THE FINITE VOLUME SCHARFETTER-GUMMEL METHOD FOR STEADY CONVECTION DIFFUSION EQUATIONS

RANDOLPH E. BANK*, W. M. COUGHRAN, JR.†, AND LAWRENCE C. COWSAR†

Abstract. *A priori* error estimates are given for the Finite Volume Scharfetter-Gummel (FVSG) discretization of the steady convection diffusion equation by showing that the FVSG method gives the same discretization as the Edge Averaged Finite Element method of Markowich and Zlámal [9] and Xu and Zikatanov [14]. The analysis also suggests a class of modifications for triangulations containing obtuse angles. Numerical results comparing the FVSG method and a modified FVSG method to other discretizations are included.

Key words. Box Methods, Finite Element Methods, Finite Volume Methods, Upwinding, Scharfetter-Gummel Methods, Convection Diffusion Equations.

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

1. Introduction. The classical Scharfetter-Gummel scheme for discretizing drift-diffusion and energy-transport models has proven itself to be the workhorse for semiconductor device modeling codes. The discretization is well defined in one spatial dimension [12], and various extensions to higher dimensions have been proposed; for example, see [11] and the references contained therein as well as [9, 5, 4, 10]. One successful extension, as partially demonstrated by the numerics in Section 7, which is used in several commercial simulators is the Finite Volume Scharfetter-Gummel (FVSG) method described by Bank, Fichtner, and Rose [3].

Here we consider the FVSG method applied to the following model convection-diffusion problem for u on a polygonal domain $\Omega \subset \Re^2$ with boundary $\partial\Omega = \overline{\Gamma}_1 \cup \overline{\Gamma}_2$, $\Gamma_1 \cap \Gamma_2 = \emptyset$ and outward normal $\mathbf{n}$:

$$-\nabla\cdot(a\nabla u + \beta u) = 0 \quad \text{in } \Omega, \tag{1.1}$$

$$u = u_0 \quad \text{on } \Gamma_1, \tag{1.2}$$

$$(a\nabla u + \beta u)\cdot \mathbf{n} = 0 \quad \text{on } \Gamma_2. \tag{1.3}$$

In the finite volume method, it is convenient to write the second order equation as a first order system. By introducing a flux $\mathbf{J}$, we rewrite (1.1) as a continuity equation

$$\nabla\cdot\mathbf{J} = 0 \quad \text{in } \Omega, \tag{1.4}$$

and a constitutive relationship

$$\mathbf{J} = -(a\nabla u + \beta u) \quad \text{in } \Omega. \tag{1.5}$$

We assume that a is a real valued function satisfying uniformly for $x \in \Omega$

$$0 < a_{\min} \le a(x) \le a_{\max}.$$

Since our model problem is driven by the Dirichlet boundary condition, we also assume that the measure of Γ_1 is nonzero.

The remainder of this paper is organized as follows. In the next section, we introduce some notation. In Section 3, we recall the FVSG method of [3]. Our main theoretical tool is developed in Section 4 in which we exhibit an equivalence between the FVSG discretization and a certain piecewise linear Galerkin discretization. In particular, it is shown that the FVSG is the same discretization as the edge based schemes of Markowich and Zlámal [9] and the more recent work of Xu and Zikatanov [14]. Using this equivalence, *a priori* error estimates for the FVSG scheme are inherited from [14]. These estimates are noteworthy since they depend only on the smoothness of the flux. Writing the FVSG method as a bilinear form suggests

*Department of Mathematics, University of California at San Diego, La Jolla, CA 92093, USA. The work of this author was supported by the U. S. National Science Foundation under grant DMS-9706090.

†Computing Sciences Research Center, Bell Labs, Lucent Technologies, Murray Hill, New Jersey 07974, USA.

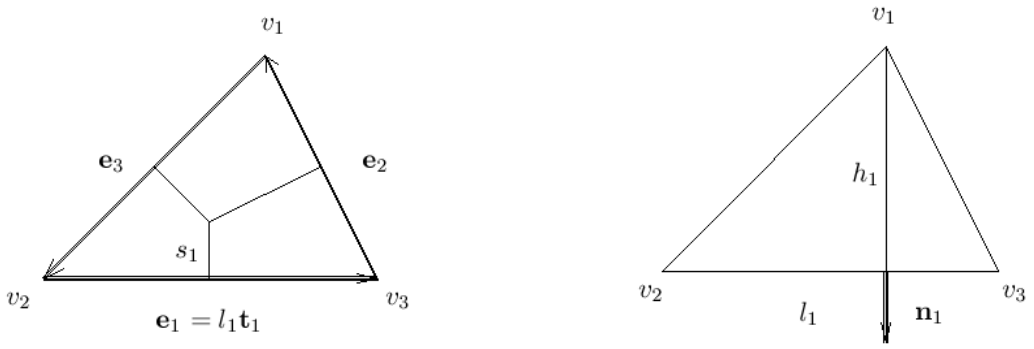


FIG. 2.1. *Parameters associated with triangle* τ

modifications of the scheme that are appropriate when the triangulations include large obtuse angles. Several authors have commented on the importance of local divergence conditions (e.g., [2]). In Section 5, we note that such properties are relatively easy to impose on a general class of discretizations; hence, they do not uniquely motivate and characterize "good" discretizations. Building on the analysis of Section 4, a particular modified FVSG scheme is given in Section 6. We conclude with some numerical experiments comparing the modified and unmodified FVSG schemes to several other advection schemes. The performance of the FVSG schemes in these experiments motivated the rest of the study.

2. Preliminaries. We introduce some local notation for triangles. As pictured in Fig. 2.1, we label the vertices v_i, $i = 1, 2, 3$ in a counterclockwise order and understand the indexing to be cyclical, e.g. $v_4 = v_1$. Let the edge opposite v_i be denoted $\mathbf{e}_i$ and oriented such that it connects v_{i+1} to v_{i-1}. Let l_i denote its length, and $\mathbf{t}_i$ denote the unit tangent vector oriented in the same direction. Let $\theta_i = \angle v_{i-1}v_iv_{i+1}$, the angle opposite $\mathbf{e}_i$. Denote the unit outward normal perpendicular to edge $\mathbf{e}_i$ by $\mathbf{n}_i$, and let h_i denote the perpendicular distance from v_i to edge $\mathbf{e}_i$. Denote the segment from the midpoint of $\mathbf{e}_i$ to the intersection of the perpendicular edge bisectors by $\mathbf{s}_i$. And let s_i denote its signed length where s_i is negative if the angle opposite edge $\mathbf{e}_i$ is obtuse. Let $\mathcal{V}(\tau)$ denote the set of vertices of a triangle τ; and more generally, for a set of triangles $\mathcal{T}$, let $\mathcal{V}(\mathcal{T}) = \bigcup_{\tau\in\mathcal{T}} \mathcal{V}(\tau)$. Finally, let ϕ_i denote the linear function on τ that is one at v_i and zero at the other vertices. Each of these quantities should be subscripted by the triangle to which it belongs, but to simplify notation we will drop this extra subscript.

Recall that the following relationships hold on an arbitrary triangle τ (e.g., [2]):

$$|\tau| = \frac{1}{2} h_i l_i, \tag{2.1}$$

$$\sum_{i=1}^{3} l_i \mathbf{t}_i = 0, \tag{2.2}$$

$$\nabla\phi_i = -\frac{\mathbf{n}_i}{h_i}, \tag{2.3}$$

$$l_i\mathbf{t}_i \cdot \nabla\phi_i = 0, \quad l_{i\pm1}\mathbf{t}_{i\pm1} \cdot \nabla\phi_i = \pm 1, \tag{2.4}$$

$$s_i = -|\tau| l_i \nabla\phi_{i+1} \cdot \nabla\phi_{i-1}. \tag{2.5}$$

We will also need a difference operator along $\mathbf{e}_i$ defined by

$$\delta_i(\psi) = \psi(v_{i-1}) - \psi(v_{i+1}).$$

For a function ψ defined on τ, we have the trivial relationship

$$\delta_i(\psi) = l_i \nabla(\mathcal{I}\psi) \cdot \mathbf{t}_i, \tag{2.6}$$

where $\mathcal{I}\psi$ is the linear interpolant of ψ agreeing at the vertices.

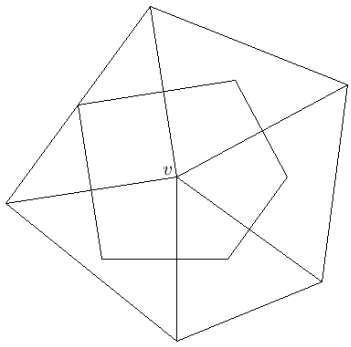


FIG. 3.1. *The control volume b_v*

3. The Finite Volume Scharfetter-Gummel Method. Let $\{\mathcal{T}_h\}_h$ denote a quasi-uniform family of triangulations of the domain Ω parameterized by h, the maximum mesh spacing (e.g., [6]). For each vertex $v \in \mathcal{V}(\mathcal{T}_h)$, let b_v denote the polygonal volume formed by the perpendicular bisectors of the triangle edges that contain the vertex v as depicted in Fig. 3.1. For $v \in \partial\Omega$, the control volumes are suitably modified. Specifically, for $v \in \overline{\Gamma}_1$, b_v is empty; for $v \in \Gamma_2$, ∂b_v contains a portion of Γ_2 as depicted in Fig. 3.2.

Let P^k denote the space of polynomials of degree at most k, and define

$$U_h = \{\phi \in C^0(\overline{\Omega}) \,|\, \phi|_\tau \in P^1 \ \forall \tau \in \mathcal{T}_h\}$$

and

$$U_{h,g} = \{\phi \in U_h \,|\, \phi = g \text{ on the nodes on } \Gamma_1\}.$$

Let W_h denote the space of piecewise constant functions on $\cup_{v\in\mathcal{V}(\mathcal{T}_h)}\partial b_v$ that are constant on each segment $\mathbf{s}_i$ of b_v. Let $\hat{\mathbf{t}}$ be the piecewise constant vector field on $\cup_{v\in\mathcal{V}(\mathcal{T}_h)}\partial b_v$ such that $\hat{\mathbf{t}}$ restricted to segment $\mathbf{s}_i$ of triangle τ is $\mathbf{t}_i$.

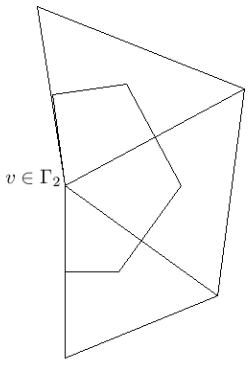


FIG. 3.2. *The modified control volume b_v for $v \in \Gamma_2$*

The FVSG [3] approximation of (1.1)–(1.3) is a pair $(j_h, u_h) \in W_h \times U_{h,g}$ such that

$$\int_{\partial b_v} j_h \hat{\mathbf{t}} \cdot \mathbf{n}_{b_v}\, ds = 0 \quad \forall v \in \mathcal{V}(\mathcal{T}_h), \tag{3.1}$$

and for each triangle $\tau \in \mathcal{T}_h$

$$j_h|_{\mathbf{s}_i} = -\frac{1}{l_i}\hat{a}_i\delta_i(e^{\psi_{\mathbf{e}_i}}u_h) \quad i = 1,2,3, \tag{3.2}$$

where $\mathbf{n}_{b_v}$ is the unit outward normal to control volume b_v,

$$\hat{a}_i = \left(\frac{1}{l_i} \int_{\mathbf{e}_i} a^{-1} e^{\psi_{\mathbf{e}_i}} \cdot \mathbf{t}_i \, ds \right)^{-1},$$

and $\psi_{\mathbf{e}_i}$ is the function defined on $\mathbf{e}_i$ such that its tangential derivative satisfies

$$\frac{d\psi_{\mathbf{e}_i}}{d\mathbf{t}_i} = a^{-1}\beta \cdot \mathbf{t}_i. \tag{3.3}$$

The function $\psi_{\mathbf{e}_i}$ is defined up to a constant which is of no consequence to the definition of j_h since any constant added to $\psi_{\mathbf{e}_i}$ is canceled by the corresponding factor in $\hat{a}_i$.

When we have at our disposal a function ψ_τ defined on the triangle consistently satisfying (3.3), then (3.2) may also be written as

$$j_h|_{\mathbf{s}_i} = -\hat{a}_i \nabla \mathcal{I}(e^{\psi_\tau} u_h)|_{\mathbf{e}_i} \cdot \mathbf{t}_i, \tag{3.4}$$

where $\mathcal{I}$ is the nodal interpolation operator into U_h. This is the case when

$$\int_{\partial\tau} a^{-1}\beta \cdot \mathbf{t} \, ds = 0;$$

for instance, when β is globally given as $\nabla\psi$ as in the drift diffusion equations, or when β and a are represented as piecewise constant functions as in some finite element codes.

Equations (3.1) and (3.2) are discretizations of the continuity equation and constitutive relationship, respectively. The finite volume discretization of the (3.1) is formally derived by first integrating (1.4) over a volume b_v and applying the divergence theorem; namely,

$$0 = \int_{b_v} \nabla\cdot\mathbf{J} \, dx = \int_{\partial b_v} \mathbf{J}\cdot\mathbf{n}_{b_v} \, ds. \tag{3.5}$$

Comparing the last integrand of (3.5) to (3.1), we see that j_h is an approximation to $\mathbf{J}\cdot\hat{\mathbf{t}}$ on the boundary of the control volumes. In practice the ancillary variable j_h may be eliminated in (3.1) using (3.2). We include it in the definition of the FVSG scheme since it is useful in Section 4.

Equation (3.2) is the discretization of the constitutive relationship using the idea of Scharfetter and Gummel [12]. Specifically, on edge $\mathbf{e}_i$, we have

$$a^{-1}\mathbf{J}\cdot\mathbf{t}_i = -(\nabla u + a^{-1}\beta u)\cdot\mathbf{t}_i = -e^{-\psi_{\mathbf{e}_i}} \frac{d(e^{\psi_{\mathbf{e}_i}} u)}{d\mathbf{t}_i}.$$

One arrives at (3.2) by assuming $\mathbf{J}\cdot\mathbf{t_i}$ is constant and integrating

$$-a^{-1} e^{\psi_{\mathbf{e}_i}} \mathbf{J}\cdot\mathbf{t}_i = \frac{d(e^{\psi_{\mathbf{e}_i}} u)}{d\mathbf{t}_i}$$

over $\mathbf{e}_i$.

4. Finite Element Formulation. To facilitate making the connection between (3.1) and the standard Galerkin discretization of (1.4), we introduce for $\tau \in \mathcal{T}_h$ (using notation local to τ) a linear map $\mathcal{J}_\tau : \Re^3 \to \Re^2$ defined by

$$\mathcal{J}_\tau(\{\gamma_i\}_{i=1}^3) = \frac{1}{|\tau|} \sum_{i=1}^3 \gamma_i l_i s_i \mathbf{t}_i. \tag{4.1}$$

LEMMA 4.1. *$\mathcal{J}_\tau$ has the following properties:*

$$\mathcal{J}_\tau(\{\mathbf{J}\cdot\mathbf{t}_i\}_{i=1}^3) = \mathbf{J} \quad \forall \mathbf{J} \in \Re^2, \tag{4.2}$$

$$\mathcal{J}_\tau(\{s_i^{-1}\}_{i=1}^3) = 0, \tag{4.3}$$

$$\int_\tau \mathcal{J}_\tau(\{\gamma_i\}_{i=1}^3) \cdot \nabla\phi_i \, dx = \gamma_{i+1} s_{i+1} - \gamma_{i-1} s_{i-1}. \tag{4.4}$$

Proof. Equation (4.3) follows directly from (2.2). Equation (4.4) is verified by a direct calculation using the definition of $\mathcal{J}_\tau$ and (2.4).

Let

$$A = \frac{1}{|\tau|}\sum_{i=1}^{3} l_i s_i \mathbf{t}_i \mathbf{t}_i^t.$$

Equation (4.2) is equivalent to A being the 2×2 identity matrix which we denote by $I_{2\times 2}$. Since the matrix A is symmetric and the normals of a triangle are pair-wise independent, it is enough to check that for $i = 1, 2, 3$

$$\nabla\phi_{i\pm1} \cdot A\nabla\phi_i = \nabla\phi_{i\pm1} \cdot \nabla\phi_i.$$

Using the definition of A, (2.4) and (2.5), we see that

$$\begin{aligned}\nabla\phi_{i\pm1} \cdot A\nabla\phi_i &= \frac{1}{|\tau|}\sum_{j=1}^{3} l_j s_j \nabla\phi_{i\pm1} \cdot \mathbf{t}_j \nabla\phi_i \cdot \mathbf{t}_j \\ &= \frac{1}{|\tau|} l_{i\mp1} s_{i\mp1} (\nabla\phi_{i\pm1} \cdot \mathbf{t}_{i\mp1})(\nabla\phi_i \cdot \mathbf{t}_{i\mp1}) = -\frac{1}{|\tau|}\frac{s_{i\mp1}}{l_{i\mp1}} = \nabla\phi_{i\pm1} \cdot \nabla\phi_i.\end{aligned}$$

Hence,

$$\frac{1}{|\tau|}\sum_{i=1}^{3} l_i s_i \mathbf{t}_i \mathbf{t}_i^t = I_{2\times 2}. \tag{4.5}$$

□

In light of (4.2), one can think of $\mathcal{J}_\tau$ as a operator that recovers a vector defined on τ given its tangential components on each of the edges. Equation (4.3) characterizes the null space of this operation, a space we note that is also annihilated by the left hand side of (3.1).

Recall that the standard finite element method applied to the continuity equation with piecewise linear test functions seeks a flux $\mathbf{J}_h$ satisfying

$$-\int_\Omega \mathbf{J}_h \cdot \nabla\phi_v \, dx = 0 \quad \forall v \in \mathcal{V}(\mathcal{T}_h), \tag{4.6}$$

where ϕ_v is the piecewise linear function that is one at v and vanishes at all the other vertices of $\mathcal{T}_h$. The following lemma exhibits an equivalence between the finite volume and Galerkin discretizations of the continuity equation. This will allow us to analyze the finite volume discretization in the more standard finite element framework.

LEMMA 4.2. *If j_h satisfies (3.1), then $\mathbf{J}_h$ satisfies (4.6) with*

$$\mathbf{J}_h|_\tau = \mathcal{J}_\tau(\{j_h|_{\mathbf{s}_i}\}_{i=1}^3). \tag{4.7}$$

Conversely, if $\mathbf{J}_h$ satisfies (4.6), then j_h defined on each segment $\mathbf{s}_i$ of the perpendicular edge bisectors of τ by

$$j_h|_{\mathbf{s}_i} = \left(\frac{1}{|\tau|}\int_\tau \mathbf{J}_h \, dx\right) \cdot \mathbf{t}_i$$

satisfies (3.1).

Proof. It is enough to show that the equivalence holds element by element. Since the support of ϕ_v consists of exactly the same set of triangles with $b_v \cap \tau \neq \emptyset$, we consider only those $\tau \in \mathcal{T}_h$ such that $b_v \cap \tau \neq \emptyset$. Using notation local to τ with $v_1 = v$ and using (4.1) and (4.4), we have

$$-\int_\tau \mathbf{J}_h \cdot \nabla\phi_{v_1} \, dx = (j_h)|_{\mathbf{s}_3} s_3 - (j_h)|_{\mathbf{s}_2} s_2 = \int_{\partial b_{v_1} \cap \tau} j_h \hat{\mathbf{t}} \cdot \mathbf{n}_{b_{v_1}} \, ds. \tag{4.8}$$

The proof of the converse is essentially identical and uses (4.2) and the fact that $\nabla\phi_v$ is constant on triangles. □

Using Lemma 4.2 and (2.6), we can write the FVSG scheme as a Galerkin scheme with piecewise linear test and trial functions.

THEOREM 4.3. *The function $(j_h, u_h) \in W_h \times U_{h,u_0}$ is a solution to the FVSG discretization (3.1)–(3.2) if and only if $u_h \in U_{h,u_0}$ satisfies*

$$\sum_{\tau\in\mathcal{T}_h}\sum_{i=1}^{3}\hat{a}_i\frac{s_i}{l_i}\delta_i(e^{\psi_{\mathbf{e}_i}}u_h)\delta_i(v_h) = 0. \tag{4.9}$$

When $\int_{\partial\tau}(a|_\tau)^{-1}\beta|_\tau\cdot\mathbf{t} = 0 \quad \forall\tau$, the equation for u_h may also be written as

$$\sum_{\tau\in\mathcal{T}_h}\int_\tau D_\tau\nabla\mathcal{I}(e^{\psi_\tau}u_h)\cdot\nabla\phi\,dx = 0 \quad \forall\phi\in U_{h,0}, \tag{4.10}$$

with

$$D|_\tau = \frac{1}{|\tau|}\sum_{i=1}^{3}\hat{a}_i l_i s_i \mathbf{t}_i\mathbf{t}_i^t. \tag{4.11}$$

Writing the finite volume method in the form of (4.9) and (4.10) makes it clear that the FVSG method is related to several finite element methods. In particular the "Inverse-average-type finite element discretization" of Markowich and Zlámal [9, Sec. 7] is the FVSG method when $a \equiv 1$ and $\beta = \nabla\psi$. For nontrivial a or for β not given as the gradient of ψ, the FVSG method is the "Edge Average Finite Element" (EAFE) method of Xu and Zikatanov [14]. The FVSG is also closely related to the "Exponential Fitting" method of Brezzi, Marini and Pietra [5, Sec. 2]. Their scheme, defined for $a \equiv 1$ and $\beta = \nabla\psi$, is precisely (4.10) with equal weights on each edge, namely

$$\hat{a}_1 = \hat{a}_2 = \hat{a}_3 = \left(\frac{1}{|\tau|}\int_\tau e^\psi\,dx\right).$$

As discussed in [4] and demonstrated in Section 7, the equal weighting on each edge causes some difficulty. A modification is discussed in [4, Sec. 4] which makes the method essentially that of [9].

Since the FVSG is identical to the EAFE method of Xu and Zikatanov, we have the following results from [14]. Let u and $\mathbf{J}$ be the solution to (1.2) – (1.5), and let $\mathcal{I}u$ be the linear interpolant of u Let $|\cdot|_{1,p,\Omega}$ denote the standard semi-norm in $W^{1,p}(\Omega)$ or $(W^{1,p}(\Omega))^2$ as appropriate.

THEOREM 4.4 ([14]). *If a and β are continuous, then the FVSG discretization gives rise to an M-matrix if and only if the triangulation is a Delaunay triangulation. Moreover, if a and β are only piecewise smooth with discontinuities aligning with the triangulations and the angles opposite the edges over which discontinuities occur are non-obtuse, the FVSG discretization gives rise to an M-matrix.*

THEOREM 4.5 ([14]). *If $\mathcal{T}_h$ is a Delaunay triangulation and a and β are continuous, or if a and β are piecewise smooth and h is sufficiently small, then the solution u_h of (4.9) exists and for $p > 2$ there exists a constant C depending on the shape regularity of the mesh, a, β and p such that following estimate holds*

$$|\mathcal{I}u - u_h|_{1,2,\Omega} < C(a,\beta,p)h|\mathbf{J}|_{1,p,\Omega}. \tag{4.12}$$

5. Local Divergence and Curl Conditions on the Flux. A number of methods based on piecewise linear approximations impose local conditions on the divergence or curl of the flux $\mathbf{J}$. A natural modification is to insist that for each triangle $\tau \in \mathcal{T}_h$ the local *divergence free* condition hold

$$\nabla\cdot(\mathbf{J}|_\tau) = 0. \tag{5.1}$$

The *divergence free upwinding scheme* [2] enforces this property.

Consider one such $\tau \in \mathcal{T}_h$; set the origin of the coordinate system to the center of mass of τ. Let U_h denote the space of piecewise linear functions on $\mathcal{T}_h$. For $\mathbf{J} = \nabla u + \beta u + D\nabla u$, $u \in U_h$, and $\beta = \nabla\psi$ with $\psi \in U_h$,

$$\nabla\cdot(\mathbf{J}|_\tau) = \beta \cdot \nabla u|_\tau + \nabla\cdot(D\nabla u|_\tau).$$

If we choose the entries of $D = \{d_{ij}\}$ such that

$$\frac{\partial d_{11}}{\partial x} + \frac{\partial d_{21}}{\partial y} = \beta_1,$$
$$\frac{\partial d_{12}}{\partial x} + \frac{\partial d_{22}}{\partial y} = \beta_2,$$

$\mathbf{J}$ automatically satisfies equation (5.1). A particular choice is

$$D = \begin{pmatrix} \psi_0 - \psi & 0 \\ 0 & \psi_0 - \psi \end{pmatrix},$$

where $\psi_0 = \psi(0,0)$. Since

$$\int_\tau x\, dx = \int_\tau y\, dx = 0,$$

the contribution of D integrated over a cell is zero. Hence, there is no contribution at all to the stiffness matrix for u!

The above argument implies that we can make a perturbation of $\mathbf{J}$ in the space of discontinuous piecewise linear functions in such a way that equation (5.1) is satisfied. Moreover, we can define a new scheme with a modified upwinding matrix D whose solution is the element-wise divergence-free $\mathbf{J}$ and the original u. A similar argument can be made regarding $\nabla \times (\mathbf{J}|_\tau)$.

Hence, local conditions like equation (5.1) may be important. However, they are binding only in so much as the finite element space associated with $\mathbf{J}$ is restricted. To say this another way, suppose we only constrain the finite element space associated with $\mathbf{J}$ to be piecewise linear functions on $\mathcal{T}_h$; in that case, the $\mathbf{J}$ can be modified to satisfy divergence conditions without changing the scalar part of the solution u.

6. Modified FVSG. Theorem 4.4 provides sufficient conditions for the discretization to yield an M-matrix which leads immediately to some stability properties. When $\mathcal{T}_h$ is not a Delaunay triangulation, we now suggest some small modifications to equation (4.11) to preserve stability. Let τ be a triangle containing an obtuse angle opposite side e_i. The modified FVSG replaces $\hat{a}_i$ in (4.9) with $\tilde{a}_i < \hat{a}_i$ in order that the element matrix remain positive semidefinite. One obvious choice would be to set $\tilde{a}_i = \min_{j=1,2,3} \hat{a}_j$. Our choice is less severe; we set $\tilde{a}_i = \gamma\hat{a}_i$, where $0 < \gamma \leq 1$ is chosen to be the largest value such that both the element matrix and the element matrix obtained by replacing β with $-\beta$ (or equivalently, rotating the element by π) are positive semidefinite. A straightforward but tedious calculation shows

$$\tilde{a}_i = \min\{1, \gamma_+, \gamma_-\}\hat{a}_i$$

where

$$L_k = s_k/l_k$$
$$\gamma_- = \frac{-L_{i+1}L_{i-1}}{L_iL_{i+1}e^{\hat{\psi}_{i+1}-\hat{\psi}_i} + L_iL_{i-1}e^{\hat{\psi}_{i-1}-\hat{\psi}_i}}$$
$$\gamma_+ = \frac{-L_{i+1}L_{i-1}}{L_iL_{i+1}e^{\hat{\psi}_i-\hat{\psi}_{i+1}} + L_iL_{i-1}e^{\hat{\psi}_i-\hat{\psi}_{i-1}}}$$

(note $L_i < 0$, while $L_{i\pm1} > 0$).

The modified scheme shares the same *a priori* error estimate given in Theorem 4.5. In fact, for h (or more properly $ha^{-1}|\beta|$) sufficiently small, it is easy to see that $\min\{1, \gamma_+, \gamma_-\} = 1$, and the modified scheme reverts to the standard FVSG discretization. To see this, note that

$$\begin{aligned}\frac{-L_{i+1}L_{i-1}}{L_iL_{i+1} + L_iL_{i-1}} &= \frac{\cos\theta_{i+1}\cos\theta_{i-1}}{-\cos\theta_i} \\ &= \frac{\cos\theta_{i+1}\cos\theta_{i-1}}{\cos\theta_{i+1}\cos\theta_{i-1} - \sin\theta_{i+1}\sin\theta_{i-1}} \\ &> 1\end{aligned}$$

where we have used $\theta_i + \theta_{i+1} + \theta_{i-1} = \pi$ and $\theta_i > \pi/2$.

7. Numerical Experiments and Implementation. In this section we compare the performance of the FVSG scheme and the modified FVSG scheme to several other methods for solving convection diffusion problems. The test problem is the "JCN" test problem in PLTMG [1, Chap. 7]. It is an idealization of an electron continuity equation in a semiconductor device model.

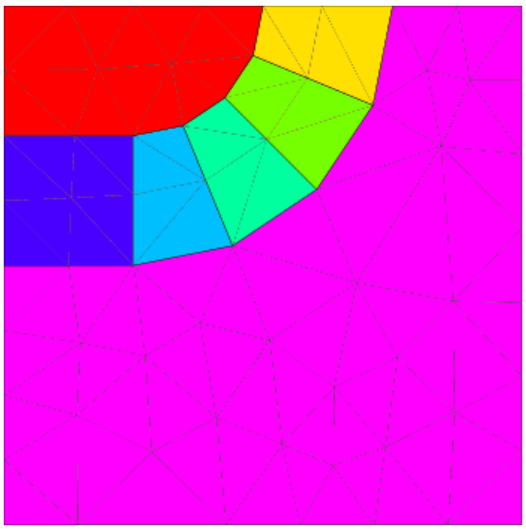


FIG. 7.1. *JCN Test Problem Geometry*

Since a version of this problem has been readily available for some time in the software package PLTMG [1], we will describe the problem in the form that is implemented in [1] instead of making a natural change of variables to transform the problem to a unit square. The test problem geometry is depicted in Fig. 7.1. The domain is $\Omega = (0, 0.03) \times (0, 0.03)$. In the polygonal approximation to the annular region, β is directed approximately radially (perpendicular to the inner and outer faces of the annular region). The magnitude of the advection in the annular region is $40 + 15 * \log(10)$; outside the annular region $\beta = 0$. We refer to the case that β is primarily flowing from upper left to lower right as the "forward biased" problem and we refer to the case that β is in the opposite direction as the "reversed biased" problem.

The discretizations considered are:

- The FVSG method defined by (4.10), (equivalently, (3.1)-(3.2));
- The Modified FVSG method of Section 6;
- A one-point upwinded finite volume method (3.1) with

$$j_h|_{\mathbf{s}_i} = (\nabla u_h + \beta u_h^+) \cdot \mathbf{t}_i, \tag{7.1}$$

where

$$u_h^+ = \begin{cases} u_h(v_{i+1}) & \text{if } \beta \cdot \mathbf{t}_i > 0 \\ u_h(v_{i-1}) & \text{if } \beta \cdot \mathbf{t}_i \le 0 \end{cases};$$

- A streamline diffusion finite element method (e.g., [8]) for $u_h \in U_{h,u_0}$ satisfying

$$\int_\Omega ((I + C\frac{\beta\beta^t}{|\beta|^2})\nabla u_h + \beta u_h) \cdot \phi \, dx = 0 \quad \forall \phi \in U_{h,0}, \tag{7.2}$$

with

$$C|_\tau = \frac{|\tau||\beta|}{(1 + |\tau||\beta|)^{1/2}};$$

- The hybrid finite element scheme of [5, Eqn. (2.12)] which is (4.10) with

$$D|_\tau = \left(\frac{1}{|\tau|}\int_\tau e^\psi \, dx\right) I,$$

without the modifications in [4].

The methods have been implemented in PLTMG by editing the routines that calculate the element stiffness matrix. For the FVSG and other schemes involving exponentials, extreme care was taken to order that calculations in a way that maintains high precision. The calculations were carried out in double precision and the sparse direct solver in PLTMG was used with iterative improvement to essentially remove questions related to convergence of the linear algebra.

Three finite element meshes are used in the experiments and are depicted in Fig. 7.2. The meshes with many obtuse angles are used primarily to test the robustness of the methods, and we do not seriously advocate their use in practice. However, in practice the discretization scheme may not be tightly coupled to the mesh generation and refinement strategy, and the assumption that the mesh is always Delaunay may not be satisfied. Thus, robustness is desirable. The "comparison solutions" depicted in Fig. 7.3 were computed on a mesh with edges 16 times finer than the good mesh in Fig. 7 using the FVSG method. The other schemes gave very similar results on this highly refined mesh. The results for the forward biased problems are in depicted in Fig. 7.4–Fig. 7.7. The reversed biased solutions are in Fig. 7.8–Fig. 7.11.

While this limited set of test problems is too small to draw any conclusions, we can make a few observations. We note that with the exception of the unmodified scheme from [5], all the schemes perform well on the good mesh (Fig. 7.4 and Fig. 7.8). The monotonicity of the FVSG method is apparent and the magnitude of the artificial diffusion in the streamline diffusion method seems appropriate. On the meshes with more obtuse angles (Fig. 7.5 and Fig. 7.9), the instability of all but the FVSG becomes apparent.

Perhaps the most striking result from this experiment is that the unmodified FVSG method performs surprisingly well, even on the meshes with obtuse angles. This provides some indication why it is used in practice and warrants further study. Note that the modified FVSG scheme performs essentially equally well and significantly better in Fig. 7.11.

REFERENCES

[1] R. E. Bank, *PLTMG: A Software Package for Solving Elliptic Parital Differential Equations, User's Guide 8.0*, vol. 5 of Software, Environments, and Tools, SIAM, 1998. See htpp://www.netlib.org/pltmg.

[2] R. E. Bank, J. F. Bürgler, W. Fichtner, and R. K. Smith, *Some upwinding techniques for finite element approximations of convection-diffusion equations*, Numer. Math., 58 (1990), pp. 185–202.

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

[4] F. Brezzi, L. D. Marini, and P. Pietra, *Numerical simulation of semiconductor devices*, Comp. Meth. Appl. Mech. Eng., 75 (1989), pp. 493–514.

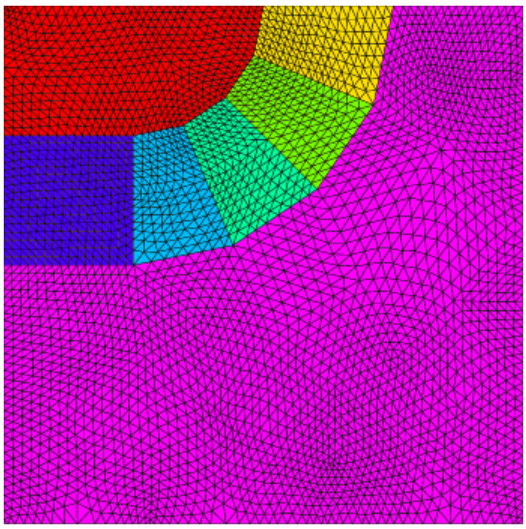
[5] ———, *Two-dimensional exponential fitting and applications to drift-diffusion models*, SIAM J. Numer. Anal., 26 (1989), pp. 1342–1355.

[6] P. Ciarlet, *Basic error estimates for elliptic problems*, in Handbook of Numerical Analysis [7].

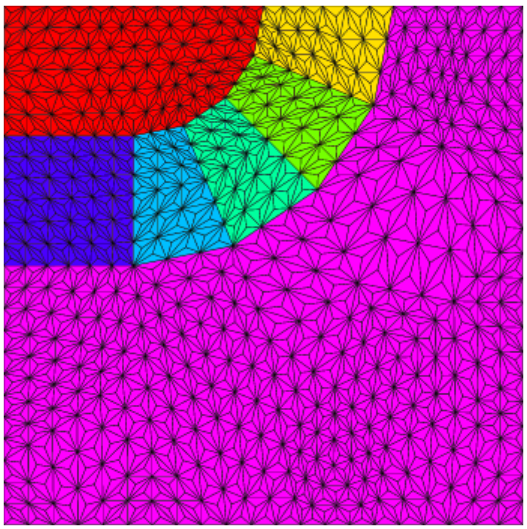
[7] P. Ciarlet and J. Lions, *Handbook of Numerical Analysis*, vol. II, Elsevier Science Publishers, 1991.

[8] T. Hughes and A. Brooks, *Streamline-upwind petrov-galerkin formulations for convection dominated flows with particular emphasis on the incmpressible navier-stokes equations*, Meth. Appl. Mech. Engng., 32 (1982), pp. 199–259.

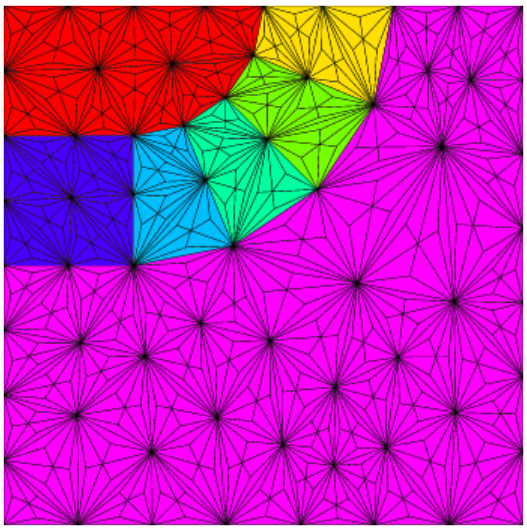
[9] P. A. Markowich and M. Zlámal, *Inverse-average-type finite element discretisations of self-adjoint second order elliptic problems*, Math. Comp., 51 (1988), pp. 431–449.



(a) Mesh 1



(b) Mesh 2 (many obtuse angles)



(c) Mesh 3 (even more obtuse angles)

FIG. 7.2. *Meshes*

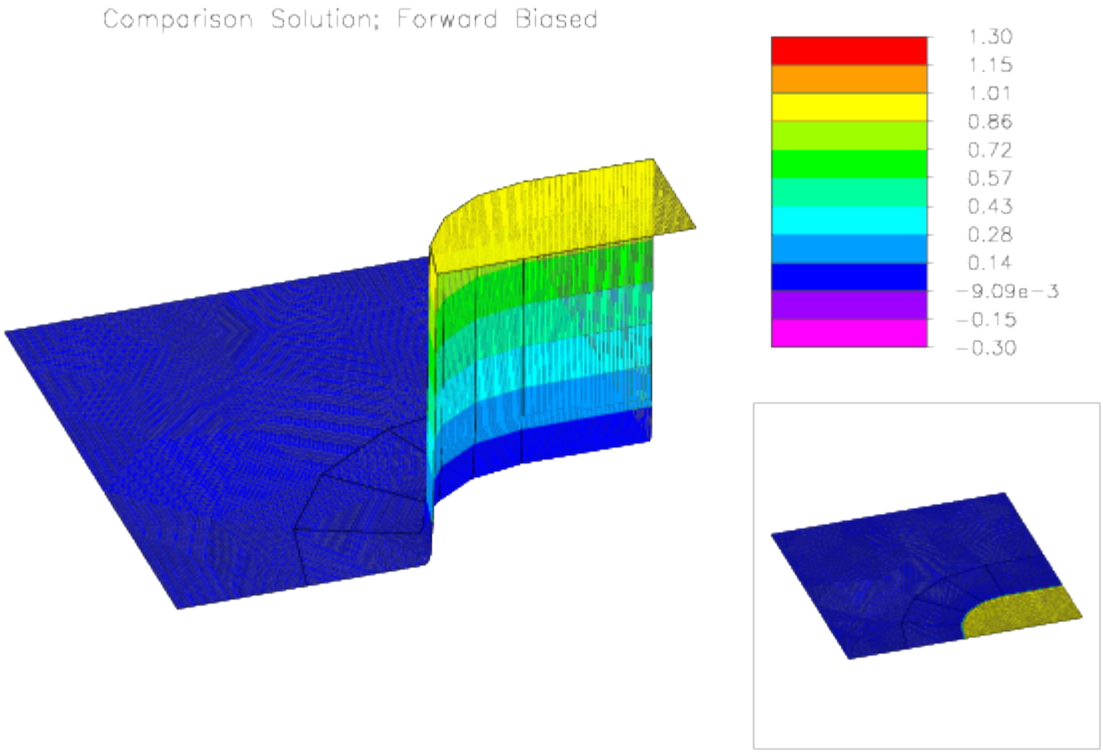
[10] J. MILLER AND S. WANG, *A triangular mixed finite element method for the stationary semiconductor device equation*, M^2AN, 25 (1991), pp. 441–463.

[11] M. MOCK, *Analysis of Mathematical Models of Semiconductor Devices*, Boole Press, Dublin, 1983.

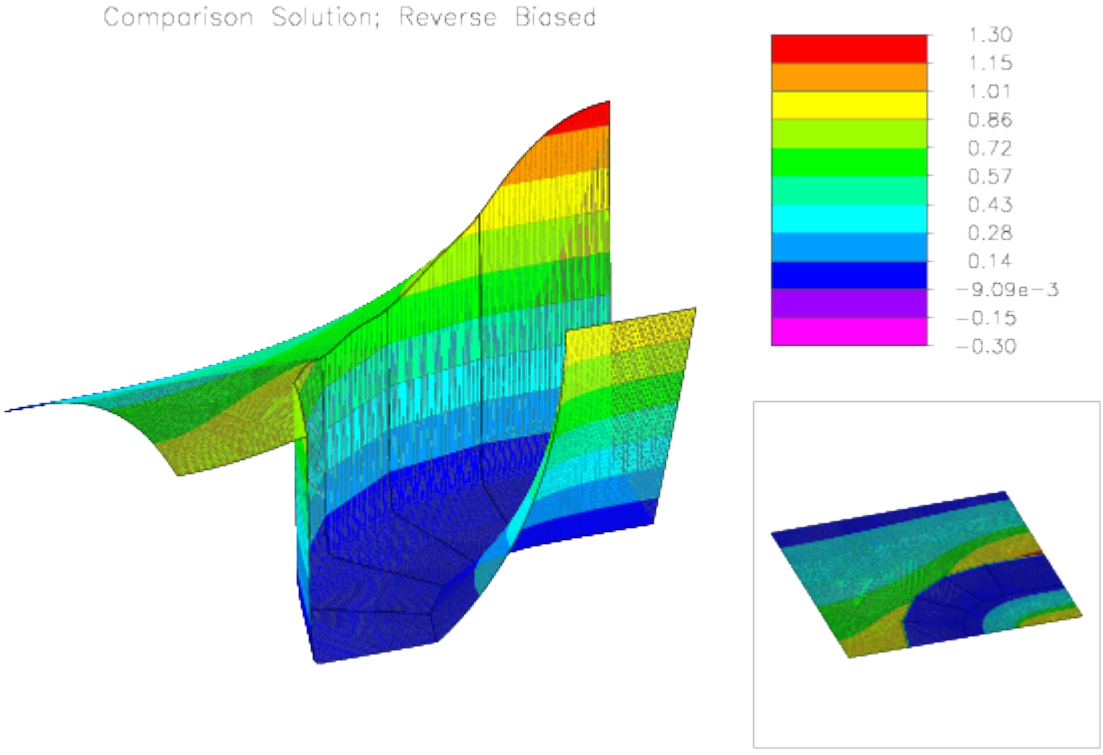
[12] D. SCHARFFETTER AND H. GUMMEL, *Large-signal analysis of a silicon Read diode oscillator*, IEEE Trans. Electron Devices, ED-16 (1969), pp. 64–77.

[13] J. XU, *The EAFE scheme and CWS method for convection dominated problems*, in Proceedings for Ninth International Conference on Domain Decomposition Methods, John Wiley & Sons, Ltd., 1996.

[14] J. XU AND L. ZIKATANOV, *A monotone finite element scheme for convection diffusion equations*. Submitted Math. Comp., http://www.math.psu.edu/xu/papers.html, 1996.

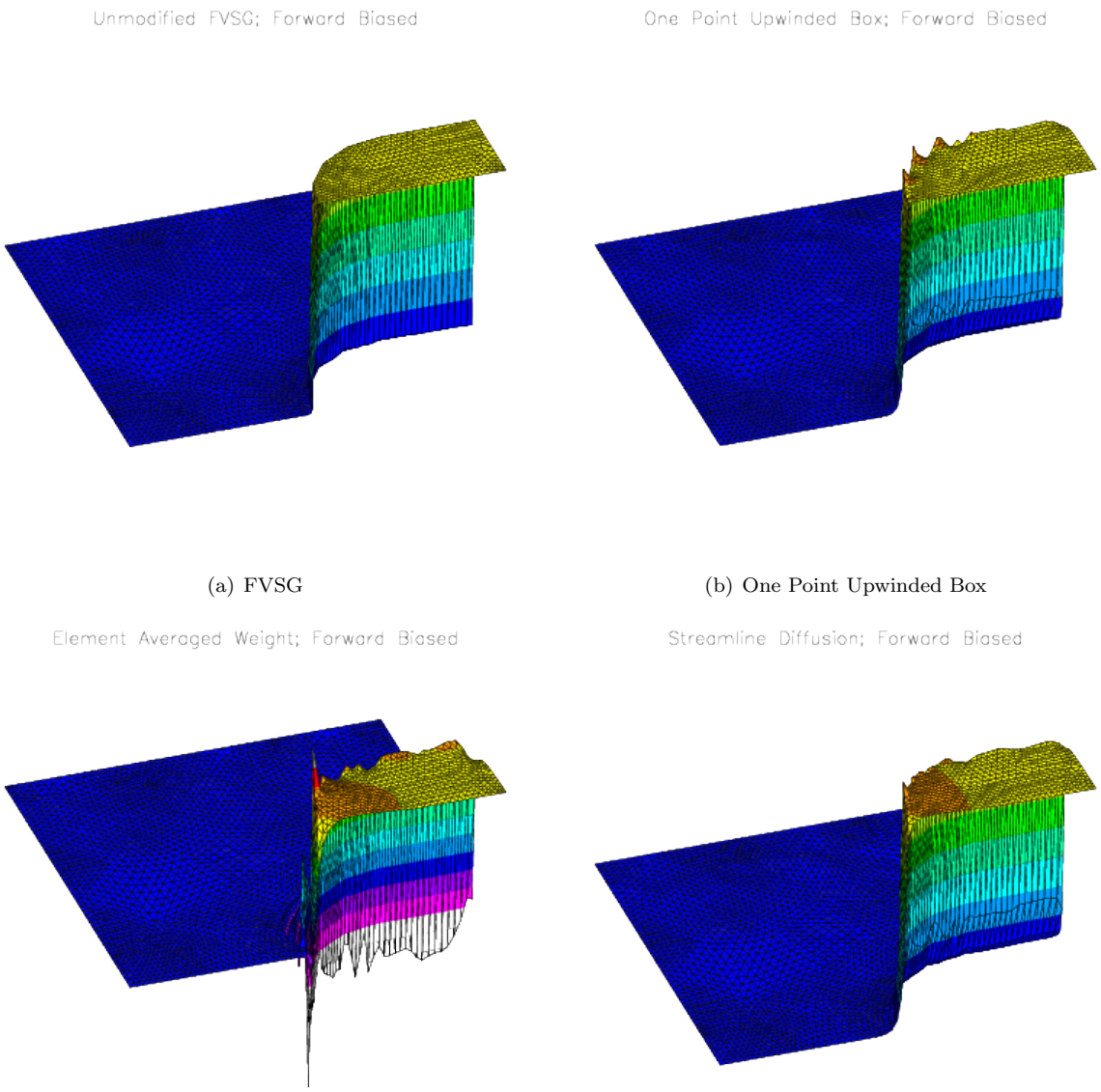


(a) Forward Biased



(b) Reverse Biased

FIG. 7.3. *Comparison Solutions*



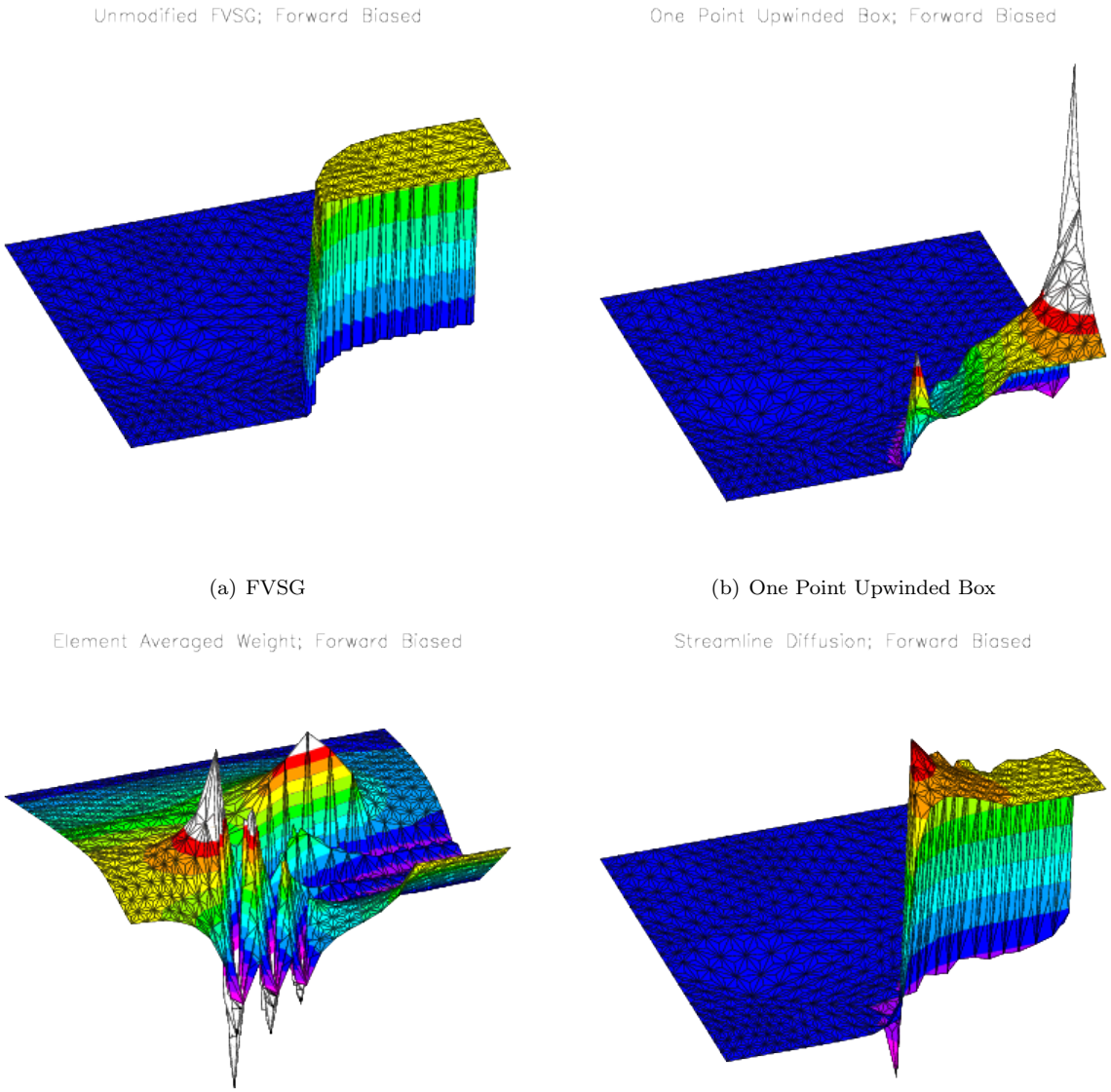
(a) FVSG

(b) One Point Upwinded Box

(c) Element Averaged Weight

(d) Streamline Diffusion

FIG. 7.4. *Forward Biased Test Problem on Mesh 1*



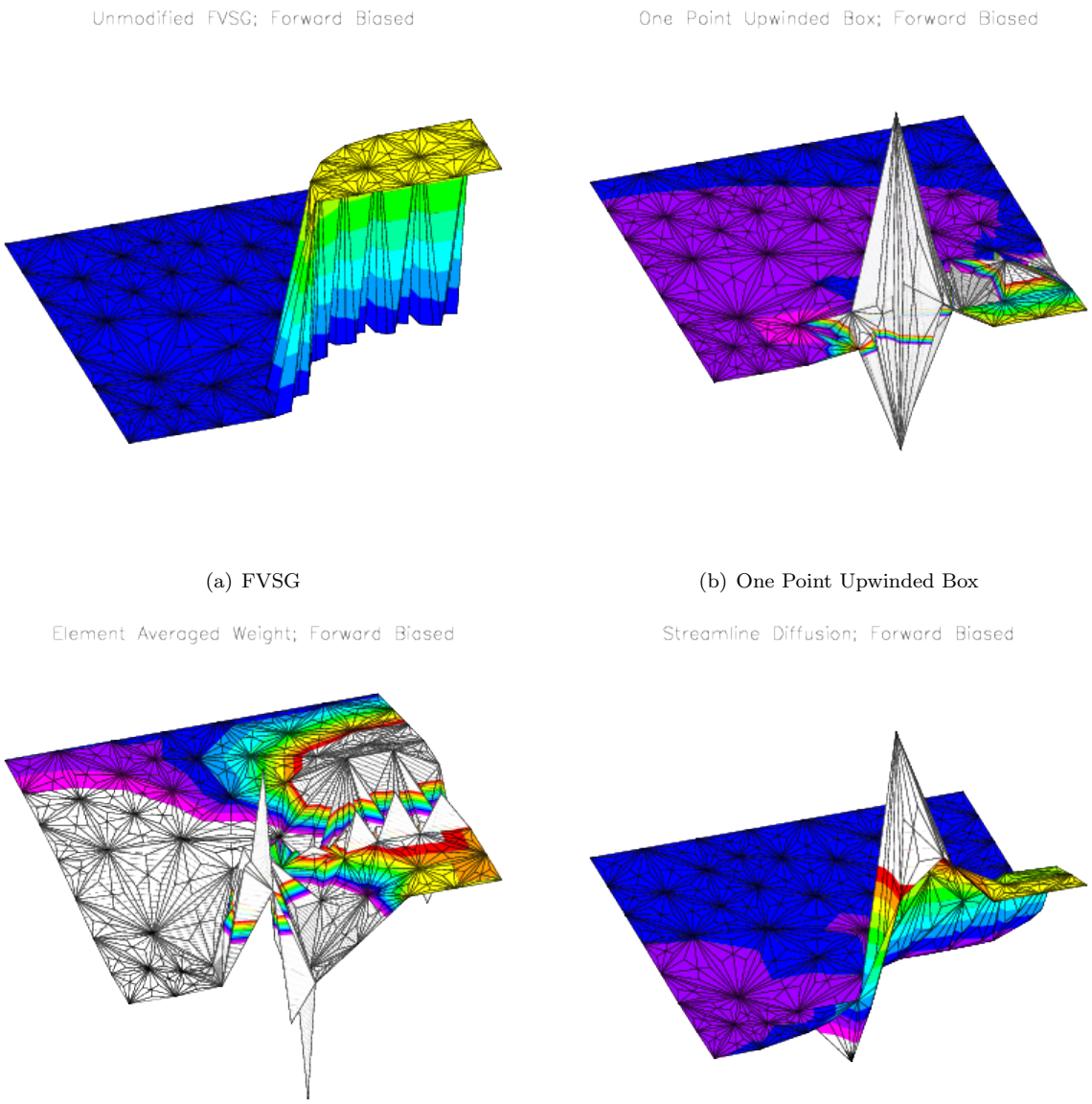
(a) FVSG

(b) One Point Upwinded Box

(c) Element Averaged Weight

(d) Streamline Diffusion

FIG. 7.5. *Forward Biased Test Problem on Mesh 2*



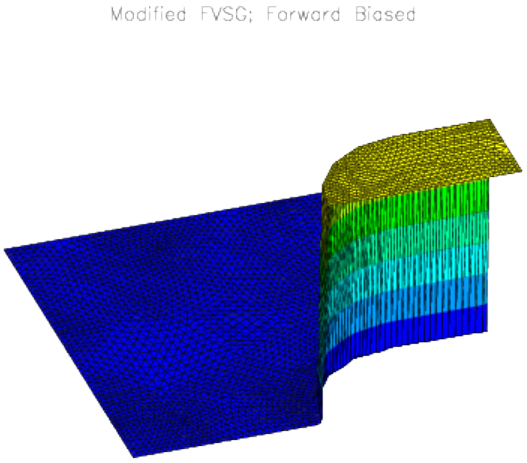
(a) FVSG

(b) One Point Upwinded Box

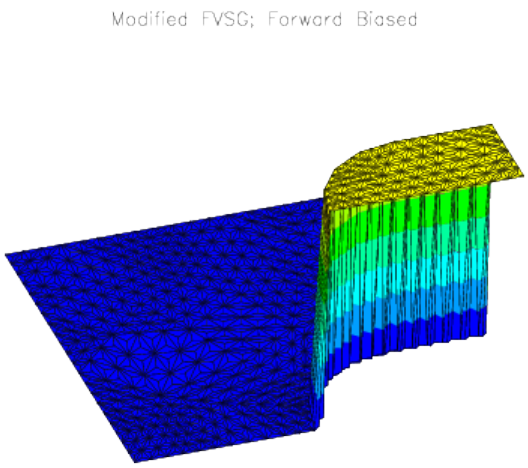
(c) Element Averaged Weight

(d) Streamline Diffusion

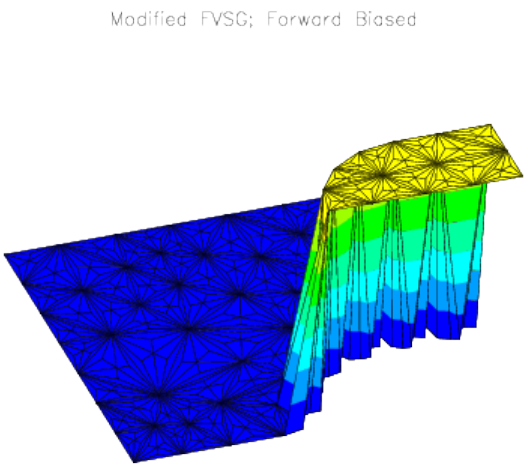
FIG. 7.6. *Forward Biased Test Problem on Mesh 3*



(a) Modified FVSG on Mesh 1

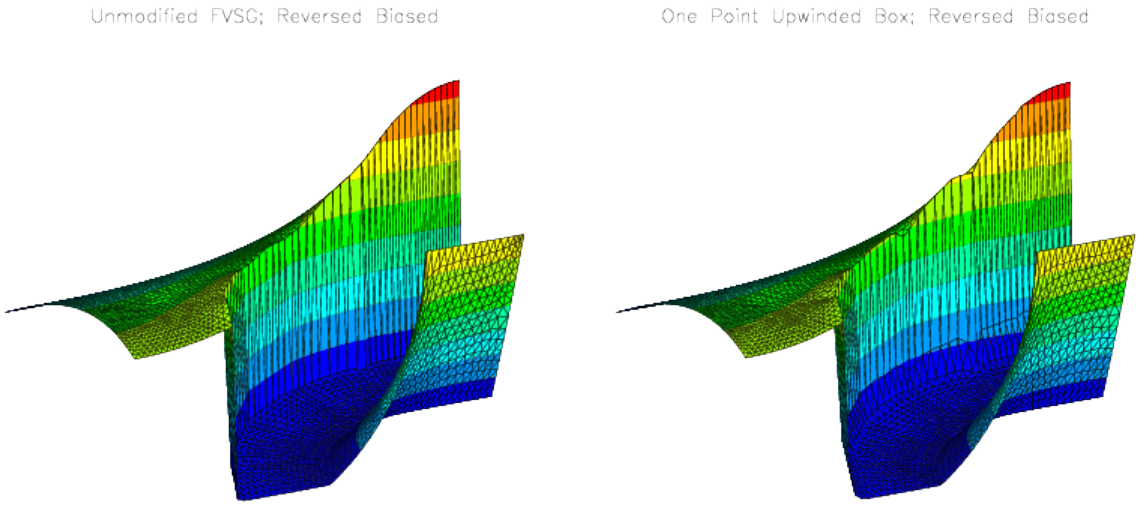


(b) Modified FVSG on Mesh 2



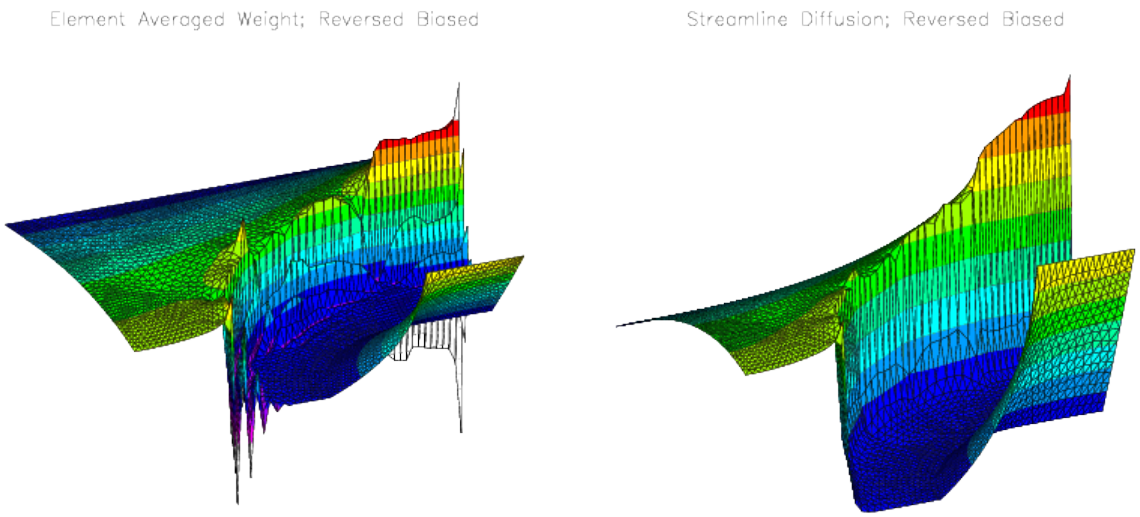
(c) Modified FVSG on Mesh 3

FIG. 7.7. *Modified FVSG Method for the Forward Biased Test Problem*



(a) FVSG

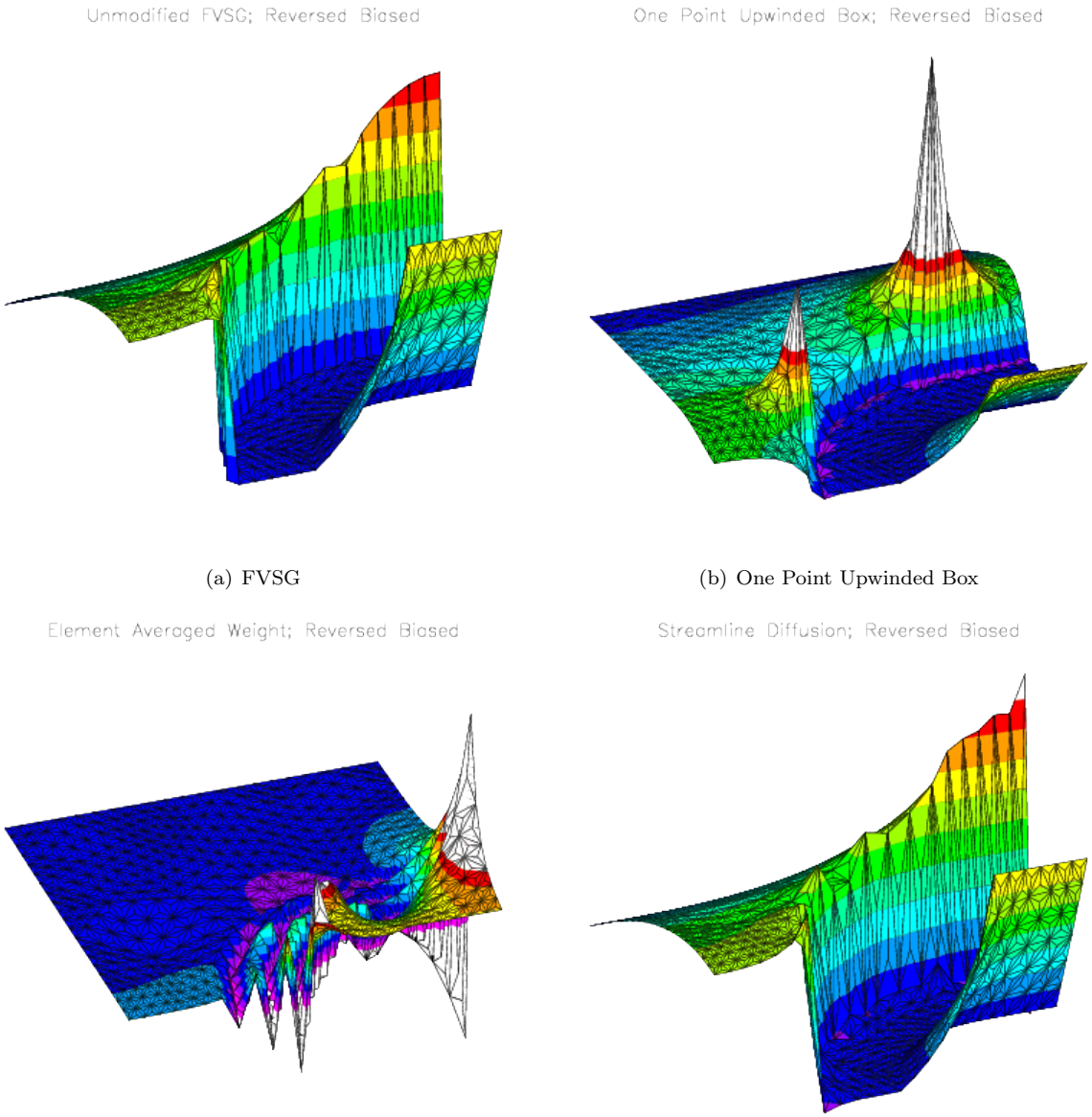
(b) One Point Upwinded Box



(c) Element Averaged Weight

(d) Streamline Diffusion

FIG. 7.8. *Reversed Biased Test Problem on Mesh 1*



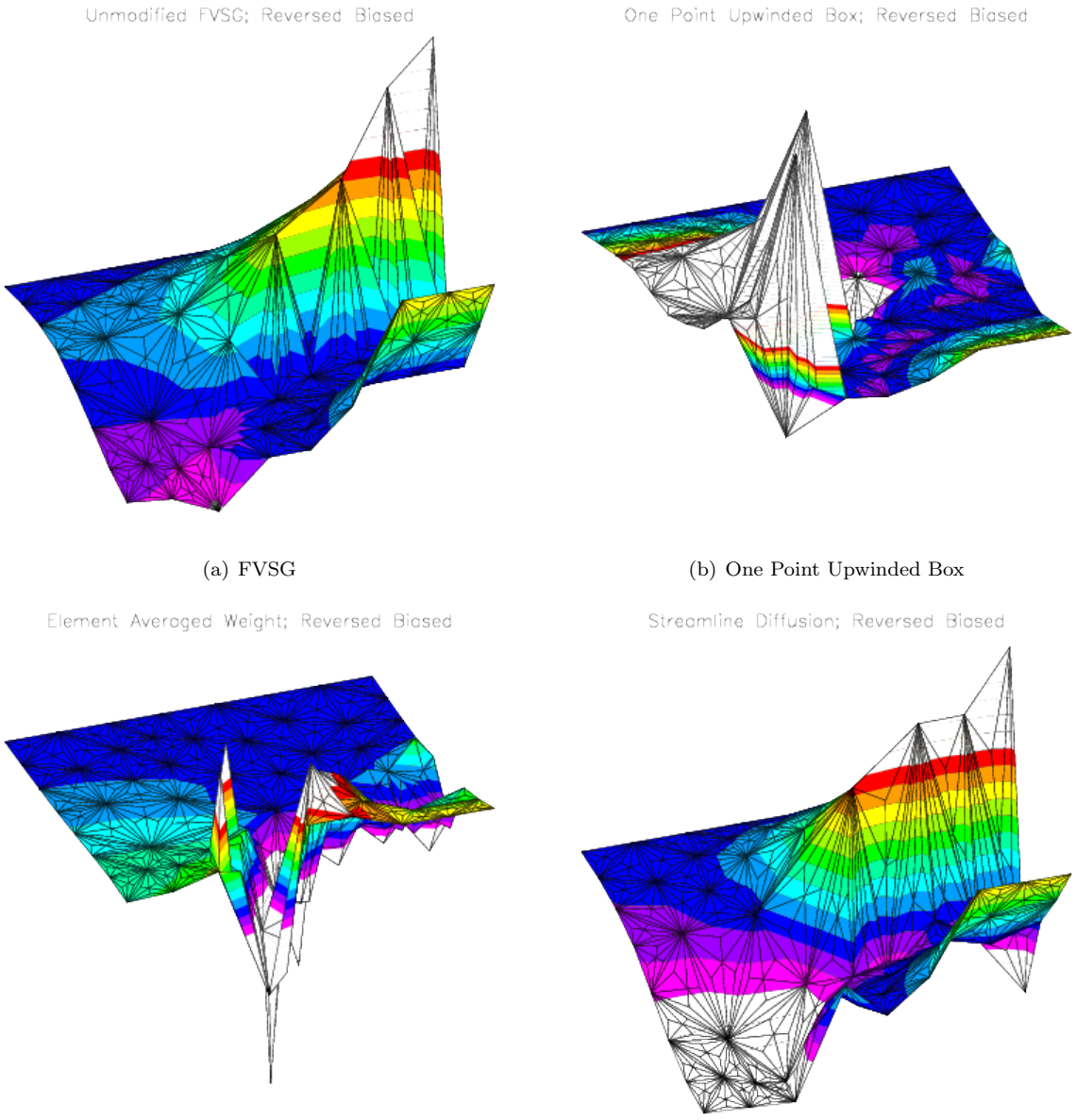
(a) FVSG

(b) One Point Upwinded Box

(c) Element Averaged Weight

(d) Streamline Diffusion

FIG. 7.9. *Reversed Biased Test Problem on Mesh 2*



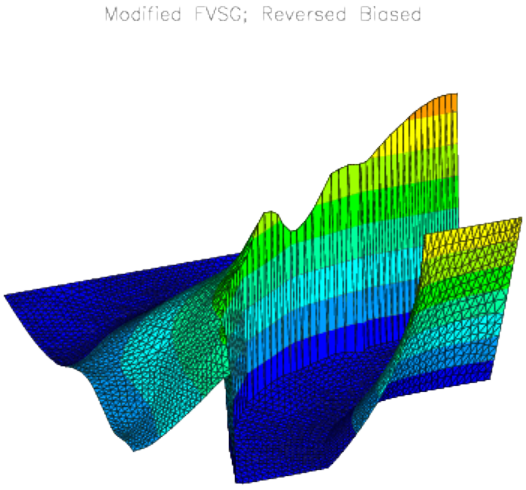
(a) FVSG

(b) One Point Upwinded Box

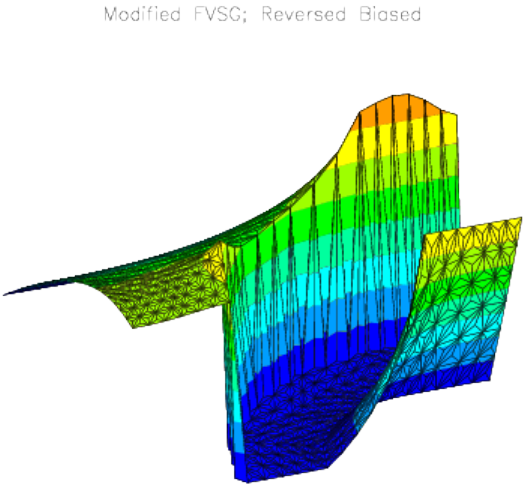
(c) Element Averaged Weight

(d) Streamline Diffusion

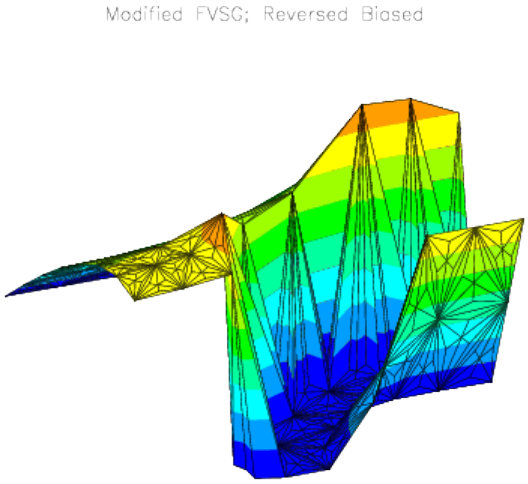
FIG. 7.10. *Reverse Biased Test Problem on Mesh 3*



(a) Modified FVSG on Mesh 1



(b) Modified FVSG on Mesh 2



(c) Modified FVSG on Mesh 3

FIG. 7.11. *Modified FVSG Method for the Reverse Biased Test Problem*