A Finite Volume Element Method for a Nonlinear Parabolic Problem

Panagiotis Chatzipantelidis and Victor Ginting

A Finite Volume Element Method for a Nonlinear Parabolic Problem

P. Chatzipantelidis and V. Ginting

Abstract We study a finite volume element discretization of a nonlinear parabolic equation in a convex polygonal domain. We show existence of the discrete solution and derive error estimates in L_2– and H^1–norms. We also consider a linearized method and provide numerical results to illustrate our theoretical findings.

Key words: nonlinear parabolic problem, finite volume element method, error estimates
MSC 2000: 65M60, 65M15

1 Introduction

We consider the nonlinear parabolic problem for $t \in [0, T]$, $T > 0$,

$$u_t + L(u)u = f, \text{ in } \Omega, \quad u = 0, \text{ on } \partial\Omega, \quad \text{with } u(0) = u^0, \text{ in } \Omega, \tag{1}$$

where Ω is a bounded convex polygonal domain in $\mathbb{R}^2$ and $L(v)w \equiv -\nabla \cdot (A(v)\nabla w)$, with $A(v) = \operatorname{diag}(a_1(v), a_2(v))$ a strictly positive definite and bounded real–valued matrix function, such that there exists $\beta > 0$

$$|x^\top A'(y)x| \leq \beta x^\top x, \quad \forall y \in \mathbb{R}, \ \forall x \in \mathbb{R}^2. \tag{2}$$

Further, we assume that A' is Lipschitz continuous, i.e. $\exists\, L > 0$

P. Chatzipantelidis
Department of Mathematics, University of Crete, Heraklion, GR–71409, Greece
e-mail: chatzipa@math.uoc.gr

V. Ginting
Department of Mathematics, University of Wyoming, Laramie, WY, 82071, USA
e-mail: vginting@uwyo.edu

$$|a_i'(y) - a_i'(\tilde{y})| \leq L|y - \tilde{y}|, \quad \forall y, \tilde{y} \in \mathbb{R},\ i = 1, 2, \tag{3}$$

and that there exists a sufficiently smooth unique solution u of (1).

Questions about the existence and regularity of solutions for (1) have been intensively investigated, for example in [11, Chapter 5]. Nonlinear parabolic problems such as (1) occur in many applied fields. To name a few, in the chemotaxis model, see Keller and Segel [10], in groundwater hydrology, see L.A. Richards [14], in modeling and simulation of oil recovery techniques in the presence of capillary pressure, see [5].

We shall study fully discrete approximations of (1) by the finite volume element method (FVEM). The FVEM, which is also called finite volume method or covolume method in some literatures, is a class of important numerical methods for solving differential equations, especially those arising from conservation laws including mass, momentum, and energy, because this method possesses local conservation property, which is crucial in many applications. It is popular in computational fluid mechanics, groundwater hydrology, reservoir simulations, and others. Many researchers have studied this method for linear and nonlinear problems. We refer to the monographs [13, 9] for the general presentation of this method and references therein for details.

The approximate solution will be sought in the space of piecewise linear functions

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

where $\mathcal{T}_h$ is a family of quasiuniform triangulations $T_h = \{K\}$ of Ω, with h denoting the maximum diameter of the triangles $K \in \mathcal{T}_h$ and $\mathcal{C} = \mathcal{C}(\Omega)$ the space of continuous functions on $\bar{\Omega}$.

The FVEM is based on a local conservation property associated with the differential equation. Namely, integrating (1) over any region $V \subset \Omega$ and using Green's formula we obtain for $t \in [0, T]$

$$\int_V u_t \, dx - \int_{\partial V} (A(u)\nabla u) \cdot n \, d\sigma = \int_V f \, dx, \tag{4}$$

where n denotes the unit exterior normal vector to ∂V. The semidiscrete FVEM approximation $u_h(t) \in \mathcal{X}_h$ will satisfy (4) for V in a finite collection of subregions of Ω called control volumes, the number of which will be equal to the dimension of the finite element space $\mathcal{X}_h$. 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. The semidiscrete FVEM for (1) is then to find $u_h(t) \in \mathcal{X}_h$, for $t \in [0, T]$, such that

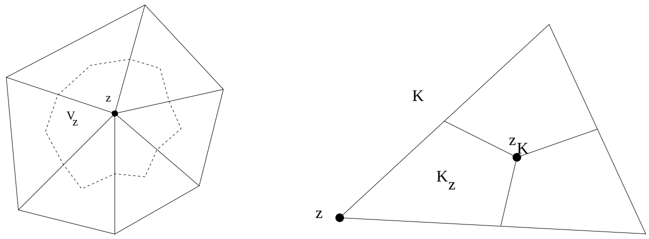

Fig. 1 *Left:* A union of triangles that have a common vertex z; the dotted line shows the boundary of the corresponding control volume V_z. *Right:* A triangle K partitioned into the three subregions K_z.

$$\int_{V_z} u_{h,t}\,dx - \int_{\partial V_z} (A(u_h)\nabla u_h)\cdot n\,ds = \int_{V_z} f\,dx, \quad \forall z \in Z_h^0, \tag{5}$$

with $u_h(0) = u_h^0$, where $u_h^0 \in \mathcal{X}_h$ is a given approximation of u^0. Note that different choices for z_K, e.g., the circumcenter of K, lead to other methods than the one considered here, see [12, 8].

In our analysis of the FVEM we use existing results associated with the finite element method approximation $\tilde{u}_h(t) \in \mathcal{X}_h$ of $u(t)$, defined by

$$(\tilde{u}_{h,t}, \chi) + a(\tilde{u}_h; \tilde{u}_h, \chi) = (f, \chi), \quad \forall \chi \in \mathcal{X}_h, \quad \text{for } t > 0, \tag{6}$$

with $(f,g) = \int_\Omega fg\,dx$, $a(w; v, g) = (A(w)\nabla v, \nabla g)$ and $\|w\| = (w,w)^{1/2}$ the norm in $L_2 = L_2(\Omega)$. Further let $H_0^1 = H_0^1(\Omega)$ be the standard Sobolev space with zero boundary conditions. Thus in order to rewrite (5) in a weak formulation, we introduce the finite dimensional space of piecewise constant functions

$$\mathcal{Y}_h = \{\eta \in L_2 :\ \eta|_{V_z} = \text{constant},\ \forall z \in Z_h^0;\ \eta|_{V_z} = 0,\ \forall z \in Z_h \setminus Z_h^0\}.$$

We now multiply (5) by $\eta(z)$ for an arbitrary $\eta \in \mathcal{Y}_h$ and sum over all $z \in Z_h^0$ to obtain the Petrov–Galerkin formulation for $t \in [0,T]$

$$(u_{h,t}, \eta) + a_h(u_h; u_h, \eta) = (f, \eta), \quad \forall \eta \in \mathcal{Y}_h, \quad \text{with } u_h(0) = u_h^0, \tag{7}$$

where $a_h(\cdot;\cdot,\cdot) : \mathcal{X}_h \times \mathcal{X}_h \times \mathcal{Y}_h \to \mathbb{R}$ is defined by

$$a_h(w; v, \eta) = -\sum_{z \in Z_h^0} \eta(z) \int_{\partial V_z} (A(w)\nabla v)\cdot n\,d\sigma, \quad \forall v, w \in \mathcal{X}_h,\ \eta \in \mathcal{Y}_h. \tag{8}$$

We shall now rewrite the Petrov–Galerkin method (7) as a Galerkin method in $\mathcal{X}_h$. For this purpose, we introduce the interpolation operator $J_h : \mathcal{C} \mapsto \mathcal{Y}_h$ by

$$J_h w = \sum_{z \in Z_h^0} w(z)\Psi_z,$$

where Ψ_z is the characteristic function of the control volume V_z. It is known that J_h is selfadjoint and positive definite, see [7], and hence the following defines an inner product $\langle\cdot,\cdot\rangle$ on $\mathcal{X}_h$,

$$\langle\chi,\psi\rangle = (\chi, J_h\psi), \quad \forall\chi,\psi\in\mathcal{X}_h. \tag{9}$$

Further, in [7] it is shown that the corresponding norm is equivalent to the L_2−norm, uniformly in h, i.e., with $C \geq c > 0$,

$$c\|\chi\| \leq |||\chi||| \leq C\|\chi\|, \quad \forall\chi\in\mathcal{X}_h, \quad \text{where } |||\chi||| \equiv \langle\chi,\chi\rangle^{1/2}.$$

With this notation, (7) may equivalently be written in Galerkin form as

$$\langle u_{h,t},\chi\rangle + a_h(u_h; u_h, J_h\chi) = (f, J_h\chi), \quad \forall\chi\in\mathcal{X}_h, \quad \text{for } t\geq 0. \tag{10}$$

Then let $N\in\mathbb{N}$, $N\geq 1$, $k = T/N$ and $t^n = n\,k$, $n = 0,\dots,N$. Discretizing in time (10), with the backward Euler method we approximate $u(t^n)$ by $U^n\in\mathcal{X}_h$, for $n = 1,\dots,N$, such that,

$$\langle\bar\partial U^n,\chi\rangle + a_h(U^n; U^n, J_h\chi) = (f^n, J_h\chi), \quad \forall\chi\in\mathcal{X}_h, \quad \text{with } U^0 = u_h^0, \tag{11}$$

where $\bar\partial U^n = (U^n - U^{n-1})/k$ and $f^n = f(t^n)$.

To show existence of the semidiscrete solution $\tilde{u}_h$ of the finite element method (6), one can employ Brouwer's fixed point theorem and the coercivity property of $a(\cdot;\cdot,\cdot)$,

$$a(w;\chi,\chi) \geq \alpha\|\nabla\chi\|^2, \quad \forall\chi\in\mathcal{X}_h,\ \forall w\in L_2, \tag{12}$$

see [15]. However, the corresponding coercivity property for $a_h(\cdot;\cdot,\cdot)$,

$$a_h(w;\chi,J_h\chi) \geq \tilde\alpha\|\nabla\chi\|^2, \quad \forall\chi\in\mathcal{X}_h, \tag{13}$$

holds for $\|\nabla w\|_{L_\infty}$ in a bounded ball, where $\|w\|_{L_\infty} = \sup_{x\in\Omega}|w(x)|$. For this reason, we will employ a different argument than the one in [15] to show existence of U^n. Note that if z_K is the circumcenter of K, it is shown in [12], that (13) is satisfied for $w\in L_2$ and thus one may show existence of the solution of the finite volume method analogously to the one for the finite element method.

We show existence and uniqueness of the solution U^n of (11), and derive error estimates in L_2− and H^1–norms, see Theorems 3.1 and 4.1. Recently in [8], a two-grid finite volume element method was considered, for circumcenter based control volumes, with suboptimal estimates in L_2− and H^1− norms.

Our analysis follows the corresponding one for the FVEM nonlinear elliptic and linear parabolic problems in [2, 4]. This is based in bounds for the error functionals $\varepsilon_h(\cdot,\cdot)$ defined by

$$\varepsilon_h(f,\chi) = (f, J_h\chi) - (f,\chi), \quad \forall f\in L_2,\ \chi\in\mathcal{X}_h, \tag{14}$$

and $\varepsilon_a(\cdot;\cdot,\cdot)$ defined by

$$\varepsilon_a(w;v_h,\chi) = a_h(w;v_h,J_h\chi) - a(w;v_h,\chi) \quad \forall v_h,\chi \in \mathcal{X}_h,\ w \in L_2. \tag{15}$$

Following [15], we introduce the projection $R_h : H_0^1 \to \mathcal{X}_h$ defined by

$$a(v;R_hv,\chi) = a(v;v,\chi), \quad \forall \chi \in \mathcal{X}_h. \tag{16}$$

In [15] optimal order error estimates in L_2- and H^1–norms were established for the difference $R_hu(t) - u(t)$. Here we combine these error estimates with bounds for the difference $\vartheta^n = U^n - R_hu^n$, which satisfies

$$\langle \bar\partial \vartheta^n, \chi\rangle + a_h(U^n;\vartheta^n, J_h\chi) = \delta(t^n;U^n,\chi), \quad \text{for } \chi \in \mathcal{X}_h, \tag{17}$$

with

$$\begin{aligned}\delta(t^n;v,\chi) &\equiv -(\omega^n, J_h\chi) - \varepsilon_h(f^n - u_t^n,\chi) + \varepsilon_a(v;R_hu^n,\chi)\\ &\quad + ((A(u^n) - A(v))\nabla R_hu^n, \nabla\chi) \equiv \sum_{j=1}^{4} I_j,\end{aligned} \tag{18}$$

and $\omega^n = (R_h - I)\bar\partial u^n + (\bar\partial u^n - u_t^n)$. Further we analyze a linearized fully discrete scheme and provide numerical examples to illustrate our results.

The rest of the paper is organized as follows. In Section 2 we recall known results and derive error bounds for the error functional δ. In Section 3 we derive error estimates and in Section 4 existence of the nonlinear fully discrete method. In Section 5 consider a linearized version of the backward Euler scheme and finally in Section 6 we present our numerical examples.

2 Preliminaries

In this section we recall known results about the projection R_h defined by (16) and the error functionals ε_h and ε_a introduced in (14) and (15). We also derive bounds for the error functional δ defined in (18).

We consider quasiuniform triangulations $\mathcal{T}_h$ for which the following inverse inequalities hold, see e.g. [15]

$$\|\nabla\chi\| \le Ch^{-1}\|\chi\|, \quad \text{and} \quad \|\nabla\chi\|_{L_\infty} \le Ch^{-1}\|\nabla\chi\|, \quad \text{for } \chi \in \mathcal{X}_h. \tag{19}$$

In such meshes, it is shown in [15, Lemma 13.2] that there exists $M_0 > 0$, independent of h, such that

$$\|\nabla u(t)\|_{L_\infty} + \|\nabla R_hu(t)\|_{L_\infty} \le M_0, \quad \text{for } t \le T, \tag{20}$$

and the following error estimates for $R_hu - u$.

Lemma 2.1. *With R_h defined by* (16) *and $\varrho = R_h u - u$, we have under the appropriate regularity assumptions on u, with $C_u > 0$ independent t,*

$$\|\nabla^s D_t^\ell \varrho(t)\| \le C_u h^{2-s}, \quad 0 < t \le T, \quad \text{and} \quad s, \ell = 0, 1, \quad \text{where } D_t = \partial/\partial t.$$

Our analysis is based on error estimates for the difference $\vartheta^n = U^n - R_h u^n$. Thus in view of the error equation (17) for ϑ^n, we recall necessary bounds for the error functionals ε_h and ε_a derived in [2, 4].

Lemma 2.2. *For the error functional ε_h, defined by* (14)*, we have*

$$|\varepsilon_h(f, \chi)| \le Ch^2 \|\nabla f\| \, \|\nabla \chi\|, \quad \forall f \in H^1, \ \chi \in \mathcal{X}_h.$$

To this end, for $M = \max(2M_0, 1)$, we consider

$$\mathcal{B}_M = \{\chi \in \mathcal{X}_h : \|\nabla \chi\|_{L_\infty} \le M\}.$$

Lemma 2.3. *For the error functional ε_a, defined in* (15)*, we have*

$$|\varepsilon_a(w_h; v_h, \chi)| \le Ch \|\nabla w_h \cdot \nabla v_h\| \, \|\nabla \chi\|, \ \forall w_h, v_h, \chi \in \mathcal{X}_h. \tag{21}$$

Further, if u is the solution of (1)*, then for $v \in \mathcal{B}_M$*

$$|\varepsilon_a(v; R_h u(t), \chi)| \le Ch^2 \|\nabla \chi\|. \tag{22}$$

Proof. The first bound is shown in [2, Lemma 2.3]. The second bound is a direct result of Lemma 2.1, [2, Lemma 2.4] and the fact that $v \in \mathcal{B}_M$. □

Then, in view of Lemma 2.3 there exists a constant $c > 0$ such that for h sufficiently small, the coercivity property (13) for a_h holds for $w \in \mathcal{B}_M$. Further, in [2] we showed the following "Lipschitz"–type estimation for ε_a.

Lemma 2.4. *For the error functional ε_a, defined in* (15)*, there exists a constant C, independent of h, such that for $\chi, \psi \in \mathcal{X}_h$*

$$|\varepsilon_a(v; \psi, \chi) - \varepsilon_a(w; \psi, \chi)| \le Ch \|\nabla \psi\|_{L_\infty} (1 + \|\nabla w\|_{L_\infty}) \|\nabla(v - w)\| \, \|\nabla \chi\|.$$

Finally, we show appropriate bounds for the functional δ, defined by (18).

Lemma 2.5. *For δ defined by* (18)*, we have for $\chi \in \mathcal{X}_h$ and $v \in \mathcal{B}_M$*

$$|\delta(t^n; v, \chi)| \le C(k + h^2)\|\chi\| + Ch^2 \|\nabla \chi\| + \left\{ \begin{array}{l} C\|v - R_h u^n\| \, \|\nabla \chi\| \\ C\|\nabla(v - R_h u^n)\| \, \|\chi\|. \end{array} \right.$$

Proof. Using the splitting in (18) we bound each of the terms I_j, $j = 1, \ldots, 4$. Recall that $\omega^n = (R_h - I)\bar{\partial} u^n + (\bar{\partial} u^n - u_t^n)$, then in view of Lemma 2.1 we have

$$\|\omega^n\| \le Ck^{-1}\int_{t^{n-1}}^{t^n}\|\varrho_t\|\,ds + C\int_{t^{n-1}}^{t^n}\|u_{tt}\|\,ds \le C(k+h^2), \tag{23}$$

and hence

$$|I_1| \le C(k+h^2)\|\chi\|. \tag{24}$$

To bound I_2+I_3, we use Lemma 2.2 and (22) to get

$$|I_2+I_3| \le Ch^2\|\nabla\chi\|. \tag{25}$$

Finally, employing (2) and (20), and adding and subtracting R_hu^n, and using Lemma 2.1, we get,

$$\begin{aligned}|I_4| &= |((A(u^n)-A(v))\nabla R_hu^n,\nabla\chi)| \le C\|v-u^n\|\,\|\nabla\chi\| \\ &\le Ch^2\|\nabla\chi\| + C\|v-R_hu^n\|\|\nabla\chi\|.\end{aligned} \tag{26}$$

Combining now (24)–(26) we get the first one of the desired bounds. To show the second estimate of this lemma, we bound I_4 differently. Using integration by parts we rewrite I_4 as

$$\begin{aligned}I_4 &= ((A(u^n)-A(R_hu^n))\nabla R_hu^n,\nabla\chi) + ((A(R_hu^n)-A(v))\nabla R_hu^n,\nabla\chi) \\ &= ((A(u^n)-A(R_hu^n))\nabla R_hu^n,\nabla\chi) + (\mathrm{div}\,[(A(R_hu^n)-A(v))\nabla R_hu^n],\chi) \\ &= I_4^i + I_4^{ii}.\end{aligned}$$

Then, in view of (2), Lemma 2.1 and (20), we have

$$|I_4^i| \le Ch^2\|\nabla\chi\|. \tag{27}$$

Further, employing (2), (3) and (20) we obtain

$$\begin{aligned}|I_4^{ii}| &\le (\|(A'(R_hu^n)-A'(v))\nabla R_hu^n\| + \|A'(v)\nabla(R_hu^n-v)\|)\|\chi\| \\ &\le C(\|v-R_hu^n\| + \|\nabla(v-R_hu^n)\|)\|\chi\|.\end{aligned} \tag{28}$$

Therefore combining (27) and (28), we have

$$|I_4| \le C\|\nabla(v-R_hu^n)\|\|\chi\| + Ch^2\|\nabla\chi\|. \tag{29}$$

Thus, combining (24), (25), (29) and (26) we obtain the second of the desired estimates of the lemma. □

3 Error estimates for the backward Euler method

In this section we derive error estimates for the FVEM (11) in L_2– and H^1–norms, under the assumption that $U^j \in \mathcal{B}_M$, for $j = 0, \dots, n$. In Section 4 we will show existence of $U^n \in \mathcal{B}_M$.

Theorem 3.1. *Let U^n and u be the solutions of* (11) *and* (1), *with $U^0 = R_h u^0$. If $U^j \in \mathcal{B}_M$, for $j = 0, \dots, n$, $n \geq 1$, and k, h be sufficiently small, then there exist $C > 0$, independent of k and h, such that*

$$\|\nabla^s (U^n - u^n)\| \leq C(k + k^{-s/2} h^{2-s}), \quad \textit{for } s = 0, 1. \tag{30}$$

Proof. Using the error splitting $U^n - u^n = (U^n - R_h u^n) + (R_h u^n - u^n) = \vartheta^n + \varrho^n$ and Lemma 2.1 it suffices to show

$$\|\nabla^s \vartheta^n\| \leq C_s(k + k^{-s/2} h^{2-s}), \quad \text{for } s = 0, 1. \tag{31}$$

We start with the estimation of $\|\vartheta^n\|$. Due to the symmetry of $\langle \chi, \psi \rangle$, we have the following identity

$$\langle \bar{\partial} \vartheta^n, \vartheta^n \rangle = \frac{1}{2k}(|||\vartheta^n|||^2 - |||\vartheta^{n-1}|||^2) + \frac{1}{2k}|||\vartheta^n - \vartheta^{n-1}|||^2. \tag{32}$$

Choosing $\chi = \vartheta^n$ in (17) and using the fact that $U^n \in \mathcal{B}_M$, (13) and (32), we get after eliminating $|||\vartheta^n - \vartheta^{n-1}|||$,

$$\frac{1}{2k}(|||\vartheta^n|||^2 - |||\vartheta^{n-1}|||^2) + \tilde{\alpha}\|\nabla \vartheta^n\|^2 \leq \delta(t^n; U^n, \vartheta^n). \tag{33}$$

Employing now the first estimate of Lemma 2.5, with $v = U^n$ and $\chi = \vartheta^n$, to bound the right hand side of (33), we obtain

$$\frac{1}{2k}(|||\vartheta^n|||^2 - |||\vartheta^{n-1}|||^2) + \tilde{\alpha}\|\nabla \vartheta^n\|^2 \leq C(k + h^2)\|\vartheta^n\| + C(k\|\vartheta^n\| + h^2)\|\nabla \vartheta^n\|.$$

Then, after eliminating $\|\nabla \vartheta^n\|^2$ and moving $|||\vartheta^n|||^2$ to the left, we have for k sufficiently small

$$|||\vartheta^n|||^2 \leq (1 + Ck)|||\vartheta^{n-1}|||^2 + CkE, \quad \text{with } E = O(k^2 + h^4).$$

Hence, using the fact that $\vartheta^0 = 0$, we obtain

$$|||\vartheta^n|||^2 \leq CkE \sum_{\ell=0}^{n} (1 + Ck)^{n-\ell+1} \leq C(k^2 + h^4).$$

Thus, there exists $C_0 > 0$, such that $|||\vartheta^n||| \leq C_0(k + h^2)$. Since $||| \cdot |||$ and $\| \cdot \|$ are equivalent norms, the first part of the proof is complete.

Next we turn to the estimation of $\|\nabla\vartheta^n\|$. Choosing this time $\chi = \bar\partial\vartheta^n$ in (17), we obtain

$$|||\bar\partial\vartheta^n|||^2 + a(U^n;\vartheta^n,\bar\partial\vartheta^n) = \delta(t^n;U^n,\bar\partial\vartheta^n) + \varepsilon_a(U^n;\vartheta^n,\bar\partial\vartheta^n). \tag{34}$$

Note now, that since $a(\cdot;\cdot,\cdot)$ is symmetric we have the identity

$$2ka(U^n;\vartheta^n,\bar\partial\vartheta^n) = a(U^n;\vartheta^n,\vartheta^n) - a(U^n;\vartheta^{n-1},\vartheta^{n-1}) + k^2a(U^n;\bar\partial\vartheta^n,\bar\partial\vartheta^n).$$

Using now this and (12) in (34) we get, after subtracting $a(U^{n-1};\vartheta^{n-1},\vartheta^{n-1})$ from both parts of (34),

$$\begin{aligned}&2k|||\bar\partial\vartheta^n|||^2 + a(U^n;\vartheta^n,\vartheta^n) - a(U^{n-1};\vartheta^{n-1},\vartheta^{n-1}) + \alpha k^2\|\nabla\bar\partial\vartheta^n\|^2\\&\le 2k\delta(t^n;U^n,\bar\partial\vartheta^n) + 2k\varepsilon_a(U^n;\vartheta^n,\bar\partial\vartheta^n)\\&\quad+\{a(U^n;\vartheta^{n-1},\vartheta^{n-1}) - a(U^{n-1};\vartheta^{n-1},\vartheta^{n-1})\} = I + II + III.\end{aligned} \tag{35}$$

Employing the second bound of Lemma 2.5, with $v = U^n$ and $\chi = \bar\partial\vartheta^n$, we have

$$\begin{aligned}|I| \le& Ck(k+h^2)\|\bar\partial\vartheta^n\| + Ckh^2\|\nabla\bar\partial\vartheta^n\| + Ck\|\nabla\vartheta^n\|\,\|\bar\partial\vartheta^n\|\\&\le k|||\bar\partial\vartheta^n|||^2 + Ck\|\nabla\vartheta^n\|^2 + \frac{\alpha k^2}{2}\|\nabla\bar\partial\vartheta^n\|^2 + CkE,\end{aligned} \tag{36}$$

with $E = O(k^2 + k^{-1}h^4)$. Next, using Lemma 2.3 and the fact that $U^n \in \mathcal{B}_M$, we obtain

$$|II| \le Ckh\|\nabla U^n\|_{L_\infty}\|\nabla\vartheta^n\|\,\|\nabla\bar\partial\vartheta^n\| \le Ch^2\|\nabla\vartheta^n\|^2 + \frac{\alpha k^2}{2}\|\nabla\bar\partial\vartheta^n\|^2. \tag{37}$$

Finally, using again (2), the fact that $\vartheta^{n-1} \in \mathcal{B}_{2M}$ and (23), we have

$$\begin{aligned}|III| &\le Ck\|\,|\nabla\vartheta^{n-1}|\,|\bar\partial U^n|\,\|\,\|\nabla\vartheta^{n-1}\|\\&\le Ck(\|\,|\nabla\vartheta^{n-1}|\,|\bar\partial\vartheta^n|\,\| + \|\,|\nabla\vartheta^{n-1}|\,|R_h\bar\partial u^n|\,\|)\,\|\nabla\vartheta^{n-1}\|\\&\le k|||\bar\partial\vartheta^n|||^2 + Ck\|\nabla\vartheta^{n-1}\|^2.\end{aligned} \tag{38}$$

Therefore applying (36)–(38), in (35), eliminating $|||\bar\partial\vartheta^n|||$ and $\|\nabla\bar\partial\vartheta^n\|$ and using (12), we obtain for k and h sufficiently small,

$$a(U^n;\vartheta^n,\vartheta^n) \le (1+Ck)a(U^{n-1};\vartheta^{n-1},\vartheta^{n-1}) + CkE.$$

Thus, using the fact that $\vartheta^0 = 0$ and A is strictly positive definite, we get

$$c\|\nabla\vartheta^n\|^2 \le a(U^n;\vartheta^n,\vartheta^n) \le CkE\sum_{\ell=0}^{n}(1+Ck)^{n-\ell+1} \le C(k^2 + k^{-1}h^4).$$

Thus, there exists $C_1 > 0$, such that

$$\|\nabla\vartheta^n\| \le C_1(k + k^{-1/2}h^2), \tag{39}$$

which completes the second part of the proof. □

4 Existence of the backward Euler approximation

Here we show existence of the solution of the nonlinear fully discrete scheme (11), if $U^0 = R_h u^0$ and the discretization parameters k and h are sufficiently small and satisfy $k = O(h^{1+\epsilon})$, with $0 < \epsilon < 1$.

Let $G_n : \mathcal{X}_h \to \mathcal{X}_h$, be defined by

$$\langle G_n v - U^{n-1}, \chi\rangle + k a_h(v; G_n v, J_h\chi) = k(f^n, J_h\chi), \quad \forall \chi \in \mathcal{X}_h. \tag{40}$$

Obviously, if G_n has a fixed point v, then $U^n = v$ is the solution of (11).

In view of (39), recall that if $U^{n-1} \in \mathcal{B}_M$, then

$$\|\nabla(U^{n-1} - R_h u^{n-1})\| \le C_1(k + k^{-1/2}h^2). \tag{41}$$

Then the following two lemmas hold.

Lemma 4.1. *Let $U^{n-1} \in \mathcal{B}_M$ such that (41) holds. Then for $k = O(h^{1+\epsilon})$ with $0 < \epsilon < 1$, there exist a constant $C_2 > 0$, independent of h, sufficiently large such that $U^{n-1} \in \widetilde{\mathcal{B}}$, where*

$$\widetilde{\mathcal{B}}_n = \{w \in \mathcal{X}_h : \ \|\nabla(w - R_h u^n)\| \le C_2 h^{1+\tilde\epsilon}\}, \quad \text{with } \tilde\epsilon = \min(\epsilon, \frac{1-\epsilon}{2}). \tag{42}$$

Proof. Using the stability property of R_h and the fact that $k = O(h^{1+\epsilon})$, we have

$$\begin{aligned}\|\nabla(U^{n-1} - R_h u^n)\| &\le \|\nabla(U^{n-1} - R_h u^{n-1})\| + k\|\nabla R_h \bar\partial u^n\| \\ &\le C_1(k + k^{-1/2}h^2) + k\|\nabla\bar\partial u^n\| \le C_2 h^{1+\tilde\epsilon}. \quad \square\end{aligned}$$

Lemma 4.2. *Let $U^{n-1}, v \in \mathcal{B}_M$ such that (41) holds and $v \in \widetilde{\mathcal{B}}_n$, with $\widetilde{\mathcal{B}}_n$ defined by (42). Then for $k = \mathcal{O}(h^{1+\epsilon})$, with $0 < \epsilon < 1$, $G_n v \in \widetilde{\mathcal{B}}_n$.*

Proof. Let us now denote by $\xi^n = G_n v - R_h u^n$, and $\xi^{n-1} = U^{n-1} - R_h u^{n-1}$. Then, using (40), (1) and (16), ξ^n satisfies a similar equation to (17), with ξ^n and v instead of ϑ^n and U^n, hence

$$\langle\bar\partial\xi^n, \chi\rangle + a_h(v; \xi^n, J_h\chi) = \delta(t^n; v, \chi), \quad \text{for } \chi \in \mathcal{X}_h, \tag{43}$$

Choosing $\chi = \bar\partial\xi^n$ in (43) and following the proof of Theorem 3.1 we obtain the corresponding inequality to (35), without the last term III, with ξ^n and v in the place of ϑ^n and U^n,

$$\begin{aligned}2k|||\bar{\partial}\xi^n|||^2 + a(v;\xi^n,\xi^n) - a(v;\xi^{n-1},\xi^{n-1}) + \alpha k^2\|\nabla\bar{\partial}\xi^n\|^2 \\ \le 2k\delta(t^n;v,\bar{\partial}\xi^n) + 2k\varepsilon_a(v;\xi^n,\bar{\partial}\xi^n) = I + II.\end{aligned} \tag{44}$$

Similarly as before we obtain the corresponding estimates to (36) and (37), with ξ^n and v in the place of ϑ^n and U^n. Thus

$$|I| \le 2k|||\bar{\partial}\xi^n|||^2 + \frac{\alpha k^2}{2}\|\nabla\bar{\partial}\xi^n\|^2 + Ck\|\nabla(v - R_h u^n)\|^2 + CkE, \tag{45}$$

with $E = O(k^2 + k^{-1}h^4)$ and

$$|II| \le Ch^2 a(v;\xi^n,\xi^n) + \frac{\alpha k^2}{2}\|\nabla\bar{\partial}\xi^n\|^2. \tag{46}$$

Then using (45) and (46) in (44) and eliminating $|||\bar{\partial}\xi^n|||^2$ and $\|\nabla\bar{\partial}\xi^n\|^2$, we get for h sufficiently small

$$a(v;\xi^n,\xi^n) \le (1 + Ck)a(v;\xi^{n-1},\xi^{n-1}) + Ck\|\nabla(v - R_h u^n)\|^2 + CkE.$$

Finally, using in this inequality, (41), the facts that $v \in \widetilde{\mathcal{B}}_n$ and $\epsilon < 1$ and (13), we obtain the desired bound for k sufficiently small. □

Theorem 4.1. *Let $\mathcal{T}_h$ satisfy the inverse assumption* (19) *and $U^{n-1}, v \in \mathcal{B}_M$ such that* (41) *holds. Then for h sufficiently small and $k = \mathcal{O}(h^{1+\epsilon})$, with $0 < \epsilon < 1$, there exists $U^n \in \mathcal{B}_M$ satisfying* (11).

Proof. Obviously, in view of Lemmas 4.1 and 4.2, starting with $v_0 = U^{n-1}$, through G_n we obtain a sequence of elements $v_{j+1} = G_n v_j \in \widetilde{\mathcal{B}}_n$, $j \ge 0$. Thus combining this with (20) and the facts that $M > M_0$ and $\tilde{\epsilon} > 0$, we get $G_n v_j \in \mathcal{B}_M$ for h sufficiently small, i.e.,

$$\|\nabla G_n v_j\|_{L_\infty} \le \|\nabla R_h u^n\|_{L_\infty} + Ch^{-1}\|\nabla(G_n v_j - R_h u^n)\| \le M, \ j \ge 0.$$

To show now the existence of $U^n \in \mathcal{B}_M$ it suffices that

$$|||G_n v - G_n w||| < L|||v - w|||, \quad \forall v, w \in \mathcal{B}_M, \quad \text{with } 0 < L < 1.$$

Employing (40) for $v, w \in \mathcal{B}_M$ and $\chi \in \mathcal{X}_h$, we obtain

$$\langle G_n v - G_n w, \chi\rangle + k a_h(v; G_n v, J_h\chi) - k a_h(w; G_n w, J_h\chi) = 0.$$

Hence, for $\chi = G_n v - G_n w$, this gives

$$\begin{aligned}|||\chi|||^2 + k a_h(w;\chi,J_h\chi) &= k(a_h(w;G_n v,J_h\chi) - a_h(v;G_n v,J_h\chi)) \\ &= k(a(w;G_n v,\chi) - a(v;G_n v,\chi)) \\ &\quad + k(\varepsilon_a(v;G_n v,\chi) - \varepsilon_a(w;G_n v,\chi)) = I + II.\end{aligned} \tag{47}$$

To bound I we use (2) and the fact that $G_n v \in \mathcal{B}_M$ to get

$$|I| \leq Ck \|\nabla G_n v\|_{L_\infty} \|v-w\| \|\nabla\chi\| \leq Ck \|v-w\| \|\nabla\chi\|. \tag{48}$$

For II, we use Lemma 2.4, the inverse inequality (19) and the fact that $v, G_n v \in \mathcal{B}_M$ to obtain

$$|II| \leq Ckh\|\nabla(v-w)\| \|\nabla\chi\| \leq Ck\|v-w\| \|\nabla\chi\|. \tag{49}$$

Employing now (13), (48), and (49) into (47) we have

$$|||\chi|||^2 + k\tilde{\alpha} \|\nabla\chi\|^2 \leq Ck \|v-w\| \|\nabla\chi\| \leq Ck\|v-w\|^2 + k\tilde{\alpha} \|\nabla\chi\|^2,$$

which in view of the fact that $\|\cdot\|$ and $|||\cdot|||$ are equivalent norms gives for sufficiently small k the desired bound. □

5 A linearized fully discrete scheme

In this section we analyze a linearized backward Euler scheme for the approximation of (1). This time for $U^0 = R_h u^0$, we define the nodal approximations $U^n \in \mathcal{X}_h$ to u^n, $n = 1, \dots, N$, by

$$\langle \bar{\partial} U^n, \chi\rangle + a_h(U^{n-1}; U^n, J_h\chi) = (f^n, J_h\chi), \quad \forall \chi \in \mathcal{X}_h,\ n \geq 1. \tag{50}$$

Theorem 5.1. *Let U^n and u be the solutions of (50) and (1), with $U^0 = R_h u^0$. Then, for $U^{n-1} \in \mathcal{B}_M$, h sufficiently small and $k = O(h^{1+\epsilon})$, with $0 < \epsilon < 1$, we have $U^n \in \mathcal{B}_M$ and*

$$\|\nabla^s (U^n - u(t^n))\| \leq C(k + k^{-s/2} h^{2-s}), \quad \textit{with } s = 0, 1.$$

Proof. Since the discrete scheme (50) is linear the existence of $U^n \in \mathcal{X}_h$ is obvious. The proof is analogous to that for Theorem 3.1, thus it suffices to bound $\|\nabla^s \vartheta^n\|$, $s = 0, 1$. This time ϑ^n satisfies a similar equation to (17) with U^{n-1} in the place of U^n,

$$\langle \bar{\partial}\vartheta^n, \chi\rangle + a_h(U^{n-1}; \vartheta^n, J_h\chi) = \delta(t^n; U^{n-1}, \chi), \quad \forall \chi \in \mathcal{X}_h.$$

We start with the estimation for $\|\vartheta^n\|$. In an analogous way to (33), we obtain the following inequality,

$$\frac{1}{2k}(|||\vartheta^n|||^2 - |||\vartheta^{n-1}|||^2) + \tilde{\alpha}\|\nabla\vartheta^n\|^2 \leq \delta(t^n; U^{n-1}, \vartheta^n).$$

To bound now the right hand side of this inequality we employ the first estimate of Lemma 2.5, with $v = U^{n-1}$ and $\chi = \vartheta^n$, use the fact that $U^{n-1} - R_h u^n = \vartheta^{n-1} - kR_h \bar{\partial} u^n$ and the stability of R_h, to get

$$\begin{aligned}
&\frac{1}{2k}(|||\vartheta^n|||^2 - |||\vartheta^{n-1}|||^2) + \tilde{\alpha}\|\nabla\vartheta^n\|^2 \\
&\quad \le C(k+h^2)\|\vartheta^n\| + C(k\|U^{n-1} - R_h u^n\| + h^2)\|\nabla\vartheta^n\| \\
&\quad \le C|||\vartheta^n|||^2 + \tilde{\alpha}\|\nabla\vartheta^n\|^2 + Ck|||\vartheta^{n-1}|||^2 + CE, \quad \text{with } E = O(k^2+h^4).
\end{aligned}$$

Next, after eliminating $\|\nabla\vartheta^n\|$, we get for k sufficiently small

$$|||\vartheta^n|||^2 \le (1+Ck)|||\vartheta^{n-1}|||^2 + CkE.$$

Hence, since $\vartheta^0 = 0$, we have by repeated application $|||\vartheta^n||| \le C(k+h^2)$, which in view of the fact that $|||\cdot|||$ and $\|\cdot\|$ are equivalent norms, completes the first part of the proof.

Next we turn to the bound for $\|\nabla\vartheta^n\|$. In an analogous way to (34) we get

$$|||\bar{\partial}\vartheta^n|||^2 + a(U^{n-1};\vartheta^n,\bar{\partial}\vartheta^n) = \delta(t^n;U^{n-1},\bar{\partial}\vartheta^n) + \varepsilon_a(U^{n-1};\vartheta^n,\bar{\partial}\vartheta^n).$$

Hence, similarly as in (35), we have

$$\begin{aligned}
&2k|||\bar{\partial}\vartheta^n|||^2 + a(U^n;\vartheta^n,\vartheta^n) - a(U^{n-1};\vartheta^{n-1},\vartheta^{n-1}) + \alpha k^2\|\nabla\bar{\partial}\vartheta^n\|^2 \\
&\quad \le 2k\delta(t^n;U^{n-1},\bar{\partial}\vartheta^n) + 2k\varepsilon_a(U^{n-1};\vartheta^n,\bar{\partial}\vartheta^n) \\
&\qquad + \{a(U^n;\vartheta^n,\vartheta^n) - a(U^{n-1};\vartheta^n,\vartheta^n)\} = I.
\end{aligned} \tag{51}$$

Thus, in a similar way that we obtained (36)–(38), we have

$$\begin{aligned}
|I| \le\ & 2k|||\bar{\partial}\vartheta^n|||^2 + Ck\|\nabla(U^{n-1} - R_h u^n)\|^2 + C(k+h^2)\|\nabla\vartheta^n\|^2 \\
&+ \alpha k^2\|\nabla\bar{\partial}\vartheta^n\|^2 + CkE,
\end{aligned}$$

with $E = O(k^2 + k^{-1}h^4)$. Combining these in (51), using the fact that $U^{n-1} - R_h u^n = \vartheta^{n-1} - kR_h\bar{\partial}u^n$ and the stability of R_h, we obtain for k sufficiently small,

$$a(U^n;\vartheta^n,\vartheta^n) \le (1+Ck)a(U^{n-1};\vartheta^{n-1},\vartheta^{n-1}) + CkE.$$

Therefore since $\vartheta^0 = 0$ we obtain

$$\alpha\|\nabla\vartheta^n\|^2 \le a(U^n;\vartheta^n,\vartheta^n) \le CkE\sum_{\ell=0}^{n}(1+Ck)^{n-\ell+1} \le C(k^2 + k^{-1}h^4),$$

which gives the desired bound. Finally, this estimate, the inverse inequality (19) and the fact that $k = O(h^{1+\epsilon})$ give, for sufficiently small h, that $U^n \in \mathcal{B}_M$, which completes the proof. $\square$

6 Numerical Examples

In this section we give numerical examples to illustrate the error estimates presented in the previous sections. Let $\{\phi_i\}_{i=1}^d$ be the standard piecewise linear basis functions of $\mathcal{X}_h$ and for $\chi \in \mathcal{X}_h$, let $\tilde{\chi} = (\tilde{\chi}_1, \dots, \tilde{\chi}_d) \in \mathbb{R}^d$ be the vector such that $\chi = \sum_{i=1}^d \tilde{\chi}_i \phi_i$. Then the backward Euler method (11) can be written as

$$(D + kS(\tilde{U}^n))\tilde{U}^n = D\tilde{U}^{n-1} + kQ^n.$$

where D is the mass matrix with elements $D_{ij} = \int_{V_i} \phi_j \, dx$, Q the vector with entries $Q_i = \int_{V_i} f \, dx$, and $S(\tilde{\chi})$ the resulting stiffness matrix for $\chi \in \mathcal{X}_h$, i.e.,

$$S_{ij}(\tilde{\chi}) = -\int_{\partial V_i} A(\chi) \nabla \phi_j \cdot n \, ds, \qquad \text{for } \chi \in \mathcal{X}_h.$$

Since, this is a nonlinear problem we employ the following iteration: Set $\tilde{\xi}^0 = \tilde{U}^{n-1}$ and for $m = 1, 2, \dots,$ we solve

$$(D + kS(\tilde{\xi}^{m-1}))\tilde{\xi}^m = D\tilde{U}^{n-1} + kQ^n,$$

until some specified convergence. We note that if the iteration is stopped at $m = 1$, we recover the linearized backward Euler method. For all examples below, we use as a stopping criteria

$$\|(D + kS(\tilde{\xi}^{m-1}))\tilde{\xi}^m - D\tilde{U}^{n-1} - kQ^n\|_{l_\infty} \le \epsilon,$$

for some preassigned small number ϵ, with $\|\tilde{\chi}\|_{l_\infty} = \max_i |\tilde{\chi}_i|$.

We consider $\Omega = [0,1] \times [0,1]$ and partition $[0,1]$ into N equidistant intervals, thus N^2 squares are formed and divide each one into two triangles, which results in a mesh with size $h = \sqrt{2}/N$. Once the spatial mesh size is determined, the time step k is computed in such way that $k = h^{1.01}$.

We consider $u(x,y,t) = 8e^{-t}(x - x^2)(y - y^2)$ and use the nonlinear coefficient $A(u) = 1/(1 - 0.8\sin^2(4u))$, with forcing function f such that u satisfies the parabolic equation (1). We compute the error at final time $T = 1$ and the results are shown in Table 1. In both methods, the error convergence rate does follow the *a priori* estimates. We also see that in the linearized backward Euler, that as we decrease h, the error contribution from k starts to dominate. This is indicated by the decrease of the convergence order in the L_2–norm.

Acknowledgements The research of P. Chatzipantelidis was partly supported by the FP7-REGPOT-2009-1 project "Archimedes Center for Modeling Analysis and Computation", funded by the European Commission. The research of V. Ginting was partially supported by the grants from DOE (DE-FE0004832 and DE-SC0004982), the Center for Fundamentals of Subsurface Flow of the School of Energy Resources of the University of Wyoming (WYDEQ49811GNTG, WYDEQ49811PER), and from NSF (DMS-1016283).

Table 1 Comparison of errors of Backward Euler (BE) and Linearized Backward Euler (LBE) methods for various h with $k = h^{1.01}$.

h	BE				LBE			
	$\|u-u_h\|$	Rate	$\|u-u_h\|_1$	Rate	$\|u-u_h\|$	Rate	$\|u-u_h\|_1$	Rate
0.125	3.6569e-03	–	8.8974e-02	–	4.9954e-03	–	8.8928e-02	–
0.0625	9.0420e-04	2.02	4.4710e-02	0.99	1.6205e-03	1.62	4.4763e-02	0.99
0.03125	2.0321e-04	2.15	2.2382e-02	1.00	6.4270e-04	1.33	2.2460e-02	1.00
0.015625	4.1362e-05	2.20	1.1194e-02	1.00	2.7213e-04	1.24	1.12480e-02	1.00
0.0078125	8.3814e-06	2.30	5.5974e-03	1.00	1.2512e-04	1.12	5.6268e-03	1.00

References

1. Chatzipantelidis, P.: Finite volume methods for elliptic PDE's: a new approach. M2AN Math. Model. Numer. Anal. 36, 307–324 2002
2. Chatzipantelidis, P., Ginting, V., Lazarov, R.D.: A finite volume element method for a non-linear elliptic problem. Numer. Linear Algebra Appl. 12, 515–546 (2005)
3. Chatzipantelidis, P., Lazarov, R.D.: Error estimates for a finite volume element method for elliptic PDE's in nonconvex polygonal domains. SIAM J. Numer. Anal. 42, 1932–1958 (2005)
4. Chatzipantelidis, P., Lazarov, R.D., Thomée, V.: Error estimates for a finite volume element method for parabolic equations in convex polygonal domains. Numer. Methods Partial Differential Equations. 20, 650–674 (2004)
5. Chavent, G., Jaffré, J.: Mathematical Models and Finite Elements for Reservoir Simulation, Elsevier Science Publisher, B.V. Amsterdam, 1986.
6. Chen, Y., Yang, M., Bi, C.: Two–grid methods for finite volume element approximations of nonlinear parabolic equations. J. Comput. Appl. Math. 228, 123–132 (2009)
7. Chou, S.-H., Li, Q.: Error estimates in L^2, H^1 and L^∞ in covolume methods for elliptic and parabolic problems: a unified approach. Math. Comp. 69, 103–120 (2000)
8. Zhang, T., Zhong, H., Zhao, J.: A fully discrete two–grid finite–volume method for a nonlinear parabolic problem. Int. J. Comput. Math. 88, 1644–1663 (2011)
9. Eymard, R., Gallouët, T., Herbin, R.: Finite Volume Methods. In: Ciarlet, P.G., Lions, J.L.(eds.) Handbook of Numerical Analysis, Vol. VII, pp. 713–1020. North-Holland, Amsterdam (2000)
10. Keller, E., Segel, L.: Initiation of slime mold aggregation viewed as an instability. J. Theor. Biol. 26, 399–415 (1970)
11. Ladyženskaja, O.A., Solonnikov, V.A., UraÍceva, N.N.: Linear and Quasilinear Equations of Parabolic Type. Translated from the Russian by S. Smith. American Mathematical Society, Providence, R.I. (1968)
12. Li, R.: Generalized difference methods for a nonlinear Dirichlet problem. SIAM J. Numer. Anal. 24, 77–88 (1987)
13. Li, R., Chen, Z., Wu, W.: Generalized Difference Methods for Differential Equations. Marcel Dekker, New York (2000).
14. Richards, L. A.: Capillary conduction of liquids through porous medius. Physics 1, 318–333 (1931)
15. Thomée, V.: Galerkin Finite Element Methods for Parabolic Problems. Springer-Verlag, Berlin (2006)