SOME UPWINDING TECHNIQUES FOR FINITE ELEMENT APPROXIMATIONS OF CONVECTION-DIFFUSION EQUATIONS

RANDOLPH E. BANK*, JOSEF F. BÜRGLER†, WOLFGANG FICHTNER‡, AND R. KENT SMITH§

Abstract. A uniform framework for the study of upwinding schemes is developed. The standard finite element Galerkin discretization is chosen as the reference discretization, and differences between other discretization schemes and the reference are written as artificial diffusion terms. These artificial diffusion terms are spanned by a four dimensional space of element diffusion matrices. Three basis matrices are symmetric, rank one diffusion operators associated with the edges of the triangle; the fourth basis matrix is skew symmetric and is associated with a rotation by $\pi/2$. While finite volume discretizations may be written as upwinded Galerkin methods, the converse does not appear to be true. Our approach is used to examine several upwinding schemes, including the streamline diffusion method, the box method, the Scharfetter-Gummel discretization, and a divergence-free scheme.

Key words. Finite Element Methods, Upwinding, Convection Diffusion Equations.

AMS subject classifications. 65N05, 65N10, 65N20

1. Introduction.

We consider the model convection diffusion problem

$$\begin{aligned} -\nabla \cdot (\nabla u + \boldsymbol{\beta} u) &= 0 && \text{in } \Omega \subset \mathcal{R}^2 \\ u &= u_0 && \text{on } \partial\Omega_1 \\ (\nabla u + \boldsymbol{\beta} u) \cdot \boldsymbol{n} &= 0 && \text{on } \partial\Omega - \partial\Omega_1 \end{aligned} \tag{1.1}$$

Here $\boldsymbol{\beta} = \nabla \psi$ and $\psi \in \mathcal{H}^1(\Omega)$. We assume that Ω is polygonal and that $\partial\Omega_1$ is composed of one or more edges of $\partial\Omega$. The function u_0 is assumed constant on each contiguous set of Dirichlet boundary edges. The outward normal direction $\boldsymbol{n}$ is defined edgewise.

The weak form of (1.1) is: find $u \in \mathcal{H}_d$ such that

$$a(u, \phi) = \int_\Omega (\nabla u + \boldsymbol{\beta} u) \cdot \nabla \phi \, dx \, dy = 0 \tag{1.2}$$

for all $\phi \in \mathcal{H}_0$, where

$$\begin{aligned} \mathcal{H}_d &= \{u \in \mathcal{H}^1(\Omega) \text{ and } u = u_0 \text{ on } \partial\Omega_1\} \\ \mathcal{H}_0 &= \{u \in \mathcal{H}^1(\Omega) \text{ and } u = 0 \text{ on } \partial\Omega_1\} \end{aligned}$$

Let $\mathcal{T}$ be a shape regular, although not necessarily quasi uniform, triangulation of Ω, characterized by a small parameter h indicating the size of the elements. Let $\mathcal{S}_h$ be the space of continuous piecewise linear polynomials with respect to $\mathcal{T}$, and define

$$\begin{aligned} \mathcal{S}_d &= \{u \in \mathcal{S}_h \text{ and } u = u_0 \text{ on } \partial\Omega_1\} \\ \mathcal{S}_0 &= \{u \in \mathcal{S}_h \text{ and } u = 0 \text{ on } \partial\Omega_1\} \end{aligned}$$

*Department of Mathematics, University of California at San Diego, La Jolla, CA 92093, USA. The work of this author was supported by the Office of Naval Research under contract N00014-82K-0197.

†Integrated Systems Laboratory, Swiss Federal Inst. of Technology, CH-8092 Zürich, Switzerland. The work of this author was supported through KWF-Landis/Gyr Grant 1496, AT&T Bell Laboratories, and Cray Research.

‡Integrated Systems Laboratory, Swiss Federal Inst. of Technology, CH-8092 Zürich, Switzerland.

§AT&T Bell Laboratories, Murray Hill, New Jersey 07974

Here we are assuming that each point at which the type of boundary condition changes from Dirichlet to Neumann is a vertex in the triangulation $\mathcal{T}$. Also, we will assume that $\boldsymbol{\beta} = \nabla\psi$, where $\psi \in \mathcal{S}_h$. In the practical application that we have in mind, where (1.1) is a current continuity equation from the semiconductor device model and $\boldsymbol{\beta}$ is the gradient of the electrostatic potential, which itself is obtained as part of the solution of a coupled system of partial differential equations [2].

The classical Galerkin finite element method for approximating (1.2) is: find $u_g \in \mathcal{S}_d$ such that

$$a(u_g, \phi) = 0 \tag{1.3}$$

for all $\phi \in \mathcal{S}_0$. The classical method roughly corresponds to the use of centered differences in the finite difference context, and is well known to be unstable when $|\boldsymbol{\beta}|h$ is large.

This has led to the use of upwind finite element techniques [4, 5, 6, 7, 8, 9], which are analogous to the use of upwind differences in the finite difference arena. In this paper, we develop a uniform framework for the study of general upwinding schemes. We choose the standard weak Galerkin form (1.3) as the reference discretization. Then differences between other discretization schemes and the weak Galerkin form are written as artificial diffusion terms; that is, we seek to write all schemes in the form:

$$a_h(u, \phi) = a(u, \phi) + \sum_{\tau \in \mathcal{T}} \int_\tau h_\tau (\rho \nabla u) \cdot \nabla\phi \, dx\, dy = 0. \tag{1.4}$$

Here $\rho \equiv \rho_\tau$ is a 2×2 diffusion matrix, defined elementwise and is characteristic of the particular scheme, and h_τ is a measure of the size of τ, for example, its diameter. Normally, one might tend to think of ρ as a symmetric, positive semidefinite matrix, but this will not be the case with many of the methods. The bilinear form $a_h(\cdot, \cdot)$ formally corresponds to the perturbed equation

$$-\nabla \cdot ((I + h_\tau \rho)\nabla u + \boldsymbol{\beta} u) = 0$$

for $\tau \in \mathcal{T}$.

For piecewise linear triangular elements, the diffusion term $h_\tau \rho$ is contained in a four dimensional space of element diffusion matrices. Three basis matrices for this space are symmetric, rank one diffusion operators that can naturally be associated with the edges of the triangle. The fourth basis matrix is skew symmetric and is associated with a rotation by $\pi/2$.

In this paper, we will first consider the streamline diffusion method, proposed and analyzed by T. Hughes *et al* [5] [6] and C. Johnson *et al* [8], among others. As this is a standard approach, we do not make a formal derivation of the method, but rather refer to the existing literature.

We then consider the box scheme [1] and the Scharfetter-Gummel scheme [2], two finite volume discretizations. Our recasting of these schemes in the form (1.4) may be regarded as an extension of [1], in which only self-adjoint problems were considered. Interestingly, while finite volume discretizations may always be written as upwinded Galerkin methods, the converse does not appear to be true, since the skew symmetric elementary diffusion operator seems to have no analogue in the standard finite volume framework.

Finally, we consider the new divergence free upwinding scheme proposed by the authors [3]. In some instances, the artificial diffusion introduced by this method

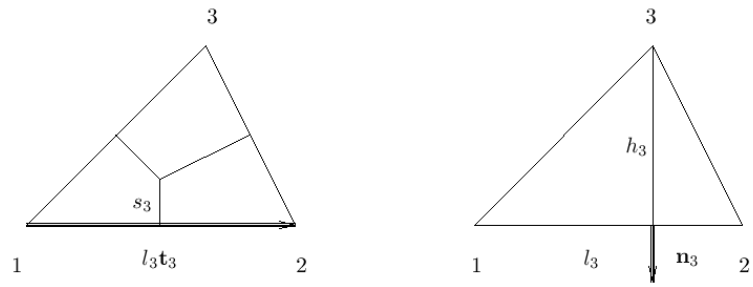

FIG. 2.1. *Parameters associated with the triangle* τ

resembles that of the streamline diffusion method. In other cases, it can lead to very nonsymmetric and indefinite artificial diffusion matrices. In extreme cases, the overall diffusion matrix $I + h_\tau \rho$ can have one positive and one negative eigenvalue. Nevertheless, the method appears to be extremely robust and stable, and remains so even in unfavorable situations where other upwinding schemes fail [3].

The remainder of this paper is organized as follows: In section 2, we describe the triangular element geometry and elemental stiffness matrix. In addition, the element diffusion matrices and there properties are presented. The next four sections are devoted to discussions of the various upwinding schemes in terms of these elemental matrices. We make some concluding remarks in the final section.

2. Preliminaries. Let $\{\phi_i\}_{i=1}^n$ denote the standard nodal basis functions for $\mathcal{S}_0$. Then the global stiffness matrix A corresponding to (1.3) is given by

$$A_{ij} = a_h(\phi_j, \phi_i) \tag{2.1}$$

The global stiffness matrix may be decomposed in terms of element stiffness matrices A_τ as

$$A = \sum_{\tau \in \mathcal{T}} A_\tau$$

where

$$\begin{aligned}(A_\tau)_{ij} &= a_\tau(\phi_j, \phi_i)\\ a_\tau(\phi_j, \phi_i) &= \int_\tau (I + h_\tau \rho)\nabla\phi_j \cdot \nabla\phi_i + \boldsymbol{\beta}\phi_j \cdot \nabla\phi_i \, dx\, dy\end{aligned}$$

Since there are only three nonzero basis functions on each element, we can characterize A_τ by a dense 3×3 element matrix. Without loss of generality, or by virtue of a local coordinate renumbering, we assume that our canonical element τ has vertices $\boldsymbol{v}_i^t = (x_i, y_i)$, for $1 \le i \le 3$, and corresponding nodal basis functions $\{\phi_i\}_{i=1}^3$.

We define $\{\boldsymbol{n}_i\}_{i=1}^3$ to be the unit outward normal vectors for τ, $\{\boldsymbol{t}_i\}_{i=1}^3$ to be the unit tangent vectors for the three edges, $\{\ell_i\}_{i=1}^3$ to be their lengths, and $\{h_i\}_{i=1}^3$ to be the perpendicular heights (see Fig.1). Let $\tilde{\boldsymbol{v}}$ be the point of intersection for the perpendicular bisectors of the three sides of τ. Let $|s_j|$ denote the distance between

$\tilde{\boldsymbol{v}}$ and side j. If τ has no obtuse angles, then the s_j will be nonnegative; otherwise, the distance to the side opposite the obtuse angle will be negative.

There are many relationships among these quantities; in particular we note the following:

$$\ell_i h_i = 2|\tau|, \qquad 1 \le i \le 3 \tag{2.2}$$

$$\nabla\phi_i = -\boldsymbol{n}_i/h_i, \qquad 1 \le i \le 3 \tag{2.3}$$

$$\phi_1 + \phi_2 + \phi_3 = 1 \tag{2.4}$$

$$\nabla\phi_1 + \nabla\phi_2 + \nabla\phi_3 = 0 \tag{2.5}$$

$$\ell_1\boldsymbol{t}_1 + \ell_2\boldsymbol{t}_2 + \ell_3\boldsymbol{t}_3 = 0 \tag{2.6}$$

$$\begin{bmatrix} \ell_1\boldsymbol{t}_1^t \\ \ell_2\boldsymbol{t}_2^t \\ \ell_3\boldsymbol{t}_3^t \end{bmatrix} \begin{bmatrix} \nabla\phi_1 & \nabla\phi_2 & \nabla\phi_3 \end{bmatrix} = \begin{bmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{bmatrix} \tag{2.7}$$

$$s_1 = -|\tau|\,\ell_1\nabla\phi_2 \cdot \nabla\phi_3 \tag{2.8}$$

Equation (2.8) is valid cyclically for s_2 and s_3. A hint for verifying (2.8) is to recall that, if the angle at vertex $\boldsymbol{v}_1$ is θ_1, then the angle at $\tilde{\boldsymbol{v}}$ between the lines joining $\tilde{\boldsymbol{v}}$ to $\boldsymbol{v}_2$ and $\tilde{\boldsymbol{v}}$ to $\boldsymbol{v}_3$ is $2\theta_1$.

The affine mapping of the reference element $\hat{\tau}$, with vertices $(\hat{x}_1, \hat{y}_1) = (0,0)$, $(\hat{x}_2, \hat{y}_2) = (1,0)$, and $(\hat{x}_3, \hat{y}_3) = (0,1)$, to our canonical element τ is given by

$$\begin{bmatrix} x \\ y \end{bmatrix} = J \begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix} + \boldsymbol{v}_1 \tag{2.9}$$

with

$$J = \begin{bmatrix} \ell_3\boldsymbol{t}_3 & -\ell_2\boldsymbol{t}_2 \end{bmatrix} \tag{2.10}$$

and

$$J^{-t} = \begin{bmatrix} \nabla\phi_2 & \nabla\phi_3 \end{bmatrix} \tag{2.11}$$

The function ∇u defined on τ is transformed to $J^{-t}\hat{\nabla}\hat{u}$ defined on the reference element $\hat{\tau}$. The local basis functions on the reference element are

$$\begin{aligned} \hat{\phi}_1 &= 1 - \hat{x} - \hat{y} \\ \hat{\phi}_2 &= \hat{x} \\ \hat{\phi}_3 &= \hat{y} \end{aligned}$$

Assuming the $\boldsymbol{\beta}$ is constant on τ, as will be the case when $\boldsymbol{\beta} = \nabla\psi$ for $\psi \in \mathcal{S}_h$, the element stiffness matrix for the standard Galerkin method is given by

$$\begin{aligned} A_g = |\tau| &\begin{bmatrix} \nabla\phi_1\cdot\nabla\phi_1 & \nabla\phi_1\cdot\nabla\phi_2 & \nabla\phi_1\cdot\nabla\phi_3 \\ \nabla\phi_1\cdot\nabla\phi_2 & \nabla\phi_2\cdot\nabla\phi_2 & \nabla\phi_2\cdot\nabla\phi_3 \\ \nabla\phi_1\cdot\nabla\phi_3 & \nabla\phi_2\cdot\nabla\phi_3 & \nabla\phi_3\cdot\nabla\phi_3 \end{bmatrix} \\ &+ \frac{|\tau|}{3} \begin{bmatrix} \boldsymbol{\beta}\cdot\nabla\phi_1 \\ \boldsymbol{\beta}\cdot\nabla\phi_2 \\ \boldsymbol{\beta}\cdot\nabla\phi_3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \end{aligned} \tag{2.12}$$

The first matrix on the right hand side of (2.12) corresponds to the contribution to A_g from the Laplace operator. This matrix is symmetric, positive semi-definite and has rank two. Its kernel is spanned by the vector $(1\;1\;1)^t$, a reflection of (2.5). The second matrix corresponds to the convection term and has rank one. Note that the column sums of both matrices are zero.

In the general setting, the contribution to the element stiffness matrix from an artificial diffusion term will be a 3×3 matrix with zero row sums and zero column sums (reflecting the fact that $\nabla c = 0$ for a constant c). It is a straightforward calculation to see that this represents five independent constraints on the nine coefficients in such a matrix. A basis for the remaining four dimensional space of element diffusion matrices is given by

$$\frac{1}{|\tau|}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \end{bmatrix} = \frac{\ell_1^2}{|\tau|}\begin{bmatrix} \nabla\phi_1^t \\ \nabla\phi_2^t \\ \nabla\phi_3^t \end{bmatrix} \boldsymbol{t}_1\boldsymbol{t}_1^t \begin{bmatrix} \nabla\phi_1 & \nabla\phi_2 & \nabla\phi_3 \end{bmatrix} \tag{2.13}$$

$$\frac{1}{|\tau|}\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix} = \frac{\ell_2^2}{|\tau|}\begin{bmatrix} \nabla\phi_1^t \\ \nabla\phi_2^t \\ \nabla\phi_3^t \end{bmatrix} \boldsymbol{t}_2\boldsymbol{t}_2^t \begin{bmatrix} \nabla\phi_1 & \nabla\phi_2 & \nabla\phi_3 \end{bmatrix} \tag{2.14}$$

$$\frac{1}{|\tau|}\begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \frac{\ell_3^2}{|\tau|}\begin{bmatrix} \nabla\phi_1^t \\ \nabla\phi_2^t \\ \nabla\phi_3^t \end{bmatrix} \boldsymbol{t}_3\boldsymbol{t}_3^t \begin{bmatrix} \nabla\phi_1 & \nabla\phi_2 & \nabla\phi_3 \end{bmatrix} \tag{2.15}$$

$$\frac{1}{|\tau|}\begin{bmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} \nabla\phi_1^t \\ \nabla\phi_2^t \\ \nabla\phi_3^t \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \nabla\phi_1 & \nabla\phi_2 & \nabla\phi_3 \end{bmatrix} \tag{2.16}$$

The 2×2 diffusion matrices

$$\hat{\rho}_i = \frac{\ell_i^2}{|\tau|}\boldsymbol{t}_i\boldsymbol{t}_i^t \tag{2.17}$$

for $1 \le i \le 3$, are symmetric, rank one diffusion operators which can naturally be associated with the three edges of τ. The skew symmetric operator

$$\hat{\rho}_s = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \tag{2.18}$$

corresponds to a rotation by $\pi/2$.

If D is a 2×2 diffusion matrix, then we may expand D in terms of this basis as

$$D = \alpha_s\hat{\rho}_s + \sum_{i=1}^{3}\alpha_i\hat{\rho}_i \tag{2.19}$$

where

$$\alpha_1 = -|\tau|\nabla\phi_2 \cdot \left(\frac{D+D^t}{2}\right)\nabla\phi_3$$

(cyclically for α_2 and α_3), and

$$\alpha_s\hat{\rho}_s = \frac{D-D^t}{2}$$

These coefficients can be computed directly using (2.2)-(2.11).

As an example, the diffusion operator corresponding to the Laplace operator $-\Delta$ is the 2×2 identity matrix, which can be decomposed as

$$I_{2\times 2} = \sum_{i=1}^{3} L_i \hat{\rho}_i \tag{2.20}$$

where

$$\begin{aligned} L_1 &= -|\tau| \nabla\phi_2 \cdot \nabla\phi_3 \\ &= \frac{s_1}{\ell_1} \end{aligned} \tag{2.21}$$

The scalars L_2 and L_3 are defined cyclically.

3. The Streamline Diffusion Method. The streamline diffusion is one of the more widely used upwinding schemes in the finite element arena. Since derivations of the method are widely available in the literature [6], [5], [8], we will merely summarize the method within the current framework.

For the streamline diffusion method, the element stiffness matrix is

$$A_s = A_g + \frac{C|\tau| h_\tau}{|\boldsymbol{\beta}|} \begin{bmatrix} \boldsymbol{\beta} \cdot \nabla\phi_1 \\ \boldsymbol{\beta} \cdot \nabla\phi_2 \\ \boldsymbol{\beta} \cdot \nabla\phi_3 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \cdot \nabla\phi_1 & \boldsymbol{\beta} \cdot \nabla\phi_2 & \boldsymbol{\beta} \cdot \nabla\phi_3 \end{bmatrix} \tag{3.1}$$

where C is a positive constant.

The artificial diffusion term is a symmetric, positive semidefinite matrix of rank one, corresponding to the diffusion term

$$\rho_s = \frac{C}{|\boldsymbol{\beta}|} \boldsymbol{\beta}\boldsymbol{\beta}^t \tag{3.2}$$

This rank one matrix adds artificial diffusion in the streamline direction (in the direction of $\boldsymbol{\beta}$).

In analogy with (2.20), the diffusion may be expanded in terms of only the edge diffusion matrices $\hat{\rho}_i$, $1 \le i \le 3$ as

$$\rho_s = \sum_{i=1}^{3} \alpha_i \hat{\rho}_i$$

where

$$\alpha_1 = -\frac{C|\tau|}{|\boldsymbol{\beta}|} \boldsymbol{\beta} \cdot \nabla\phi_2 \; \boldsymbol{\beta} \cdot \nabla\phi_3$$

and α_2 and α_3 are defined cyclically.

Upwinding in the crosswind direction involves contributions perpendicular to the streamline direction. These terms are also symmetric and therefore involve only the edge diffusion operators. Thus both the streamline and the crosswind upwinding terms do not involve the skew symmetric operator given by (2.18).

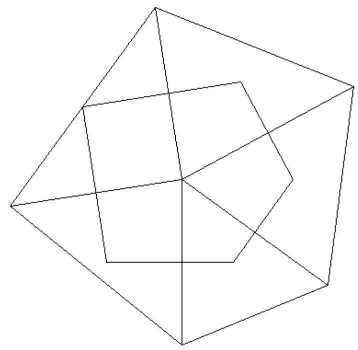

FIG. 4.1. *The box b_i*

4. The Box Method. The box method is formally derived as a finite volume approximation of (1.1). Assume, for the moment, that $\mathcal{T}$ is such that all triangles have interior angles that are not obtuse. This is nonessential to the definition, but will simplify our initial derivation. Indeed, once the box method has been cast in the form (1.4), such a restriction will obviously not be required. In any event, for each vertex $\boldsymbol{v}_i$, we can associate a *box* b_i, generated by the perpendicular bisectors of the triangle edges incident on that vertex, as illustrated in Fig. 2 (although we could allow a more general definition of boxes, as in [1]).

A given triangle τ contains parts of three boxes; thus one can easily develop the concept of an element stiffness matrix for the box method. This matrix will contain the contributions to the global matrix arising from integrals on the portions of box boundaries lying within τ. See [1] for a complete discussion of this point with respect to the Laplace operator.

We now integrate equation (1.1) over the box b_i, and then apply the divergence theorem to get

$$-\int_{\partial b_i} (\nabla u + \boldsymbol{\beta} u) \cdot \boldsymbol{n}\, ds = 0 \tag{4.1}$$

where $\boldsymbol{n}$ is the outward normal for the box b_i, defined edgewise.

Let η_i be the index set of vertices in $\mathcal{T}$ connected via a triangle edge to vertex $\boldsymbol{v}_i$. Then (4.1) is approximated by

$$\sum_{j \in \eta_i} \left\{ \left(\frac{u_i - u_j}{\ell_{ij}} \right) s_{ij} - (\boldsymbol{\beta} \cdot \boldsymbol{n}_{ij})\, u_k s_{ij} \right\} = 0 \tag{4.2}$$

where

$$k = \begin{cases} i & \text{if } \boldsymbol{\beta} \cdot \boldsymbol{n}_{ij} < 0 \\ j & \text{if } \boldsymbol{\beta} \cdot \boldsymbol{n}_{ij} \geq 0 \end{cases}$$

u_i is the approximate solution at vertex $\boldsymbol{v}_i$, ℓ_{ij} is the length of the triangle edge connecting vertices $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$, and s_{ij} is the length of the box edge corresponding to the perpendicular bisector of the edge connecting $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$. The normal directions $\boldsymbol{n}_{ij}$ for the box b_i correspond to (plus or minus) tangent directions for triangle edges.

To simplify our indices, we will write $\boldsymbol{\beta} \cdot \boldsymbol{n}_{ij} u_k$ as

$$-\boldsymbol{\beta} \cdot \boldsymbol{n}_{ij} u_k = \frac{1}{2}\left\{|\boldsymbol{\beta} \cdot \boldsymbol{n}_{ij}| - \boldsymbol{\beta} \cdot \boldsymbol{n}_{ij}\right\} u_i - \frac{1}{2}\left\{|\boldsymbol{\beta} \cdot \boldsymbol{n}_{ij}| + \boldsymbol{\beta} \cdot \boldsymbol{n}_{ij}\right\} u_j$$

so then (4.2) becomes

$$\sum_{j \in \eta_i} \left(1 + \frac{\ell_{ij}|\boldsymbol{\beta} \cdot \boldsymbol{n}_{ij}|}{2}\right)\left(\frac{u_i - u_j}{\ell_{ij}}\right) s_{ij} - (\boldsymbol{\beta} \cdot \boldsymbol{n}_{ij})\left(\frac{u_i + u_j}{2}\right) s_{ij} = 0 \tag{4.3}$$

It should be noted that the effect of the upwinding is to add a diffusion term to each triangle edge of strength $\frac{1}{2}\ell_{ij}|\boldsymbol{\beta} \cdot \boldsymbol{n}_{ij}|$.

A straightforward calculation shows that the element stiffness matrix for the box method is given by

$$\begin{aligned} A_b = {} & \left(1 + \frac{\ell_1|\boldsymbol{\beta} \cdot \boldsymbol{t}_1|}{2}\right)\frac{s_1}{\ell_1}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \end{bmatrix} + \left(1 + \frac{\ell_2|\boldsymbol{\beta} \cdot \boldsymbol{t}_2|}{2}\right)\frac{s_2}{\ell_2}\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix} \\ & + \left(1 + \frac{\ell_3|\boldsymbol{\beta} \cdot \boldsymbol{t}_3|}{2}\right)\frac{s_3}{\ell_3}\begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \frac{s_1\boldsymbol{\beta} \cdot \boldsymbol{t}_1}{2}\begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & -1 \\ 0 & 1 & 1 \end{bmatrix} \\ & + \frac{s_2\boldsymbol{\beta} \cdot \boldsymbol{t}_2}{2}\begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & -1 \end{bmatrix} + \frac{s_3\boldsymbol{\beta} \cdot \boldsymbol{t}_3}{2}\begin{bmatrix} -1 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{aligned} \tag{4.4}$$

The element stiffness matrix has zero column sums, with nonnegative diagonal and nonpositive off diagonal entries, if we assume no obtuse angles for each element. For elements with vertices on the boundary, the rows and columns corresponding to Dirichlet vertices are ignored in computing the global stiffness matrix. Thus, the global stiffness matrix will be an irreducible, diagonally dominant M-matrix with respect to its columns. This leads to a number of desirable properties, including a discrete maximum principle associated with the columns. We remark that the assumption of no obtuse angles is necessary for the condition of nonnegative diagonal and nonpositive off diagonal entries to hold element by element. It is *not* a necessary condition (but certainly sufficient) for the global stiffness matrix to inherit these properties [11].

The first three terms on the right hand side of (4.4) correspond to the Laplace term and the upwinding, which can be written as

$$|\tau| \begin{bmatrix} \nabla\phi_1^t \\ \nabla\phi_2^t \\ \nabla\phi_3^t \end{bmatrix} \left\{\sum_{i=1}^{3} L_i\hat{\rho}_i + \frac{L_i\ell_i|\boldsymbol{\beta} \cdot \boldsymbol{t}_i|\hat{\rho}_i}{2}\right\} \begin{bmatrix} \nabla\phi_1 & \nabla\phi_2 & \nabla\phi_3 \end{bmatrix}$$

where we have used (2.13)-(2.16), and (2.20)-(2.21).

The last three terms of (4.4) correspond to the centered difference approximation to the convective term by the finite volume method. To analyze these terms, we begin by defining

$$\begin{aligned} \beta_i &= \frac{1}{|\tau|}\ell_i s_i \boldsymbol{t}_i \cdot \boldsymbol{\beta} \\ &= \frac{\ell_i^2}{|\tau|} L_i \boldsymbol{t}_i \cdot \boldsymbol{\beta} \end{aligned} \tag{4.5}$$

With this definition, we have, from (2.20) and (2.21)

$$\boldsymbol{\beta} = \sum_{i=1}^{3} \beta_i \boldsymbol{t}_i \tag{4.6}$$

This decomposes $\boldsymbol{\beta}$ into components lying along the tangent directions of each edge of τ. Using (2.7), we next observe that

$$\begin{aligned}
\begin{bmatrix} -1 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} &= (\boldsymbol{e}_2 - \boldsymbol{e}_1)(\boldsymbol{e}_1 + \boldsymbol{e}_2)^t \\
&= \begin{bmatrix} \nabla\phi_1^t \\ \nabla\phi_2^t \\ \nabla\phi_3^t \end{bmatrix} \ell_3 \boldsymbol{t}_3 (\boldsymbol{e}_1 + \boldsymbol{e}_2)^t
\end{aligned}$$

where $\boldsymbol{e}_i$ is the i-th column of the 3×3 identity matrix.

Thus, the last three terms of (4.4) can be written as

$$\frac{|\tau|}{2} \begin{bmatrix} \nabla\phi_1^t \\ \nabla\phi_2^t \\ \nabla\phi_3^t \end{bmatrix} \left\{ \beta_1 \boldsymbol{t}_1 (\boldsymbol{e}_2 + \boldsymbol{e}_3)^t + \beta_2 \boldsymbol{t}_2 (\boldsymbol{e}_3 + \boldsymbol{e}_1)^t + \beta_3 \boldsymbol{t}_3 (\boldsymbol{e}_1 + \boldsymbol{e}_2)^t \right\}$$

Our next task is to compute the form of the artificial diffusion associated with the box method, and then to recast the box method in the form (1.4). We begin by finding the matrix corresponding to the upwinding (relative to the standard Galerkin method) given by $A_b - A_g$. To simplify the resulting expressions, we will need

$$\begin{aligned}
\left(\frac{\boldsymbol{e}_2 + \boldsymbol{e}_3}{2}\right) - \left(\frac{\boldsymbol{e}_1 + \boldsymbol{e}_2 + \boldsymbol{e}_3}{3}\right) &= \frac{1}{6}\left(\boldsymbol{e}_2 - \boldsymbol{e}_1 + \boldsymbol{e}_3 - \boldsymbol{e}_1\right) \\
&= \frac{1}{6} \begin{bmatrix} \nabla\phi_1^t \\ \nabla\phi_2^t \\ \nabla\phi_3^t \end{bmatrix} (\ell_3 \boldsymbol{t}_3 - \ell_2 \boldsymbol{t}_2) \\
&\equiv \frac{1}{2} \begin{bmatrix} \nabla\phi_1^t \\ \nabla\phi_2^t \\ \nabla\phi_3^t \end{bmatrix} \boldsymbol{d}_1
\end{aligned} \tag{4.7}$$

The vectors $\boldsymbol{d}_2$ and $\boldsymbol{d}_3$ are defined cyclically. Thus using (4.5)-(4.7), as well as (2.20), we have

$$A_b - A_g = \frac{|\tau|}{2} \begin{bmatrix} \nabla\phi_1^t \\ \nabla\phi_2^t \\ \nabla\phi_3^t \end{bmatrix} \left\{ \sum_{i=1}^{3} \ell_i |\beta_i| \boldsymbol{t}_i \boldsymbol{t}_i^t + \beta_i \boldsymbol{t}_i \boldsymbol{d}_i^t \right\} \begin{bmatrix} \nabla\phi_1 & \nabla\phi_2 & \nabla\phi_3 \end{bmatrix} \tag{4.8}$$

Note that $\beta_i = O(|\boldsymbol{\beta}|)$ and $\boldsymbol{d}_i = O(h_\tau)$. Thus

$$h_\tau \rho_b = \frac{1}{2} \sum_{i=1}^{3} \ell_i |\beta_i| \boldsymbol{t}_i \boldsymbol{t}_i^t + \beta_i \boldsymbol{t}_i \boldsymbol{d}_i^t \tag{4.9}$$

is the artificial diffusion term (1.4) for the box method. Note that there are two types of terms on the right hand side of (4.9). The first type comes from upwinding along a single edge; these terms contribute symmetric, positive semidefinite artificial diffusion terms to $h_\tau \rho_b$. The second set of terms arise from the differences in approximating the convection term using centered differences; the box method considers only approximations along each edge, while the standard Galerkin method develops approximations within the triangle as a whole. This generally contributes a nonsymmetric artificial diffusion term to the overall upwinding.

Having defined the form of the artificial diffusion, we can now interpret the box method as a finite element method, which remains well defined even when some elements have obtuse angles, and when $\boldsymbol{\beta}$ is no longer assumed to be constant on each element.

5. The Scharfetter-Gummel Method. A second finite volume scheme, similar to the box method of section 4, but making explicit use of the assumption that $\boldsymbol{\beta} = \nabla\psi$, is the Scharfetter-Gummel discretization. Originally proposed for the one dimensional discretization of the current continuity equation in the semiconductor device model [10], it has been generalized to two dimensions [2], and is a widely used discretization in contemporary device simulators. The Scharfetter-Gummel discretization is an exponential upwinding scheme which will produce the exact values at the vertices for a one dimensional problem in the special case where ψ is linear.

We define the Bernoulli function $\mathcal{B}(x)$ by

$$\mathcal{B}(x) = \frac{x}{e^x - 1} \tag{5.1}$$

We will use the identity

$$\mathcal{B}(-x) = \mathcal{B}(x) + x$$

in the forms

$$\begin{aligned} \mathcal{B}(x) &= \mathcal{C}(x) - \frac{x}{2} \\ \mathcal{B}(-x) &= \mathcal{C}(x) + \frac{x}{2} \end{aligned}$$

where

$$\mathcal{C}(x) = \frac{\mathcal{B}(x) + \mathcal{B}(-x)}{2} = \mathcal{B}(|x|) + \frac{|x|}{2} \tag{5.2}$$

Along the triangle edge connecting vertices $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ in element τ, the flux term

$$-(\nabla u + \boldsymbol{\beta} u) \cdot \boldsymbol{n} = -e^{-\psi} \nabla(e^{\psi} u) \cdot \boldsymbol{n}$$

is approximated along the box boundary by

$$e^{-\tilde{\psi}} \left(\frac{e^{\psi_1} u_1 - e^{\psi_2} u_2}{\ell_3} \right) \tag{5.3}$$

where $\psi_i \equiv \psi(\boldsymbol{v}_i)$. The value of $\tilde{\psi}$ is given by [2]

$$e^{-\tilde{\psi}} = \left(\frac{\int_{\psi_2}^{\psi_1} e^{\psi}\, d\psi}{\psi_1 - \psi_2} \right)^{-1} = \frac{\psi_1 - \psi_2}{e^{\psi_1} - e^{\psi_2}}$$

This allows us to write (5.3) as

$$\frac{u_1\mathcal{B}(\psi_1-\psi_2)-u_2\mathcal{B}(\psi_2-\psi_1)}{\ell_3} \tag{5.4}$$

Assuming that ψ is linear, we have

$$\psi_2-\psi_1=\boldsymbol{\beta}\cdot\boldsymbol{t}_3\,\ell_3$$

Setting $\mathcal{C}_3\equiv\mathcal{C}(\boldsymbol{\beta}\cdot\boldsymbol{t}_3\,\ell_3)$, our flux approximation becomes

$$\mathcal{C}_3\left(\frac{u_1-u_2}{\ell_3}\right)+\boldsymbol{\beta}\cdot\boldsymbol{t}_3\left(\frac{u_1+u_2}{2}\right) \tag{5.5}$$

Notice that the second term in (5.5) is identical to the corresponding term for the box method.

Using this approximation to the flux, the element stiffness matrix for the Scharfetter-Gummel discretization can be found in a fashion, completely analogous to (4.1)-(4.4), to be

$$\begin{aligned} A_{sg}=|\tau|&\begin{bmatrix}\nabla\phi_1^t\\ \nabla\phi_2^t\\ \nabla\phi_3^t\end{bmatrix}\left\{\sum_{i=1}^{3}L_i\mathcal{C}_i\hat{\rho}_i\right\}\begin{bmatrix}\nabla\phi_1 & \nabla\phi_2 & \nabla\phi_3\end{bmatrix}\\ &+\frac{|\tau|}{2}\begin{bmatrix}\nabla\phi_1^t\\ \nabla\phi_2^t\\ \nabla\phi_3^t\end{bmatrix}\left\{\beta_1\boldsymbol{t}_1(\boldsymbol{e}_2+\boldsymbol{e}_3)^t+\beta_2\boldsymbol{t}_2(\boldsymbol{e}_3+\boldsymbol{e}_1)^t+\beta_3\boldsymbol{t}_3(\boldsymbol{e}_1+\boldsymbol{e}_2)^t\right\}\end{aligned} \tag{5.6}$$

Similarly, the upwinding operator ρ_{sg} can be found, by forming $A_{sg}-A_g$, to be

$$h_\tau\rho_{sg}=\sum_{i=1}^{3}L_i(\mathcal{C}_i-1)\hat{\rho}_i+\beta_i\boldsymbol{t}_i\boldsymbol{d}_i^t \tag{5.7}$$

where the $\boldsymbol{d}_i$ are defined as in (4.7).

We point out here that the term $\mathcal{C}_i-1$ is formally of order $O(h_\tau^2|\boldsymbol{\beta}|^2)$ as $h_\tau\boldsymbol{\beta}\to 0$, whereas the corresponding term in the standard box method is $O(h_\tau|\boldsymbol{\beta}|)$. Thus, while in some regimes we can expect the two discretizations to behave quite similarly, there can be cases where there are significant differences.

6. Divergence-Free Upwinding. Our new discretization [3] is defined in terms of a single element τ and the corresponding element stiffness matrix A_d. Let the *current* $\boldsymbol{\mathcal{J}}$ be defined by

$$\boldsymbol{\mathcal{J}}=\nabla u+\boldsymbol{\beta}u \tag{6.1}$$

so that (1.2) becomes

$$\int_\Omega \boldsymbol{\mathcal{J}}\cdot\nabla\phi\,dx\,dy=0 \tag{6.2}$$

for all $\phi\in\mathcal{H}_0$.

Since $\boldsymbol{\beta}=\nabla\psi$, we may write (6.1) as

$$\boldsymbol{\mathcal{J}}=e^{-\psi}\nabla(e^{\psi}u) \tag{6.3}$$

For the case $\psi \in \mathcal{S}_h$, we can replace $e^{\pm\psi}$ by $e^{\pm\boldsymbol{\beta}\cdot\boldsymbol{v}}$, where $\boldsymbol{v}^t = (x\ y)$.

For our approximation, we seek a discrete current $\boldsymbol{\mathcal{J}}_h$ in the form

$$\begin{aligned}\boldsymbol{\mathcal{J}}_h &= e^{-\psi}\nabla(e^{\psi}\eta) \\ &= \nabla\eta + \boldsymbol{\beta}\eta\end{aligned} \tag{6.4}$$

where η is a linear polynomial in τ. Over all of Ω, η will be a discontinuous piecewise linear polynomial.

The consistency of our approximation is determined by the *edge conditions*

$$\int_{\boldsymbol{v}_i}^{\boldsymbol{v}_j} e^{\psi}\boldsymbol{\mathcal{J}}\cdot \boldsymbol{ds} \equiv \int_{\boldsymbol{v}_i}^{\boldsymbol{v}_j} e^{\psi}\boldsymbol{\mathcal{J}}_h\cdot \boldsymbol{ds} \tag{6.5}$$

where $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$ are two vertices of τ. Since the integrations can be carried out exactly, we may write (6.5) as

$$e^{\psi(\boldsymbol{v}_j)}u(\boldsymbol{v}_j) - e^{\psi(\boldsymbol{v}_i)}u(\boldsymbol{v}_i) = e^{\psi(\boldsymbol{v}_j)}\eta(\boldsymbol{v}_j) - e^{\psi(\boldsymbol{v}_i)}\eta(\boldsymbol{v}_i) \tag{6.6}$$

Although there are three edge conditions, only two represent independent constraints on η. In any event, the edge conditions imply that

$$\eta = u_h + \alpha\mathcal{I}(e^{-\psi}) \tag{6.7}$$

where u_h is the finite element solution, α is a scalar, and $\mathcal{I}(e^{-\psi})$ is the linear polynomial interpolating $e^{-\psi}$ at the vertices of τ. Note that since $u_h \in \mathcal{S}_h$, the discontinuities in η can arise only from α having different values in different elements.

The scalar α, and the stability of the discretization, is determined by the *divergence condition*

$$\nabla\cdot\boldsymbol{\mathcal{J}}_h = 0 \tag{6.8}$$

on τ, which implies, for $\psi \in \mathcal{S}_h$,

$$\alpha = -\frac{\boldsymbol{\beta}\cdot\nabla u_h}{\boldsymbol{\beta}\cdot\nabla\mathcal{I}(e^{-\psi})} \tag{6.9}$$

Setting $z = \mathcal{I}(e^{-\psi})$, we have

$$\begin{aligned}\boldsymbol{\mathcal{J}}_h\cdot\nabla\phi &= (\nabla\eta + \boldsymbol{\beta}\eta)\cdot\nabla\phi \\ &= (\nabla u_h + \boldsymbol{\beta}u_h)\cdot\nabla\phi - \frac{\boldsymbol{\beta}\cdot\nabla u_h}{\boldsymbol{\beta}\cdot\nabla z}(\nabla z + \boldsymbol{\beta}z)\nabla\phi \\ &= (\nabla u_h + \boldsymbol{\beta}u_h)\cdot\nabla\phi + \nabla u_h\cdot(\boldsymbol{\beta}\boldsymbol{d}^t)\nabla\phi\end{aligned} \tag{6.10}$$

where

$$\boldsymbol{d} = \frac{\nabla z + \boldsymbol{\beta}z}{-\boldsymbol{\beta}\cdot\nabla z} \tag{6.11}$$

The first term on the right hand side of the last line in (6.10) corresponds to the standard Galerkin method; thus the artificial diffusion for the divergence-free upwinding scheme is

$$h_\tau\rho_d = \boldsymbol{\beta}\boldsymbol{d}^t \tag{6.12}$$

which is a generally nonsymmetric, rank one diffusion matrix.

By noting that

$$\nabla e^{-\psi} + \boldsymbol{\beta} e^{-\psi} = 0$$

we can set

$$\begin{aligned} \varepsilon &= e^{-\psi} - \mathcal{I}(e^{-\psi}) \\ &= e^{-\psi} - z \end{aligned}$$

and write (6.11) as

$$\boldsymbol{d} = -\frac{\nabla \varepsilon + \boldsymbol{\beta}\varepsilon}{\boldsymbol{\beta}\cdot\boldsymbol{\beta} e^{-\psi} + \boldsymbol{\beta}\cdot\nabla\varepsilon} \tag{6.13}$$

Since ε is the interpolation error for linear interpolation of $e^{-\psi}$, we can see (formally) that $|\boldsymbol{d}| = O(h_\tau)$.

An interesting special case occurs whenever $\boldsymbol{\beta}$ is perpendicular to one of the edges of τ. Then $\boldsymbol{d}$ and $\boldsymbol{\beta}$ are parallel vectors, and the divergence-free upwinding scheme is similar to the streamline diffusion method, in terms of the added artificial diffusion. However, unlike the streamline diffusion method, there is no constant to be adjusted; in effect, the constant was chosen to satisfy the divergence condition.

For the case $\psi \in \mathcal{S}_h$, the element stiffness matrix for the divergence-free upwinding scheme is given by

$$A_d = A_g + |\tau| \begin{bmatrix} \nabla\phi_1 \cdot \boldsymbol{d} \\ \nabla\phi_2 \cdot \boldsymbol{d} \\ \nabla\phi_3 \cdot \boldsymbol{d} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}\cdot\nabla\phi_1 & \boldsymbol{\beta}\cdot\nabla\phi_2 & \boldsymbol{\beta}\cdot\nabla\phi_3 \end{bmatrix} \tag{6.14}$$

An important consideration for the divergence-free upwinding scheme is the question of whether it is always well defined. In particular, we must examine conditions under which $\boldsymbol{\beta}\cdot\nabla z = 0$, since this term is in the denominator of (6.11). We can begin by observing that

$$\begin{aligned} -\boldsymbol{\beta}\cdot\nabla z &= \boldsymbol{\beta}\cdot\boldsymbol{\beta} e^{-\psi} + \boldsymbol{\beta}\cdot\nabla\varepsilon \\ &= |\boldsymbol{\beta}|^2 e^{-\psi} + O(|\boldsymbol{\beta}|^2 h_\tau e^{-\psi}) \\ &> 0 \qquad \text{as } h_\tau \to 0 \end{aligned}$$

so that the method is certainly well defined for h sufficiently small. On the other hand, it is possible on a coarse mesh, with proper element geometry and a certain element orientation with respect to $\boldsymbol{\beta}$, that $-\boldsymbol{\beta}\cdot\nabla z \le 0$.

To see how this can occur, assume for the moment that our element τ has vertices $\boldsymbol{v}_1^t = (0\ 0)$, $\boldsymbol{v}_2^t = (1\ 0)$, $\boldsymbol{v}_3^t = (\bar{x}\ \bar{y})$, and that $\psi \in \mathcal{S}_h$. The Jacobian matrix J for this element is

$$\begin{aligned} J &= \begin{bmatrix} 1 & \bar{x} \\ 0 & \bar{y} \end{bmatrix} \\ J^{-1} &= \frac{1}{\bar{y}} \begin{bmatrix} \bar{y} & -\bar{x} \\ 0 & 1 \end{bmatrix} \\ J^{-1}J^{-t} &= \frac{1}{\bar{y}^2} \begin{bmatrix} \bar{x}^2 + \bar{y}^2 & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix} \end{aligned}$$

Let

$$J^t\boldsymbol{\beta} = \begin{bmatrix} q_2 \\ q_3 \end{bmatrix}$$

Then

$$z = \phi_1 + e^{-q_2}\phi_2 + e^{-q_3}\phi_3$$

and

$$\nabla z = (e^{-q_2} - 1)\nabla\phi_2 + (e^{-q_3} - 1)\nabla\phi_3$$

Without loss of generality, assume that $q_2 \geq q_3 \geq 0$ and $q_2 > 0$. Then let

$$r = \frac{q_3}{q_2}$$
$$s = \frac{e^{-q_3} - 1}{e^{-q_2} - 1}$$

Clearly

$$0 \leq r \leq s \leq 1$$

and

$$-\nabla z \cdot \boldsymbol{\beta} = -\frac{q_2(e^{-q_2} - 1)}{\bar{y}^2} \begin{bmatrix} 1 & r \end{bmatrix} \begin{bmatrix} \bar{x}^2 + \bar{y}^2 & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix} \begin{bmatrix} 1 \\ s \end{bmatrix}$$

The condition $-\nabla z \cdot \boldsymbol{\beta} = 0$ implies

$$\bar{x}^2 + \bar{y}^2 - \bar{x}(r+s) + rs = 0$$

or

$$\bar{y}^2 + \left(\bar{x} - \frac{r+s}{2}\right)^2 = \left(\frac{s-r}{2}\right)^2 \tag{6.15}$$

Equation (6.15) is the equation of a circle with center $((r+s)/2,\, 0)$ and radius $(s-r)/2$.

The properties of this upwinding scheme have a nice geometrical interpretation as illustrated in Fig. 3. The outer circle C_1 separates acute from obtuse triangles. All triangles with $(\bar{x}, \bar{y})$ lying outside this circle are acute, those with $(\bar{x}, \bar{y})$ inside are obtuse, and those with $(\bar{x}, \bar{y})$ lying on C_1 are right triangles.

The inner circle C_0, corresponding to (6.15), always lies inside the circle C_1, and separates triangles of positive and negative $-\boldsymbol{\beta}\cdot\nabla z$. Triangles with $(\bar{x}, \bar{y})$ lying outside this circle are have $-\boldsymbol{\beta} \cdot \nabla z > 0$. Clearly, $-\nabla z \cdot \boldsymbol{\beta} \leq 0$ requires τ to have an obtuse angle.

For triangles with $(\bar{x}, \bar{y})$ lying on this circle, $\boldsymbol{\beta} \cdot \nabla z = 0$, and the discretization is not defined. The chance of this condition being met in practice is very small. Indeed, we don't even check for this in our code, since roundoff error will almost certainly produce nonzero values of $\boldsymbol{\beta} \cdot \nabla z$.

On coarse meshes containing many badly shaped elements, it may be possible to have triangles with $(\bar{x}, \bar{y})$ lying inside this circle, in which case, $-\boldsymbol{\beta} \cdot \nabla z < 0$. When this occurs, it is analogous to *subtracting* a one dimensional artificial diffusion from

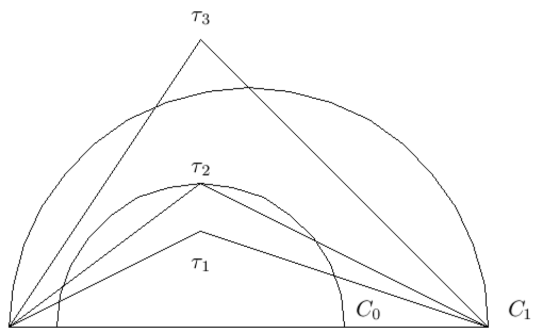

FIG. 6.1. *Geometrical interpretation of upwinding term*

the system, which seems rather counter intuitive (and dangerous). In particular, the eigenvalues of the of the 2×2 diffusion matrix $I + \boldsymbol{\beta} \boldsymbol{d}^t$ are

$$\lambda = 1$$

and

$$\lambda = \frac{\boldsymbol{\beta} \cdot \boldsymbol{\beta} z}{-\boldsymbol{\beta} \cdot \nabla z}$$

so that the overall diffusion term ceases to be elliptic whenever $-\boldsymbol{\beta} \cdot \nabla z < 0$.

If we scale τ to be an element with the same geometry but with diameter h, we note that q_2 and q_3 will scale to be of size $|\boldsymbol{\beta}|h$. Thus, as $h \to 0$, $r \to s$ and the (relative) radius of the circle tends to zero, which is consistent with our earlier remarks. Also, $r = s$ if $q_3 = 0$; this implies that $\boldsymbol{\beta}$ is perpendicular to one side of τ. In general, if $\boldsymbol{\beta}$ is perpendicular to *any* side of τ, then $-\nabla z \cdot \boldsymbol{\beta} \neq 0$, since then ∇z is in the direction $\boldsymbol{\beta}$ as in the streamline diffusion method.

Given the above comments, one might naturally approach this method with a great deal of skepticism with respect to its usefulness in general and its stability in particular (we certainly did). At present, we do not have any *a priori* error estimates for the method, except in the case when it reduces to the streamline diffusion method and existing estimates for that method apply. Nevertheless, the method is extremely stable, even under unfavorable geometric conditions. This stability comes from the divergence condition, as can be seen from the following line of reasoning. Let ϕ_i be the piecewise linear nodal basis function associated with vertex $\boldsymbol{v}_i$ in the triangulation. Then, using integration by parts, element by element, we have from (1.4)

$$\int_\Omega \boldsymbol{\mathcal{J}}_h \cdot \nabla \phi_i \, dx = \sum_{e_{ij}} \int_{e_{ij}} \{\boldsymbol{\mathcal{J}}_h \cdot \boldsymbol{n}_{ij}\} \phi_i \, ds = 0$$

where e_{ij} is the triangle edge connecting vertices $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$ and $\{\boldsymbol{\mathcal{J}}_h \cdot \boldsymbol{n}_{ij}\}$ is the jump in the normal component of $\boldsymbol{\mathcal{J}}_h$ across e_{ij}. By simple geometry, it seems clear that in order to have a massive overshoot or undershoot (a "spike") at $\boldsymbol{v}_i$, the sum of the normal components of these jumps must be correspondingly large in magnitude, a circumstance which is prohibited by the divergence condition.

In effect, the divergence condition prevents the creation of any numerical sources or sinks within element interiors. The edge conditions guarantee good approximation along element edges, in particular at the vertices. The situation is entirely analogous to the finite element approximation of the Laplacian using piecewise linear elements; there $\Delta u = 0$ within each element and it is the jumps in the normal components of ∇u across the triangle edges that support the approximation. Thus we can expect, at least with hindsight, that this method will provide a stable and accurate approximation to (1.2).

We end this section by noting that this method and its derivation remain well defined for three dimensional meshes based on tetrahedral elements. Indeed, it was our desire to have an upwinding procedure for tetrahedral meshes that remains stable even in the presence of unfavorable element geometries, which motivated our current work.

7. Summary. A uniform framework is developed for the study of general upwinding schemes. The standard finite element weak Galerkin discretization is chosen as the reference. Differences between other discretization schemes and the weak Galerkin form are written as artificial diffusion terms. These artificial diffusion terms are spanned by a four dimensional space of element diffusion matrices. Three basis matrices are symmetric, rank one diffusion operators which can naturally be associated with the edges of the triangle. The fourth basis matrix is skew symmetric and is associated with a rotation by $\pi/2$.

The streamline diffusion method is one of the more widely used upwinding schemes in the finite element arena. Both the streamline and the crosswind upwinding terms are symmetric, positive semidefinite matrices of rank one and involve only the edge diffusions operators.

Two finite volume discretizations, the box method and the Scharfetter-Gummel method, are then analyzed. Finite volume methods involve only approximations along each triangle edge, while the standard Galerkin method uses approximations within the triangle as a whole. Discretizations of convection diffusion problems give rise to two types of contributions to the element stiffness matrices. The first type corresponds to the upwinding terms, which contribute symmetric, positive semidefinite artificial edge diffusion terms. The second type arises from the centered difference approximation of the convective term. When viewed as a finite element method, these terms contribute nonsymmetric artificial diffusion upwinding terms. While finite volume discretizations may always be written as upwinded Galerkin methods, the converse does not appear to be true, since the skew symmetric elementary diffusion operator seems to have no analogue in the standard finite volume framework.

Finally, the divergence-free upwinding scheme is analyzed. In general, the artificial diffusion introduced by this method leads to both symmetric and nonsymmetric diffusion terms. However, whenever the velocity is perpendicular to one of the triangle edges, the streamline diffusion method is recovered. In some extreme cases, the overall diffusion matrix has both positive and negative eigenvalues. Nevertheless, the method appears to be extremely robust and stable, and remains so even in unfavorable situations where other upwinding schemes fail.

REFERENCES

[1] R. E. Bank and D. J. Rose, *Some error estimates for the box method*, SIAM J. Numer. Anal., 24 (1987), pp. 777–787.

[2] R. E. Bank, D. J. Rose, and W. Fichtner, *Numerical methods for semiconductor device simulation*, SIAM J. Sci. Stat, 4 (1983), pp. 416–435.

[3] J. F. Bürgler, R. E. Bank, W. Fichtner, and R. K. Smith, *A new discretization for the semiconductor continunity equations*, IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems, 8 (1989), pp. 479–489.

[4] P. Caussignac and R. Touzani, *Linear conforming and nonconforming upwind finite elements for the convection-diffusion equation*, IMA J. Numer. Anal., 8 (1988), pp. 85–103.

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

[6] ———, *A theoretical framework for Petrov-Galerkin methods with discontinuous weighting functions: Application to the streamline upwind procedure*, in Finite Elements in Fluids, Vol. 4, Wiley, New York, 1984.

[7] T. Ikeda, *Maximum principle in finite element models for convection-diffusion phenomena*, in Lecture Notes in Numerical and Applied Analysis, Vol 4., North-Holland, Amsterdam, 1983.

[8] C. Johnson, U. Navert, and J. Pitkaranta, *Finite element methods for linear hyperbolic problems*, Comp. Meth. in Appl. Mech. Engr., 45 (1984), pp. 285–312.

[9] K. Ohmori and T. Ushijima, *A technique of upstream type applied to a linear conforming finite element approximation of convection-diffusion equations*, RAIRO Anal. Numer., 18 (1984), pp. 309–332.

[10] D. Scharfetter and H. K. Gummel, *Large-signal analysis of a silicon Read diode oscillator*, IEEE Trans. Electron Devices, ED-16 (1969), pp. 64–77.

[11] G. Strang and G. Fix, *An Analysis of the Finite Element Method*, Prentice-Hall, Englewood Cliffs, New Jersey, 1973.