Virtual element stabilization of convection-diffusion equation with shock capturing

M. Arrutselvi, E. Natarajan
Department of Mathematics, Indian Institute of Space Science and Technology, Thiruvananthapuram, Kerala
E-mail: arrut.m@gmail.com, thanndavan@iist.ac.in

Abstract. Streamline upwind Petrov-Galerkin (SUPG) stabilization for the virtual element discretization of the convection-diffusion equation produces local oscillations in the crosswind direction. To overcome such a shortcoming, an additional shock-capturing term is added into the formulation. In this paper, we propose a nonlinear shock-capturing technique and prove the existence and stability of the discrete solution. The resulted nonlinear system of equations is solved using a simple iterative technique. Numerical experiments are conducted to show the efficiency of shock-capturing term in reducing the spurious oscillations along the crosswind direction.

1. Introduction

Steady scalar convection-diffusion equation

\[-\epsilon \Delta u + \mathbf{b} \cdot \nabla u = f \quad \text{in } \Omega, \quad u = 0 \quad \text{on } \partial \Omega,\]

are often studied in various physical and environmental applications. It is still a challenging task to provide accurate numerical solutions to this equation. In (1), the term \(\epsilon \Delta u\) denotes the diffusion and \(\mathbf{b} \cdot \nabla u\) denotes convection. When the problem is convection dominated, i.e. \(\epsilon \ll |\mathbf{b}|\), then the standard numerical discretization such as finite difference or finite element methods generate non-physical oscillations. In order to reduce the oscillations several stabilization techniques are proposed in the literature such as Streamline upwind Petrov-Galerkin (SUPG) [1, 2], local projection stabilization [3, 4], edge stabilization [5]. However, these stabilizations do not reduce the local oscillations along sharp layers. To overcome this, a suitable artificial diffusion term is added to the SUPG stabilization consistently. We call these approaches as shock-capturing methods [6]. Several problems are studied in this direction, e.g. [6, 7, 8, 9].

Recently, a new technique has been developed to obtain numerical solutions of partial differential equations on arbitrary polygonal/polyhedral meshes known as the Virtual element method (VEM). VEM has evolved from the mimetic finite difference [10, 11] method and is seen as a natural generalization of the finite element method. Virtual element space is defined such that the discrete solution is easily computed over polygons with the help of specific degrees of freedom. Using the polynomial projection operators, the bilinear form is split into its polynomial, and non-polynomial counterparts. Then the virtual element matrices are evaluated solely based on the chosen degrees of freedom.

VEM has been studied successfully for several problems such as linear elasticity [12], conforming and nonconforming VEM for elliptic equation [13, 14], parabolic problem [15], hyperbolic problem [16], semilinear and quasilinear problems [17, 18, 19, 20], mixed VEM [21] for elliptic problems, acoustic vibration problem [22], Stokes problem [23], 2D magnetostatic problems [24], posteriori error estimation for the elliptic problems [25]. Mesh generation problems pose a major challenge in the numerical simulation of transport problems such as subsurface fluid flows. These issues can be circumvented by using the virtual element discretization. More recent work on VEM applied to
models of fluid flows can be found in [26]. Virtual element stabilization of convection dominated transport problems are studied in [27, 28]. In [29] the spurious oscillations present along the sharp layers are addressed. However, the stabilizers introduced in this technique do not address the mesh generation problems that are inherently present in some of the fluid dynamics models in particular underground fluid flow simulations. To overcome this deficiency, we propose in this article a shock-capturing technique for the virtual element discretization of convection-diffusion equation.

The paper is organized as follows. In section 2 we describe our model problem with assumptions and the weak formulation. A brief description of virtual element spaces is provided in section 3 by considering the finite-dimensional space with suitable degrees of freedom and section 4 discusses the VEM discretization with the Streamline upwind Petrov Galerkin (SUPG) stabilization. In section 5 we derive the well-posedness of the discrete formulation by proving the coercivity and continuity of the bilinear form. Then the nonlinear shock-capturing stabilization term is added to the bilinear form and we prove the existence of the solution in section 6. This reduces the problem to a nonlinear algebraic system of equations which is solved with the help of a simple iterative technique. We perform the numerical simulations in section 7 to show the efficiency of the shock-capturing technique in reducing the spurious oscillations developed along the sharp layers. Finally, the concluding remarks are discussed in section 8.

2. Governing equations and weak formulation

For a measurable set $D \subset \mathbb{R}^2$, let $L^2(D)$ denote the usual Lebesgue space with the standard $L^2$ inner product and norm, $(\cdot, \cdot)_D$ and $\| \cdot \|_D$ respectively.

Consider the following convection-diffusion equation with homogeneous Dirichlet boundary condition:

$$-\nabla \cdot (K \nabla u) + b \cdot \nabla u + \alpha u = f \quad \text{in } \Omega, \quad u = 0 \quad \text{on } \partial \Omega. \quad (2)$$

Here $u(x)$ denotes the unknown where $x \in \Omega \subset \mathbb{R}^2$, $K > 0$, $b \in W^{1,\infty}(\Omega)^2$ is the velocity field, $\alpha \geq 0$ and $f \in L^2(\Omega)$. We also assume $(\nabla \cdot b)(x) = 0$ a.e in $\Omega$ and $K \geq K_0 > 0$.

The bilinear form of equation (2) is defined as $B : H^1_0(\Omega) \times H^1_0(\Omega) \to \mathbb{R}$ such that

$$B(w, v) = (K \nabla w, \nabla v)_\Omega + (b \cdot \nabla w, v)_\Omega + (\alpha w, v)_\Omega \quad \forall w, v \in H^1_0(\Omega).$$

Using integration by parts on the convective term $(b \cdot \nabla w, v)$ and the condition $\nabla \cdot b = 0$, the bilinear form $B$ can be equivalently redefined as,

$$B(w, v) = (K \nabla w, \nabla v)_\Omega + \frac{1}{2}[(b \cdot \nabla w, v)_\Omega - (w, b \cdot \nabla v)_\Omega] + (\alpha w, v)_\Omega \quad \forall w, v \in H^1_0(\Omega). \quad (3)$$

The weak formulation of (2) is: Find $u \in H^1_0(\Omega)$ such that

$$B(u, v) = (f, v)_\Omega \quad \forall v \in H^1_0(\Omega). \quad (4)$$

The existence and uniqueness of the solution of weak formulation (4) follows from the Lax-Milgram lemma [30].

3. Overview of VEM

Let $\{E_h\}_{h>0}$ be a family of partitions of $\Omega$ into polygonal elements $E$ with $h$ being the maximum diameter over the polygons. Each element $E \in E_h$ satisfies the following assumptions:

(i) every element $E$ is star-shaped with respect to a disc $D_h$ of radius $\gamma > 0 h_E$, where $h_E$ being the element diameter,

(ii) for each edge $e \subset E$, the length $|e| \geq c h_E$ for a positive constant $c$ independent of $h_E$,

(iii) for the elements $E \in E_h$, the boundary is made up of finite number of edges.

Let $P_k(E)$ denote the set of polynomials of degree $\leq k$ on $E$. We define the projection operator $\Pi_k^E : H^1(E) \to P_k(E)$ by,

$$\begin{cases}
(\nabla (\Pi_k^E u - u), \nabla q_k)_E = 0 & \forall q_k \in P_k(E), \\
P_k^E(\Pi_k^E u - u) = 0,
\end{cases} \quad (5)$$
where

\[
P^0 u = \begin{cases} \frac{1}{|\partial E|} \int_{\partial E} u \, ds & \text{for } k = 1, \\ \frac{1}{|E|} \int_E u \, ds & \text{for } k \geq 2, \end{cases}
\]

and define \( \Pi^k \) which is the \( L^2 \) projection onto \( P_k(E) \) by,

\[
(u - \Pi^k u, q_k)_E = 0 \quad \forall q_k \in P_k(E).
\]

Consider the following space \( W^k_E \) for each \( E \in T_h \) by,

\[
W^k_E = \{ u \in H^1(E) \cap C^0(\partial E) : v_e \in P_k(e) \forall \text{edge } e \subset \partial E, \Delta u \in P_k(E) \}.
\]

Now we define the local virtual element space \( V^k_E \) as follows,

\[
V^k_E = \{ u \in W^k_E \ s.t. \ (u - \Pi^k u, q)_E = 0 \quad \forall q \in P_k(E)/P_{k-2}(E) \}.
\]

where \( P_k(E)/P_{k-2}(E) \) denotes the polynomials of degree \( k \) that are \( L^2 \) orthogonal to polynomials of degree \( k - 2 \) on \( E \). Along with this, we provide the following set of degrees of freedom on \( V^k_E \) by,

\begin{align*}
(\text{G}_1) & 	ext{ the values of } u \text{ at the } \alpha(\partial E) \text{ vertices of polygon } E, \\
(\text{G}_2) & \text{ the values of } u \text{ at } k - 1 \text{ internal Gauss-Lobatto quadrature points on every edge } e \subset \partial E, \\
(\text{G}_3) & \text{ the moments up to order } k - 2 \text{ of } u \text{ in } E, \text{ i.e.,}
\end{align*}

\[
\int_E u q_{k-2} \, dx \quad \forall q_{k-2} \in P_{k-2}(E).
\]

We note that the degrees of freedom mentioned above determines \( u \) uniquely on the polygon \( E \). Now we define the global virtual element space \( V^k_h \) by,

\[
V^k_h = \{ u \in H^1_0(\Omega) \ s.t. \ u|_E \in V^k_E \ \forall E \in T_h \}.
\]

4. VEM-SUPG formulation

When the problem is singularly perturbed i.e., \( K \ll 1 \) the standard Galerkin approximation produces spurious oscillations. One of the possible remedy is to use stabilization techniques to circumvent such situations. Streamline upwind Petrov-Galerkin (SUPG) stabilization [31] tackles the problem by adding artificial diffusion term along the streamline. However, in order to formulate the problem with respect to VEM discretization the terms that appear in the discrete formulation should be suitably redefined so as to ensure its VEM computability.

Now we proceed to define the SUPG stabilised virtual element discretisation of the formulation (4) as follows. Find \( u_h \in V^k_h \) such that

\[
B_{\text{ca}}(u_h, v_h) = F_{\text{ca}}(v_h) \quad \forall v_h \in V^k_h,
\]

where the bilinear form \( B_{\text{ca}} : V^k_h \times V^k_h \to \mathbb{R} \) is such that,

\[
B_{\text{ca}}(u_h, v_h) := a_h(u_h, v_h) + b_h(u_h, v_h) + c_h(u_h, v_h) + d_h(u_h, v_h),
\]

with,

\begin{align*}
a_h(u_h, v_h) & := \sum_{E \in T_h} \left( (K \Pi^0_{k-1} \nabla w_h, \Pi^0_{k-1} \nabla v_h)_E + \tau_E (b \cdot \Pi^0_{k-1} \nabla w_h, b \cdot \Pi^0_{k-1} \nabla v_h)_E \right), \\
b_h(u_h, v_h) & := \sum_{E \in T_h} \left( (\alpha \Pi^0_{k} w_h, \Pi^0_{k} v_h)_E + \alpha S^E_h (I - \Pi^0_{k}) \nabla w_h, (I - \Pi^0_{k}) \nabla v_h)_E \right), \\
c_h(u_h, v_h) & := \frac{1}{2} \sum_{E \in T_h} \left( (b \cdot \Pi^0_{k-1} \nabla w_h, \Pi^0_{k} v_h)_E - (\Pi^0_{k} w_h, b \cdot \Pi^0_{k-1} \nabla v_h)_E \right), \\
d_h(u_h, v_h) & := \sum_{E \in T_h} \tau_E \left( - \nabla \cdot K \Pi^0_{k-1} \nabla w_h + \alpha \Pi^0_{k} w_h, b \cdot \Pi^0_{k-1} \nabla v_h \right)_E.
\end{align*}
and the linear form \( F_{\alpha} : V_h^N \rightarrow \mathbb{R} \) is defined as

\[
F_{\alpha}(v_h) := \sum_{E \in \mathcal{T}_h} \left[ (f, \Pi^i_0 v_h)_E + \tau_E (f, b \cdot \Pi^i_{E-1} \nabla v_h)_E \right].
\]

(15)

where \( b_E = \sup_{x \in E} \|b(x)\|_{\mathbb{R}^2} \), \( \tau_E \) is the stabilization parameter that is chosen accordingly and, \( S^E_i(\cdot, \cdot) \) and \( S^E_2(\cdot, \cdot) \) denotes the symmetric positive bilinear forms defined on \( V_h^N \times V_h^N \) by the following,

\[
S^E_i(u_h, v_h) = \sum_{i=1}^{N} \text{dof}(u_h) \text{dof}(v_h) \quad \text{and} \quad S^E_2(u_h, v_h) = h_E^2 \sum_{i=1}^{N} \text{dof}(u_h) \text{dof}(v_h),
\]

(16)

where \( \text{dof}(u_h) \) denotes the \( i \)th degree of freedom of \( u_h \) with \( N \) denoting the total degrees of freedom. Let there exists non-zero positive constants \( \beta_\alpha, \beta_\eta, \eta_\eta \) and \( \eta^* \) independent of \( h_E \), such that

\[
\begin{align*}
\beta_\alpha(\nabla u_h, \nabla u_h)_E & \leq S^E_i(u_h, u_h)_E \leq \beta(\nabla u_h, \nabla u_h)_E \quad \forall u_h \in \ker(\Pi^i_0), \quad (17) \\
\eta_\eta(u_h, u_h)_E & \leq S^E_2(u_h, u_h) \leq \eta^*(u_h, u_h)_E \quad \forall u_h \in \ker(\Pi^i_0). \quad (18)
\end{align*}
\]

We introduce the norm \( ||| \cdot ||| \) to be used in our error analysis,

\[
|||v|||^2 := \sum_{E \in \mathcal{T}_h} \left( K \|\nabla v\|_E^2 + \|\sqrt{\alpha} v\|_E^2 + \tau_E \|b \cdot \nabla v\|_E^2 \right). \quad (19)
\]

We also state the local inverse inequality to be used later, there exists a constant \( C_I \) such that

\[
\|\nabla \cdot K \nabla v_h\|_E \leq C_I h^{-1} E \|K \nabla v_h\|_E \quad \forall v_h \in V_h \quad \text{and} \quad E \in \mathcal{T}_h. \quad (20)
\]

In the sequel we assume the following,

\[
(G1) \ \exists \rho \in (0, 3) \ \text{independent of} \ \mathcal{T}_h \ \text{such that} \ (i) \ K \tau_E C_I^2 \leq \frac{3}{2} \rho h_E^2 \quad \text{and} \quad (ii) \ \tau_E \alpha \leq \frac{1}{2} \rho \quad \text{a.e. in} \ \Omega, \quad \text{where} \ \mathcal{C}_I \ \text{is the same constant used in} \ (20).
\]

5. Well-posedness of VEM-SUPG formulation

In this section we will show the well-posedness of the VEM-SUPG formulation (9) by first showing the coercivity and then the continuity of bilinear form \( B_{\alpha} \).

Lemma 5.1. (Coercivity) The bilinear form \( B_{\alpha}(\cdot, \cdot) \) satisfies the following estimate,

\[
B_{\alpha}(v_h, v_h) \geq C_p \|v_h\|^2 \quad \forall v_h \in V_h, \quad (21)
\]

with \( C_p = \min \left\{ \beta_\alpha, \eta_\eta, (1 - \frac{\sqrt{\eta^*}}{2}) \right\} > 0. \)

Proof. We estimate the terms of \( B_{\alpha}(\cdot, \cdot) \) one by one. We have,

\[
\begin{align*}
ah(v_h, v_h) &= \sum_{E \in \mathcal{T}_h} \left[ (K \Pi^i_{E-1} \nabla v_h, \Pi^i_{E-1} \nabla v_h)_E + \tau_E (b \cdot \Pi^i_{E-1} \nabla v_h, b \cdot \Pi^i_{E-1} \nabla v_h)_E \right] \\
&\geq \sum_{E \in \mathcal{T}_h} \left( K \|\nabla v_h\|^2_E + \tau_E \|b \cdot \nabla v_h\|^2_E + \beta_\alpha (K + \tau_E b_E^2)(1 - \Pi^i_0 v_h, (I - \Pi^i_0) v_h) \right) \\
&\geq \sum_{E \in \mathcal{T}_h} \left( K \|\nabla v_h\|^2_E + \tau_E \|b \cdot \nabla v_h\|^2_E + \beta_\alpha(K + \tau_E b_E^2)(1 - \Pi^i_0 v_h, (I - \Pi^i_0) v_h) \right).
\end{align*}
\]

Similarly,

\[
b_h(u_h, v_h) \geq \sum_{E \in \mathcal{T}_h} \alpha \left( \|\Pi^i_0 v_h\|^2_E + \eta_\eta (1 - \Pi^i_0) v_h, (I - \Pi^i_0) v_h \right), \quad \text{and} \quad c_h(v_h, v_h) = 0. \quad (22)
\]
Now, estimating the last term of $B_{\alpha}$, we have for some $\lambda > 0$,

$$|d_h(v_h, v_h)| \leq \sum_{E \in T_h} \tau_E \left| \langle -\nabla \cdot K \Pi_{E}^{0}, \nabla v_h + \alpha \Pi_{E}^{0} h, b \cdot \Pi_{E}^{0} \nabla v_h \rangle_E \right|$$

$$\leq \sum_{E \in T_h} \tau_E \| -\nabla \cdot K \Pi_{E}^{0} \nabla v_h + \alpha \Pi_{E}^{0} v_h \|_E \| b \cdot \Pi_{E}^{0} \nabla v_h \|_E$$

$$\leq \sum_{E \in T_h} \left( \tau_E \| -\nabla \cdot K \Pi_{E}^{0} \nabla v_h + \alpha \Pi_{E}^{0} v_h \|_E^2 + \frac{\lambda \tau_E}{2} \| b \cdot \Pi_{E}^{0} \nabla v_h \|_E^2 \right)$$

$$\leq \sum_{E \in T_h} \left( \tau_E \| -\nabla \cdot K \Pi_{E}^{0} \nabla v_h \|_E^2 + \frac{\tau_E}{2} \| \alpha \Pi_{E}^{0} v_h \|_E^2 + \frac{\lambda \tau_E}{2} \| b \cdot \Pi_{E}^{0} \nabla v_h \|_E^2 \right). \quad (24)$$

Applying the inverse inequality estimate (20), the assumptions in (G1) and later choosing $\lambda = \sqrt{\rho}$ in (24), we get,

$$|d_h(v_h, v_h)| \leq \sum_{E \in T_h} \left( \frac{\rho}{2} K \| \Pi_{E}^{0} \nabla v_h \|_E^2 + \frac{\rho}{2} \sqrt{\alpha} \| \Pi_{E}^{0} v_h \|_E^2 + \frac{\lambda \tau_E}{2} \| b \cdot \Pi_{E}^{0} \nabla v_h \|_E^2 \right)$$

$$\leq \sum_{E \in T_h} \left( \frac{\sqrt{\rho}}{2} K \| \Pi_{E}^{0} \nabla v_h \|_E^2 + \sqrt{\alpha} \| \Pi_{E}^{0} v_h \|_E^2 + \frac{\sqrt{\rho}}{2} \| b \cdot \Pi_{E}^{0} \nabla v_h \|_E^2 \right). \quad (25)$$

Combining all the above estimates (22), (23) and (25) of the terms of $B_{\alpha}(v_h, v_h)$, we have,

$$B_{\alpha}(v_h, v_h) \geq \left( 1 - \frac{\sqrt{\rho}}{2} \right) \sum_{E \in T_h} \left[ \langle K \Pi_{E}^{0} \nabla v_h \rangle_E^2 + \tau_E \| b \cdot \Pi_{E}^{0} \nabla v_h \|_E^2 + \alpha \Pi_{E}^{0} v_h \|_E^2 \right] +$$

$$\sum_{E \in T_h} \left[ \beta_s \langle K + \tau_E \rho E \rangle \| (I - \Pi_{E}^{0}) \nabla v_h \|_E^2 \right] + \eta_s \langle (I - \Pi_{E}^{0}) v_h \rangle^2 \right]$$

$$\geq \min \left\{ \beta_s, \eta_s, (1 - \frac{\sqrt{\rho}}{2}) \right\} \sum_{E \in T_h} \left[ K \| \nabla v_h \|_E^2 + \tau_E \| b \cdot \nabla v_h \|_E^2 + \alpha \| v_h \|_E^2 \right].$$

Thus, we obtain the estimate (21), proving the coercivity. \qed

**Lemma 5.2.** (Continuity) For $u \in H^1_0(\Omega)$ with $(\nabla \cdot K \nabla u)_{\mathcal{E}} \in L^2(\mathcal{E})$, $\forall \mathcal{E} \in T_h$ and $v_h \in V_h$ we have,

$$|B_{\alpha}(u, v_h)| \leq C_{\alpha} \gamma(u)^{\frac{1}{2}} \| v_h \|_E. \quad (26)$$

$$\gamma(u) := \| u \| + \left( \sum_{E \in T_h} \min \left\{ \frac{1}{\tau_E}, \frac{\rho}{K_0} \right\} \| u \|_E^2 \right)^{\frac{1}{2}}, \quad (27)$$

where $C_{\alpha}$ is a constant depending on $K$, $b$, and $\alpha$, but independent of $h$ and $\tau_E$.

**Proof.** Using the triangle inequality, $|B_{\alpha}(u, v_h)| \leq |a_h(u, v_h)| + |b_h(u, v_h)| + |c_h(u, v_h)| + |d_h(u, v_h)|$.

Using the inequality (17), we estimate,

$$|a_h(u, v_h)| \leq \sum_{E \in T_h} \left[ \langle K \Pi_{E}^{0} \nabla u, \Pi_{E}^{0} \nabla v_h \rangle_E \right]$$

$$+ \beta(K + \tau_E \rho E) \| (I - \Pi_{E}^{0}) \nabla u \|_E \| (I - \Pi_{E}^{0}) \nabla v_h \|_E$$

$$\leq \sum_{E \in T_h} \left( 1 + \beta^* \right) K \| \nabla u \|_E \| \nabla v_h \|_E +$$

$$+ \left( \frac{\sqrt{\rho}}{2} \right) \| \tau_E \| \| \nabla u \|_E \| \nabla v_h \|_E$$

$$\leq (1 + \beta^*) \sum_{E \in T_h} \left[ K \| \nabla u \|_E \| \nabla v_h \|_E +$$

$$+ \left( \frac{\sqrt{\rho}}{2} \right) \| \tau_E \| \| \nabla u \|_E \| \nabla v_h \|_E \right)$$

$$\leq (1 + \beta^*) \left( 1 + \max_{E \in T_h} \left( \frac{\sqrt{\rho}}{K_0 \alpha} \right) \right) \sum_{E \in T_h} \left( K \| \nabla u \|_E \| \nabla v_h \|_E \right) \quad (\text{use (ii) of (G1))}$$
Using Holder’s inequality, we get,
\[
\leq (1 + \beta^*) \left(1 + \max_{E \in \mathcal{T}_h} \left( \frac{b_E^2 \mu}{K_0 \alpha} \right) \right) \left( \sum_{E \in \mathcal{T}_h} K \|\nabla u\|_E^2 \right)^{\frac{1}{2}} \left( \sum_{E \in \mathcal{T}_h} K \|\nabla v_h\|_E^2 \right)^{\frac{1}{2}}
\]
Similarly using the inequality (18), we have,
\[
|b_h(u, v_h)| \leq (1 + \eta^*) \|u\| \|v_h\|.
\]
(28)
Consider the third term \(|c_h(u, v_h)|\), we note,
\[
|c_h(u, v_h)| \leq \frac{1}{2} \sum_{E \in \mathcal{T}_h} |(b \cdot \Pi_{k-1}^0 \nabla u, \Pi_{k-1}^0 v_h)_E| + \frac{1}{2} \sum_{E \in \mathcal{T}_h} |(\Pi_{k-1}^0 u, b \cdot \Pi_{k-1}^0 \nabla v_h)_E|.
\]
(30)
Estimating the first term of (30), we get,
\[
\frac{1}{2} \sum_{E \in \mathcal{T}_h} |(b \cdot \Pi_{k-1}^0 \nabla u, \Pi_{k-1}^0 v_h)_E| \leq \frac{1}{2} \sum_{E \in \mathcal{T}_h} \left( \frac{b_E}{\sqrt{K_0 \alpha}} \right) \sqrt{K} \|\nabla u\|_E \sqrt{\alpha} \|v_h\|_E
\]
\leq \max_{E \in \mathcal{T}_h} \left( \frac{b_E}{\sqrt{K_0 \alpha}} \right) \|u\| \|v_h\|_E.
\]
(31)
The second term of (30) is estimated in two different ways namely,
\[
\frac{1}{2} \sum_{E \in \mathcal{T}_h} |(\Pi_{k-1}^0 u, b \cdot \Pi_{k-1}^0 \nabla v_h)_E| \leq \frac{1}{2} \sum_{E \in \mathcal{T}_h} \left( \frac{1}{\sqrt{\tau_E}} \|u\|_E \frac{b_E}{\sqrt{K_0 \alpha}} \sqrt{K} \|v_h\|_E \right)
\leq \left( \max_{E \in \mathcal{T}_h} \frac{b_E}{\sqrt{K_0 \alpha}} \right) \left( \sum_{E \in \mathcal{T}_h} \left( \frac{1}{\tau_E} \|u\|_E^2 \right)^{\frac{1}{2}} \|v_h\|_E \right).
\]
(32)
and
\[
\frac{1}{2} \sum_{E \in \mathcal{T}_h} |(\Pi_{k-1}^0 u, b \cdot \Pi_{k-1}^0 \nabla v_h)_E| \leq \frac{1}{2} \sum_{E \in \mathcal{T}_h} \|u\|_E \frac{b_E}{\sqrt{K_0 \alpha}} \sqrt{K} \|v_h\|_E
\leq \left( \sum_{E \in \mathcal{T}_h} \frac{b_E^2}{K_0 \alpha} \|u\|_E^2 \right)^{\frac{1}{2}} \|v_h\|_E.
\]
(33)
Combining the estimates in (31),(32) and (33) we have,
\[
|c_h(u, v_h)| \leq \max_{E \in \mathcal{T}_h} \left( \frac{b_E}{\sqrt{K_0 \alpha}} \right) \|u\| \|v_h\|_E + \left( \max_{E \in \mathcal{T}_h} \frac{b_E}{\sqrt{K_0 \alpha}} \right) \left( \sum_{E \in \mathcal{T}_h} \left( \frac{1}{\tau_E} \right) \frac{b_E^2}{K_0 \alpha} \|u\|_E^2 \right)^{\frac{1}{2}} \|v_h\|_E.
\]
(34)
Now, we estimate,
\[
|d_h(u, v_h)| \leq \sum_{E \in \mathcal{T}_h} (\tau_E \|\nabla \cdot K \Pi_{k-1}^0 \nabla u\|_E + \tau_E \|\alpha \Pi_{k-1}^0 u\|_E)|b \cdot \Pi_{k-1}^0 \nabla v_h|_E
\leq \sum_{E \in \mathcal{T}_h} (\tau_E \|\nabla \cdot K \Pi_{k-1}^0 \nabla u\|_E \|b \cdot \Pi_{k-1}^0 \nabla v_h\|_E) + (\tau_E \|\alpha \Pi_{k-1}^0 u\|_E \|b \cdot \Pi_{k-1}^0 \nabla v_h\|_E)
\]
Using the inverse inequality (20) and the assumptions in (G1), we get
\[
\leq \sum_{E \in \mathcal{T}_h} \left( \frac{\rho b_E}{2\sqrt{K_0 \alpha}} \right) \left( K \|\nabla u\|_E \|\nabla v_h\|_E + \sqrt{\alpha} \|u\|_E \sqrt{K} \|v_h\|_E \right)
\leq \max_{E \in \mathcal{T}_h} \left( \frac{\rho b_E}{\sqrt{K_0 \alpha}} \right) \|u\| \|v_h\|_E.
\]
(35)
Let $B := \max_{E \in \mathcal{T}_h} b_E$ and $C := \frac{1}{\sqrt{K_0 \alpha}}$. Then, combining the estimates (28), (29), (34) and (35) we obtain the result (26) with

$$C_{us} = \max \left\{ \left[ (1 + \beta')(1 + \rho B^2 c^2) + (1 + \eta') + \beta C + \rho BC \right] : \max(1, \sqrt{\beta} B C) \right\}.$$

Hence the lemma is proved. \hfill \square

6. VEM-SUPG with shock-capturing

In this section we formulate the shock-capturing technique for the VEM discretization of our model problem (2). We add a nonlinear artificial diffusion term along the crosswind direction in the VEM-SUPG discrete formulation provided in section 4 and discuss the results concerning the proof of existence of solution.

Consider the following term,

$$T_{sc}(w; u, v) = \sum_{E \in \mathcal{T}_h} (\delta_E(w) N_{sc} \Pi_{k-1} \nabla u, \Pi_{k-1} \nabla v)_E. \quad (36)$$

where $\delta_E(w)$ is chosen satisfying the following condition:

(G2) We suppose that $\delta_E(w)$ depends continuously on $w$ and

$$0 \leq \delta_E(w) \leq M_E(h_E) \quad \text{with} \quad \lim_{h \to 0} M_E(h) = 0 \quad (37)$$

and $N_{sc}$ is a symmetric positive definite matrix function chosen such that $\| (N_{sc})_{ij} \|_{L^\infty(\Omega)} \leq 1$.

We consider the following shock-capturing formulation, find $u_h \in V_h^k$ such that,

$$B_{us}(u_h, v_h) + T_{sc}(u_h, u_h, v_h) = F_{us}(v_h) \quad \forall v_h \in V_h^k. \quad (38)$$

In this paper we take two choices for $\delta_E$ and $N_{sc}$ considering the isotropic and anisotropic diffusion.

Case I : Anisotropic diffusion

$$\delta_E(w) = \frac{\sigma_E(w) \| L_{sc}(w) - f \|_E}{\kappa + \| \nabla \Pi_k w \|_E}, \quad N_{sc} := \begin{cases} I - \frac{b \otimes b}{|b|^2}, & b \neq 0 \\ 0, & b = 0 \end{cases} \quad (39)$$

Case II : Isotropic diffusion

$$\delta_E(w) = \frac{\sigma_E(w) \| L_{sc}(w) - f \|_E^2}{\kappa + \left( \| \Pi_k w \|_E^2 + \| \nabla \Pi_k w \|_E^2 \right)^{\frac{1}{2}}} \quad N_{sc} := I \quad (40)$$

where

$$L_{sc}(w) := -\nabla \cdot (K_0 \Pi_{k-1} \nabla w) + b \cdot \Pi_{k-1} \nabla w + \alpha \Pi_k w \quad (41)$$

and $\sigma_E(w) \geq 0$, $\kappa \geq 0$ are chosen such that $\delta_E$ satisfies (37).

We make a particular choice for $\sigma_E$ as follows,

$$\sigma_E(w) := l_0 h E \max \left\{ 0, \beta - \frac{2K E}{h E R_E(w)} \right\}. \quad (42)$$

where

$$R_E(w) := \frac{\| L_{sc}(w) - f \|_E}{\kappa + \left( \| \Pi_k w \|_E^2 + \| \nabla \Pi_k w \|_E^2 \right)^{\frac{1}{2}}}. \quad (43)$$

$l_0$, $\kappa$ and $\beta$ are positive constants less than 1.

Remark 1. The effect of $\delta_E(w)$ becomes significant only when the residual of VEM-SUPG formulation is very large.
Remark 2. It is noted that, $\delta_E(u_h)$ depends non-linearly on $u_h$. Thus, the discrete formulation with shock-capturing term in equation (36) reduces to nonlinear system of equations. This increases the computational cost significantly.

Now, we proceed to prove the existence of a solution for the equation (38) by the following theorem.

**Theorem 6.1.** The shock-capturing scheme (38) has at least one solution $u_h \in V_h$ satisfying the condition,

$$\|u_h\|^2 + T_{sc}(u_h; u_h, u_h) \leq C \|f\|^2$$

with the dual norm $\|f\|^* := \sup_{v_h \in V_h} \frac{F_{sc}(v_h)}{\|v_h\|}$.

**Proof.** We use a variant of Brouwer’s fixed point theorem (see [32], II, Lemma 1.4) to show the existence of a solution.

For this, let us define an inner product on $V_h$ as $\langle v_h, v_h \rangle := \langle \nabla v_h, \nabla v_h \rangle$ and let $P : V_h \rightarrow V_h$ be an operator, such that,

$$\langle Pu_h, v_h \rangle = \langle \nabla Pu_h, \nabla v_h \rangle = B_{sc}(u_h, v_h) + T_{sc}(u_h; u_h, v_h) - F_{sc}(v_h).$$

Using lemma (5.1) and Young’s inequality we get,

$$\langle Pu_h, v_h \rangle \geq T_{sc}(v_h; v_h, v_h) + B_{sc}(v_h, v_h) - F_{sc}(v_h)$$

$$\geq T_{sc}(v_h; v_h, v_h) + C_{sc} \|v_h\|^2 - \|f\| \|v_h\|$$

$$\geq T_{sc}(v_h; v_h, v_h) + \frac{C_{sc}}{2} \|v_h\|^2 - \frac{1}{2C_{sc}} \|f\|^2.$$  

We conclude that $\langle Pu_h, v_h \rangle > 0$ for all $v_h \in V_h$ with $\|v_h\| \geq \|\nabla v_h\|^2 > \tilde{C} C_{sc} \|f\|$, for some constant $\tilde{C} > 0$. Clearly, $F_{sc}$ is continuous. Also lemma 5.2 and the assumption (G2) imply the continuity of $B_{sc}$ and $T_{sc}$. Thus, we get that the operator $P$ is continuous. Then, using a variant of Brouwer’s fixed point theorem (see [32]) we get atleast one solution $u_h$ satisfying $P(u_h) = 0$. This inturn imposes the existence of a solution of the discrete problem and finally the estimate (44) is obtained by using $P(u_h) = 0$ in the inequality (46).

**Remark 3.** We have shown the existence of atleast one solution for the shock-capturing technique, but unfortunately the uniqueness result is still open. If we assume that $\delta_E$ is Lipschitz continuous then using Banach fixed point theorem we can prove the uniqueness. But this condition restricts the choice of $\delta_E$ for practical applications. On the otherhand, using the result of Schauder fixed point theorem [33] a corresponding result using Brouwer’s fixed point theorem with specific assumptions on $\delta_E$ the uniqueness result can be proved. Once again this imposes severe restrictions on $\delta_E$.

**Remark 4.** We would also like to mention that the choice of $\delta_E$ given in equation (39) does not satisfy the Lipschitz continuity.

7. Numerical experiments

In this section we illustrate the performance of shock-capturing technique with an example. We would like to make the following choice for the stabilization parameter $\tau_E$ proposed in [34],

$$\tau_E = \min \left\{ \frac{h_E}{|\mathbf{b}|}, \frac{1}{|\mathbf{a}|}, \frac{h_E^2}{K} \right\}.$$  

The reduced nonlinear algebraic system of equations can be solved by the application of inexact Newton-GMRES algorithm [35]. Since this approach is very expensive we consider solving the scheme (38) using the following simple iterative technique,

$$n \in \mathbb{N}, \quad B_{sc}(U^{n+1}, v) + T_{sc}(U^n; U^{n+1}, v) = F_{sc}(v) \quad \forall v \in V_h$$

The well-posedness of this iterative technique is discussed in [7].
For our numerical experiment we consider four different type of meshes namely, smoothed Voronoi, nonconvex polygons, regular hexagons and distorted hexagons respectively shown in figure 1. We use VEM of order $k = 1$ and $k = 2$ for our computations.

![Polygonal meshes](image)

**Figure 1: Polygonal meshes**

### 7.1. Example 1

We consider a stationary linear convection-diffusion problem. Let $\Omega = (0,1)^2$, $K = 10^{-6}$, $b = (-y,x)$, $\alpha = 1$, and $f \equiv 0$, in equation (2). We specify the discontinuous boundary conditions as follows: the Dirichlet condition $u(x,y) = 1$ for $x \in (\frac{1}{3}, \frac{2}{3})$, $y = 0$ and $u(x,y) = 0$ on the remaining parts of lower boundary as well as on the right and upper boundary; assume the homogeneous Neumann condition on the left boundary, i.e. $\frac{\partial u(x,y)}{\partial n} = 0$ for $x = 0$, $y \in (0,1)$, where $n$ is the unit outerward normal. The discontinuous profile specified on the boundary is carried over to the characteristic curves and the solution develops interior layers.

To present our numerical results we denote SUPG with shock-capturing and without shock-capturing as SUPG-SC and SUPG respectively. We choose the following values in the equation (39) as $l_0 = 0.2$, $\beta = 0.7$ and $\kappa = 10^{-4}$. The iterative scheme (48) is used for solving the nonlinear system with tolerance $10^{-7}$. We note that the solution has two interior layers that are efficiently damped by the VEM-SUPG with shock capturing method on both the orders $k = 1$ and $k = 2$. The cross-section plots of the solution at the left outflow boundary for both SUPG and SUPG-SC are shown in figures 2-6.

![Cross-section plots](image)

**Figure 2: Smoothed Voronoi: The cross-section plots of the solution at the left outflow boundary.**

In order to show the removal of spurious oscillations along the crosswind direction we provide the numerical solutions with and without shock-capturing in figure 3.
Figure 3: Surface plots of numerical solution of SUPG (left) and SUPG-SC (right) for $k = 2$ on nonconvex polygons with $h=1/80$.

Figure 4: Nonconvex polygons: The cross-section plots of the solution at the left outflow boundary.

Figure 5: Regular hexagons: The cross-section plots of the solution at the left outflow boundary.

Figure 6: Distorted hexagons: The cross-section plots of the solution at the left outflow boundary.
To compare the results with finite element method we consider the mesh of structured triangles shown in figure 7. We show the cross-section plots of both FEM and VEM at the left outflow boundary for order $k = 1$ in figure 8. We can observe that VEM performs similar to FEM in reducing the oscillations along the sharp layers.

Figure 7: Sample structured triangle mesh.

7.2. Example 2
In this example we consider the problem (see Example 4.1) discussed in [29]. Let $\Omega = (0,1)^2$, $K = 10^{-6}$, $b = \frac{1}{\sqrt{5}} (1, 2)^T$ with added nonlinear reaction term $u^4$. We consider the exact solution as $u(x) = \frac{1}{2} \left(1 - \tanh \left(\frac{2x_1 - x_2 - 1}{\sqrt{K}}\right)\right)$. This solution exhibits an interior layer with thickness $O(\sqrt{K} |\ln K|)$. We use Dirichlet boundary values prescribed by the solution. In order to make a comparison with finite element method we have considered regular triangular meshes for the numerical computation. We present a result (see table 1) depicting the errors evaluated in $||·||$ along with roc i.e., the rate of convergence. From this we observe that our proposed method performs better than the method discussed in the paper [29].

Table 1: Comparison of errors in $||·||$ and the rate of convergence (roc).

| Order $k = 2$ | SC-CD (Table 1,[29]) | SUPG-SC (VEM) |
|--------------|----------------------|----------------|
| $h$          | $||·||$              | roc | $||·||$ | roc |
| $\frac{1}{4}$| 1.70e-01             | *   | 1.43e-01| *   |
| $\frac{1}{8}$| 1.32e-01             | 0.36| 1.02e-01| 0.48|
| $\frac{1}{16}$| 1.13e-01             | 0.22| 7.14e-02| 0.52|
| $\frac{1}{32}$| 9.05e-02             | 0.32| 5.33e-02| 0.42|
| $\frac{1}{64}$| 7.05e-02             | 0.36| 3.85e-02| 0.47|
| $\frac{1}{128}$| 5.37e-02             | 0.39| 2.67e-02| 0.53|

8. Concluding remarks
In this paper we have proposed VEM-SUPG formulation with shock-capturing technique. The shock-capturing term is added to the VEM-SUPG formulation appropriately to make the discrete
scheme VEM computable. In order to overcome the cost for solving nonlinear system of algebraic equations we have used the simple iterative technique. We have observed that the oscillations gets reduced for the shock-capturing technique for the meshes considered and in particular this is much more evident for the VEM of order $k = 2$. However, the nonlinear transport problems occur more often in scientific/engineering applications where the shock-capturing technique would be interesting to study which will be considered in our future work.

References

[1] Brooks A and Hughes T 1982 Comput. Methods Appl. Mech. Eng. 32 199–259
[2] Burman E 2010 Comput. Methods Appl. Mech. Eng. 199 1114–1123
[3] Braack M and Burman E 2006 SIAM J. Numer. Anal. 43 2544–2566
[4] Matthies G, Skryptycz P and Tobiska L 2007 Math. Model. Numer. Anal. 41 713–742
[5] Burman E and Hansbo P 2004 Comput. Methods Appl. Mech. Eng. 193 1437–1453
[6] Codina R 1993 Comp. Meth. Appl. Mech. Engrg. 110 325–342
[7] Knopp T, Lube G and Rapin G 2002 Comput. Methods Appl. Mech. Engrg. 191 2997–3013
[8] John V and Schmeyer E 2007 Int. J. Comput. Sci. Math. 1 245–258
[9] John V and Knobloch P 2007 Comput. Methods Appl. Mech. Engrg. 196 2197–2215
[10] Beirao da Veiga L 2010 M2AN: Math. Model. Numer. Anal. 2010 2331–250
[11] Antonietti P F, Bigoni N and Verani M 2013 J. Sci. Comput. 56 14–27
[12] Beirao Da Veiga L, Brezzi F and Marini L D 2013 SIAM J Numer Anal 51 794–812
[13] Beirao Da Veiga L, Brezzi F, Marini L D and Russo A 2016 Math Models Methods Appl Sci 26 729–750
[14] Cangiani A, Manzini G and Sutton O J 2017 IMA Journal of Numerical Analysis 37 1317–1354
[15] Vacca G and Beirao Da Veiga L 2015 Numer Methods Partial Differ Eqs 31 2190–2194
[16] Beirao da Veiga L, Brezzi F and Marini L 2016 Math. Mod. Numer. Anal. 50 727–747
[17] Beirao da Veiga L, Brezzi F, Marini L and Russo A 2017 Numer. Math. 136 725–763
[18] Beirao da Veiga L, Lovadina C and Vacca G 2017 Math. Mod. Numer. Anal. 51 509–535
[19] Beirao da Veiga L and Marini L D 2018 Comp. Meth. Appl. Mech. Engrg. 327 173–195
[20] Beirao da Veiga L and Marini L D 2015 Math. Mod. Numer. Anal. 49 577–599
[21] Benedetto M F, Berrone S, Pieraccini S and Scialo S 2014 Comput. Meth. Appl. Mech. Engrg. 280 135–156
[22] Cangiani A, Manzini G and Sutton O 2014 The conforming virtual element method for the convection-diffusion-reaction equation with variable coefficients Report Los Alamos National Laboratory
[23] Benedetto M F, Berrone S, Borio A, Pieraccini S and Scialo S 2016 Comput Methods Appl Mech Eng 311 18–40
[24] Beirao da Veiga L and Marini L D 2017 Comput. Methods Appl. Mech. Eng 209-212 184–196
[25] Brenner S C and Ridgway Scott L 2002 The Mathematical Theory of Finite Element Methods (Springer-Verlag, New York)