Numer. Math. (1999) 83: 161–175

Numerische
Mathematik

Finite volume element methods for non-definite problems

Ilya D. Mishev[⋆]

Mobil E&P Technical Center, 13777 Midway Rd, Dallas, TX 75244-4390, USA

Received August 27, 1997 / Revised version received May 12, 1998

Summary. The error estimates for finite volume element method applied to 2 and 3-D non-definite problems are derived. A simple upwind scheme is proven to be unconditionally stable and first order accurate.

Mathematics Subject Classification (1991): 65N30

1. Introduction

The purpose of this note is three fold. We would like to extend the results due to Bank and Rose [4], Hackbusch [9], Cai and McCormick [6,7] and Jianguo and Shitong [11] to 3-D problems and provide a theory for non-definite equations. Finally we give a more flexible way to obtain a priori estimates with the flavor of the first Fix lemma in the finite element theory and generalize the technique used by Cai [6] to analyze the effects of numerical integration. We will demonstrate this approach on a simple upwind scheme, although the technique can handle more sophisticated upwind strategies (see [2] for example).

We consider the following boundary value problem:

$$\nabla \cdot (-A(x)\nabla u + \boldsymbol{b}(x)u) + c(x)u = f(x) \quad \text{in } \Omega, \tag{1a}$$

$$u(\mathrm{x}) = 0 \qquad \text{on } \partial\Omega, \tag{1b}$$

where Ω is a open subset of $\mathbb{R}^d$, $d = 2$ or 3. We refer for the extensive discussion of solvability of the problem (1) to the monograph by Ladyzhenskaya and Ural'tseva [12].

⋆ The author was partially supported by funding from DOE grant #DE-FG05-92ER25143 and ARO grant # DAAH04-96-1-0069

Our approach is based on the generalization of Lax–Milgram lemma due to Nečas [13] and modified by Babuška and Aziz [3]. First we introduce some notations.

Let $\mathcal{U}$ and $\mathcal{V}$ be two real Hilbert spaces equipped with the norms $\|.\|_{\mathcal{U}}$ and $\|.\|_{\mathcal{V}}$ respectively, and let $\mathcal{A} : \mathcal{U} \times \mathcal{V} \to \mathbb{R}$ be a bilinear form. We define the following variational problem:

Find an element $u \in \mathcal{U}$ such that

$$\mathcal{A}(u, v) = f(v) \qquad \forall\, v \in \mathcal{V}. \tag{2}$$

Theorem 1 (Babuška and Aziz [3]). *Assume that there exist positive constants C and α such that the bilinear form $\mathcal{A} : \mathcal{U} \times \mathcal{V} \to R$ satisfies*

$$|\mathcal{A}(u, v)| \le C\|u\|_{\mathcal{U}}\|v\|_{\mathcal{V}} \quad \forall\, u \in \mathcal{U},\, \forall\, v \in \mathcal{V}, \tag{3a}$$

$$\sup_{v\in\mathcal{V}} \frac{|\mathcal{A}(u, v)|}{\|v\|_{\mathcal{V}}} \ge \alpha\|u\|_{\mathcal{U}} \quad \forall\, u \in \mathcal{U},\, v \ne 0 \tag{3b}$$

$$\sup_{u\in\mathcal{U}} |\mathcal{A}(u, v)| > 0 \qquad \forall\, v \in \mathcal{V},\, v \ne 0, \tag{3c}$$

and that $f(.) : \mathcal{V} \to \mathbb{R}$ is a continuous linear form. Then the variational problem (2) *has one and only one solution and the following stability estimate holds:*

$$\|u\|_{\mathcal{U}} \le \frac{1}{\alpha}\|f\|_{\mathcal{V}'}.$$

We use the standard notation for Sobolev spaces [1]. Let $\mathcal{U} = \mathcal{V} = H_0^1(\Omega)$, $\mathcal{V}' = H^{-1}(\Omega)$, let the bilinear form $\mathcal{A}$ be defined by

$$\mathcal{A}(u, v) = \mathcal{A}^{(2)}(u, v) + \mathcal{A}^{(1)}(u, v) + \mathcal{A}^{(0)}(u, v), \tag{4a}$$

$$\mathcal{A}^{(2)}(u, v) = \int_\Omega (A\nabla u, \nabla v)\, dx, \tag{4b}$$

$$\mathcal{A}^{(1)}(u, v) = -\int_\Omega (\boldsymbol{b}, \nabla v) u\, dx, \tag{4c}$$

$$\mathcal{A}^{(0)}(u, v) = \int_\Omega cuv\, dx, \tag{4d}$$

and let the linear form be given by

$$f(v) = \int_\Omega fv\, dx.$$

Suppose that the boundary value problem (1) poses a unique solution. Then $\mathcal{A}(.,.)$ defined by (4) satisfies the conditions (3) (see [3] for a proof).

Note that the solution u of (1a) satisfies the "weak" form:

$$\int_{\partial V_i} (-A\nabla u + \boldsymbol{b}u, \boldsymbol{n})\, ds + \int_{V_i} cu\, dx = \int_{V_i} f\, dx, \tag{5}$$

where V_i is a given control volume. This observation provides the motivation to reformulate (5) as a Petrov–Galerkin method on given finite dimensional spaces.

Let $\mathcal{V}^h$ be a finite dimensional space of piecewise constants defined on the control volumes V_i and denote $v_i = v(x_i)$ for $v \in \mathcal{V}^h$ and $x_i \in \omega$. Let $\mathcal{U}^h$ be a piecewise polynomial subspace of $\mathcal{U}$. Consider the problem:
Find $u_h \in \mathcal{U}^h$ such that

$$\mathcal{B}(u_h, v) = f(v) \qquad \forall\, v \in \mathcal{V}^h, \tag{6}$$

where

$$f(v) = \sum_{x_i \in \omega} \int_{V_i} f\, dx\, v_i,$$

and $\mathcal{B}(.,.)$ is a bilinear form defined in $\mathcal{U}^h \times \mathcal{V}^h$

$$\mathcal{B}(u_h, v) = \mathcal{B}^{(2)}(u_h, v) + \mathcal{B}^{(1)}(u_h, v) + \mathcal{B}^{(0)}(u_h, v), \tag{7a}$$

$$\mathcal{B}^{(2)}(u_h, v) = -\sum_{x_i \in \omega} \int_{\partial V_i} (A\nabla u_h, \boldsymbol{n})\, ds\, v_i, \tag{7b}$$

$$\mathcal{B}^{(1)}(u_h, v) = \sum_{x_i \in \omega} \int_{\partial V_i} (\boldsymbol{b}, \boldsymbol{n}) u_h\, ds\, v_i, \tag{7c}$$

$$\mathcal{B}^{(0)}(u_h, v) = \sum_{x_i \in \omega} \int_{V_i} cu_h\, dx\, v_i. \tag{7d}$$

We eventually will replace the bilinear form $\mathcal{B}(.,.)$ with a certain approximation $\mathcal{B}_h(.,.)$, i.e., we solve the discrete problem:

Find $u_h \in \mathcal{U}^h$ such that

$$\mathcal{B}_h(u_h, v) = f(v) \qquad \forall\, v \in \mathcal{V}^h. \tag{8}$$

We describe the control volumes V_i, piecewise polynomial spaces $\mathcal{U}^h$ and $\mathcal{V}^h$, and the corresponding norms $\|.\|_{1,1}$ and $\|.\|_{1,\mathrm{c}}$ in the next section.

We use Theorem 1 to prove uniqueness and existence of the solution of (8). The second step is to show that the following a priori estimate holds

$$|I_h^1 u - u_h|_{1,1} \le C(\|\eta\|_{*,1} + \|\mu\|_{*,1} + \|\zeta\|_{**,1}),$$

where I_h^1 is a linear interpolant and $\|\eta\|_{*,1}$ is the error due to the approximation of the diffusion term (second derivatives), $\|\mu\|_{*,1}$ - convection term (first derivatives) and $\|\zeta\|_{**,1}$ - reaction term (zero derivatives). Finally we estimate these terms and obtain the bound for the error of approximation.

2. Grids, control volumes and discrete norms

We consider a family of triangulations $\mathcal{F}_h$ of Ω into finite elements K regular in sense of Ciarlet [8, p. 132]. We use the standard symbols

$$h_i = \operatorname{diam}(K_i)\,, \qquad h = \max_i h_i.$$

Here we describe a general way to construct grids starting from a finite element triangulation. The vertices of the finite element triangulation uniquely determine the grid, which we call the primary grid $\overline{\omega}$,

$$\overline{\omega} = \left\{ x_i \in \overline{\Omega} : x_i \text{ is a vertex in a finite element } K \right\} ,$$

split into the set of interior grid points ω and the boundary grid points γ;

$$\omega = \overline{\omega} \cap \Omega, \quad \gamma = \overline{\omega} \backslash \omega.$$

We define the secondary grid ω_S in the following way. Choose one interior point $S_K \in \overset{\circ}{K}$ in every finite element $K \in \mathcal{F}_h$. Then

$$\omega_S = \{S_K : K \in \mathcal{F}_h\}.$$

Given a primary grid vertex x_i we define by $\Pi(i)$ the index set of all neighbors of x_i in ω, i.e.,

$$\Pi(i) = \{j : \text{ there is an edge between } x_i \text{ and } x_j \text{ in } \mathcal{F}_h\}.$$

Consider a particular finite element K with vertices $x_{i_1}, \ldots , x_{i_k}$ and let I_K be the index set $\{i_1, \ldots , i_k\}$. Denote by $\{Z_{ij}\}_{i,j \in I_K}$ the edges and by $\{Z_{j_1 \ldots j_l}\}_{j_1,\ldots,j_l \in I_K}$ the faces of a given finite element (the polygons with vertices $x_{j_1}, \ldots , x_{j_l} \in K$). To describe vertex–centered control volumes we select one interior point on each face of every finite element K_i, $M_{j_1 \ldots j_l} \in Z_{j_1 \ldots j_l}$. The points on the edges are selected in the same manner. Connect a given point from the secondary grid S_{K_i}, $K_i \in \mathcal{F}_h$ with $M_{j_1 j_2}$, $j_1, j_2 \in I_{K_i}$ and $M_{i_1 \ldots i_l}$, $i_1, \ldots , i_l \in I_{K_i}$. These lines and the planes that they span form a polygonal (polyhedral) domain around each vertex of the primary grid and are called vertex–centered control volumes. There is one–to–one correspondence of nodes in primary grid with vertex–centered control volumes. If $x_i \in \omega$ we denote the corresponding vertex–centered control volume with V_i and with

$$\gamma_{ij} = V_i \cap V_j\,, \quad j \in \Pi(i)$$

the face between them.

To specify a particular primary and secondary grid we have to choose the finite elements, secondary grid points and the points $M_{j_1 j_2}$ on the edges, $M_{j_1 \ldots j_l}$ on the faces.

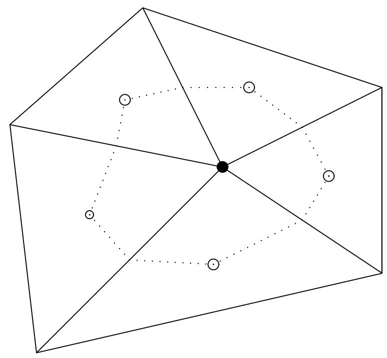

Fig. 1. Vertex–centered control volume

We choose finite elements to be triangles in 2–D and tetrahedra in 3–D. The secondary mesh consists of the barycenters (centers of mass) of the finite elements and the points M are barycenters of the edges and faces, correspondingly. A specific 2–D example is shown on Fig. 1, where the primary node is displayed with a filled circle and the secondary nodes are shown with empty circles. The control volume corresponding to the primary node is depicted by a dotted line. Note that in general γ_{ij} is not a straight line.

We show how a 3-D finite element (tetrahedron) is split by the control volumes on Fig. 2.

The theory presented in Sects. 3 and 4 works also for more general positions of the points of the secondary grid and the points M, but in practice the barycenters are the most frequently used.

We introduce a piecewise linear finite element space for the simplex triangulation

$$\mathcal{U}^h = \{v \in C^0(\Omega) : \ v_{|K} \text{ is linear for all } K \in \mathcal{F}_h, \ v_{|\partial\Omega} = 0\},$$

where $v_{|K}$ is the restriction of v to K. Functions defined for $x \in \omega$ are called grid functions and the space of such functions is $G(\omega)$. To emphasize their dependence of the triangulation we use the subscript h. Denote by χ_i the characteristic functions that corresponds to the vertex–centered control volume V_i and with $\mathcal{V}^h$ the space spanned by $\{\chi_i\}_{x_i \in \omega}$. Let $\{\varphi_i\}_{x_i \in \omega}$ be the basis of $\mathcal{U}^h$. We define the linear interpolant $I_h^1 : G(\omega) \to \mathcal{U}^h$ and the

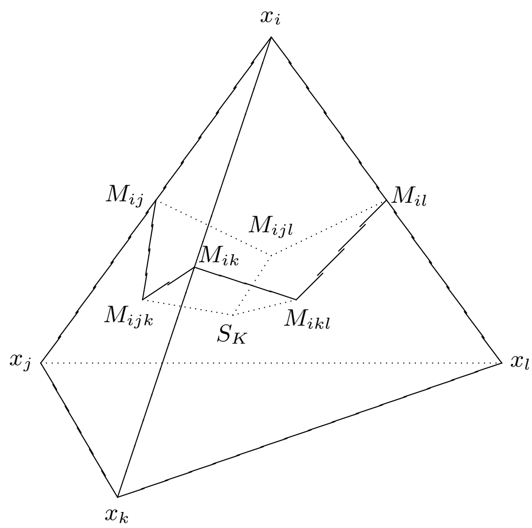

Fig. 2. Finite element K

"box" interpolant (constant interpolant) $I_h^{\mathrm{c}} : G(\omega) \to \mathcal{V}^h$ by

$$I_h^{\mathrm{l}} u_h(x) = \sum_{x_i \in \omega} u_h(x_i) \varphi_i(x), \quad I_h^{\mathrm{c}} u_h(x) = \sum_{x_i \in \omega} u_h(x_i) \chi_i(x). \tag{9}$$

It is clear how to modify (9) to get the mappings $\bar{I}_h^{\mathrm{l}} : \mathcal{V}^h \to \mathcal{U}^h$, $\bar{I}_h^{\mathrm{c}} : \mathcal{U}^h \to \mathcal{V}^h$ and $\tilde{I}_h^{\mathrm{l}} : H^s(\Omega) \to \mathcal{U}^h$, $\tilde{I}_h^{\mathrm{c}} : H^s(\Omega) \to \mathcal{V}^h$ for $s > 3/2$. When there is no danger of ambiguity we will skip the bars and tildes.

We define discrete inner products and norms in the following way:

$$(u_h, v_h)_l = (I_h^{\mathrm{l}} u_h, I_h^{\mathrm{l}} v_h)_{L^2}, \qquad \|u_h\|_{0,\mathrm{l}}^2 = (u_h, u_h)_l \,,$$

$$|u_h|_{1,\mathrm{l}} = |I_h^{\mathrm{l}} u_h|_{1,\Omega}, \qquad \|u_h\|_{1,\mathrm{l}}^2 = \|u_h\|_{0,\mathrm{l}}^2 + |u_h|_{1,\mathrm{l}}^2.$$

We also use the norms and seminorms associated with the constant interpolant I_h^{c}:

$$\|u_h\|_{0,\mathrm{c}}^2 = \sum_{x_i \in \omega} m(V_i) u_h^2(x_i),$$
$$|u_h|_{1,\mathrm{c}}^2 = \frac{1}{2} \sum_{x_i \in \omega} m(V_i) \sum_{j \in \Pi(i)} \left(\frac{u_h(x_i) - u_h(x_j)}{d(x_i, x_j)} \right)^2,$$

where $d(x, y)$ is the Euclidean distance between x and y.

The following result is well known (see for example [14] for the 2–D case and regular geometry, [4] for the 2–D case and general geometry, and [10] for the discussion of the finite difference case).

Lemma 1. *Assume the triangulations $\mathcal{F}_h$ are regular and the secondary mesh consists of the barycenters of the finite elements. Then the seminorms $|.|_{1,\mathrm{l}}$, $|.|_{1,\mathrm{c}}$ and the norms $\|.\|_{0,\mathrm{l}}$, $\|.\|_{0,\mathrm{c}}$ are equivalent on $G(\omega)$, i.e., there exist positive constants C_1, C_2, C_3 and C_4, independent of h, such that for any $u_h \in G(\omega)$*

$$C_1|u_h|_{1,\mathrm{l}} \leq |u_h|_{1,\mathrm{c}} \leq C_2|u_h|_{1,\mathrm{l}}, \tag{10}$$

$$C_3\|u_h\|_{0,\mathrm{l}} \leq \|u_h\|_{0,\mathrm{c}} \leq C_4\|u_h\|_{0,\mathrm{l}}. \tag{11}$$

Remark 1. If the secondary grid is arbitrary the norms $\|.\|_{0,\mathrm{l}}$ and $\|.\|_{0,\mathrm{c}}$ are not equivalent. This is seen by the following simple example. Consider one control volume V_i, such that $m(V_i) \to 0$, i.e., the secondary points around x_i go to x_i. Pick a function $u_h = (0, \dots, 1, \dots, 0)$, where the only nonzero element is on the i^{th} position. Then $\|u_h\|_{0,\mathrm{c}} \to 0$, but $\|u_h\|_{0,\mathrm{l}}$ is bounded from below.

The seminorms $|.|_{1,\mathrm{c}}$ and $|.|_{1,\mathrm{l}}$ are equivalent without any restriction on the secondary grid.

3. Diffusion dominated problem

First we elaborate the finite volume element theory for the compact perturbation of a symmetric problem. In this case we define $\mathcal{B}_h(.,.)$ by

$$\mathcal{B}_h^{(2)}(u,v) = \mathcal{B}^{(2)}(u,v), \quad \mathcal{B}_h^{(1)}(u,v) = \mathcal{B}^{(1)}(u,v),$$
$$\mathcal{B}_h^{(0)}(u,v) = \mathcal{B}^{(0)}(u,v).$$

We prove (3) via comparing with the bilinear forms for the finite element method (4b), (4c) and (4d). The first result is due to Jianguo and Shitong [11].

Lemma 2. *For every $u,\ v \in \mathcal{U}^h$ the following estimate holds:*

$$|\mathcal{B}^{(2)}(u, I_h^{\mathrm{c}}v) - \mathcal{A}^{(2)}(u,v)| \leq Ch\|A\|_{1,\infty.\Omega}|u|_{1,\Omega}|v|_{1,\Omega}.$$

We compare $\mathcal{B}^{(1)}(u, I_h^{\mathrm{c}}v)$ and $\mathcal{A}^{(1)}(u,v)$ in the following lemma.

Lemma 3. *For every $u,\ v \in \mathcal{U}^h$ the following estimate holds:*

$$|\mathcal{B}^{(1)}(u, I_h^{\mathrm{c}}v) - \mathcal{A}^{(1)}(u,v)| \leq Ch\|\boldsymbol{b}\|_{1,\infty,\Omega}|u|_{1,\Omega}|v|_{1,\Omega}.$$

Proof. Consider the contribution of one particular element K in the computation of $B_h^{(1)}(u, I_h^{\rm c} v)$ corresponding to the $i^{\rm th}$ node

$$\begin{aligned}\int_{\partial V_i \cap K} (\boldsymbol{b}.\boldsymbol{n}) u \, ds \, v_i &= \left[\int_{(\partial V_i \cap K) \cup M_i} (\boldsymbol{b}.\boldsymbol{n}) u \, ds - \int_{M_i} (\boldsymbol{b}.\boldsymbol{n}) u \, ds \right] v_i \\ &= \int_{V_i \cap K} \text{div}(\boldsymbol{b} u) \, \text{dx} \, v_i - \int_{M_i} (\boldsymbol{b}.\boldsymbol{n}) u \, ds \, v_i \\ &= \int_K \text{div}(\boldsymbol{b} u) v_i \chi_i \, \text{dx} - \int_{\partial K} (\boldsymbol{b}.\boldsymbol{n}) u v_i \chi_i \, ds,\end{aligned}$$

where $M_i = \partial K \cap V_i$. Then, the contribution of the element K is equal to

$$\mathcal{B}^{(1)}(u, I_h^{\rm c} v)_{|K} = \int_K \text{div}(\boldsymbol{b} u) I_h^{\rm c} v \, \text{dx} - \int_{\partial K} (\boldsymbol{b}.\boldsymbol{n}) u I_h^{\rm c} v \, ds$$

and

$$\mathcal{B}^{(1)}(u, I_h^{\rm c} v) = \sum_{K \in T_h} \int_K \text{div}(\boldsymbol{b} u) I_h^{\rm c} v \, dx.$$

because the surface integrals vanish. Therefore,

$$\begin{aligned}|\mathcal{B}^{(1)}(u, I_h^{\rm c} v) - \mathcal{A}^{(1)}(u, v)| &\le \sum_{K \in T_h} \left| \int_K \text{div}(\boldsymbol{b} u)(I_h^{\rm c} v - v) \, dx \right| \\ &\le \|\boldsymbol{b}\|_{1,\infty,\Omega} \sum_{K \in T_h} |u|_{1,K} \|v - I_h^{\rm c} v\|_{0,K} \\ &\le Ch \|\boldsymbol{b}\|_{1,\infty,\Omega} |u|_{1,\Omega} |v|_{1,\Omega}.\end{aligned}$$

□

Finally, the difference between $\mathcal{B}^{(0)}(u, I_h^{\rm c} v)$ and $\mathcal{A}^{(0)}(u, v)$ is estimated in the lemma below.

Lemma 4. *For every* $u,\ v \in \mathcal{U}^h$ *the following estimate holds:*

$$|\mathcal{B}^{(0)}(u, I_h^{\rm c} v) - \mathcal{A}^{(0)}(u, v)| \le Ch \|c\|_{0,\Omega} \|u\|_{0,\Omega} |v|_{1,\Omega}.$$

Proof. The estimate follows from the chain of inequalities:

$$\begin{aligned}\left| \int_\Omega cuv \, dx - \sum_{x_i \in \omega} \int_{V_i} cu \, dx \, v_i \right| &= \left| \sum_{x_i \in \omega} \left[\int_{V_i} cu(v - I_h^{\rm c} v) \, dx \right] \right| \\ &\le Ch \|c\|_{0,\Omega} \|u\|_{0,\Omega} |v|_{1,\Omega}.\end{aligned}$$

□

The stability of problem (8) is established in the following theorem. We use that for sufficiently small h the finite element approximation of (1) is well defined, i.e., Theorem 1 holds for the bilinear form $\mathcal{A}(.,.)$ (4) and $\mathcal{U} = \mathcal{V} = \mathcal{U}^h$. (See [3] and Schatz [15] for another approach).

Theorem 2. *There exists h_0 such that for any $h < h_0$ the bilinear form $\mathcal{B}(.,.)$ satisfies* (3) *and the problem* (6) *has one and only one solution and the following stability estimates holds:*

$$|u_h|_{1,1} \le C\|f\|_{-1,\mathrm{c}}.$$

Proof. The continuity of the bilinear form $\mathcal{B}(.,.)$ for sufficiently small h follows from Lemmas 2, 3, and 4:

$$|\mathcal{B}(u, I_h^{\mathrm{c}}v) - \mathcal{A}(u,v)| \le C(a,\boldsymbol{b},c)h|u|_{1,\Omega}|v|_{1,\Omega}, \tag{12}$$

and the continuity of $\mathcal{A}(.,.)$ (3a).

From (12) and the equivalence of the norms (Lemma 1) we get

$$\frac{\mathcal{B}(u, I_h^{\mathrm{c}}v)}{|I_h^{\mathrm{c}}v|_{1,\mathrm{c}}} \ge C_1 \frac{\mathcal{A}(u,v)}{|v|_{1,\Omega}} - C_2 h|u|_{1,\Omega}, \tag{13}$$

therefore

$$\begin{aligned}\sup_{w\in\mathcal{V}^h} \frac{|\mathcal{B}(u,w)|}{|w|_{1,\mathrm{c}}} &\ge \sup_{I_h^{\mathrm{c}}v\in\mathcal{V}^h} \frac{|\mathcal{B}(u, I_h^{\mathrm{c}}v)|}{|I_h^{\mathrm{c}}v|_{1,\mathrm{c}}}\\ &= \sup_{v\in\mathcal{U}^h} \frac{|\mathcal{B}(u, I_h^{\mathrm{c}}v)|}{|I_h^{\mathrm{c}}v|_{1,\mathrm{c}}} \ge \alpha_1|u|_{1,1} \quad \forall\, u \in \mathcal{U}^h.\end{aligned}$$

We prove the condition (3c) for the bilinear form $\mathcal{B}(.,.)$ in the same way as (3b) using the fact that (3c) is equivalent to:

$$\sup_{u\in\mathcal{U}} \frac{|\mathcal{A}(u,v)|}{\|v\|_{\mathcal{V}}} \ge C\|u\|_{\mathcal{U}} \quad \forall\, u \in \mathcal{U} \text{ such that } \|u\|_{\mathcal{U}} \le 1,\ v \ne 0.$$

(See [3]) □

Let u be the solution of (2) with the bilinear forms defined by (4). Define the local truncation error ψ via:

$$(\psi, v) = \mathcal{B}(u,v) - \mathcal{B}_h(I_h^1 u, v)$$

and the components of ψ due to different terms by:

$$\eta_{i,j}(u) = \int_{\gamma_{ij}} (-A(\nabla(u - I_h^1 u), \boldsymbol{n})\, ds, \tag{14a}$$

$$\mu_{i,j}(u) = \int_{\gamma_{ij}} (\boldsymbol{b},\boldsymbol{n})(u - I_h^1 u)\, ds, \tag{14b}$$

$$\zeta_i(u) = \int_{V_i} c(u - I_h^1 u)\, dx. \tag{14c}$$

Note that

$$\mathcal{B}_h(u_h, v) = (f, v) \quad \text{and} \quad \mathcal{B}(u, v) = (f, v),$$

and therefore

$$\mathcal{B}_h(u_h - I_h^1 u, v) = (\psi, v).$$

We prove the a priori estimate in the following lemma.

Lemma 5. *The following a priori estimate holds:*

$$|I_h^1 u - u_h|_{1,w} \le C \left(\|\eta\|_{*,1} + \|\mu\|_{*,1} + \|\zeta\|_{**,1} \right). \tag{15}$$

(The definition of $\|.\|_{*,1}$ *and* $\|.\|_{**,1}$ *will become clear from the proof.)*

Proof.

$$\begin{aligned}
(\psi, v) &= \mathcal{B}(u, v) - \mathcal{B}_h(I_h^1 u, v) \\
&= \sum_{x_i \in \omega} \sum_{j \in \Pi(i)} \int_{\gamma_{ij}} (-A\nabla(I_h^1 u - u) + \boldsymbol{b}(I_h^1 u - u)).\mathbf{n}\, ds\, v_i \\
&\qquad + \sum_{x_i \in \omega} \int_{V_i} c(I_h^1 u - u)\, dx\, v_i \\
&= \left[\sum_{x_i \in \omega} \sum_{j \in \Pi(i)} \eta_{ij}(u) v_i \right] + \left[\sum_{x_i \in \omega} \sum_{j \in \Pi(i)} \mu_{ij}(u) v_i \right] + \left[\sum_{x_i \in \omega} \zeta_i v_i \right] \\
&= I_d + I_c + I_r
\end{aligned}$$

Denote $k_{i,j}^2 = d(x_i, x_j)^2 / m(V_i)$. The term due to the diffusion discretization I_d is estimated as follows:

$$\begin{aligned}
I_d &= \sum_{x_i \in \omega} \sum_{j \in \Pi(i)} \eta_{ij}(u) v_i \\
&= \frac{1}{2} \sum_{x_i \in \omega} \sum_{j \in \Pi(i)} [\eta_{ij}(u) v_i + \eta_{ji}(u) v_j] = \frac{1}{2} \sum_{x_i \in \omega} \sum_{j \in \Pi(i)} \eta_{ij}(u)(v_i - v_j) \\
&\le C \left(\sum_{x_i \in \omega} \sum_{j \in \Pi(i)} k_{i,j}^2 \, |\eta_{i,j}(u)|^2 \right)^{1/2} \\
&\quad \times \left(\sum_{x_i \in \omega} m(V_i) \sum_{j \in \Pi(i)} \left(\frac{v_j - v_i}{d(x_i, x_j)} \right)^2 \right)^{1/2} \\
&\le C \|\eta\|_{*,\omega} |v|_{1,\mathrm{c}}.
\end{aligned}$$

Similarly, we prove the estimate

$$I_c \le C\|\mu\|_{*,\omega}|v|_{1,\mathrm{c}}.$$

Finally, we estimate I_r:

$$\begin{aligned} I_r &= \sum_{x_i\in\omega}\int_{V_i} c(x)(u - I_h^1)\,dx\cdot v_i = \sum_{x_i\in\omega}\zeta_i(u)v_i \\ &\le \left(\sum_{x_i\in\omega}\frac{1}{m(V_i)}|\zeta_i(u)|^2\right)^{1/2}\left(\sum_{x_i\in\omega} m(V_i)v_i^2\right)^{1/2} \\ &\le \|\zeta\|_{**,\omega}|v|_{1,\mathrm{c}}. \end{aligned}$$

In the last inequality we used (11). We can prove the estimate without the equivalence of zero norms with more elaborate argument.

The a priori estimate (15) follows from

$$\beta|I_h^1 u - u_h|_{1,1} \le \sup_{v\in\mathcal{V}^h}\frac{|\mathcal{B}_h(I_h^1 u - u_h, v)|}{|v|_{1,\mathrm{c}}} \le C(\|\eta\|_{*,1} + \|\mu\|_{*,1} + \|\zeta\|_{**,1}).$$

□

Now, we are ready to prove our main result.

Theorem 3. *Let u denote the solution of (1) and u_h be the solution of FVE (5). Then we have the following estimate*

$$|u - u_h|_{1,\Omega} \le Ch[\|A\|_{0,\infty,\Omega} + h(\|\boldsymbol{b}\|_{0,\infty,\Omega} + \|c\|_{0,\infty,\Omega})]|u|_{2,\Omega}.$$

Proof. We have to estimate the functionals $|\eta_{ij}(u)|$, $|\mu_{ij}(u)|$ and $|\zeta_i(u)|$ on a given face γ_{ij} and control volume V_i, respectively. Let $\gamma = \gamma_{ij}\cap K$, K be a finite element. Using the affine transformation $F:\hat{K}\to K$, $x = F(\hat{x}) = B_K\hat{x} + \boldsymbol{d}$ such that $K = F(\hat{K})$ and Bramble-Hilbert lemma argument we obtain for the contribution in $\eta_{ij}(u)$ from γ:

$$\begin{aligned} |\eta_{ij}(u)_{|\gamma}| = |\eta_{ij}(\tilde{u})_{|\gamma}| &= \left|\int_{\tilde{\gamma}} |\det B_K|\left(\tilde{A}B_K^{-\mathrm{T}}\nabla(I_h^1\tilde{u} - \tilde{u}).B_K^{-\mathrm{T}}\tilde{n}\right)\,d\tilde{s}\right| \\ &\le \|A\|_{0,\infty,\gamma}.\|B_K^{-1}\|^2.|\det B_K|.|\tilde{\gamma}|.\|\tilde{u}\|_{2,\tilde{K}} \\ &\le C\|A\|_{0,\infty,\Omega}\|B_K\|^2\|B_K^{-1}\|^2.|\det B_K|^{1/2}.|u|_{2,K} \\ &\le Ch^{d/2}\|A\|_{0,\infty,\Omega}|u|_{2,K}. \end{aligned}$$

Similarly, for $|\mu_{ij}(u)_{|\gamma}|$ we have

$$
\begin{aligned}
|\mu_{ij}(u)_{|\gamma}| = |\mu_{ij}(\tilde{u})_{|\gamma}| &= \left| \int_{\tilde{\gamma}} |\det B_K| \left(\tilde{\boldsymbol{b}}(I_h^1 \tilde{u} - \tilde{u}).B_K^{-\mathrm{T}} \tilde{n} \right) \, d\tilde{s} \right| \\
&\le \|B_K^{-1}\|.|\det B_K|.\|\boldsymbol{b}\|_{0,\infty,\tilde{K}}.|\tilde{\gamma}|.\|\tilde{u}\|_{2,\tilde{K}} \\
&\le C\|B_K\|^2 \|B_K^{-1}\| |\det B_K|^{1/2} \|\boldsymbol{b}\|_{0,\infty,\Omega} |u|_{2,K} \\
&\le C h^{d/2+1} \|\boldsymbol{b}\|_{0,\infty,\Omega} |u|_{2,K}.
\end{aligned}
$$

We use the bound for the interpolation error in a uniform norm [5] to estimate the term $\zeta_i(u)$ (see also [11] for another application):

$$
\begin{aligned}
|\zeta_i(u)| = |\zeta_i(\tilde{u})| &= \left| \int_{\tilde{V} \cap \tilde{K}} |\det B_K| \tilde{c}(\tilde{x}) (I_h^1 \tilde{u} - \tilde{u}) \, d\tilde{x} \right| \\
&\le \|c\|_{0,\infty,\tilde{K}} \cdot |\det B_K| \cdot \|I_h^1 \tilde{u} - \tilde{u})\|_{0,\infty,\tilde{K}} \\
&\le \|c\|_{0,\infty,\Omega} \cdot h^d h^{2-d/2} |u|_{2,K}.
\end{aligned}
$$

Taking into account that $k_{i,j}^2 = O(h^{2-d})$ we find that

$$
\begin{aligned}
\|\eta\|_{*,1} &\le C_1 h^{1-d/2} h^{d/2} \left(\sum_{x_i \in \omega} \sum_{j \in \Pi(i)} \|A\|_{0,\infty,\Omega}^2 |u|_{2,K}^2 \right)^{1/2} \\
&\le C h \|A\|_{0,\infty,\Omega} |u|_{2,\Omega} \\
\|\mu\|_{*,1} &\le C_1 h^{2-d/2} h^{d/2} \left(\sum_{x_i \in \omega} \sum_{j \in \Pi(i)} \|\boldsymbol{b}\|_{0,\infty,\Omega}^2 |u|_{2,K}^2 \right)^{1/2} \\
&\le C h^2 \|\boldsymbol{b}\|_{0,\infty,\Omega} |u|_{2,\Omega} \\
\|\zeta\|_{**,1} &\le C h^{-d/2} h^{2+d/2} \left(\sum_{x_i \in \omega} \|c\|_{0,\infty,\Omega}^2 |u|_{2,K}^2 \right)^{1/2} \\
&\le C h^2 \|c\|_{0,\infty,\Omega} |u|_{2,\Omega}.
\end{aligned}
$$

Finally the result follows from the triangle inequality and the standard estimate for the linear interpolant. □

4. Upwind finite volume element method

In this section we modify the definition of $\mathcal{B}_h(.,.)$ (7c) in order to obtain a stable approximation for convection dominated problems.

We define the bilinear form $\mathcal{B}_h^{(1)}(.,.)$ in an upwind manner:

$$\mathcal{B}_h^{(1)}(u,v) = \sum_{x_i \in \omega} \sum_{j \in \Pi(i)} (\beta_{ij}^+ u_i + \beta_{ij}^- u_j). \tag{16}$$

Here

$$\beta_{ij}^+ = \frac{\beta_{ij} + |\beta_{ij}|}{2}, \qquad \beta_{ij}^- = \frac{\beta_{ij} - |\beta_{ij}|}{2}.$$

Let β_{ij} be an approximation of $\int_{\gamma_{ij}} (\boldsymbol{b}, \boldsymbol{n})\, ds$ with the properties

$$(i) \quad \beta_{i,j} + \beta_{j,i} = 0. \tag{17a}$$

$$(ii) \quad |\beta_{i,j}| \le C\, m(\gamma_{ij}) \|\boldsymbol{b}\|_{d/2+\alpha,\infty,\Omega}, \tag{17b}$$

$$(iii) \quad \left| \int_{\gamma_{ij}} (\boldsymbol{b}, \boldsymbol{n})\, ds - \beta_{i,j} \right| \le C h^{d+\alpha} |\boldsymbol{b}|_{1+\alpha,\infty,\Omega}, \tag{17c}$$

where C is a positive constant and $\alpha > 0$.

Lemma 6. *Let the bilinear form* $\mathcal{B}_h^{(1)}(.,.)$ *be defined by* (16) *and let the approximations* $\beta_{i,j}$ *fulfill the conditions* (17). *Then for every* $u,\ v \in \mathcal{U}^h$ *the following estimate holds:*

$$\left| \mathcal{B}^{(1)}(u, I_h^{\mathrm{c}} v) - \mathcal{B}_h^{(1)}(u, I_h^{\mathrm{c}} v) \right| \le C h^{\delta} \|\boldsymbol{b}\|_{1+\alpha,\infty,\Omega} |u|_{1,\omega} |I_h^{\mathrm{c}} v|_{1,\mathrm{c}},$$

where $\delta = \min(\alpha, 1)$.

Proof. Note that by the definition of $\beta_{ij}^{\pm}$

$$\beta_{ij}^+ u_i + \beta_{ij}^- u_j = \beta_{ij} u_S,$$

where $S \equiv S(i,j) = i$ if $\beta_{ij} > 0$ and $S(i,j) = j$ otherwise. We have for the difference of interest

$$\left| \mathcal{B}^{(1)}(u, I_h^{\mathrm{c}} v) - \mathcal{B}_h^{(1)}(u, I_h^{\mathrm{c}} v) \right| \le \sum_{x_i \in \omega} \sum_{j \in \Pi(i)} \left| \int_{\gamma_{ij}} (\boldsymbol{b}, \boldsymbol{n}) u\, ds\, v_i - \beta_{ij} u_S v_i \right|.$$

We estimate the term $\left|\int_{\gamma_{ij}}(\boldsymbol{b},\boldsymbol{n})u\,ds\,v_i - \beta_{ij}u_S v_i\right|$ below.

$$\begin{aligned}
&\left|\int_{\gamma_{ij}}(\boldsymbol{b},\boldsymbol{n})u\,ds\,v_i - \beta_{ij}u_S v_i\right| \\
&\quad = \left|\int_{\gamma_{ij}}(\boldsymbol{b},\boldsymbol{n})(u-u_S)\,ds\,v_i + \int_{\gamma_{ij}}(\boldsymbol{b},\boldsymbol{n}) - \beta_{ij}u_S v_i\right| \\
&\quad \le \left|\int_{\gamma_{ij}}(\boldsymbol{b},\boldsymbol{n})\,ds\right| C_1 h|u|_{1,K}|v_i| \\
&\quad + C_2 h^{d+\alpha}|\boldsymbol{b}|_{1+\alpha,\infty,\Omega}|u_{S(i,j)}v_i| \\
&\quad \le C_1 h^{d/2+1}|\boldsymbol{b}|_{1,\infty,\Omega}|u|_{1,K}\cdot|h^{d/2}v_i| \\
&\quad + C_2 h^{\alpha}|\boldsymbol{b}|_{1+\alpha,\infty,\Omega}|h^{d/2}u_{S(i,j)}|\cdot|h^{d/2}v_i|
\end{aligned}$$

□

The existence and uniqueness of the solution of the upwind finite volume element method follows from Lemmas 2, 4 and 6. It is identical with Theorem 2 and we skip it.

We redefine $\mu_{ij}(u)$:

$$\mu_{i,j} = \int_{\gamma_{ij}}(\boldsymbol{b},\boldsymbol{n})u\,ds - \left[\beta^+_{i,j}u_{h,i} + \beta^- u_{h,j}\right]. \tag{18}$$

Note that in the proof of the a priori estimate (15) we did not use the particular form of $\mu_{ij}(u)$. Therefore (15) holds for the upwind finite volume element method as well. The final step is to find an error bound for $\mu_{ij}(u)$:

$$|\mu_{i,j}| \le Ch^{d/2}\left[|\boldsymbol{b}|_{0,\infty,\Omega}|u|_{1,e_{ij}} + h\|\boldsymbol{b}\|_{d/2+\alpha,\infty,\Omega}\|u\|_{2,e_{ij}}\right] \tag{19}$$

in a similar way as in Theorem 3 and Lemma 6. The final result for the upwind method is:

Theorem 4. *If the solution $u(x)$ of the problem (1) is H^2-regular, then the upwind finite volume element method has first order of convergence*

$$|u-u_h|_{1,\Omega} \le Ch\|u\|_{2,\Omega}.$$

References

1. Adams, R.A. (1975): Sobolev spaces. Academic Press, New York
2. Angermann, L. (1991): Numerical solution of second-order elliptic problems on plane domain. RAIRO Model. Math. Anal. Numer. 25(2):169–191
3. Babuška, I., Aziz, A.K. (1972): Foundations of the finite element method. In Aziz, A.K., editor, The Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations, pages 3–362. Academic Press, New York

4. Bank, R.E., Rose, D.J. (1987): Some error estimates for the box method. SIAM J.Numer. Anal. 24:777–787
5. Brenner, S.C., Scott, L.R. (1994): The mathematical theory of finite element methods. Springer, New York
6. Cai, Z. (1991): On the finite volume element method. Numer. Math. 58:713–735
7. Cai, Z., McCormick S.F. (1990): On the accuracy of the finite volume element method for diffusion equations on composite grids. SIAM J.Numer. Anal. 27(3):636–655
8. Ciarlet, P.G. (1978): The Finite Element Method for Elliptic Problems. North-Holland Publishers, Amsterdam
9. Hackbusch, W. (1989): On first and second order box schemes. Computing 41:277–296
10. Heinrich, B. (1987): Finite Difference Methods on Irregular Networks. Akademie-Verlag, Berlin
11. Jianguo, H., Shitong, X. (1995): On the finite volume element method for general self–adjoint elliptic problem. SIAM J.Numer. Anal. 35(5):1762–1774
12. Ladyzhenskaya, O.A., Ural'tseva, N.N. (1968): Linear and Quasilinear Elliptic Equations. Academic Press, New York
13. Nečas, J. (1967): Les méthodes directes dans la théorie des equations elliptiques. Academia, Prague
14. Oganesyan, L.A., Rukhovets, L.A. (1979): Variational–difference methods for solving elliptic equations. Izdatelstvo Akad. Nauk Arm. SSR, Jerevan, (Russian)
15. Schatz, A.H. (1974): An observation concerning Ritz-Galerkin methods with indefinite bilinear forms. Math. Comp. 28(128):959–962