \sum_{T \in \mathcal{T}_h} \left\{ \sum_{E \subset T} \omega_E^T \tilde{\alpha}_E^T(\beta) \delta_E(e^{\psi_E^T} u_h) \delta_E v_h + \gamma_T(u_h v_h) \right\}.$$

The last term in the above equation comes from a standard “mass-lumping” quadrature on each triangle:

$$\gamma_T(u_h v_h) = \frac{|T|}{n+1} \sum_{i=1}^{n+1} \gamma(q_i) u_h(q_i) v_h(q_i),$$

where we recall that  $q_i$  are vertices of  $T$ . The resulting finite element scheme is then: Find  $u_h \in V_h$  such that

$$(5.7) \quad a_h(u_h, v_h) = f(v_h) \quad \text{for any } v_h \in V_h.$$

It is worth noting that (3.14) is no longer valid. But the analogue of (3.16) remains true, namely

$$(5.8) \quad a_h(u_I, v_h) = \sum_{T \in \mathcal{T}_h} \left\{ \sum_{E \subset T} \omega_E^T \left[ \frac{\tilde{\alpha}_E^T(\beta)}{|\tau_E|} \int_E \frac{e^{\psi_E^T}}{\alpha} J(u) \cdot \tau_E ds \right] \delta_E v_h + \gamma_T(u_I v_h) \right\}$$

for  $u \in H_D^1(\Omega) \cap C^0(\bar{\Omega})$ .

The  $M$ -matrix property also holds under some slightly stronger assumptions when the coefficients are only piecewise smooth. In fact, by an argument analogous to that in the proof of Lemma 3.2 we have the following result.

**Lemma 5.1.** *Let  $\gamma \geq 0$ . The stiffness matrix corresponding to the bilinear form (5.6) is an  $M$ -matrix for any piecewise smooth functions  $\alpha > 0$  and  $\beta$ , if for any edge  $E$  where the coefficients  $\alpha$  and  $\beta$  have discontinuity, the angles  $\theta_E$  satisfy  $0 < \theta_E^T \leq \frac{\pi}{2}$  for all  $T \supset E$  and (2.5) is satisfied for all other edges.*

### 6. ERROR ANALYSIS

In this section we present some error estimates for the EAFE scheme using the more general problem in Section 5. As we shall see, in comparison with other upwinding type schemes, one distinctive feature of our EAFE scheme is that its error analysis appears to be completely straightforward. Of course the error analysis we are talking about here is a standard formal analysis if we assume that the solution has a certain regularity.

In the convection dominated case, like any other schemes, an analysis taking into account some singular behavior of the solution is much more elaborate. We will report such an analysis in our future work (cf. Xu and Ying [32]).

**6.1. An estimate for the discrete bilinear form.** Our estimate will be based on the assumptions made in Section 5. In addition, we also assume that, for all  $T \in \mathcal{T}_h$ , the solution of the problem (5.1) satisfies  $J(u) \equiv \alpha(x)\nabla u + \beta(x)u \in [W^{1,p}(T)]^n$  and  $\gamma(x) \in C(\bar{T})$ ,  $\gamma u \in W^{1,r}(T)$  with  $r > n$ ,

$$(6.1) \quad p = 2 \text{ for } n = 1, 2 \text{ and } p > n - 1 \text{ for } n > 2.$$

As a first step we give an estimate for the difference between continuous and approximating bilinear forms.

**Lemma 6.1.** *Let  $w \in C(\bar{\Omega})$ . If for any  $T \in \mathcal{T}_h$  we have  $J(w) \in [W^{1,p}(T)]^n$  and  $\gamma w \in W^{1,r}(T)$ , then the following inequality holds for every  $v_h \in V_h$ :*

$$(6.2) \quad |a(w, v_h) - a_h(w_I, v_h)| \leq Ch \left\{ \sum_{T \in \mathcal{T}_h} |J(w)|_{1,p,T}^2 + \sum_{T \in \mathcal{T}_h} |\gamma w|_{1,r,T}^2 \right\}^{\frac{1}{2}} \|v_h\|_{1,\Omega}$$

for  $p$  satisfying (6.1)

*Proof.* By (5.8) we have

$$(6.3) \quad a(w, v_h) - a_h(w_I, v_h) = \sum_{T \in \mathcal{T}_h} \mathcal{E}_T(J, v_h) + \mathcal{Q}_T(\gamma w, v_h),$$

where

$$(6.4) \quad \begin{aligned} \mathcal{E}_T(J(w), v_h) &= \int_T J(w) \cdot \nabla v_h dx \\ &\quad - \sum_{E \subset T} \omega_E^T \left[ \frac{\tilde{\alpha}_E(\beta)}{|\tau_E|} \int_E \frac{e^{\psi_E}}{\alpha} J(w) \cdot \tau_E ds \right] \delta_E v_h, \end{aligned}$$

and

$$(6.5) \quad \mathcal{Q}_T(\gamma w, v_h) = \int_T (\gamma w v_h - (\gamma w v_h)_I) dx.$$

We first consider  $\mathcal{E}_T(J(w), v_h)$  and apply the standard scaling from  $T$  to the reference element  $\hat{T}$ . The scaled bilinear functional is properly bounded:

$$\hat{\mathcal{E}}_{\hat{T}}(\widehat{J(w)}, \hat{v}_h) \leq \begin{cases} C_0 (\|\widehat{J(w)}\|_{0,1,\partial\hat{T}} + \|\widehat{J(w)}\|_{0,\hat{T}}) \|\hat{v}_h\|_{1,\hat{T}}, \\ C_1 \|\widehat{J(w)}\|_{0,\infty,\hat{T}} \|\hat{v}_h\|_{1,\hat{T}}, \end{cases}$$

where  $C_0$  might depend on  $\alpha$  and  $\beta$  but  $C_1$  is independent of  $\alpha, \beta$ . Let us first assume that  $p > n$ . By the Sobolev embedding theorem (see [1]),  $W^{1,p}(\hat{T}) \hookrightarrow W^{0,\infty}(\hat{T})$ , we get

$$\|\widehat{J(w)}\|_{0,\infty,\hat{T}} \leq C \|\widehat{J(w)}\|_{1,p,\hat{T}}.$$

By the trace theorem, if  $p = 2$  for  $n = 2$  or  $p > n - 1$  for  $n > 2$ , then

$$\|\widehat{J(w)}\|_{0,1,\partial\hat{T}} \leq C \|\widehat{J(w)}\|_{1,p,\hat{T}}.$$

From the derivation of the EAFE scheme, it is clear that  $\mathcal{E}_T(J(w), v_h) = 0$  if  $J(w) \equiv \text{const}$  on  $T$ . With this simple fact in hand the rest of the analysis is completely routine. By applying the Bramble-Hilbert lemma on  $\hat{T}$  and scaling back to  $T$ , we get

$$(6.6) \quad |\mathcal{E}_T(J, v_h)| \leq Ch |J(w)|_{1,p,T} |v_h|_{1,T}.$$

The estimate for  $\mathcal{Q}_T$  can be done using the similar and simpler argument (note that  $\mathcal{Q}_T(\gamma w, v_h) = 0$  if  $\gamma w \equiv \text{const}$ ). We have

$$\mathcal{Q}_T(\gamma w, v_h) \leq Ch|\gamma w|_{1,r,T}\|v_h\|_{1,T}.$$

The proof is then completed by summing over all elements and using the Schwarz inequality.  $\square$

**6.2. Well-posedness of the discrete problem.** In this section we shall consider the conditions for existence and uniqueness of the discrete solution. In what follows we take  $\gamma = 0$ . The reason is that a positive lower order term does not add any difficulties to the analysis.

**Lemma 6.2.** *Let  $\alpha \in W^{1,\infty}(T)$  and  $\beta \in [W^{1,\infty}(T)]^n$  for all  $T \in \mathcal{T}_h$ . Then for sufficiently small  $h$*

$$(6.7) \quad \sup_{v_h \in V_h} \frac{a_h(w_h, v_h)}{\|v_h\|_{1,\Omega}} \geq c_0\|w_h\|_{1,\Omega} \quad \forall w_h \in V_h.$$

*Proof.* It is well-known (see Schatz [26] or Xu [31]) that if the discrete problem is defined by the original bilinear form, then the following estimates hold for sufficiently small  $h$ :

$$(6.8) \quad \sup_{v_h \in V_h} \frac{a(w_h, v_h)}{\|v_h\|_{1,\Omega}} \geq 2c_0\|w_h\|_{1,\Omega}, \quad \forall w_h \in V_h.$$

Let  $v_h, w_h \in V_h$ . Then obviously

$$a_h(w_h, v_h) = a(w_h, v_h) + [a_h(w_h, v_h) - a(w_h, v_h)].$$

The first term is estimated using the condition (6.8). By Lemma 6.1,

$$|a(w_h, v_h) - a_h(w_h, v_h)| \leq Ch \left\{ \sum_{T \in \mathcal{T}_h} |J(w_h)|_{1,p,T}^2 \right\}^{\frac{1}{2}} \|v_h\|_{1,\Omega}.$$

Observing that  $|w_h|_{2,T} = 0$  for any  $w_h \in V_h$  and  $T \in \mathcal{T}_h$ , we get

$$|J(w_h)|_{1,p,T} \leq C(\|\alpha\|_{1,\infty,T} + \|\beta\|_{1,\infty,T})\|w_h\|_{1,T}.$$

Summing over all the elements of the partition, we have

$$(6.9) \quad |a(w_h, v_h) - a_h(w_h, v_h)| \leq C \max_{T \in \mathcal{T}_h} (\|\alpha\|_{1,\infty,T} + \|\beta\|_{1,\infty,T}) h \|w_h\|_{1,\Omega} \|v_h\|_{1,\Omega}.$$

The desired result is easily obtained if

$$h \leq h_0 \equiv c_0 \left[ C \max_{T \in \mathcal{T}_h} (\|\alpha\|_{1,\infty,T} + \|\beta\|_{1,\infty,T}) \right]^{-1}$$

$\square$

As a consequence of the previous lemmata we get the following convergence result.

**Theorem 6.3.** *Let  $u$  be the solution of the problem (5.1). Assume that for all  $T \in \mathcal{T}_h$  we have  $\alpha \in W^{1,\infty}(T)$ ,  $\beta \in [W^{1,\infty}(T)]^n$ ,  $J(u) \equiv \alpha(x)\nabla u + \beta(x)u \in (W^{1,p}(T))^n$ ,  $\gamma(x) \in C(\bar{T})$  and  $\gamma u \in W^{1,r}(T)$ . Then the following estimate holds:*

$$(6.10) \quad \|u_I - u_h\|_{1,\Omega} \leq Ch \left\{ \sum_{T \in \mathcal{T}_h} |J(u)|_{1,p,T}^2 + \sum_{T \in \mathcal{T}_h} |\gamma u|_{1,r,T}^2 \right\}^{\frac{1}{2}}$$

for sufficiently small  $h$ .

*Remark 6.1.* It is clear that under the assumptions of Lemma 5.1 the discrete problem has a unique solution without assuming  $h$  to be sufficiently small, since the resulting stiffness matrix is an  $M$ -matrix. Therefore, we may conclude that, under the assumptions in Lemma 5.1, (6.7) in fact holds for any feasible mesh size  $h$  with a constant  $c_0$  independent of  $h$ . Indeed this conclusion can be rigorously justified, but we will not get into the details here.

## 7. NUMERICAL EXAMPLES

Our EAFE scheme is a type of upwinding scheme, and hence its numerical behavior is similar to other upwinding schemes. Here we report two simple, but not trivial examples of convection dominated problems. The computational domain is the square  $\Omega = (0, 1) \times (0, 1)$ . As our first example we consider the equation

$$(7.1) \quad -\nabla \cdot (\varepsilon \nabla u + (y, -x)u) = 1,$$

subject to the homogeneous Dirichlet boundary conditions.

As our second example we consider a symmetrizable convection diffusion problem, similar to the one presented in [9]. The differential equation is

$$(7.2) \quad \begin{aligned} -\nabla \cdot (\varepsilon \nabla u + \nabla \psi u) &= 0, & x \in \Omega, \\ u &= g, & x \in \bar{\Gamma}_D, \\ \frac{\partial u}{\partial \nu} + \frac{\partial \psi}{\partial \nu} u &= 0, & x \in \Gamma_N. \end{aligned}$$

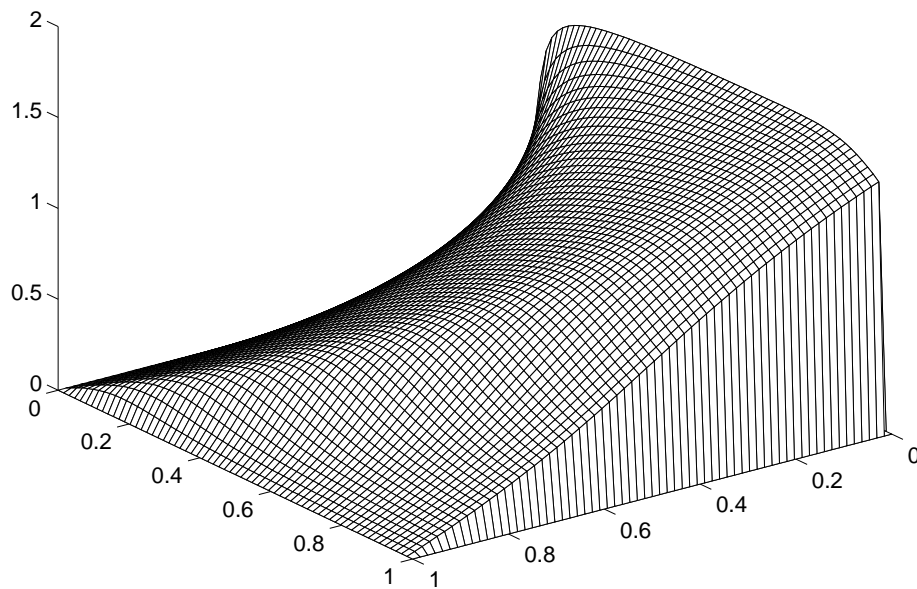


FIGURE 7.1. Surface plot of the discrete solution to (7.1)

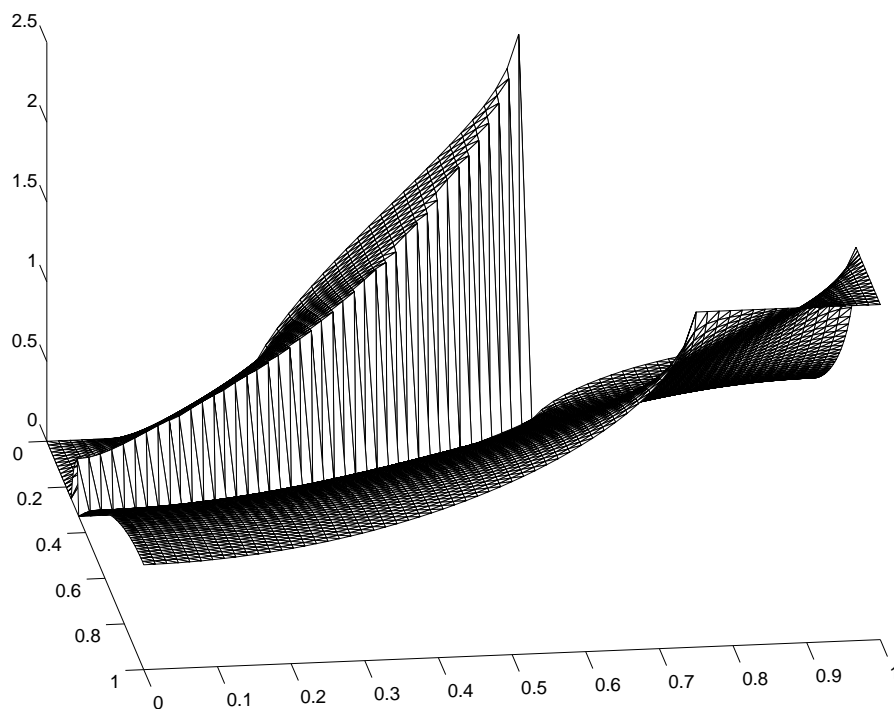


FIGURE 7.2. Surface plot of the EAFE solution to (7.2)

The potential function  $\psi$  is defined as

$$\psi = \begin{cases} 0, & 0 \leq \rho + x < 0.55, \\ 2(\rho - 0.55), & 0.55 \leq \rho + x < 0.65, \\ 0.2, & 0.65 \leq \rho + x, \end{cases}$$

where  $\rho = (x^2 + y^2)^{1/2}$ . Dirichlet boundary conditions are prescribed on the part of the boundary as follows

$$g = \begin{cases} 0 & \{x = 0, y \in [0, 0.25]\} \cup \{x \in [0, 0.25], y = 0\}, \\ -2.1 & \{x = 1, y \in [0.75, 1]\} \cup \{x \in [0.75, 1], y = 1\}. \end{cases}$$

In Figures 7.1 and 7.2 we have plotted the solutions. In both examples we have taken  $\varepsilon = 10^{-6}$  and  $h = 2^{-6}$ . Compared to the characteristic mesh size  $h$ , the ratio  $h/\varepsilon = 15625$  is rather large. In Figures 7.1 and 7.2, it is clearly seen that there are no spurious oscillations or smearing near boundary or internal layers. Our second numerical example also shows that in the subdomain where the gradient of  $\psi$  is well behaved, namely, for  $\rho(x, y) + x > 0.65$ , the discrete solution is smooth, as expected.

APPENDIX. A PROOF OF (2.4)

We shall give a proof of (2.4). Let us introduce some notation (see Figure A.1). Given an  $m$ -dimensional simplex  $S$ , let  $\tilde{S}$  denote the hyperplane containing it. Let  $\nu_k$  denote the outward unit normal vector to the face  $F_k$ ,  $k = i, j$ . Define

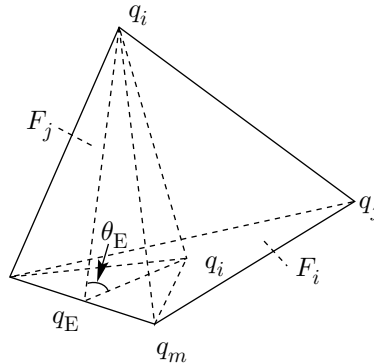


FIGURE A.1

the projections  $q_i^*$  and  $q_E^{**}$  as follows:

$$\begin{aligned} q_i^* \in \tilde{F}_i : & \quad (q_i^* - q_i) \cdot (s - q_j) = 0, \quad \forall s \in \tilde{F}_i, \quad i = 1, \dots, n + 1, j \neq i, \\ q_E^{**} \in \tilde{\kappa}_E : & \quad (q_E^{**} - q_i) \cdot (s - q_m) = 0, \quad \forall s \in \tilde{\kappa}_E. \end{aligned}$$

By definition  $\nu_i = \frac{q_i^* - q_i}{|q_i - q_i^*|}$ . The vectors  $\nu_i, \nu_j$  and  $\nu_E = \frac{q_E^{**} - q_i}{|q_E^{**} - q_i|}$  are all orthogonal to  $\tilde{\kappa}_E$ , which has dimension  $n - 2$ . Hence they must be linearly dependent. It follows then that they are congruent with the sides of a planar right triangle. Consequently

$$(A.1) \quad \frac{|q_i - q_i^*|}{|q_i - q_E^{**}|} = \sin \theta_E.$$

For  $\varphi_k$  we have

$$\varphi_k = \frac{(x - q_k^*) \cdot (q_k - q_k^*)}{|q_k - q_k^*|^2}, \quad k = i, j.$$

To prove (2.4) we apply the formula for the volume of the simplex  $|T| = \frac{1}{n} |F_k| |q_k^* - q_k|$  twice (first for  $|T|$ , then for  $|F_j|$ ) and we get

$$\int_T \nabla \varphi_i \cdot \nabla \varphi_j dx = -|T| \frac{\cos \theta_E}{|q_i - q_i^*| |q_j - q_j^*|} = -\frac{\cos \theta_E}{n |q_i - q_i^*|} |F_j| = -\frac{\cot \theta_E}{n(n-1)} |\kappa_E|.$$

In the last equality we have used (A.1). This completes the proof.

ACKNOWLEDGMENTS

The first author’s work was completed while both authors were visiting the Department of Mathematics at UCLA, and thanks go to the Applied Math Group at UCLA and especially to Professors Tony Chan, Bjorn Engquist and Stanley Osher for helpful discussions on related topics. Thanks also go to Professor Randolph Bank for relevant remarks on the results reported in this paper, to Feng Wang for his help on numerical experiments, and to David Keyes for his hospitality and discussions during the first author’s visit at ICASE, while this work was initiated.

## REFERENCES

1. R. Adams, *Sobolev spaces*, Academic Press Inc., 1975. MR **56**:9247
2. I. Babuška and J. Osborn, *Generalized finite element methods: their performance, and their relation to the mixed methods*, SIAM J. Num. Anal. **20** (1983), no. 3, 510–536. MR **84h**:65076
3. R. Bank, J. Bürger, W. Fichtner, and R. Smith, *Some up-winding techniques for finite element approximations of convection diffusion equations*, Numer. Math. **58** (1990), 185–202. MR **91i**:65175
4. R. Bank and D. Rose, *Some error estimates for the box method*, SIAM J. Num. Anal. **24** (1987), 777–787. MR **88j**:65235
5. T. Barth, *Aspects of unstructured grids and finite-volume solvers for the Euler and Navier-Stokes equations*, Tech. Report AGARD Report 787, AGARD, 1992, Special course on unstructured grids methods for advection dominated flows.
6. M. Bern and D. Eppstein, *Mesh generation and optimal triangulation*, Computing in Euclidian Geometry, World Scientific, 1992, 23–90. MR **94i**:68034
7. J. Bey and G. Wittum, *Downwind Gauß-Seidel smoothing for convection dominated problems*, Numer. Linear Algebra Appl. **4** (1997), no. 2, 85–102.
8. ———, *Downwind numbering: A robust multigrid method for convection diffusion problems on unstructured grids*, Applied Numerical Mathematics **23** (1997), no. 1, 177–192. MR **97j**:65192
9. F. Brezzi, L. Marini, and P. Pietra, *2-dimensional exponential fitting and applications to drift-diffusion models*, SIAM J. Num. Anal. **26** (1989), no. 6, 1342–1355. MR **90m**:65194
10. ———, *Numerical simulation of semiconductor devices*, Comp. Meth. Appl. Mech. Eng. **75** (1989), no. 3, 493–514. MR **91b**:78023
11. A. Brooks and T. Hughes, *Streamline upwind/Petrov-Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equations*, Comp. Meth. in Appl. Mech. Eng. **32** (1982), 199–259. MR **83k**:76005
12. P. Ciarlet, *The finite element method for elliptic problems*, North-Holland, 1978. MR **58**:25001
13. L. Dorlofsky, B. Engquist, and S. Osher, *Triangle based adaptive stencils for the solution of hyperbolic conservation laws*, J. Comp. Phys. **98** (1992), no. 1.
14. D. Gilbarg and N. Trudinger, *Elliptic partial differential equations of second order*, Springer-Verlag, 1983. MR **86c**:35035
15. M.D. Huang, *The constant-flow patch test—a unique guideline for the evaluation of discretization schemes for the current continuity equations*, IEEE Trans. CAD **4** (1985).
16. T. Hughes, *Multiscale phenomena: Greens functions, the Dirichlet-to-Neumann formulation, subgrid scale models, bubbles and the origins of stabilized methods*, Comp. Meth. in Appl. Mech. Eng. **127** (1995), 387–401. MR **96h**:65135
17. C. Johnson, *Numerical solution of partial differential equations by the finite element method*, Cambridge University Press, Cambridge, 1987. MR **89b**:65003a
18. L. D. Marini and P. Pietra, *New mixed finite element schemes for current continuity equations*, COMPEL **9** (1990), 257–268. MR **91m**:78002
19. P. Markowich and M. Zlamal, *Inverse-average-type finite element discretizations of self-adjoint second order elliptic problems*, Math. Comp. **51** (1989), 431–449. MR **89a**:65171
20. J.J.H. Miller and S. Wang, *A triangular mixed finite element method for the stationary semiconductor device equations*, RAIRO Modél. Math. Anal. Numér. **25** (1991), 441–463. MR **92c**:65138
21. J.J.H. Miller, S. Wang, and C. Wu, *A mixed finite element method for the stationary semiconductor device equations*, Engineering Computations **5** (1988), 285–288.
22. M.S. Mock, *Analysis of a discretization algorithm for stationary continuity equations in semiconductor device models*, COMPEL **2** (1983), 117–139.
23. K. W. Morton, *Numerical solution of convection-diffusion problems*, Chapman & Hall, 1996. MR **98b**:65004
24. H. Roos, M. Stynes, and L. Tobiska, *Numerical methods for singularly perturbed differential equations*, Springer, 1996. MR **99a**:65134
25. D. Scharfetter and H. Gummel, *Large-signal analysis of a silicon read diode oscillator*, IEEE Trans. Electron Devices **ED-16** (1969), 64–77.
26. A. Schatz, *An observation concerning Ritz-Galerkin methods with indefinite bilinear forms*, Math. Comp. **28** (1974), no. 205, 959–962. MR **51**:9526

27. S. Selberherr, *Analysis and simulation of semiconductor devices*, Springer-Verlag, New York, 1984.
28. G. Strang and G. Fix, *An analysis of the finite element method*, Prentice Hall, 1973. MR **56**:1747
29. F. Wang and J. Xu, *A cross-wind strip block iterative method for convection-dominated problems*, SIAM J. Comput. (submitted).
30. J. Xu, *The EAFE scheme and CWDD method for convection dominated problems*, The Proceedings for Ninth International Conference on Domain Decomposition Methods (P. Bjørstad, M. Espedal, and D. Keyes, eds.), Domain Decomposition Press, Bergen, Norway, 1998, 619–625.
31. ———, *Two-grid discretization techniques for linear and nonlinear PDE*, SIAM J. of Numer. Anal. **33** (1996), no. 5, 1759–1777. MR **97i**:65169
32. J. Xu and L. Ying, *Convergence of an explicit upwinding schemes to conservation laws in any dimensions*, (1997), preprint.
33. L. Zikatanov, *Generalized finite element method and inverse-average-type discretisation for selfadjoint elliptic boundary value problems*, Num. Meth. for PDEs (to appear).

CENTER FOR COMPUTATIONAL MATHEMATICS AND APPLICATIONS, DEPARTMENT OF MATHEMATICS, PENNSYLVANIA STATE UNIVERSITY, UNIVERSITY PARK, PENNSYLVANIA 16802

*E-mail address*: [xu@math.psu.edu](mailto:xu@math.psu.edu)

*URL*: <http://www.math.psu.edu/xu>

*E-mail address*: [ltz@math.psu.edu](mailto:ltz@math.psu.edu)

*URL*: <http://www.math.psu.edu/ltz>