NUMERICAL LINEAR ALGEBRA WITH APPLICATIONS
*Numer. Linear Algebra Appl.* 2004; **00**:1–26 *Prepared using* nlaauth.cls *[Version: 2002/09/18 v1.02]*

# A Finite Volume Element Method for a Nonlinear Elliptic Problem

P. Chatzipantelidis, V. Ginting and R. D. Lazarov*

*Department of Mathematics, Texas A&M University, College Station, TX, 77843*

*Dedicated to Owe Axelsson on the occasion of his 70th birthday*

SUMMARY

We consider a finite volume discretization of second order nonlinear elliptic boundary value problems on polygonal domains. Using relatively standard assumptions we show the existence of the finite volume solution. Furthermore, for a sufficiently small data the uniqueness of the finite volume solution may also be deduced. We derive error estimates in $H^1$–, $L_2$– and $L_\infty$–norm for small data and convergence in $H^1$–norm for large data. In addition a Newton's method is analyzed for the approximation of the finite volume solution and numerical experiments are presented. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS: finite volume element method, nonlinear elliptic equation, error estimates, fixed point iterations, Newton's method

## 1. INTRODUCTION

We analyze a finite volume element method for the discretization of second order nonlinear elliptic partial differential equations on a polygonal domain $\Omega \subset \mathbb{R}^2$. Namely, for a given function $f$ we seek $u$ such that

$$L(u)u \equiv -\nabla \cdot (A(u)\nabla u) = f \quad \text{in } \Omega, \quad \text{and} \quad u = 0, \quad \text{on } \partial\Omega, \tag{1.1}$$

with $A : \mathbb{R} \to \mathbb{R}$ sufficiently smooth such that there exist constants $\beta_i$, $i = 1, 2, 3$, satisfying

$$0 < \beta_1 \le A(x) \le \beta_2, \quad |A'(x)| \le \beta_3, \quad \text{for } x \in \mathbb{R}. \tag{1.2}$$

The study of the solution of (1.1) have been done by many authors. Since there is nonlinearity involves, various assumptions on the coefficient $A$ have been proposed to prove the existence of solution of (1.1). A typical practice is to assume that the coefficient $A$ follows a polynomial growth condition with respect to the nonlinearity. Together with the ellipticity of the coefficient

*Correspondence to: R. D. Lazarov, Department of Mathematics, Texas A&M University, College Station, TX, 77843 (e-mail:lazarov@math.tamu.edu)

$A$, an existence of the weak solution associated with (1.1) may be deduced using theories of monotone operators [25]. A different approach that does not rely on the monotonicity of the problem was proposed in [2, 3]. In these papers, the continuity of coefficient $A$ was assumed to follow Lipschitz-like condition. In particular, the absolute difference of the $A$ is bounded from above by an Osgood's function. Using this condition, the uniqueness of the weak solution was established even in the absence of the monotonicity of the weak variational problem. In the finite dimensional setting, the authors of [16] studied the existence of Galerkin formulation associated with (1.1). The corresponding asymptotic error estimates were established using essentially standard approach.

Finite volume approximations rely on the local conservation property expressed by the differential equation. Namely, integrating (1.1) over any region $V \subset \Omega$ and using Green's formula, we obtain

$$-\int_{\partial V} (A(u)\nabla u) \cdot n \, ds = \int_V f \, dx, \tag{1.3}$$

where $n$ denotes the unit exterior normal to $\partial V$.

In general, the finite volume method has the ability to recover fluxes that are numerically conservative. For many physical and engineering applications, such as heat transfer, transport phenomena, and flows in porous media, this numerical conservation property is crucial. There are various approaches in deriving finite volume approximations of nonlinear elliptic equations. One, often called finite volume element method, uses a finite element partition of $\Omega$, where the solution space consists of continuous piecewise linear functions, a collection of vertex centered control volumes and a test space of piecewise constant functions over the control volumes, cf., e.g., [7, 23, 22]. A second approach, usually called finite volume difference method, uses cell-centered grids and approximates the derivatives in the balance equation by finite differences, cf., e.g., [19]. A third, uses mixed reformulation of the problem, [27]. The first approach is quite close to the finite element method. The second approach is closer to the classical finite difference method and extends it to more general meshes other than rectangular shape. It is used mostly on perpendicular bisector (PEBI) or Voronoi type of meshes. The third approach is close to mixed and hybrid finite element methods and can deal for example with irregular quadrilateral and hexahedral cells. Finite volume discretizations for more general nonlinear convection–diffusion–reaction problems were studied by many authors, cf., e.g., [15, 20].

We shall use the standard notation for the Sobolev spaces $W_p^s$ and $H^s = W_2^s$ [1]. Namely, $L_p(V)$, $1 \le p < \infty$, denotes the $p$–integrable real–valued functions over $V \subset \mathbb{R}^2$, $(\cdot,\cdot)_V$ the inner product in $L_2(V)$, and $\|\cdot\|_{W_p^s(V)}$ the norm in the Sobolev space $W_p^s(V)$, $s \ge 0$. If $V = \Omega$ we suppress the index $V$, and if $p = 2$ we write $H^s = W_2^s$ and $\|\cdot\| = \|\cdot\|_{L_2}$. Further we shall denote with $p'$ the adjoint of $p$, i.e., $\frac{1}{p} + \frac{1}{p'} = 1$, $p > 1$.

It is well known that for domains with smooth boundary, for $f \in C^r$, with $r \in (0,1)$, there exists a unique solution $u \in C^{2+r}$, cf., e.g., [17]. Also for $\|f\|$ sufficiently small, there exists a unique solution $u \in H^2 \cap H_0^1$. However, since in this study we assume the domain $\Omega$ to be polygonal, we do not expect the solution $u$ to have such regularity. We shall assume that for $f \in L_2$, problem (1.1) has a solution $u \in W_q^2 \cap H_0^1$, with $4/3 < q \le 2$. Note that in order (1.3) to be well defined, $u \in H^{1+s}$ with $s > 1/2$. Using a standard Sobolev embedding we see that this is true for $u \in W_q^2$ with $q > 4/3$.

In this paper, we shall study approximations of (1.1) by the finite volume element method, which for brevity we shall refer to as the finite volume method below. The approximate solution

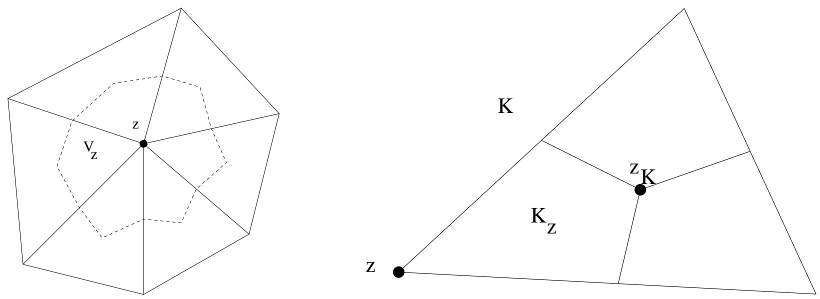


Figure 1. *Left:* A union of triangles that have a common vertex $z$; the dotted line shows the boundary of the corresponding control volume $V_z$. *Right:* A triangle $K$ partitioned into the three subregions $K_z$.

will be sought in the piecewise linear finite element space

$$X_h \equiv X_h(\Omega) = \{\chi \in C(\Omega) : \ \chi|_K \ \text{ linear}, \ \forall K \in \mathcal{T}_h; \ \chi|_{\partial\Omega} = 0\},$$

where $\{\mathcal{T}_h\}_{0<h<1}$ is a family of quasi-uniform triangulations of $\Omega$, $h$ denotes the maximum diameter of the triangles of $\mathcal{T}_h$.

The discrete finite volume problem will satisfy a relation similar to (1.3) $V$ is in a finite collection of subregions of $\Omega$ which are called control volumes. We note that for the discrete system to be well defined, the number these control volumes has to be equal to the dimension of the finite element space $X_h$. These control volumes are constructed in the following way. Let $z_K$ be the barycenter of $K \in \mathcal{T}_h$. We connect $z_K$ with line segments to the midpoints of the edges of $K$, thus partitioning $K$ into three quadrilaterals $K_z$, $z \in Z_h(K)$, where $Z_h(K)$ are the vertices of $K$. Then with each vertex $z \in Z_h = \cup_{K \in \mathcal{T}_h} Z_h(K)$ we associate a control volume $V_z$, which consists of the union of the subregions $K_z$, sharing the vertex $z$ (see Figure 1). We denote the set of interior vertices of $Z_h$ by $Z_h^0$.

Then, the finite volume method is to find $u_h \in X_h$ such that

$$-\int_{\partial V_z} (A(u_h)\nabla u_h) \cdot n \, ds = \int_{V_z} f \, dx, \quad \forall z \in Z_h^0. \tag{1.4}$$

The analysis of the finite volume method that we perform in this paper will use existing results associated with the finite element method applied to (1.1). The Galerkin formulation of finite element method for (1.1) is to find $\underline{u}_h \in X_h$ such that

$$a(\underline{u}_h; \underline{u}_h, \chi) = (f, \chi), \quad \forall \chi \in X_h, \tag{1.5}$$

with $a(\cdot;\cdot,\cdot)$ the form defined by

$$a(v; w, \phi) = \int_\Omega A(v)\nabla w \cdot \nabla\phi \, dx.$$

It is known that the solution $\underline{u}_h$ of (1.5) satisfies

$$\begin{aligned} \|\underline{u}_h - u\| + h\|\nabla(\underline{u}_h - u)\| &\le C(u,f)h^2 \\ \|\underline{u}_h - u\|_{L_\infty} &\le C_p \inf_{\chi \in X_h} \|\nabla(u-\chi)\|_{W_p^1}, \text{ with } p > 2. \end{aligned} \tag{1.6}$$

Numerical methods for this type and more general problems has been considered by many authors, cf., e.g, [6, 16, 21, 24].

To the best of our knowledge, there has not been many investigations conducted to analyze the finite volume method for general nonlinear elliptic problems. Li [23] considers a variation of the finite volume method under investigation here. The method differs in the construction of the control volumes. Instead of the barycenter $z_K$, the circumcenter is selected. For this finite volume method, the $H^1$-norm error estimate is similar to one in the finite element method.

The paper is organized as follows. We will first introduce the finite volume method in Section 2. Then in Section 3, we establish existence of the finite volume solution $u_h$ of (2.3), using a fixed point iteration method. In particular, in Theorem 3.1 we show that the iterations remain inside a fixed ball with a radius that depends only on $f$. Then in Theorem 3.2 we show that for a sufficiently small data, $f$, the fixed point iteration operator is Lipschitz continuous with Lipschitz constant less that 1.

In Section 4 we derive optimal order $H^1$–, $L_2$– and almost optimal $L_\infty$–norm error estimates, under the assumption that $\|f\|$ is small. Note that for the $L_2$ estimate we assume that $A'$ is also Lipschitz continuous, $A'' \in L_1(\mathbb{R})$ and $f \in H^1$. Also, for large data, assuming a unique solution $u \in H_0^1 \cap L_\infty$ we show $u_h \to u$, in $H_0^1$.

Section 5 gives an analysis of a Newton's method for the approximation of the finite volume solution $u_h$. A similar approach for the finite element method was analyzed by Douglas and Dupont in [16]. As already well established, one has to start the Newton iteration with an initial approximation $u_h^0$ sufficiently close to $u_h$. Following the framework presented in [16], we show that the Newton iterations converge to $u_h$ with order 2.

Discussion on several numerical experiments is presented in Section 6. We compare two iterative methods, namely, the fixed point iteration and the Newton iteration. For the implementation, we employ an inexact Newton iteration, a variant of the Newton iteration for nonlinear systems of equations, where the Jacobian of the system is solved approximately, cf., e.g., [4, 5, 14].

## 2. PRELIMINARIES–THE FINITE VOLUME METHOD

There has been a tendency of analyzing finite volume element method using the existing results from its finite element counterpart, cf., e.g., [10, 11, 12, 13]. The investigations recorded in all these references were concentrated on elliptic and/or parabolic problems with coefficients independent of the solution, i.e., the function $A$ is only spatially varied. The finite volume element method is viewed as a perturbation of standard Galerkin finite element method with the help of an interpolation operator $I_h : C(\Omega) \to Y_h$, defined by

$$I_h v = \sum_{z \in Z_h^0} v(z)\Psi_z, \tag{2.1}$$

where

$$Y_h = \{\eta \in L_2(\Omega) : \ \eta|_{V_z} = \text{constant}, \ \forall z \in Z_h^0; \ \eta|_{V_z} = 0, \ \forall z \in \partial\Omega\},$$

and $\Psi_z$ is characteristic function of $V_z$. We note that $I_h : X_h \to Y_h$ is a bijection and bounded with respect to the $L_2$−norm, i.e., there exist $c_1, c_2 > 0$, such that

$$c_1\|\chi\| \le \|I_h\chi\| \le c_2\|\chi\|, \quad \forall \chi \in X_h. \tag{2.2}$$

The finite volume problem (1.4) can be rewritten in a variational form. For an arbitrary $\eta \in Y_h$, we multiply the integral relation in (1.4) by $\eta(z)$ and sum over all $z \in Z_h^0$ to obtain the Petrov–Galerkin formulation, to find $u_h \in X_h$ such that

$$a_h(u_h; u_h, \eta) = (f, \eta), \quad \forall \eta \in Y_h, \tag{2.3}$$

where the form $a_h(\cdot;\cdot,\cdot) : X_h \times X_h \times Y_h \to \mathbb{R}$ is defined by

$$a_h(w; v, \eta) = -\sum_{z \in Z_h^0} \eta(z) \int_{\partial V_z} (A(w)\nabla v) \cdot n \, ds, \quad v, w \in X_h, \ \eta \in Y_h. \tag{2.4}$$

Obviously, $a_h(w; v, \eta)$ may also be defined by (2.4) for $v, w \in W_p^1(\Omega) \cap H_0^1(\Omega)$, $p > 2$. In particular, using Green's formula we easily see that

$$a_h(w; v, \eta) = (L(w)v, \eta), \quad \text{for } v, w \in W_p^1(\Omega) \cap H_0^1(\Omega), \ \eta \in Y_h. \tag{2.5}$$

In the case of $w \in L_\infty$, the bilinear form $a_h(w; \cdot, \cdot)$ in (2.4) may be equivalently written as

$$a_h(w; v, \eta) = \sum_K \Big\{ (L(w)v, \eta)_K + (A(w)\nabla v \cdot n, \eta)_{\partial K} \Big\}, \quad \forall v \in X_h, \ \eta \in Y_h. \tag{2.6}$$

Indeed, using integration by parts, the identity

$$\int_{K_z} L(w)v \, dx = -\int_{\partial K_z \cap \partial K} (A(w) \, \nabla v) \cdot n \, ds - \int_{\partial K_z \cap \partial V_z} (A(w) \, \nabla v) \cdot n \, ds, \tag{2.7}$$

holds for $z \in Z_h^0$ and $K \in \mathcal{T}_h$, and hence (2.6) follows from multiplication of (2.7) by $\eta(z)$ and by summing it up first over the triangles that have $z$ as a vertex and then over the vertices $z \in Z_h^0$.

The interpolation operator $I_h$ has the following properties, cf., e.g., [10],

$$\int_K I_h \chi \, dx = \int_K \chi \, dx, \ \ \forall \chi \in X_h, \ \ \text{for any } K \in \mathcal{T}_h, \tag{2.8}$$

$$\int_e I_h \chi \, ds = \int_e \chi \, ds, \ \ \forall \chi \in X_h, \ \ \text{for any side } e \text{ of } K \in \mathcal{T}_h, \tag{2.9}$$

$$\|I_h \chi\|_{L_\infty(e)} \le \|\chi\|_{L_\infty(e)}, \ \ \forall \chi \in X_h, \ \ \text{for any side } e \text{ of } K \in \mathcal{T}_h, \tag{2.10}$$

$$\|\chi - I_h \chi\|_{L_p(K)} \le h \|\nabla \chi\|_{L_p(K)}, \ \ \forall \chi \in X_h, \ \ 1 \le p < \infty. \tag{2.11}$$

In addition in [10, Lemma 6.1, Remark 6.1, Lemma 5.1] the following lemma was derived.

**Lemma 2.1.** *Let $e$ be a side of a triangle $K \in \mathcal{T}_h$. Then for $v \in W_p^1(K)$ there exists a constant $C_1 > 0$ independent of $h$ such that*

$$|\int_e v(\chi - I_h \chi) \, ds| \le C_1 h \|\nabla v\|_{L_p(K)} \|\nabla \chi\|_{L_{p'}(K)}, \quad \forall \chi \in X_h, \ \text{with } \frac{1}{p} + \frac{1}{p'} = 1. \tag{2.12}$$

*Moreover, introducing $\varepsilon_h : L_2 \times X_h \to \mathbb{R}$ which is defined by*

$$\varepsilon_h(f, \chi) = (f, \chi - I_h \chi), \tag{2.13}$$

*then for $f \in W_p^i$, $i = 0, 1$ and $\chi \in X_h$,*

$$|\varepsilon_h(f, \chi)| \le C h^{i+j} \|f\|_{W_p^i} \|\chi\|_{W_{p'}^j}, \ \ f \in W_p^i, \ i, j = 0, 1, \ \text{with } \frac{1}{p} + \frac{1}{p'} = 1. \tag{2.14}$$

**Lemma 2.2.** *Let $v \in W_q^2$, $4/3 < q \le 2$. The following identities hold.*

$$\sum_K \int_{\partial K} A(\bar{w}) \nabla v \cdot n \, \chi \, ds = 0, \quad \sum_K \int_{\partial K} A(\bar{w}) \nabla v \cdot n \, I_h \chi \, ds = 0, \quad \forall \chi \in X_h. \tag{2.15}$$

*where $\bar{w}$ can be either an element of $X_h$ or the value of an element of $X_h$ at the midpoint of the edge $e$ of triangle $K$.*

*Proof.* Note, that for $v \in W_q^2$, the trace $\nabla v \cdot n$ on $\partial K$ exists for $q > 4/3$. The left identity is obvious by rewriting the sum as integrals of jump terms over the interior edges of $\mathcal{T}_h$. These jumps obviously vanish because of the continuity of $A(\bar{w}) \nabla v \cdot n$ (in the trace sense). A similar argument gives the second identity. □

Our analysis will use arguments similar to the corresponding linear problems, cf., e.g, [10, 11]. In those papers, the error estimates are derived by bounding the error between the bilinear forms of the finite element, $a$, and the finite volume methods, $a_h$. This is shown to be $O(h)$ uniformly in $X_h$. Then for sufficiently small $h$ the finite volume bilinear form $a_h$ is coercive in $X_h$, which leads to the existence and uniqueness of the finite volume approximation.

In this paper, we will show that a similar estimate for the error functional $\varepsilon_a$,

$$\varepsilon_a(w; v_h, \chi) = a(w; v_h, \chi) - a_h(w; v_h, I_h \chi) \quad \forall v_h, \chi \in X_h, \ w \in L_\infty, \tag{2.16}$$

is no longer $O(h)$ uniformly in $X_h$. This is due to the fact that the bound of $\varepsilon_a(w_h; v_h, \chi)$, will depend on $\|\nabla w_h\|_{L_\infty}$. We note that inverse inequalities of the form, cf., e.g., [8],

$$\|\nabla \chi\|_{L_s} \le C h^{2/s - 2/t} \|\nabla \chi\|_{L_t}, \quad \forall \chi \in X_h, \text{ with } 1 \le t \le s \le \infty, \tag{2.17}$$

are true in a quasi-uniform mesh. Applying this inequality for $\|\nabla w_h\|_{L_\infty}$ give a uniform estimate of $O(h^{1-2/t})$ for (2.16) in a ball of $X_h$ with respect to $W_t^1$–norm, for $t > 2$.

The basic estimate result for (2.16) is stated in the following lemma.

**Lemma 2.3.** *There exists a constant $C_2 > 0$, independent of $h$, such that*

$$|\varepsilon_a(w_h; v_h, \chi)| \le C_2 \beta_3 h \|\nabla w_h \cdot \nabla v_h\|_{L_p} \|\nabla \chi\|_{L_{p'}}, \quad \forall w_h, v_h, \chi \in X_h, \ \frac{1}{p} + \frac{1}{p'} = 1. \tag{2.18}$$

*Proof.* In view of Green's formula and (2.6), we may write $\varepsilon_a$ in the following form:

$$\begin{aligned} \varepsilon_a(w_h; v_h, \chi) &= \sum_K \left\{ (L(w_h) v_h, \chi - I_h \chi)_K + (A(w_h) \nabla v_h \cdot n, \chi - I_h \chi)_{\partial K} \right\} \\ &= \sum_K \left\{ I_K + II_K \right\}. \end{aligned} \tag{2.19}$$

Applying Hölder's inequality to $I_K$, and using the fact that $w_h$ and $v_h$ are linear in $K$, and using (1.2) and (2.11), we have

$$|I_K| \le \beta_3 \|\nabla w_h \cdot \nabla v_h\|_{L_p(K)} \|\chi - I_h \chi\|_{L_{p'}(K)} \le \beta_3 h \|\nabla w_h \cdot \nabla v_h\|_{L_p(K)} \|\nabla \chi\|_{L_{p'}(K)}. \tag{2.20}$$

For $II_K$, we break the integration over the boundary of each triangle $K$, into the sum of integrations over its sides, and thus may use (2.12), and follow the same steps as in estimating $I_K$. Hence,

$$|II_K| \le C_1 h |A(w_h) \nabla v_h|_{W_p^1(K)} \|\nabla \chi\|_{L_{p'}(K)} \le C_1 \beta_3 h \|\nabla w_h \cdot \nabla v_h\|_{L_p(K)} \|\nabla \chi\|_{L_{p'}(K)}. \tag{2.21}$$

Finally, (2.20) and (2.21) establish the desired estimate with $C_2 = C_1 + 1$. $\square$

To establish an optimal error estimate in the $L_2$–norm, Lemma 2.3 is no longer sufficient. In other words, we need a stronger result than the one stated in that lemma. For this, we will need to assume that $A'$ is Lipschitz continuous with constant $L$, i.e.

$$|A'(x) - A'(y)| \le L|x - y|, \quad \forall x, y \in \mathbb{R}. \tag{2.22}$$

This stronger result is presented in the following lemma which will be used in Section 4 to estimate the error in the $L_2$–norm.

**Lemma 2.4.** *Assume that $A'$ is Lipschitz continuous and $v \in W_q^2 \cap H_0^1$, for $4/3 < q \le 2$. Then there exists a constant $C > 0$ independent of $h$ such that for $w_h, v_h, \chi \in X_h$,*

$$\begin{aligned} |\varepsilon_a(w_h; v_h, \chi)| \le & C\big\{h^2 \|\nabla w_h\|_{L_\infty} (\|\nabla w_h \cdot \nabla v_h\| + \|v\|_{W_q^2}) \\ & + h\|\nabla w_h \cdot \nabla(v_h - v)\|_{L_q}\big\} \|\nabla \chi\|_{L_{q'}}, \end{aligned} \tag{2.23}$$

*with* $1/q + 1/q' = 1$.

*Proof.* Let $w_K$ and $w_e$ denote the average value of a function $w$ over triangle $K$ and the edge $e$, respectively. Since $v \in W_q^2$, Lemma 2.2 gives the identity

$$\big((A(w_h) - A(w_{h,e}))\nabla v \cdot n, \chi - I_h\chi\big)_{\partial K} = 0, \quad \forall \chi \in X_h.$$

Employing this identity, the fact that $v_h$ is linear in $K$, Green's formula, and (2.8) we get

$$\begin{aligned} \varepsilon_a(w_h; v_h, \chi) = & \sum_K \big((A'(w_h) - A'(w_{h,K}))\nabla w_h \cdot \nabla v_h, \chi - I_h\chi\big)_K \\ & + \sum_K \big((A(w_h) - A(w_{h,e}))\nabla(v_h - v) \cdot n, \chi - I_h\chi\big)_{\partial K} = \sum_K \{I_K + II_K\}. \end{aligned}$$

Using now Hölder's inequality, the fact that $w_h$ is linear in $K$, and (2.11), we can bound $I_K$,

$$\begin{aligned} |I_K| & \le C \int_K |w_h - w_{h,K}|\, |\nabla w_h \cdot \nabla v_h|\, |\chi - I_h\chi|\, dx \\ & \le Ch^2 \|\nabla w_h\|_{L_\infty} \|\nabla w_h \cdot \nabla v_h\|_{L_2(K)} \|\nabla \chi\|_{L_2(K)}. \end{aligned} \tag{2.24}$$

To estimate $II_K$, we apply (2.12) and obtain,

$$|II_K| \le Ch\, |(A(w_h) - A(w_{h,e}))\nabla(v_h - v)|_{W_q^1(K)} \|\nabla \chi\|_{L_{q'}(K)}. \tag{2.25}$$

Furthermore, a simple calculation gives

$$|(A(w_h) - A(w_{h,e}))\nabla(v_h - v)|_{W_q^1(K)} \le C(\|\nabla w_h \cdot \nabla(v_h - v)\|_{L_q(K)} + h\|\nabla w_h\|_{L_\infty} \|v\|_{W_q^2(K)}\big).$$

Summing up (2.24) and (2.25) over all triangles, and using the fact that $q' > 2$, we obtain (2.23). $\square$

Next we will derive a "Lipschitz"-type estimate for $\varepsilon_a$.

**Lemma 2.5.** *Let $v \in H^1 \cap L_\infty$, $w \in W_p^1$ with $p > 2$ and $A'$ be Lipschitz continuous with constant L, cf.* (2.22). *There exists $C_2 > 0$ such that*

$$\begin{aligned} & |\varepsilon_a(v; \phi_h, \chi) - \varepsilon_a(w; \phi_h, \chi)| \\ & \quad \le C_2 h \|\nabla \phi_h\|_{L_\infty} (\beta_3 + L\|\nabla w\|_{L_p}) \|\nabla(v - w)\| \, \|\nabla \chi\|, \quad \forall \phi_h, \chi \in X_h, \end{aligned} \tag{2.26}$$

*where $\beta_3$ is the upper bound of $A'$, cf.,* (1.2).

*Proof.* We can easily see that

$$\varepsilon_a(v;\phi_h,\chi)-\varepsilon_a(w;\phi_h,\chi)=\sum_K\Big\{\int_K \operatorname{div}((A(v)-A(w))\nabla\phi_h)(\chi-I_h\chi)\,dx + \int_{\partial K}(A(v)-A(w))\nabla\phi_h\cdot n(\chi-I_h\chi)\,ds\Big\}.$$

Also, since $\phi_h$ is linear in $K$, $\operatorname{div}(\nabla\phi_h)=0$, therefore,

$$\operatorname{div}((A(v)-A(w))\nabla\phi_h)=\big\{A'(v)\nabla(v-w)+(A'(v)-A'(w))\nabla w\big\}\cdot\nabla\phi_h,\quad \text{in } K.$$

Then, this equality together with (2.11), (2.12), the Hölder inequality

$$\|vw\|_{L_s}\le\|v\|_{L_t}\|w\|_{L_{\bar t}},\ \text{with } t>s,\quad \frac{s}{t}+\frac{s}{\bar t}=1, \tag{2.27}$$

for $s=2$ and $t=p$, and the Sobolev inequality, cf. e.g., [8, 4.x.11],

$$\|v\|_{L_s}\le\|\nabla v\|,\quad \forall s<\infty, \tag{2.28}$$

give the following

$$\begin{aligned}|\varepsilon_a(v;\phi_h,\chi)-\varepsilon_a(w;\phi_h,\chi)| &\le C_2h(\beta_3\|\nabla(v-w)\|+L\|\,|v-w|\,|\nabla w|\,\|)\|\nabla\chi\|\,\|\nabla\phi_h\|_{L_\infty}\\ &\le C_2h(\beta_3\|\nabla(v-w)\|+L\|v-w\|_{L_{\bar p}}\|\nabla w\|_{L_p})\|\nabla\chi\|\,\|\nabla\phi_h\|_{L_\infty}\\ &\le C_2h(\beta_3+L\|\nabla w\|_{L_p})\|\nabla(v-w)\|\,\|\nabla\chi\|\,\|\nabla\phi_h\|_{L_\infty},\end{aligned}$$

with $C_2=C_1+1$. □

## 3. EXISTENCE OF FVE APPROXIMATIONS

In this section, we prove the existence of the finite volume solution $u_h$ of (2.3) under some standard assumptions. A fixed point iteration is applied to (2.3) and in turn the existence is shown within certain subspace of $X_h$. To be specific, we are seeking the solution of (2.3) in a ball $\mathcal{B}_M$ which is defined as

$$\mathcal{B}_M=\{\chi\in X_h:\ \|\nabla\chi\|_{L_p}\le M\},\qquad \text{with } p>2,$$

for some $M>0$. We note that there is no restriction required on $M$ to show the existence of the finite volume solution. Moreover, we can establish the uniqueness of the solution of (2.3) by imposing further requirements on the coefficient and data, namely that $A'$ is Lipschitz continuous and $M$ is sufficiently small.

A crucial ingredient for proving the existence of solution of (2.3) is the inf–sup condition. Here we will need an inf–sup condition that holds for $L_p$–norm with $p\ge 2$. This is attributed to the fact that we seek the solution in $\mathcal{B}_M$. In view of the Sobolev imbedding, $\|v\|_{L_\infty}\le C\|v\|_{W_p^1}$ for $p>2$, we shall assume the following inf–sup condition, cf., e.g., [8, Chapter 7]: There exist constants $\alpha=\alpha(A,\Omega)>0$, $h_\alpha>0$ and $\epsilon=\epsilon(A,\Omega)>0$ such that for all $0<h\le h_\alpha$ and $v_h\in X_h$ and $w\in L_\infty$,

$$\|\nabla v_h\|_{L_p}\le\alpha\sup_{0\ne\chi\in X_h}\frac{a(w;v_h,\chi)}{\|\nabla\chi\|_{L_{p'}}}, \tag{3.1}$$

with $2 \le p \le 2+\epsilon$ and $\frac{1}{p}+\frac{1}{p'}=1$.

Using this assumption, our task is show that similar inf–sup condition holds for the bilinear form $a_h(w;\cdot,\cdot)$ used in (2.3). In view of the identity $a(w_h;v_h,\chi) = a_h(w_h;v_h,I_h\chi) + \varepsilon_a(w_h;v_h,\chi)$, and using Lemma 2.3, and (2.17), we get

$$|\varepsilon_a(w_h;v_h,\chi)| \le Ch\|\nabla w_h\cdot\nabla v_h\|_{L_p}\|\nabla\chi\|_{L_{p'}} \le Ch^{1-2/p}\|\nabla w_h\|_{L_p}\|\nabla v_h\|_{L_p}\|\nabla\chi\|_{L_{p'}}.$$

It is then straight forward to see that given $M>0$, there exists $h_M>0$ such that for all $0<h\le h_M\le h_\alpha$

$$\|\nabla v_h\|_{L_p} \le \tilde\alpha \sup_{0\ne\chi\in X_h}\frac{a_h(w_h;v_h,I_h\chi)}{\|\nabla\chi\|_{L_{p'}}}, \quad \forall v_h\in X_h, w_h\in\mathcal{B}_M,\ 2<p\le 2+\epsilon, \tag{3.2}$$

where $\tilde\alpha$ depends only $\alpha$ and $h_M$. We note that (3.2) holds also for $p=2$ and $w_h\in\mathcal{B}_M = \{\chi\in X_h:\ \|\nabla\chi\|_{L_{\tilde p}}\le M\}$, with $\tilde p>2$.

Having established the inf–sup condition, we are now ready to show the existence of the solution of (2.3). As mentioned earlier, we will devise a fixed point iteration in (2.3), namely, for a fixed $f\in L_2$, we consider an iteration map $T_h: X_h\to X_h$ given by

$$a_h(v_h;T_hv_h,\eta) = (f,\eta), \quad \forall\eta\in Y_h. \tag{3.3}$$

It is easy to see that by the inf–sup condition (3.2) $T_hv_h$ is well defined for $h<h_M$ and $v_h\in\mathcal{B}_M$. The following theorem states the existence of (2.3).

**Theorem 3.1.** *Let $f\in L_2$ be given and assume that the standard requirements* (1.2) *on the coefficient A hold. Choose $M>0$ such that $\|f\|\le M\tilde\alpha^{-1}$, where as before $\tilde\alpha$ is the coefficient in the inf–sup condition* (3.2)*. Then there exists a solution of* (2.3) *in $\mathcal{B}_M$.*

*Proof.* Let $v_h\in\mathcal{B}_M$ then in view of (3.2) we have

$$\|\nabla T_hv_h\|_{L_p} \le \tilde\alpha\sup_{0\ne\chi\in X_h}\frac{a_h(v_h;T_hv_h,I_h\chi)}{\|\nabla\chi\|_{L_{p'}}} \le \tilde\alpha\sup_{0\ne\chi\in X_h}\frac{(f,I_h\chi)}{\|\nabla\chi\|_{L_{p'}}}. \tag{3.4}$$

Then, using (2.2) and the Sobolev inequality $\|v\|\le\|v\|_{W_p^1}$, for $p>1$, cf. [8, 4.x.11], we get

$$\|\nabla T_hv_h\|_{L_p}\le\tilde\alpha\|f\|\le M. \tag{3.5}$$

Thus the operator $T_h$ maps the ball $v_h\in\mathcal{B}_M$ into itself. By the Brouwer fixed point theorem, we know that $T_h$ has a fixed point, and this implies that (2.3) has a solution in $\mathcal{B}_M$. □

Some comments related to the theorem above worth mentioning. It is clear from the above proof, that up to existence of the finite volume solution, the regularity on the data $f$ and the coefficient $A$ are standard. Given the data $f$, we have a freedom to choose $M>0$ that satisfies the requirement in the theorem. As for the coefficient $A$, we have used relatively different constraint. Instead of imposing Lipschitz continuity of $A$, we have assumed that its derivative is bounded from above, cf. the second part of (1.2). This kind of requirement is more due to the construction of the finite volume variational problem (2.3). In particular, the corresponding bilinear form $a_h(w;\cdot,\cdot)$ on which our analysis is relied upon involves a stronger form.

In the next theorem we show that in a sufficiently small ball $\mathcal{B}_M$ and data $f$, there exists a unique solution $u_h\in X_h$ of (2.3).

Next, we will show that the iteration map $T_h$ is Lipschitz continuous. For $M$ sufficiently small, $T_h$ is a contraction in $\mathcal{B}_M$ in $H^1$-norm, which gives the uniqueness of the solution $u_h$ of (2.3)

**Theorem 3.2.** *Let $A'$ be Lipschitz continuous with constant $L$, cf.* (2.22). *Then there exists a constant $C_L = C_L(A,\Omega) > 0$ and $h'_M > 0$, such that for $\|f\| \le M\tilde{\alpha}^{-1}$, $M < {C_L}^{-1}$ and all $0 < h \le h'_M$, $T_h$ is a contraction, with constant $\ell = C_L M < 1$,*

$$\|\nabla(T_h v - T_h w)\| \le \ell \|\nabla(v-w)\|, \quad \forall v, w \in \mathcal{B}_M. \tag{3.6}$$

*Proof.* Let $v, w \in \mathcal{B}_M$. Then, in view of the definition of $T_h$, (3.3) we have

$$a_h(v; T_h v, \eta) - a_h(w; T_h w, \eta) = 0, \quad \forall \eta \in Y_h.$$

Therefore, we can easily see that for $\eta = I_h\chi$, $\chi \in X_h$,

$$\begin{aligned} a_h(v; T_h v - T_h w, I_h\chi) &= a_h(w; T_h w, I_h\chi) - a_h(v; T_h w, I_h\chi) \\ &= \varepsilon_a(v; T_h w, \chi) - \varepsilon_a(w; T_h w, \chi) + ((A(w) - A(v))\nabla T_h w, \nabla\chi). \end{aligned} \tag{3.7}$$

Using now the fact that for sufficiently small $h$, $T_h w \in \mathcal{B}_M$, cf., Theorem 3.1, the Hölder inequality (2.27) with $s = 2$ and $t = p$ and the Sobolev (2.28), the last term of the right–hand side of (3.7) can be bounded for any $\chi \in X_h$,

$$\begin{aligned} |(A(w) - A(v))\nabla T_h w, \nabla\chi)| &\le \beta_3 \|(w-v)|\nabla T_h w|\| \, \|\nabla\chi\| \\ &\le \beta_3 \|w - v\|_{L_{\bar{p}}} \|\nabla T_h w\|_{L_p} \|\nabla\chi\| \le \beta_3 M \|\nabla(v-w)\| \, \|\nabla\chi\|. \end{aligned} \tag{3.8}$$

Also, in view of Lemma 2.5 the remaining two terms in the right–handside of (3.7), give

$$|\varepsilon_a(v; T_h w, \chi) - \varepsilon_a(w; T_h w, \chi)| \le C_2 h^{1-2/p} M(\beta_3 + LM)\|\nabla(v-w)\| \, \|\nabla\chi\|. \tag{3.9}$$

Since, $a_h(v_h; \cdot, \cdot)$ is coercive for $v_h \in \mathcal{B}_M$ and $h$ sufficiently small, choosing $\chi = T_h v - T_h w$ in the above relation and in (3.7) and (3.8) gives that there exists a constant $C_L = C_L(A,\Omega) > 0$ such that

$$\|\nabla(T_h v - T_h w)\| \le C_L M \|\nabla(v-w)\|.$$

Therefore, for $M < C_L^{-1}$, $T_h$ is a contraction with constant $0 < \ell = C_L M < 1$. □

Finally, Theorems 3.1 and 3.2 give the following corollary,

**Corollary 3.3.** *Assume that $A'$ is Lipschitz continuous with a constant $L$. Then there exist constants $C_L = C_L(A,\Omega) > 0$ and $h_0 > 0$ such that if $\|f\| \le \tilde{\alpha}^{-1} C_L^{-1}$, with $2 < p < 2+\epsilon$ then for $h$ sufficiently small the problem* (2.3)*, i.e., find $u_h \in X_h$ such that*

$$a_h(u_h; u_h, I_h\chi) = (f, I_h\chi), \quad \forall \chi \in X_h,$$

*has a unique solution, with $\epsilon$ given in* (3.1).

## 4. ERROR ESTIMATES

In this section we shall derive the error estimates for the finite volume solution, i.e., the estimate for $u_h - u$ in the $W^1_s$–, with $2 \le s < p$, $L_2$– and $L_\infty$-norm. We assume that $f \in L_2$ and that the nonlinear problem (1.1) has a unique solution $u \in W^2_q \cap H^1_0$, with $4/3 < q \le 2$. Recall that in Section 3 we show that a finite volume solution $u_h$ of (2.3) exists and is unique.

First, we will derive an a priori error estimate $\|\nabla(u_h - u)\|_{L_s}$, $2 \le s < p$. For $s = 2$ we get the usual $H^1$–norm error bound. Moreover, for $s > 2$ the estimate combined with a standard Sobolev imbedding gives an $L_\infty$–norm error estimate, cf. Theorem 4.2.

**Theorem 4.1.** *Let $u_h$ and $u$ be the solutions of* (2.3) *and* (1.1)*, respectively, with $f \in L_2$. Then, if $\gamma = \alpha\beta_3 M < 1$ there exists a constant $C = C(u, f)$, independent of $h$, such that for $0 < h \le h_M$*

$$\|\nabla(u_h - u)\|_{L_s} \le C(u,f)h^{1+2/s-2/q}, \quad \text{with } 2 \le s < p < 2+\epsilon, \quad \frac{4}{3} < q \le 2, \tag{4.1}$$

*where $\alpha$ is the constant appeared in* (3.1).

*Proof.* Using the triangle inequality we get

$$\|\nabla(u_h - u)\|_{L_s} \le \|\nabla(u-\chi)\|_{L_s} + \|\nabla(u_h - \chi)\|_{L_s}, \quad \forall \chi \in X_h. \tag{4.2}$$

In view of the approximation property of $X_h$,

$$\inf_{\chi \in X_h} \|\nabla(v-\chi)\|_{L_s} \le Ch^{1+2/s-2/q}\|v\|_{W_q^2}, \quad \text{with } 4/3 < q \le 2 \le s, \tag{4.3}$$

the first term on the right side of (4.2) is bounded as desired. Also, we can easily see that

$$a(u; u_h - \chi, \psi) = a(u; u_h - u, \psi) + a(u; u-\chi, \psi) \le a(u; u_h - u, \psi) + \beta_2\|\nabla(u-\chi)\|_{L_s}\|\nabla\psi\|_{L_{s'}},$$

with $1/s + 1/s' = 1$. Hence, in view of (3.1), we may write for $2 \le s < p$,

$$\begin{aligned}\|\nabla(u_h - \chi)\|_{L_s} &\le \alpha \sup_{0\ne\psi\in X_h} \frac{a(u; u_h-\chi, \psi)}{\|\nabla\psi\|_{L_{s'}}} \\ &\le \alpha \sup_{0\ne\psi\in X_h} \frac{a(u; u_h-u, \psi)}{\|\nabla\psi\|_{L_{s'}}} + \alpha\beta_2\|\nabla(u-\chi)\|_{L_s}.\end{aligned} \tag{4.4}$$

Then in view of (4.3), it suffices to estimate the first term of the right–handside in the relation above. We can easily see for any $\psi \in X_h$,

$$\begin{aligned}a(u; u_h - u, \psi) &= a(u; u_h, \psi) - (f, \psi) \\ &= \{a(u; u_h, \psi) - a(u_h; u_h, \psi)\} + \{\varepsilon_a(u_h; u_h, \psi) - \varepsilon_h(f, \psi)\} = I + II.\end{aligned} \tag{4.5}$$

Then using the fact that $u_h \in \mathcal{B}_M$, the Hölder inequality (2.27), with $t = p$, and the Sobolev inequality (2.28), we have for any $\chi, \psi \in X_h$,

$$\begin{aligned}|I| &= |a(u; u_h, \psi) - a(u_h; u_h, \psi)| \le \beta_3\|(u_h - u)|\nabla u_h|\|_{L_s}\|\nabla\psi\|_{L_{s'}} \\ &\le \beta_3\|u_h - u\|_{L_{\bar{p}}}\|\nabla u_h\|_{L_p}\|\nabla\psi\|_{L_{s'}} \le \beta_3 M\|\nabla(u_h-u)\|\,\|\nabla\psi\|_{L_{s'}} \\ &\le \beta_3 M(\|\nabla(u_h-\chi)\|_{L_s} + \|\nabla(u-\chi)\|_{L_s})\|\nabla\psi\|_{L_{s'}}.\end{aligned} \tag{4.6}$$

The remaining term $II$ can be bounded using Lemma 2.3 and (2.14), the inverse inequality (2.17) and the Hölder inequality (2.27), with $t = 2q/(2-q)$ and $\bar{t} = st/(t-s)$,

$$|\varepsilon_h(f,\psi)| \le Ch\|f\|\|\nabla\psi\| \le Ch^{2-2/s'}\|f\|\,\|\nabla\psi\|_{L_{s'}} = Ch^{2/s}\|f\|\,\|\nabla\psi\|_{L_{s'}}, \tag{4.7}$$

and

$$\begin{aligned}|\varepsilon_a(u_h; u_h, \psi)| &\le Ch(\|\nabla u_h \cdot \nabla(u_h-u)\|_{L_s} + \|\nabla u_h \cdot \nabla u\|_{L_s})\|\nabla\psi\|_{L_{s'}} \\ &\le C\big(h^{1-2/p}M\|\nabla(u_h-u)\|_{L_s} + h\|\nabla u_h\|_{L_{\bar{t}}}\|\nabla u\|_{L_t}\big)\|\nabla\psi\|_{L_{s'}} \\ &\le C\big(h^{1-2/p}M\|\nabla(u_h-u)\|_{L_s} + h^{1+2/\bar{t}-2/p}M\|\nabla u\|_{L_t}\big)\|\nabla\psi\|_{L_{s'}}.\end{aligned} \tag{4.8}$$

Further, in view of the Sobolev imbedding, cf., e.g., [1]

$$\|v\|_{L_t} \leq C\|v\|_{W_r^1}, \quad \forall v \in W_r^1,\ r \leq 2,\ \text{and } t \leq 2r/(2-r), \tag{4.9}$$

and

$$1 + \frac{2}{\bar{t}} - \frac{2}{p} = 1 - \frac{2}{t} + \frac{2}{s} - \frac{2}{p} = 2 - \frac{2}{q} + \frac{2}{s} - \frac{2}{p} > 1 + \frac{2}{s} - \frac{2}{q},$$

relation (4.8) becomes

$$|\varepsilon_a(u_h; u_h, \psi)| \leq C\big(h^{1-2/p} M \|\nabla(u_h - u)\|_{L_s} + h^{1+2/s-2/q} M \|u\|_{W_q^2}\big)\|\nabla\psi\|_{L_{s'}}. \tag{4.10}$$

Thus (4.4)–(4.10) and the fact that $1 - 2/q \leq 0$, give

$$\begin{aligned}(1-\gamma)\|\nabla(u_h - \chi)\|_{L_s} \leq (\gamma + \alpha\beta_2)\|\nabla(u-\chi)\|_{L_s} + Ch^{1-2/p}\|\nabla(u_h - u)\|_{L_s} \\ + Ch^{1+2/s-2/q}(\|u\|_{W_q^2} + \|f\|).\end{aligned} \tag{4.11}$$

Finally, for sufficiently small $h$, the estimate above combined with (4.2) and (4.3) give the desired result. □

**Corollary 4.2.** *Let $u_h$ and $u$ be the solutions of (2.3) and (1.1), respectively, with $f \in L_2$. Then, if $\gamma = \alpha\beta_3 M < 1$ there exists a constant $C_s = C_s(u, f)$, independent of $h$, such that for $0 < h \leq h_M$*

$$\|u - u_h\|_{L_\infty} \leq C_s(u, f) h^{1+2/s-2/q}, \quad \text{with } 2 < s < p < 2 + \epsilon. \tag{4.12}$$

*Proof.* In view of the Sobolev imbedding $\|v\|_{L_\infty} \leq C_s\|\nabla v\|_{L_s}$, $s > 2$ and Theorem 4.1 we can easily see that (4.12) holds. □

Note that the constant $C_s$ in Corollary 4.2 gets larger as $s \to 2$. Later, in Theorem 4.6, we will show an almost optimal order $L_\infty$ error estimate.

One drawback in the proof of Theorem 4.1 is that we require the data $f$ to be sufficiently small. This is evident from (4.11) in which we need $\gamma < 1$, where $\gamma$ depends on $M$. We can actually discard this restriction provided that the solution $u$ of (1.1) belongs in $W_p^1 \cap H_0^1$ and is the unique weak solution in $H_0^1 \cap L_\infty$. In the next theorem we show a different version of Theorem 4.1. The result is based on similar estimate for the finite element method in [24].

**Theorem 4.3.** *If the solution $u$ of (1.1) belongs to $H_0^1 \cap W_p^1$, with $p > 2$ and is unique in $H_0^1 \cap L_\infty$ then $u_h \to u$ in $H_0^1$*

*Proof.* We will follow the proof of Theorem 4.1 for $s = 2$ and [24]. Repeating the arguments in (4.6) we get

$$|I| = |a(u; u_h, \psi) - a(u_h; u_h, \psi)| \leq \beta_3 M \|u_h - u\|_{L_{\bar{p}}} \|\nabla\psi\|_{L_2}. \tag{4.13}$$

Also, we can reevaluate the bounds in (4.7) and (4.8) by

$$\begin{aligned}|II| &\leq Ch\big(\|f\|_{L_2} + \|\nabla u_h \cdot \nabla(u_h - u)\|_{L_2} + \|\nabla u_h \cdot \nabla u\|_{L_2}\big)\|\nabla\psi\|_{L_2} \\ &\leq C\big(h\|f\|_{L_2} + h^{1-2/p} M \|\nabla(u_h - u)\|_{L_2} + h\|\nabla u_h\|_{L_{\bar{p}}}\|\nabla u\|_{L_p}\big)\|\nabla\psi\|_{L_2} \\ &\leq C\big(h\|f\|_{L_2} h^{2/\bar{p}} M \|\nabla(u_h - u)\|_{L_2} + h^{4/\bar{p}} M \|\nabla u\|_{L_p}\big)\|\nabla\psi\|_{L_2} \\ &\leq \big(o(h)\|\nabla(u_h - u)\|_{L_2} + o(h)\big)\|\nabla\psi\|_{L_2}.\end{aligned}$$

Then, combining these with the approximation property of $X_h$

$$\inf_{\chi \in X_h} \|\nabla(v-\chi)\|_{L_2} \le Ch^{1-2/p}\|v\|_{W_p^1}, \quad \text{with } p>2,$$

and (4.2), (4.4) and (4.5), we get

$$\|\nabla(u_h-u)\|_{L_2} \le \alpha\beta_3 M\|u_h-u\|_{L_{\bar{p}}} + o(h).$$

Then we use the Gaglardo–Nirenberg estimate

$$\|v\|_{L_{\bar{p}}} \le C\|\nabla v\|_{L_2}^k \|v\|_{L_2}^{1-k}, \quad \text{with } k = 1-2/\bar{p},$$

and Young's inequality to obtain

$$\|\nabla(u_h-u)\|_{L_2} \le C\|u_h-u\|_{L_2} + o(h),$$

with a large constant $C$. Therefore in order to prove this theorem it suffices to show $u_h \to u$ in $L_2$.

In view of the fact that $u_h \in \mathcal{B}_M$, we have $\|\nabla u_h\|_{L_2} \le K_1$ and $\|u_h\|_{L_\infty} \le K_2$. Therefore, for a subsequence $u_{h_k} \to w$ weakly in $H_0^1$ and $u_{h_k} \to w$ $\star$–weakly in $L_\infty$, cf., e.g., [9]. This means that $u_h \to w$ strongly in $L_{\bar{p}}$. Finally we will show that $w$ is the weak solution of (1.1), and since we have assumed that the weak solution is unique in $H_0^1 \cap L_\infty$, we have $w = u$. Thus $u_h \to u$ in $H^1$.

In order to show that $w$ is a weak solution we consider an arbitrary $v \in C_0^\infty$ and a sequence $\{v_h\}$ in $X_h$ satisfying $\|\nabla(v_h - v)\|_{L_p} \to 0$. This implies

$$(f, v_h - v) \le \|f\|_{L_2}\|\nabla(v_h-v)\|_{L_p} \to 0.$$

Also

$$\begin{aligned} (f, v_h) &= a(u_h; u_h, v_h) + \varepsilon_a(u_h; u_h, v_h) && (4.14)\\ &= ((A(u_h) - A(w))\nabla u_h, \nabla v_h) + a(w; u_h, v_h - v) + a(w; u_h, v). && (4.15)\end{aligned}$$

Since $u_h \to w$ weakly in $H_0^1$,

$$a(w; u_h, v) \to a(w; w, v).$$

Also

$$((A(u_h) - A(w))\nabla u_h, \nabla v_h) \le L\|u_h - w\|_{L_{\bar{p}}}\|\nabla u_h\|_{L_p}\|\nabla v_h\|_{L_2}$$

Since $\|u_h - w\|_{L_{\bar{p}}} \to 0$, $((A(u_h) - A(w))\nabla u_h, \nabla v_h) \to 0$. Finally,

$$a(w; u_h, v_h - v) \le M\|\nabla u_h\|_{L_2}\|\nabla(v_h - v)\|_{L_2} \to 0$$

Therefore, passing to the limits we obtain

$$(f, v) = a(w; w, v), \quad \forall v \in C_0^\infty.$$

This gives that $w$ is the weak solution of (1.1) and completes the proof. □

Next, we will show that $\|\nabla u_h\|_{L_{\bar{q}}}$, with $2/q + 2/\bar{q} = 1$, is also bounded . This will be used later in the proof of error estimate in $L_2$–norm.

**Theorem 4.4.** *Let $u_h$ and $u$ be the solutions of* (2.3) *and* (1.1)*, respectively, with $u \in W_q^2 \cap H_0^1$, $4/3 < q \le 2$. Then $u_h \in W_{\bar{q}}^1$, uniformly for all $0 < h \le h_M$, i.e.,*

$$\|\nabla u_h\|_{L_{\bar{q}}} \le C(u,f), \quad \textit{with } \frac{2}{q} + \frac{2}{\bar{q}} = 1. \tag{4.16}$$

*Proof.* Let $R_h : H_0^1 \to X_h$ be the elliptic projection operator defined by

$$a(u; R_h u, \chi) = a(u; u, \chi), \quad \forall \chi \in X_h,$$

and $\Pi_h : C(\Omega) \to X_h$ be the standard nodal interpolant. Then we have

$$\begin{aligned} \|\nabla u_h\|_{L_{\bar q}} &\le \|\nabla(u_h - R_h u)\|_{L_{\bar q}} + \|\nabla R_h u\|_{L_{\bar q}} \\ &\le \|\nabla(u_h - R_h u)\|_{L_{\bar q}} + \|\nabla(R_h u - \Pi_h u)\|_{L_{\bar q}} + \|\nabla \Pi_h u\|_{L_{\bar q}}. \end{aligned} \tag{4.17}$$

Using the approximation property (4.3), $\Pi_h$ satisfies

$$\|\nabla(\Pi_h v - v)\|_{L_s} \le C h^{1+2/s-2/q} \|v\|_{W_q^2}, \quad 4/3 < q \le 2 \le s. \tag{4.18}$$

In view of (4.9) and (4.18), the last term in (4.17) can easily be estimated as

$$\|\nabla \Pi_h u\|_{L_{\bar q}} \le C \|u\|_{W_q^2}. \tag{4.19}$$

Also, we can easily see that the identity

$$a(u; R_h u - u, R_h u - u) = a(u; R_h u - u, \Pi_h u - u),$$

gives

$$\|\nabla(R_h u - u)\| \le C \|\nabla(\Pi_h u - u)\|.$$

Thus, using the inverse inequality (2.17), (4.18) and the fact that $2 - 2/q = 1 - 2/\bar q$, we can bound the second term in (4.17) by

$$\begin{aligned} \|\nabla(R_h u - \Pi_h u)\|_{L_{\bar q}} &\le C h^{2/\bar q - 1} \|\nabla(R_h u - \Pi_h u)\| \\ &\le C h^{2/\bar q - 1} (\|\nabla(R_h u - u)\| + \|\nabla(\Pi_h u - u)\|) \le C \|u\|_{W_q^2} \end{aligned} \tag{4.20}$$

Finally, the first term in (4.17) can be estimated similarly. Applying Theorem 4.1, (2.17) and the fact that $2 - 2/q = 1 - 2/\bar q$ we have

$$\begin{aligned} \|\nabla(u_h - R_h u)\|_{L_{\bar q}} &\le C h^{2/\bar q - 1} \|\nabla(u_h - R_h u)\| \\ &\le C h^{2/\bar q - 1} (\|\nabla(u_h - u)\| + \|\nabla(R_h u - u)\| + \|\nabla(\Pi_h u - u)\|) \\ &\le C(u, f). \end{aligned} \tag{4.21}$$

All these estimates establish the claim in the theorem. □

For the proof of the $L_2$–norm error estimate we will employ a similar duality argument as the one used in [16]. Let us consider the following auxiliary problem. Let $\varphi \in H_0^1$ be such that

$$a(u; \varphi, v) + (A'(u) \nabla u \nabla \varphi, v) = (u - u_h, v), \quad \forall v \in H_0^1. \tag{4.22}$$

If $A(u)$ is Lipschitz continuous and $A'(u)\nabla u \in L_\infty$, then the solution $\varphi$ of (4.22) satisfies the following elliptic regularity estimate,

$$\|\varphi\|_{W_{q_0}^2} \le C \|u_h - u\|, \quad \text{with } 4/3 < q_0 \le 2, \tag{4.23}$$

where $q_0$ depends on the biggest interior angle of $\Omega$ and the coefficients $A(u), A'(u)\nabla u$. If $\Omega$ is convex then $q_0 = 2$, and if it is nonconvex then $q_0 < 2$.

**Theorem 4.5.** *Let $u_h$ and $u$ be the solutions of* (2.3) *and* (1.1), *respectively, with $u \in W_q^2 \cap H_0^1 \cap W_\infty^1$, $4/3 < q \leq 2$. Then, if $u$ and $A'$ are Lipschitz continuous, $A'' \in L_1(\mathbb{R})$, $f \in H^1$ and $\gamma = \beta_1^{-1}\beta_3 M < 1$ there exists a constant $C$, independent of $h$, such that for sufficiently small $h$,*

$$\|u_h - u\| \leq C(u,f)h^{4-2/q-2/q_0}. \tag{4.24}$$

*Proof.* Before we begin the proof we note the following Taylor expansions

$$\begin{aligned} A(u_h) - A(u) &= (u_h - u)\int_0^1 A'(u - t(u-u_h))\,dt \equiv (u_h - u)\bar{A}', \\ A(u_h) - A(u) - A'(u)(u_h - u) &= (u_h - u)^2\int_0^1 A''(u - t(u-u_h))(1-t)\,dt \\ &\equiv (u_h - u)^2\bar{A}''. \end{aligned} \tag{4.25}$$

Then, in view of (4.22), we have

$$\begin{aligned} \|u - u_h\|^2 &= a(u; u - u_h, \varphi) + (A'(u)(u-u_h)\nabla u, \nabla\varphi) \\ &= a(u;u,\varphi) - a(u_h;u_h,\varphi) + ((A(u_h) - A(u))\nabla u_h, \nabla\varphi) \\ &\quad - ((A(u_h) - A(u))\nabla u, \nabla\varphi) + ((A(u_h) - A(u))\nabla u, \nabla\varphi) - (A'(u)(u_h - u)\nabla u, \nabla\varphi) \\ &= a(u;u,\varphi) - a(u_h;u_h,\varphi) + ((A(u_h) - A(u))\nabla(u_h - u), \nabla\varphi) \\ &\quad + ((A(u_h) - A(u) - A'(u)(u_h - u))\nabla u, \nabla\varphi). \end{aligned}$$

Furthermore, using (2.3) and (4.25), the relation above gives for any $\chi \in X_h$,

$$\begin{aligned} \|u - u_h\|^2 &= a(u;u,\varphi - \chi) - a(u_h;u_h,\varphi-\chi) + \varepsilon_h(f,\chi) - \varepsilon_a(u_h;u_h,\chi) \\ &\quad + ((u_h - u)\bar{A}'\nabla(u_h - u) + (u_h - u)^2\bar{A}''\nabla u, \nabla\varphi) \\ &= \{a(u_h; u - u_h, \varphi - \chi) + ((u_h - u)\bar{A}'\nabla u, \nabla(\varphi - \chi)) + \varepsilon_h(f,\chi)\} \\ &\quad - \varepsilon_a(u_h;u_h,\chi) + \{(u_h - u)\bar{A}'\nabla(u_h - u) + ((u_h - u)^2\bar{A}''\nabla u, \nabla\varphi)\} \\ &= I_1 + I_2 + I_3. \end{aligned} \tag{4.26}$$

Choosing now $\chi = \Pi_h\varphi$ in (4.26) and using (2.14) and Lemma 2.4 we get

$$\begin{aligned} |I_1| &\leq C(\|\nabla(u_h - u)\| + \|\nabla u\|_{L_\infty}\|u_h - u\|)\|\nabla(\varphi - \Pi_h\varphi)\| + Ch^2\|f\|_{H^1}\|\nabla\Pi_h\varphi\|, \\ |I_2| &\leq C\{h^2\|\nabla u_h\|_{L_\infty}(\|\,|\nabla u_h|^2\| + \|u\|_{W_q^2}) + h\|\nabla u_h \cdot \nabla(u_h - u)\|_{L_q}\}\|\nabla\Pi_h\varphi\|_{L_{q'}}. \end{aligned} \tag{4.27}$$

Since $2 < \bar{q} = 2q/(2-q)$, (4.19), the approximation property (4.18) and the fact that $2 \geq 3 - 2/q_0$, now give

$$|I_1| \leq Ch^{2-2/q_0}(\|\nabla(u - u_h)\| + \|\nabla u\|_{L_\infty}\|u - u_h\| + h\|f\|_{H^1})\|\varphi\|_{W_{q_0}^2}. \tag{4.28}$$

Using then Theorem 4.1 and (4.23), we obtain

$$\begin{aligned} |I_1| &\leq C(u)h^{2-2/q_0}\{\|\nabla(u_h - u)\| + h\|f\|_{H^1} + \|u_h - u\|\}\|u_h - u\| \\ &\leq C(u,f)h^{4-2/q-2/q_0}\|u_h - u\| + C(u,f)h^{2-2/q_0}\|u_h - u\|^2. \end{aligned} \tag{4.29}$$

Also, using the fact that $q, q_0 > 4/3$ we get $q' \leq 2q_0/(2-q_0)$, thus in view of (4.9) and (4.18),

$$\|\nabla\Pi_h\varphi\|_{L_{q'}} \leq C\|\varphi\|_{W_{q_0}^2}.$$

Using this relation, along with the inverse inequality (2.17), the Hölder inequality (2.27), with $s = 2$, $t = \bar{q}$ and $s = q$, $t = 2$, and the fact that $2\bar{q}/(\bar{q}-2) \le \bar{q}$, for $q > 4/3$, we get

$$\begin{aligned}|I_2| &\le C\big\{h^{2-2/\bar{q}}\|\nabla u_h\|_{L_{\bar{q}}}(\|\nabla u_h\|_{L_{\bar{q}}}\|\nabla u_h\|_{L_{2\bar{q}/(\bar{q}-2)}} + \|u\|_{W^2_q}) \\ &\quad + h\|\nabla u_h\|_{L_{\bar{q}}}\|\nabla(u_h - u)\|)\big\}\|\nabla \Pi_h \varphi\|_{L_{q'}} \\ &\le C\|\nabla u_h\|_{L_{\bar{q}}}\big\{h^{2-2/\bar{q}}(\|\nabla u_h\|^2_{L_{\bar{q}}} + \|u\|_{W^2_q}) + h\|\nabla(u_h-u)\|\big\}\|\varphi\|_{W^2_{q_0}}.\end{aligned}$$

Next using Theorems 4.1 and 4.4 and (4.23), we obtain

$$|I_2| \le C(u,f)(h^{2-2/\bar{q}} + h\|\nabla(u_h - u)\|)\|u_h - u\| \le C(u,f)h^{3-2/q}\|u-u_h\|. \tag{4.30}$$

Next, we turn to the estimate of the term $I_3$ in (4.26). For this we use the Hölder inequality (2.27) with $t = q_0$; hence

$$|I_3| \le C\|\nabla(u_h-u)\|\,\|(u-u_h)\nabla\varphi\| \le C\|\nabla(u_h-u)\|\,\|u_h-u\|_{L_{q_0}}\|\nabla\varphi\|_{L_{\bar{q}_0}}. \tag{4.31}$$

Then the interpolation inequality, cf., e.g., [18, Appendix B],

$$\|v\|_{L_{q_0}} \le \|v\|^{1/2}\|v\|^{1/2}_{L_s}, \text{ with } s = 2q_0/(4-q_0),$$

and the Sobolev inequality (2.28) give

$$\|u_h - u\|_{L_{q_0}} \le C\|\nabla(u_h-u)\|^{1/2}\|u_h-u\|^{1/2}.$$

Therefore, using the above relation and Theorem 4.1 in (4.31) give

$$\begin{aligned}|I_3| &\le C\|\nabla(u-u_h)\|^{3/2}\|u-u_h\|^{1/2}\|\varphi\|_{W^2_{q_0}} \le (C\|\nabla(u-u_h)\|^3 + \frac{1}{2}\|u-u_h\|)\|u-u_h\| \\ &\le C(u,f)h^{3(2-2/q)}\|u-u_h\| + \frac{1}{2}\|u-u_h\|^2.\end{aligned}$$

We can easily see that $3(2-2/q) > 4 - 2/q - 2/q_0$. Therefore, combining the relation above with (4.26), (4.29) and (4.30), we get

$$\begin{aligned}\|u-u_h\|^2 &\le |I_1| + |I_2| + |I_3| \\ &\le C(u,f)h^{4-2/q-2/q_0}\|u_h-u\| + C(u,f)h^{2-2/q_0}\|u_h-u\|^2 + C(u,f)h^{3-2/q}\|u-u_h\| \\ &\quad + C(u,f)h^{3(2-2/q)}\|u-u_h\| + \frac{1}{2}\|u-u_h\|^2,\end{aligned}$$

which for sufficiently small $h$ gives the desired estimate. □

**Theorem 4.6.** *Let $u_h$ and $u$ be the solutions of* (2.3) *and* (1.1)*, respectively. Then, if $\Omega$ is convex, $\gamma = C_\Omega \beta_1^{-1}\beta_2\beta_3\|u\|_{W^1_p} < 1$, with $C_\Omega > 0$ a constant depending only on $\Omega$, $u \in W^2_\infty$ and $f \in L_\infty$, then there exists a constant $C$ independent of $h$, such that for sufficiently small $h$,*

$$\|u-u_h\|_{L_\infty} \le C(u,f)h^2\log(\frac{1}{h}). \tag{4.32}$$

*Proof.* Using again a triangle inequality we get

$$\|u_h - u\|_{L_\infty} \leq \|\underline{u}_h - u\|_{L_\infty} + \|u_h - \underline{u}_h\|_{L_\infty},$$

where $\underline{u}_h$ is the Galerkin finite element approximation of $u$, i.e.,

$$a(\underline{u}_h; \underline{u}_h, \chi) = (f, \chi), \quad \forall \chi \in X_h. \tag{4.33}$$

In the case of the linear problem $-\mathrm{div}\,(A(x)\nabla w) = f$, we have

$$\|w_h - w\|_{L_\infty} \leq Ch^2 \log(\frac{1}{h})\|w\|_{W^2_\infty},$$

for $A \in W^2_\infty$, cf., eg., [8]. Here $w_h$ is the finite element approximation of $w$. Since $f \in L_\infty$ and $u \in W^2_\infty$, then $A(u) \in W^2_\infty$. Therefore,

$$\|R_h u - u\|_{L_\infty} \leq C(u)h^2 \log(\frac{1}{h}). \tag{4.34}$$

Moreover, it has been shown in [24] that

$$\|\underline{u}_h - R_h u\|_{L_\infty} \leq \gamma \|\underline{u}_h - u\|_{L_\infty}, \tag{4.35}$$

with $\gamma = C_\Omega \beta_1^{-1}\beta_2\beta_3\|u\|_{W^1_p}$. Thus (4.34) and (4.35) give

$$(1-\gamma)\|\underline{u}_h - u\|_{L_\infty} \leq C(u)h^2 \log(\frac{1}{h}). \tag{4.36}$$

Next, we turn our attention to the estimate of $\|\underline{u}_h - u_h\|_{L_\infty}$. Let $x_0 \in K_0 \in \mathcal{T}_h$ such that $\|\underline{u}_h - u_h\|_{L_\infty} = |(\underline{u}_h - u_h)(x_0)|$ and $\delta_{x_0} = \delta \in C_0^\infty(\Omega)$ a regularized Dirac $\delta$–function satisfying

$$(\delta, \chi) = \chi(x_0), \quad \forall \chi \in X_h.$$

For such a function $\delta$, cf., e.g., [8], we have

$$\begin{aligned} &\mathrm{supp}\,\delta \subset B = \{x \in \Omega : |x - x_0| \leq h/2\}, \\ &\int_\Omega \delta = 1, \quad 0 \leq \delta \leq Ch^{-2}, \quad \|\delta\|_{L_p} \leq Ch^{2(1-p)/p}, \quad 1 < p < \infty. \end{aligned}$$

Also let us consider the corresponding regularized Green's function $G \in H^1_0$, defined by

$$a(\underline{u}_h; G, v) = (\delta, v), \quad \forall v \in H^1_0. \tag{4.37}$$

Then, we have

$$\begin{aligned} \|\underline{u}_h - u_h\|_{L_\infty} &= (\delta, \underline{u}_h - u_h) = a(\underline{u}_h; G, \underline{u}_h - u_h) = a(\underline{u}_h; G_h, \underline{u}_h - u_h) \\ &= (f, G_h) - a(\underline{u}_h; u_h, G_h) \\ &= \varepsilon_h(f, G_h) - \varepsilon_a(u_h; u_h, G_h) + \{a(\underline{u}_h; u_h, G_h) - a(u_h; u_h, G_h)\}, \end{aligned} \tag{4.38}$$

where $G_h \in X_h$ is the finite element approximation of $G$, i.e.,

$$a(\underline{u}_h; G, \chi) = a(\underline{u}_h; G_h, \chi), \quad \forall \chi \in X_h.$$

Since $u \in W^2_\infty$, we have $u \in H^2$. Thus, it follows from Theorem 4.4 that $\|\nabla u_h\|_{L_\infty} \leq C$. Furthermore, using Lemma 2.4 and (2.14), (1.2) and Theorem 4.5, we obtain

$$\begin{aligned}\|\underline{u}_h - u_h\|_{L_\infty} &\leq C\big\{h^2(\|f\|_{H^1} + \|\nabla u_h\|^2_{L_\infty}\|\nabla u_h\| + \|\nabla u_h\|_{L_\infty}\|u\|_{H^2}) \\ &\quad + h\|\nabla u_h\|_{L_\infty}\|\nabla(u_h - u)\| + \|(\underline{u}_h - u_h)|\nabla u_h|\|\big\}\|\nabla G_h\| \\ &\leq Ch^2(\|f\|_{H^1} + \|u\|_{H^2} + \|\underline{u}_h - u\|)\|\nabla G_h\|.\end{aligned} \tag{4.39}$$

The last term can be estimated by, cf., e.g., [16],

$$\|\underline{u}_h - u\| \leq C(u, f)h^2. \tag{4.40}$$

In addition in view of [26, Lemma 3.1] we get

$$\|G_h\|_{H^1} \leq C\|\nabla G\|_{L_2} \leq C\frac{1}{(s-1)^{1/2}}\|\delta\|_{L_s}, \tag{4.41}$$

with $s \downarrow 1$. Choosing now $s = 1 + (\log(1/h))^{-1}$ we have

$$\|G_h\|_{H^1} \leq C(\log(\frac{1}{h}))^{1/2}. \tag{4.42}$$

Combining (4.38)–(4.42), we obtain

$$\|\underline{u}_h - u_h\|_{L_\infty} \leq C(u, f)h^2 \log(\frac{1}{h})^{1/2}. \tag{4.43}$$

¿From this and (4.36) for $\gamma < 1$ we get (4.32).

## 5. NEWTON'S METHOD

In this section we shall analyze Newton's method for the computation of the finite volume solution $u_h$ of (2.3). Our analysis is based on a similar approach used in the finite element method, studied by Douglas and Dupont in [16].

Also here, we will assume that (1.1) has a unique solution $u \in H^2 \cap H^1_0$. For $\phi \in H^1$ we define the bilinear form $N(\phi;\cdot,\cdot)$ on $H^1_0 \times H^1_0$ by

$$N(\phi; v, w) = a(\phi; v, w) + d(\phi; v, w), \tag{5.1}$$

where $d$ is given by

$$d(\phi; v, w) = (A'(\phi)v\nabla\phi, \nabla w). \tag{5.2}$$

Furthermore, let $N_h$ be the form corresponding to N, associated with the finite volume method defined as

$$N_h(\phi; v, w) = a_h(\phi; v, w) + d_h(\phi; v, w), \tag{5.3}$$

for $\phi \in H^2 \cap H^1_0$ on $(H^2 \cap H^1_0) + X_h \times (H^2 \cap H^1_0) + X_h$, and $d_h$ is given by

$$d_h(\phi; v, w) = -\sum_K \int_K \operatorname{div}(A'(\phi)v\nabla\phi)I_h w\,dx + \int_{\partial K}(A'(\phi)v\nabla\phi)\cdot nI_h w\,ds. \tag{5.4}$$

For $u_h^0 \in X_h$, the Newton approximations to the solution $u_h$ forms a sequence $\{u_h^k\}_{k=0}^{\infty}$ in $X_h$ satisfying

$$N_h(u_h^k; u_h^{k+1} - u_h^k, \chi) = (f, I_h\chi) - a_h(u_h^k; u_h^k, I_h\chi), \quad \forall \chi \in X_h. \tag{5.5}$$

We will show that $u_h^k \to u_h$ in $H^1$–norm as $k \to \infty$, with order two, provided that $u_h^0$ is sufficiently close to $u_h$. For this we will assume that $u_h$ converges to $u$ sufficiently fast,

$$\|u - u_h\|_{L_\infty} + \sigma_h \|u - u_h\|_{H^1} \to 0, \quad \text{as } h \to 0, \tag{5.6}$$

where

$$\sigma_h \equiv \sup\{\|\chi\|_{L_\infty} / \|\chi\|_{H^1} : 0 \neq \chi \in X_h\}. \tag{5.7}$$

Since $\mathcal{T}_h$ is a quasi-uniform mesh, there exists a constant $C$, independent of $h$ such that

$$|\sigma_h| \leq C \log(\frac{1}{h}). \tag{5.8}$$

Further, let $C_3$ be another constant, independent of $h$, satisfying

$$\|u_h\|_{W_\infty^1} \leq C_3. \tag{5.9}$$

Note that this assumption holds, for $u \in H^2$, cf. Section 3. In addition we assume that $A''$ is bounded and is Lipschitz continuous, i.e.,

$$|A''(x)| \leq \beta_4, \quad |A''(x) - A''(y)| \leq L_2|x - y|, \quad \forall x, y \in \mathbb{R}. \tag{5.10}$$

Next, we will show various auxiliary results which will be used in the proof of Theorem 5.1. We start by stating the following lemma of Douglas and Dupont, [16].

**Lemma 5.1.** *Given $\tau > 0$, there exists positive constants $\delta$, $h_0$ and $C_4$ such that the following holds. If $0 < h < h_0$, if $\phi \in W_\infty^1$ satisfies*

$$\|\phi\|_{W_\infty^1} \leq \tau \quad \textit{and} \quad \sigma_h \|\phi - u\|_{H^1} \leq \delta,$$

*and if $G$ is a linear functional on $H_0^1$ with*

$$|||G||| = \sup_{0 \neq \chi \in X_h} \frac{|G(\chi)|}{\|\chi\|_{H^1}},$$

*then there exists a unique $v \in X_h$ satisfying the equations*

$$N(\phi; v, \chi) = G(\chi), \quad w \in X_h. \tag{5.11}$$

*Furthermore, $v$ satisfies the bound*

$$\|v\|_{H^1} \leq C_4 |||G|||. \tag{5.12}$$

We shall also use the error functional $\epsilon_N$, defined by $\varepsilon_N = N - N_h$, and we derive similar estimates to $\varepsilon_a$, cf. Section 2.

**Lemma 5.2.** *For $\phi \in X_h$ the error functional $\varepsilon_N$ satisfies*

$$|\varepsilon_N(\phi; \psi, \chi)| \leq Ch\|\nabla\phi\|_{L_\infty}(1 + \sigma_h\|\phi\|_{H^1})\|\psi\|_{H^1}\|\chi\|_{H^1}, \quad \forall \chi, \psi \in X_h.$$

*Proof.* ¿From the definition of $\varepsilon_N$ we can easily see that, $\varepsilon_N = \varepsilon_a + (d - d_h)$. Therefore in view of Lemma 2.3, it suffices to bound $d - d_h$. Following the proof of Lemma 2.3 we have,

$$\begin{aligned} d(\phi;\psi,\chi) &- d_h(\phi;\psi,\chi) \\ &= \sum_K \left\{ (\operatorname{div}\big((A'(\phi)\psi)\nabla\phi\big), \chi - I_h\chi)_K + \big((A'(\phi)\psi)\nabla\phi\big)\cdot n, \chi - I_h\chi)_{\partial K} \right\} \\ &= \sum_K \{I_K + II_K\}. \end{aligned} \tag{5.13}$$

Applying Hölder's inequality to $I_K$, and using the fact that $\phi$ is linear in $K$, (1.2), (5.10) and (2.11), we have

$$\begin{aligned} |I_K| &\le (\beta_3\|\nabla\phi\cdot\nabla\psi\|_{L_2(K)} + \beta_4\|\,|\nabla\phi|^2\psi\|_{L_2(K)})\|\chi - I_h\chi\|_{L_2(K)} \\ &\le Ch(\beta_3\|\nabla\phi\cdot\nabla\psi\|_{L_2(K)} + \beta_4\|\,|\nabla\phi|^2\psi\|_{L_2(K)})\|\nabla\chi\|_{L_2(K)}. \end{aligned} \tag{5.14}$$

For the $II_K$, we break the integration over the boundary of each triangle $K$, into the sum of integrations over its sides, and thus may use (2.12), and follow the same steps as in estimating $I_K$. Hence,

$$\begin{aligned} |II_K| &\le Ch|(A'(\phi)\psi)\nabla\phi|_{H^1(K)}\|\nabla\chi\|_{L_2(K)} \\ &\le Ch(\beta_3\|\nabla\phi\cdot\nabla\psi\|_{L_2(K)} + \beta_4\|\,|\nabla\phi|^2\psi\|_{L_2(K)})\|\nabla\chi\|_{L_2(K)}. \end{aligned}$$

Then combining this with Lemma 2.3 and (5.14), we get

$$|\varepsilon_N(\phi;\psi,\chi)| \le Ch\big(\|\nabla\phi\|_{L_\infty}\|\nabla\psi\| + \|\nabla\phi\|_{L_\infty}\|\psi\|_{L_\infty}\|\nabla\phi\|\big)\|\chi\|_{H^1}.$$

Finally, in view of the definition of $\sigma_h$ we get the desired estimate. □

Next, we derive a "Lipchitz"–type estimate for $\varepsilon_N$.

**Lemma 5.3.** *Let $v, w, \phi, \chi \in X_h$ then*

$$\begin{aligned} |\varepsilon_N(v;\phi,\chi) - \varepsilon_N(w;\phi,\chi)| \le Ch\big\{&\|\nabla(v-w)\cdot\nabla\phi\| + \|\nabla w\|_{L_\infty}\|(v-w)\nabla\phi\| \\ &+ \|(|\nabla v|^2 - |\nabla w|^2)\phi\| + \|\nabla w\|^2_{L_\infty}\|(v-w)\phi\|\big\}\|\nabla\chi\|. \end{aligned} \tag{5.15}$$

*Proof.* Similarly as in the proof of the previous lemma, we can easily see that $\varepsilon_N = \varepsilon_a + (d - d_h)$. Thus in view of Lemma 2.5, it suffices to estimate $d(v;\phi,\chi) - d_h(w;\phi,\chi)$. Using a similar decomposition as in (5.13) and then applying (2.11) and (2.12) we get

$$\begin{aligned} |d(v;\phi,\chi) - d_h(w;\phi,\chi)| \le Ch\big\{&\|\operatorname{div}\big((A'(v)\nabla v - A'(w)\nabla w)\phi\big)\| \\ &+ |(A'(v)\nabla v - A'(w)\nabla w)\phi|_{H^1}\big\}\|\nabla\chi\|. \end{aligned} \tag{5.16}$$

Next, since $\phi \in X_h$, we have

$$\begin{aligned} &\operatorname{div}((A'(v)\nabla v - A'(w)\nabla w)\phi) \\ &= (A''(v)|\nabla v|^2 - A''(w)|\nabla w|^2)\phi + (A'(v)\nabla v - A'(w)\nabla w)\cdot\nabla\phi \\ &= (A''(v)(|\nabla v|^2 - |\nabla w|^2)\phi + (A''(v) - A''(w))|\nabla w|^2\phi \\ &\quad + (A'(v)(\nabla v - \nabla w)\cdot\nabla\phi + (A'(v) - A'(w))\nabla w\cdot\nabla\phi. \end{aligned} \tag{5.17}$$

Therefore, (5.16) gives

$$\begin{aligned}|d(v;\phi,\chi)-d_h(w;\phi,\chi)| &\le Ch(\|\nabla(v-w)\cdot\nabla\phi\|+\|\nabla w\|_{L_\infty}\|(v-w)\nabla\phi\|)\|\nabla\chi\| \\ &\quad + Ch(\|(|\nabla v|^2-|\nabla w|^2)\phi\|+\|\nabla w\|^2_{L_\infty}\|(v-w)\phi\|)\|\nabla\chi\|.\end{aligned} \tag{5.18}$$

Finally, the above inequality and Lemma 2.5 give (5.15). □

Next, we show an error bound that will be employed in the proof of Theorem 5.1.

**Lemma 5.4.** *For $v_h, w_h, \chi \in X_h$, we have*

$$\begin{aligned}&|\varepsilon_N(v_h;w_h-v_h,\chi)+\varepsilon_a(v_h;v_h,\chi)-\varepsilon_a(w_h;w_h,\chi)| \\ &\quad \le Ch\big(\sigma_h(\|\nabla v_h\|^2_{L_\infty}+\|\nabla(w_h+v_h)\|_{L_\infty})+h^{-1}\big)\|w_h-v_h\|^2_{H^1}\|\chi\|_{H^1}.\end{aligned} \tag{5.19}$$

*Proof.* In view of the definition of $\varepsilon_N$ and $\varepsilon_a$ we have

$$\begin{aligned}&\varepsilon_N(v_h;w_h-v_h,\chi)+\varepsilon_a(v_h;v_h,\chi)-\varepsilon_a(w_h;w_h,\chi) \\ &= \sum_K\int_K \operatorname{div}\Big(A(v_h)\nabla(w_h-v_h)+A'(v_h)(w_h-v_h)\nabla v_h+A(v_h)\nabla v_h \\ &\qquad\qquad - A(w_h)\nabla w_h\Big)(\chi-I_h\chi)\,dx \\ &+ \sum_K\int_{\partial K}\Big(A(v_h)\nabla(w_h-v_h)+A'(v_h)(w_h-v_h)\nabla v_h+A(v_h)\nabla v_h \\ &\qquad\qquad - A(w_h)\nabla w_h\Big)\cdot n(\chi-I_h\chi)\,ds.\end{aligned}$$

Then, since $v_h, w_h$ are linear in $K\in\mathcal{T}_h$, we get

$$\begin{aligned}&\operatorname{div}\Big(A(v_h)\nabla(w_h-v_h)+A'(v_h)(w_h-v_h)\nabla v_h+A(v_h)\nabla v_h-A(w_h)\nabla w_h\Big) \\ &= 2A'(v_h)\nabla v_h\cdot\nabla(w_h-v_h)+A''(v_h)(w_h-v_h)|\nabla v_h|^2+A'(v_h)|\nabla v_h|^2-A'(w_h)|\nabla w_h|^2 \\ &= A''(v_h)(w_h-v_h)|\nabla v_h|^2+A'(v_h)|\nabla v_h|^2-A'(w_h)|\nabla v_h|^2 \\ &\quad + A'(w_h)|\nabla v_h|^2-A'(w_h)|\nabla w_h|^2+2A'(v_h)\nabla v_h\cdot\nabla(w_h-v_h).\end{aligned}$$

We consider now similar Taylor expansions as in (4.25) and denoting this $\tilde{A}'$ and $\tilde{A}''$ the expressions corresponding to $\bar{A}'$ and $\bar{A}''$, where we substitute $A$ with $A'$. Then the previous relation gives

$$\begin{aligned}&\operatorname{div}\Big(A(v_h)\nabla(w_h-v_h)+A'(v_h)(w_h-v_h)\nabla v_h+A(v_h)\nabla v_h-A(w_h)\nabla w_h\Big) \\ &= -(w_h-v_h)^2\tilde{A}''|\nabla v_h|^2-A'(w_h)\nabla v_h\cdot\nabla(w_h-v_h)-A'(w_h)\nabla w_h\cdot\nabla(w_h-v_h) \\ &\quad + 2A'(v_h)\nabla v_h\cdot\nabla(w_h-v_h) \\ &= -(w_h-v_h)^2\tilde{A}''|\nabla v_h|^2+(A'(v_h)-A'(w_h))\nabla v_h\cdot\nabla(w_h-v_h) \\ &\quad + (A'(v_h)-A'(w_h))\nabla w_h\cdot\nabla(w_h-v_h)-A'(v_h)|\nabla(w_h-v_h)|^2 \\ &= -(w_h-v_h)^2\tilde{A}''|\nabla v_h|^2+(A'(v_h)-A'(w_h))\nabla(w_h+v_h)\cdot\nabla(w_h-v_h)-A'(v_h)|\nabla(w_h-v_h)|^2 \\ &= -(w_h-v_h)^2\tilde{A}''|\nabla v_h|^2-(w_h-v_h)\tilde{A}'\nabla(w_h+v_h)\cdot\nabla(w_h-v_h)-A'(v_h)|\nabla(w_h-v_h)|^2.\end{aligned}$$

Finally, this combined with (2.11) and (2.12) give the desired estimate

$$\begin{aligned}
&|\varepsilon_N(v_h; w_h - v_h, \chi) + \varepsilon_a(v_h; v_h, \chi) - \varepsilon_a(w_h; w_h, \chi)| \\
&\le Ch\big(\|w_h - v_h\|_{L_\infty} \|\nabla v_h\|^2_{L_\infty} + \|w_h - v_h\|_{L_\infty} \|\nabla(w_h + v_h)\|_{L_\infty} \\
&\qquad + \|w_h - v_h\|_{L_\infty}\big) \|w_h - v_h\|_{H^1} \|\chi\|_{H^1} \\
&\le Ch\big(\sigma_h(\|\nabla v_h\|^2_{L_\infty} + \|(w_h + v_h)\|_{L_\infty}) + h^{-1}\big) \|w_h - v_h\|^2_{H^1} \|\chi\|_{H^1}. \qquad \square
\end{aligned}$$

Next, we show that the Newton sequence obtained by (5.5) is well defined and it converges to the finite volume approximation $u_h$ of (2.3) with order 2.

**Theorem 5.1.** *There exists positive constants $h_0$, $\delta$ and $C_5$ such that if $0 < h \le h_0$ and $\sigma_h \|u_h^0 - u_h\|_{H^1} \le \delta$ then $\{u_h^k\}_{k=0}^\infty$ exists and $\nu_k = \|u_h^k - u_h\|_{H^1}$ is a decreasing sequence satisfying*

$$\nu_{k+1} \le C_5 \sigma_h \nu_k^2. \tag{5.20}$$

*Proof.* The proof is based on a similar result of Douglas and Dupont, [16], for the finite element method. First we show that for $h_0$ and $\delta$ are sufficiently small, and $\sigma_h \|u_h^k - u_h\|_{H^1} = \sigma_h \nu_k \le \delta$, with $0 < h \le h_0$, there exists a unique $u_h^{k+1}$, given by (5.5). It suffices to show that if

$$N_h(u_h^k; v, \chi) = 0, \quad \forall \chi \in X_h,$$

then $v \equiv 0$, or else $\|v\|_{H^1} \le 0$. For this we will employ Lemma 5.1 and demonstrate that $C_4 |||G||| < \|v\|_{H^1}$, for an appropriately defined functional $G$. We can easily see that

$$N(u_h; v, \chi) = G(\chi),$$

where $G$ is given by

$$G(\chi) = N(u_h; v, \chi) - N_h(u_h^k; v, \chi) = \{N(u_h; v, \chi) - N(u_h^k; v, \chi)\} + \varepsilon_N(u_h^k; v, \chi) = I + II,$$

Following the proof in [16] we have that

$$|I| \le C\sigma_h \|u_h - u_h^k\|_{H^1} \|v\|_{H^1} \|\chi\|_{H^1} = C\sigma_h \nu_k \|v\|_{H^1} \|\chi\|_{H^1}. \tag{5.21}$$

For the estimation of $II$ we use the inverse inequality, (2.17), (5.9), Lemma 5.2 and the fact that induction hypothesis and (5.6) give

$$\|u_h^k\|_{H^1} \le \nu_k + \|u_h\|_{H^1} \le \sigma_h^{-1}\delta + \|u_h\|_{H^1} \le C, \tag{5.22}$$

to get

$$\begin{aligned}
|II| &\le C(\nu_k(1 + \sigma_h \|u_h^k\|_{H^1}) + h(1 + \sigma_h \|u_h^k\|_{H^1}) \|u_h\|_{W^1_\infty}) \|v\|_{H^1} \|\chi\|_{H^1} \\
&\le C\sigma_h \nu_k \|v\|_{H^1} \|\chi\|_{H^1} + Ch\sigma_h \|v\|_{H^1} \|\chi\|_{H^1}.
\end{aligned} \tag{5.23}$$

Hence, since $\sigma_h \le C\log(1/h)$, (5.21) and (5.23) give for $\delta$ and $h$ sufficiently small, $\|v\|_{H^1} \le C_0 \sigma_h (\nu_k + h\log(1/h)) \|v\|_{H^1} < \|v\|_{H^1}$; thus $v = 0$.

In order to show (5.20) we will employ again Lemma 5.1 for a different functional $G$. This time let

$$N(u_h; u_h^{k+1} - u_h, \chi) = G(\chi), \quad \forall \chi \in X_h,$$

where $G$ is defined by

$$\begin{aligned} G(\chi) &= N(u_h; u_h^k - u_h, \chi) + N(u_h^k; u_h^{k+1} - u_h^k, \chi) \\ &\quad + N(u_h; u_h^{k+1} - u_h^k, \chi) - N(u_h^k; u_h^{k+1} - u_h^k, \chi) \\ &= \{N(u_h; u_h^k - u_h, \chi) + a(u_h; u_h, \chi) - a(u_h^k; u_h^k, \chi)\} \\ &\quad + \{\varepsilon_N(u_h^k; u_h^{k+1} - u_h^k, \chi) - \varepsilon_a(u_h; u_h, \chi) + \varepsilon_a(u_h^k; u_h^k, \chi)\} \\ &\quad + \{N(u_h; u_h^{k+1} - u_h^k, \chi) - N(u_h^k; u_h^{k+1} - u_h^k, \chi)\} = I + II + III. \end{aligned} \tag{5.24}$$

We will show that

$$|||G||| \le C\sigma_h \nu_k(\nu_k + \nu_{k+1}) + Ch\sigma_h\nu_{k+1}. \tag{5.25}$$

Then Lemma 5.1, and $\sigma_h \nu_k \le \delta$, give

$$\begin{aligned} \nu_{k+1} &\le C_4 |||G||| \le C\sigma_h \nu_k(\nu_k + \nu_{k+1}) + Ch\sigma_h\nu_{k+1} \\ &\le C\sigma_h \nu_k^2 + C(\delta + h\log(\frac{1}{h}))\nu_{k+1}. \end{aligned} \tag{5.26}$$

Finally for sufficiently small $\delta$ and $h$, the desired estimate, (5.20), follows easily.

Let us turn now to the estimation of $|||G|||$, for $G$ given by (5.24). The terms $I$ and $III$ are similar to the ones that appear in the analysis of the finite element method in [16], thus using the same arguments we get

$$|I + III| \le C\sigma_h \nu_k(\nu_k + \nu_{k+1}) \|\chi\|_{H^1}. \tag{5.27}$$

Then, we can easily see that $II$ can be rewritten in the following way,

$$\begin{aligned} II &= \varepsilon_N(u_h^k; u_h^{k+1} - u_h^k, \chi) - \varepsilon_N(u_h; u_h^{k+1} - u_h^k, \chi) \\ &\quad + \varepsilon_N(u_h; u_h^{k+1} - u_h^k, \chi) - \varepsilon_a(u_h; u_h, \chi) + \varepsilon_a(u_h^k; u_h^k, \chi) \\ &= \{\varepsilon_N(u_h^k; u_h^{k+1} - u_h^k, \chi) - \varepsilon_N(u_h; u_h^{k+1} - u_h^k, \chi)\} \\ &\quad + \varepsilon_N(u_h; u_h^{k+1} - u_h, \chi) \\ &\quad - \{\varepsilon_N(u_h; u_h^k - u_h, \chi) + \varepsilon_a(u_h; u_h, \chi) - \varepsilon_a(u_h^k; u_h^k, \chi)\} = II_1 + II_2 + II_3. \end{aligned} \tag{5.28}$$

Using Lemma 5.3, (5.9), inverse inequality, (2.17), (5.6) and (5.22), we can bound $II_1$ in the following way,

$$\begin{aligned} |II_1| &\le Ch\big\{(\|\nabla(u_h^k - u_h)\|_{L_\infty} + \|\nabla u_h\|_{L_\infty}\|u_h^k - u_h\|_{L_\infty})\|\nabla(u_h^{k+1} - u_h^k)\| \\ &\quad + \big(\|\nabla(u_h^k + u_h)\|_{L_\infty}\|\nabla(u_h^k - u_h)\| \\ &\qquad + \|\nabla u_h\|_{L_\infty}^2 \|u_h^k - u_h\|\big)\|u_h^{k+1} - u_h^k\|_{L_\infty}\big\}\|\chi\|_{H^1} \\ &\le C\big((1 + (\|u_h^k + u_h\|_{H^1} + h)\sigma_h)\nu_k(\nu_k + \nu_{k+1})\big)\|\chi\|_{H^1} \\ &\le C\sigma_h\nu_k(\nu_k + \nu_{k+1})\|\chi\|_{H^1}. \end{aligned} \tag{5.29}$$

Further, using Lemma 5.2, (5.9) and (5.6), we can easily bound $II_2$,

$$\begin{aligned} |II_2| &\le Ch(\|\nabla u_h\|_{L_\infty} + \sigma_h\|\nabla u_h\|_{L_\infty}\|u_h\|_{H^1})\|u_h^{k+1} - u_h\|_{H^1}\|\chi\|_{H^1} \\ &\le Ch(1 + \sigma_h)\nu_{k+1}\|\chi\|_{H^1}. \end{aligned} \tag{5.30}$$

Finally using, Lemma 5.4 and the fact that $\|\nabla u_h\|_{L_\infty} \le C_3$ and $h\|\nabla u_h^k\|_{L_\infty} \le C\|u_h^k\|_{H^1} \le C$, $II_3$ can be estimated by

$$\begin{aligned} |II_3| &\le C(h\sigma_h\|\nabla u_h\|^2_{L_\infty} + h\sigma_h\|\nabla(u_h^k + u_h)\|_{L_\infty} + 1)\|u_h^k - u_h\|^2_{H^1}\|\chi\|_{H^1} \\ &\le C(\sigma_h + 1)\nu_k^2\|\chi\|_{H^1}. \end{aligned} \tag{5.31}$$

Therefore combining (5.27) and (5.29)–(5.31), we get the desired (5.25). □

## 6. NUMERICAL IMPLEMENTATIONS

In this section we present procedures for implementing the finite volume method for the nonlinear problem. A series of numerical examples is given to further assess the theories that were preceedingly deduced. Following the previous mathematical works, we implement two iterative methods to solve the nonlinear finite volume problems, namely the fixed point iteration and the Newton iteration. As will be clear in the following subsection, these two methods are built in the finite dimensional setting, i.e., using the finite element space $X_h$. We denote $\{\phi_i\}_{i=1}^d$ to be the standard piecewise linear basis functions of $X_h$. Then the finite volume element solution may be written as

$$u_h = \sum_{i=1}^d \alpha_i\,\phi_i \quad \text{for some} \quad \alpha = (\alpha_1, \alpha_2, \cdots, \alpha_d)^T$$

### *6.1. Fixed Point Iteration vs Newton Iteration*

To describe the method, we begin with several notations, noting that some of them have already been mentioned. Let $Z_h$ be the collection of vertices $z_i$ that belong to all triangles $K \in \mathcal{T}_h$ and $Z_h^0 = \{z_i \in Z_h : z_i \notin \Gamma_D\}$. Let $I = \{i : z_i \in Z_h^0\}$, $I_K = \{m : z_m \text{ is a vertex of } K\}$, $\mathcal{T}_{h,i} = \{K \in \mathcal{T}_h : i \in I_K\}$, and $I_i = \{m \in I : z_m \text{ is a vertex of } K \in \mathcal{T}_{h,i}\}$. Let $V_i$ be the control volume surrounding the vertex $z_i$.

Now we may write this finite volume problem as to find $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_d)^T$ such that

$$F(\alpha) = 0, \tag{6.1}$$

where $F : \mathbb{R}^d \to \mathbb{R}^d$ is a nonlinear operator with

$$F_i(\alpha) = -\int_{\partial V_i} A(u_h)\,\nabla\,u_h \cdot n\,ds - \int_{V_i} f\,dx \qquad \forall\, i \in I. \tag{6.2}$$

The fixed point iteration is derived from the linearization of (6.1) on the coefficient $A(u)$ in (6.2). Thus, given an initial iterate $\alpha^0$ (i.e., equivalently $u_h^0 = \sum_{i=0}^d \alpha_i^0\phi_i$), for $k = 0, 1, 2, \cdots$ until convergence solve the linear system $M(\alpha^k)\,\alpha^{k+1} = q$, where $M(\alpha^k)$ is the resulting stiffness matrix evaluated at $u_h^k = \sum_{i=0}^d \alpha_i^k\phi_i$, whose entries are

$$M_{ij}^k = -\int_{\partial V_i} A(u_h^k)\,\nabla\,\phi_j \cdot n\,ds.$$

On the other hand, the classical Newton iteration relies on the first order Taylor expansion of $F(\alpha)$. It results in solving a linear system of the Jacobian of $F(\alpha)$. An inexact-Newton iteration

is a variation of Newton iteration for nonlinear system of equations in that the system Jacobian is only solved approximately, cf. e.g., [4, 5, 14]. To be specific, given an initial iterate $\alpha^0$, for $k = 0, 1, 2, \cdots$ until convergence do the following:
(a) Solve $F'(\alpha^k)\delta^k = -F(\alpha^k)$ until $\|F(\alpha^k) + F'(\alpha^k)\delta^k\| \le \beta_k \|F(\alpha^k)\|$;
(b) Update $\alpha^{k+1} = \alpha^k + \delta^k$.

In this algorithm $F'(\alpha^k)$ is the Jacobian matrix evaluated at iteration $k$. For iterative technique solving a linear system such as the Krylov method we only need the action of the Jacobian to a vector. It has been common practice to use the following finite difference approximation for such an action:

$$F'(\alpha^k)\, v \;\approx\; \frac{F(\alpha^k + \sigma v) - F(\alpha^k)}{\sigma}, \tag{6.3}$$

where $\sigma$ is a small number computed as follows:

$$\sigma = \frac{\mathrm{sign}(\alpha^k \cdot v)\, \sqrt{\epsilon}\, \max(|\alpha^k \cdot v|, \|v\|_1)}{v \cdot v},$$

with $\epsilon$ being the machine unit round-off number. We note that when $\beta_k = 0$ then we have recovered the classical Newton iteration. One common used relation is

$$\beta_k = 0.001 \left( \frac{\|F(\alpha^k)\|}{\|F(\alpha^{k-1})\|} \right)^2,$$

with $\beta_0 = 0.001$. Choosing $\beta_k$ this way we avoid oversolving the Jacobian system when $\alpha^k$ is still considerably far from the exact solution.

Instead of using (6.3), we will present below an explicit construction of the Jacobian matrix. We note that we may decompose $F_i(\alpha)$ as follows:

$$F_i(\alpha) = \sum_{K \in \mathcal{T}_{h,i}} F_{i,K}(\alpha), \quad \text{where} \quad F_{i,K}(\alpha) = -\int_{K \cap \partial V_i} A(u_h)\, \nabla u_h \cdot n\, ds - \int_{K \cap V_i} f\, dx.$$

¿From the above description it is apparent that $F_i(\alpha)$ is not fully dependent on all $\alpha_1, \alpha_2, \cdots, \alpha_d$. Consequently, $\frac{\partial F_i(\alpha)}{\partial \alpha_j} = 0$ for $j \notin I_i$. Next we want to find an explicit form of $\frac{\partial F_i(\alpha)}{\partial \alpha_j}$ for $j \in I_i$.

Now suppose the edge $\overline{z_i z_j}$ is shared by triangles $K_l, K_r \in \mathcal{T}_{h,i}$. Then

$$\frac{\partial F_i}{\partial \alpha_j} = -\sum_{e=l,r} \int_{K_e \cap \partial V_i} \left( A'(u_h)\, \phi_j \nabla u_h \cdot n + A(u_h)\, \nabla \phi_j \cdot n \right)\, ds.$$

Furthermore,

$$\frac{\partial F_i}{\partial \alpha_i} = -\sum_{K \in \mathcal{T}_{h,i}} \int_{K \cap \partial V_i} \left( A'(u_h)\, \phi_i \nabla u_h \cdot n + A(u_h)\, \nabla\, \phi_i \cdot n \right)\, ds.$$

¿From this derivation it is obvious that the Jacobian matrix is not symmetric but sparse. Computation of this Jacobian matrix is similar to computing the stiffness matrix resulting from standard finite volume element, in that each entry is formed by accumulation of element by element contribution. Once we have the matrix stored in memory, then its action to a vector is straightforward. Since it is a sparse matrix, devoting some amount of memory for entries storage is not very expensive.

*6.2. Numerical Examples*

In this subsection we present several numerical experiments to verify the theoretical investigations. We solve a set of Dirichlet boundary value problems in $\Omega = [0,1] \times [0,1]$. We compare the fixed point iteration and the Newton iteration. In both methods, the iteration is stopped once $\|u_h^k - u_h^{k-1}\|_{L_\infty} < 10^{-10}$. In all examples below, the initial iteration is taken to be $\alpha = (0, 0, \cdots, 0)^T$.

The first example is solving $-\nabla\cdot(k(u)\nabla u) = f$ in $\Omega$ where the function $f$ is chosen such that the known solution is $u(x, y) = (x - x^2)(y - y^2)$ The nonlinearity comes from the coefficient with $k(u) = \frac{1}{(1+u)^2}$. The results are listed in Table I. First column represents the mesh size. The domain is discretized into $N$ numbers of rectangle in each direction. Each of these rectangle is divided into two triangles. Second and third columns correspond to the number of iterations performed until the stopping criteria is reached for fixed point iteration (FP) and Newton iteration (NW), respectively. The table shows that a superconvergence is observed in $H^1$-norm due to the smoothness of the solution. Number of iterations in both methods do not depend on the mesh size. The numerical results for the second example are presented in Table II. Here

Table I. Error of FVEM for nonlinear elliptic BVP, with $u = (x - x^2)(y - y^2)$ and $k(u) = 1/(1 + u)^2$

| h | # iter | | $H^1$-seminorm | | $L_2$-norm | | $L_\infty$-norm | |
|---|---|---|---|---|---|---|---|---|
| | FP | NW | Error $\times 10^{-5}$ | Rate | Error $\times 10^{-5}$ | Rate | Error $\times 10^{-5}$ | Rate |
| 1/16 | 7 | 5 | 17.1931 | - | 3.73555 | - | 7.51200 | - |
| 1/32 | 7 | 5 | 4.31635 | 1.99 | 0.94094 | 1.99 | 1.88100 | 1.99 |
| 1/64 | 7 | 5 | 1.08075 | 1.99 | 0.23568 | 1.99 | 0.47000 | 2.00 |
| 1/128 | 7 | 5 | 0.27778 | 1.96 | 0.05894 | 2.00 | 0.01180 | 1.99 |

the exact solution is chosen to be $u = 40(x - x^2)(y - y^2)$ and $k(u) = 0.125(-u^3 + 4u^2 - 7u + 8)$ if $u < 1$ and $k(u) = 1/(1 + u)$ if $u \geq 1$. Again a superconvergence is observed for this example. Furthermore, number of iterations needed are slightly higher than the previous example, which may be due to larger source term $f$. In this case the Newton iteration is shown to converge in fewer iterations than the fixed point iteration. Next we consider a problem with known solution

Table II. Error of FVEM for nonlinear elliptic BVP, with $u = 40(x - x^2)(y - y^2)$ and $k(u) = 0.125(-u^3 + 4u^2 - 7u + 8)$ if $u < 1$ and $k(u) = 1/(1 + u)$ if $u \geq 1$

| N | # iter | | $H^1$-seminorm | | $L_2$-norm | | $L_\infty$-norm | |
|---|---|---|---|---|---|---|---|---|
| | FP | NW | Error $\times 10^{-2}$ | Rate | Error $\times 10^{-2}$ | Rate | Error $\times 10^{-2}$ | Rate |
| 1/16 | 16 | 10 | 33.65484 | - | 7.33022 | - | 13.3000 | - |
| 1/32 | 15 | 8 | 9.10047 | 1.89 | 1.98347 | 1.89 | 3.57150 | 1.90 |
| 1/64 | 15 | 7 | 2.32645 | 1.97 | 0.50708 | 1.97 | 0.91120 | 1.97 |
| 1/128 | 15 | 7 | 0.58451 | 1.99 | 0.12740 | 1.99 | 0.22880 | 1.99 |

$u(x, y) = x^{1.6}$ with $k(u) = 1 + u$. Obviously, this solution is an element of $H^2(\Omega)$ but not in $H^3(\Omega)$. Also the resulting source term $f$ only belongs to $L^2(\Omega)$. The results are presented in

Table III. These experiments show that the $H^1$-norm of the error decreases at first order. The $L_2$-norm of the error decreases slower than second order.

Table III. Error of FVEM for nonlinear elliptic BVP with $u(x, y) = x^{1.6}$ and $k(u) = 1 + u$

| N | # iter | | $H^1$-seminorm | | $L_2$-norm | | $L_\infty$-norm | |
|---|---|---|---|---|---|---|---|---|
| | FP | NW | Error $\times 10^{-4}$ | Rate | Error $\times 10^{-4}$ | Rate | Error $\times 10^{-4}$ | Rate |
| 1/16 | 11 | 6 | 34.1671 | - | 3.71216 | - | 8.97536 | - |
| 1/32 | 11 | 7 | 17.5558 | 0.96 | 1.44873 | 1.36 | 3.53674 | 1.34 |
| 1/64 | 11 | 7 | 8.68644 | 1.02 | 0.53714 | 1.43 | 1.33414 | 1.40 |
| 1/128 | 11 | 8 | 4.20084 | 1.05 | 0.19272 | 1.48 | 0.48582 | 1.46 |

Related to number of nonlinear iterations in the two methods are the CPU times used to solve the problem. Table IV shows the CPU time of the fixed point iteration and the Newton iteration, corresponding to nonlinear elliptic problem whose results presented earlier in Table II. It can be seen that the CPU times of the two methods are relatively comparable for coarser mesh. However, as the mesh size gets smaller, the Newton iteration is more efficient than the fixed point iteration.

Table IV. Comparion of CPU time for nonlinear elliptic BVP, with $u = 40(x - x^2)(y - y^2)$ and $k(u) = 0.125(-u^3 + 4u^2 - 7u + 8)$ if $u < 1$ and $k(u) = 1/(1 + u)$ if $u \geq 1$

| N | CPU time (sec) | |
|---|---|---|
| | FP | NW |
| 1/16 | 1.1 | 1.5 |
| 1/32 | 5.3 | 5.5 |
| 1/64 | 30.1 | 24.7 |
| 1/128 | 192.7 | 145.3 |
| 1/256 | 1490.3 | 945.8 |

Tables V and VI illustrate Theorem 5.1. In this theorem, it has been shown that there exists a sequence of solutions in the Newton iteration such that their errors with respect to the finite volume solution $u_h$ are a decreasing sequence. Using the notation in that theorem, $\nu_k = \|u_h^k - u_h\|_{H^1}$ is a decreasing sequence satisfying

$$\nu_{k+1} \leq C_5 \sigma_h \nu_k^2, \qquad k = 0, 1, 2, \cdots .$$

We would like to examine the numerical behavior of this sequence for a fixed mesh size $h$. It is obvious that given $\nu_0$ we have

$$\nu_k \leq (C_5 \sigma_h)^{2^k - 1} \nu_0^{2^k}, \qquad k = 1, 2, \cdots ,$$

which after dividing by $\nu_0^{2^k}$ and taking logarithm on both sides give

$$|\log(\nu_k / \nu_0^{2^k})| \leq C_5 \sigma_h \, (2^k - 1), \qquad k = 1, 2, \cdots .$$

Table V. Results for case 2

| k | $h = 1/32$ | | $h = 1/64$ | | $h = 1/128$ | |
|---|---|---|---|---|---|---|
| | $\lvert\log(\nu_k/\nu_0^{2^k})\rvert$ | m | $\lvert\log(\nu_k/\nu_0^{2^k})\rvert$ | m | $\lvert\log(\nu_k/\nu_0^{2^k})\rvert$ | m |
| 1 | 1.13 | | 1.13 | | 1.13 | |
| 2 | 3.40 | 3.02 | 3.40 | 3.02 | 3.40 | 3.02 |
| 3 | 7.97 | 7.08 | 7.96 | 7.06 | 7.96 | 7.05 |
| 4 | 16.8 | 15.0 | 16.8 | 14.9 | 16.6 | 14.7 |

Hence we should expect that the sequence $\nu_k$ would decrease exponentially as $k \to \infty$.

The Tables V and VI show the decreasing behavior of the sequence resulting from the Newton iteration for last two model problems described above. In each table, $k$ represents the iteration level, $h$ is the mesh size, and $m$ is the value of row $k$ divided by the value of row $k-1$.

For case 2 presented in Table V, in which the problem has a piecewise continuous coefficient and larger source term, we see that the decreasing behavior of the sequence is approximately exponential, and it is independent of the mesh size. Similar trends are also evident for case 3 shown in Table VI.

Table VI. Results for case 3

| k | $h = 1/32$ | | $h = 1/64$ | | $h = 1/128$ | |
|---|---|---|---|---|---|---|
| | $\lvert\log(\nu_k/\nu_0^{2^k})\rvert$ | m | $\lvert\log(\nu_k/\nu_0^{2^k})\rvert$ | m | $\lvert\log(\nu_k/\nu_0^{2^k})\rvert$ | m |
| 1 | 1.17 | | 1.32 | | 1.45 | |
| 2 | 3.57 | 3.05 | 3.86 | 2.93 | 4.19 | 2.89 |
| 3 | 8.04 | 6.85 | 8.72 | 6.63 | 9.26 | 6.37 |
| 4 | 16.8 | 14.3 | 18.2 | 13.9 | 19.6 | 13.4 |
| 5 | 32.7 | 27.9 | 36.9 | 28.1 | 40.1 | 27.6 |

## REFERENCES

1. R. A. Adams. *Sobolev spaces.* Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London, 1975. Pure and Applied Mathematics, Vol. 65.
2. M. Artola. Sur une classe de problèmes paraboliques quasi-linéaires. *Boll. Un. Mat. Ital. B (6)*, 5(1):51–70, 1986.
3. M. Artola and L. Boccardo. Positive solutions for some quasilinear elliptic equations. *Comm. Appl. Nonlinear Anal.*, 3(4):89–98, 1996.
4. O. Axelsson. On global convergence of iterative methods. In *Iterative solution of nonlinear systems of equations (Oberwolfach, 1982)*, volume 953 of *Lecture Notes in Math.*, pages 1–19. Springer, Berlin, 1982.
5. O. Axelsson and A. T. Chronopoulos. On nonlinear generalized conjugate gradient methods. *Numer. Math.*, 69(1):1–15, 1994.
6. O. Axelsson and W. Layton. A two-level discretization of nonlinear boundary value problems. *SIAM J. Numer. Anal.*, 33(6):2359–2374, 1996.

7. A. Bergam, Z. Mghazli, and R. Verfürth. Estimations a posteriori d'un schéma de volumes finis pour un problème non linéaire. *Numer. Math.*, 95(4):599–624, 2003.
8. S. C. Brenner and L. R. Scott. *The mathematical theory of finite element methods*, volume 15 of *Texts in Applied Mathematics*. Springer-Verlag, New York, second edition, 2002.
9. H. Brezis *Analyse fonctionnelle*, Masson, Paris, 1983.
10. P. Chatzipantelidis. Finite volume methods for elliptic PDE's: a new approach. *M2AN Math. Model. Numer. Anal.*, 36(2):307–324, 2002.
11. P. Chatzipantelidis and R. D. Lazarov. Error estimates for a finite volume element method for elliptic pde's in nonconvex polygonal domains. *SIAM J. Numer. Anal.*, 2004. To appear.
12. P. Chatzipantelidis, R. D. Lazarov, and V. Thomée. Error estimates for a finite volume element method for parabolic equations in convex polygonal domains. *Numer. Methods Partial Differential Equations*, 20(5):650–674, 2004.
13. S.-H. Chou and Q. Li. Error estimates in $L^2$, $H^1$ and $L^\infty$ in covolume methods for elliptic and parabolic problems: a unified approach. *Math. Comp.*, 69(229):103–120, 2000.
14. R. S. Dembo, S. C. Eisenstat, and T. Steihaug. Inexact Newton methods. *SIAM J. Numer. Anal.*, 19(2):400–408, 1982.
15. V. Dolejší, M. Feistauer, and C. Schwab. A finite volume discontinuous Galerkin scheme for nonlinear convection-diffusion problems. *Calcolo*, 39(1):1–40, 2002.
16. J. Douglas, Jr. and T. Dupont. A Galerkin method for a nonlinear Dirichlet problem. *Math. Comp.*, 29:689–696, 1975.
17. J. Douglas, Jr., T. Dupont, and J. Serrin. Uniqueness and comparison theorems for nonlinear elliptic equations in divergence form. *Arch. Rational Mech. Anal.*, 42:157–168, 1971.
18. L. C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1998.
19. R. Eymard, T. Gallouët, and R. Herbin. Finite volume methods. In *Handbook of numerical analysis, Vol. VII*, Handb. Numer. Anal., VII, pages 713–1020. North-Holland, Amsterdam, 2000.
20. R. Eymard, T. Gallouët, R. Herbin, M. Gutnic, and D. Hilhorst. Approximation by the finite volume method of an elliptic-parabolic equation arising in environmental studies. *Math. Models Methods Appl. Sci.*, 11(9):1505–1528, 2001.
21. J. Frehse and R. Rannacher. Asymptotic $L^\infty$-error estimates for linear finite element approximations of quasilinear boundary value problems. *SIAM J. Numer. Anal.*, 15(2):418–431, 1978.
22. R. Li, Z. Chen, and W. Wu. *Generalized difference methods for differential equations*, volume 226 of *Monographs and Textbooks in Pure and Applied Mathematics*. Marcel Dekker Inc., New York, 2000. Numerical analysis of finite volume methods.
23. R. H. Li. Generalized difference methods for a nonlinear Dirichlet problem. *SIAM J. Numer. Anal.*, 24(1):77–88, 1987.
24. Y. Matsuzawa. Finite element approximation for some quasilinear elliptic problems. *J. Comput. Appl. Math.*, 96(1):13–25, 1998.
25. Jindřich Nečas. *Introduction to the theory of nonlinear elliptic equations.* A Wiley-Interscience Publication. John Wiley & Sons Ltd., Chichester, 1986. Reprint of the 1983 edition.
26. R. H. Nochetto. Pointwise a posteriori error estimates for elliptic problems on highly graded meshes. *Math. Comp.*, 64(209):1–22, 1995.
27. R. Sacco and F. Saleri. Mixed finite volume methods for semiconductor device simulation. *Numer. Methods Partial Differential Equations*, 13(3):215–236, 1997.