Advances in Applied Mathematics and Mechanics
Adv. Appl. Math. Mech., Vol. **15**, No. 6, pp. 1407-1427

DOI: 10.4208/aamm.OA-2021-0313
December 2023

A Quadratic Finite Volume Method for Parabolic Problems

Yuanyuan Zhang and Xiaoping Liu*

School of Mathematics and Information Sciences, Yantai University, Yantai, Shandong 264005, China

Received 5 October 2021; Accepted (in revised version) 10 March 2022

Abstract. In this paper, a quadratic finite volume method (FVM) for parabolic problems is studied. We first discretize the spatial variables using a quadratic FVM to obtain a semi-discrete scheme. We then employ the backward Euler method and the Crank-Nicolson method respectively to further disctetize the time vatiable so as to derive two full-discrete schemes. The existence and uniqueness of the semi-discrete and full-discrete FVM solutions are established and their optimal error estimates are derived. Finally, we give numerical examples to illustrate the theoretical results.

AMS subject classifications: 65N15, 65N30

Key words: Higher-order finite volume method, parabolic problems, error estimate.

1 Introduction

Lots of scientific and engineering processes can be described by parabolic equations, such as diffusion, biomechanics, environmental protection, etc. Finite element methods (FEMs) for solving parabolic problems have been deeply studied, see e.g., [1, 4, 7, 16, 18, 23, 24, 30, 32]. Compared with the FEM, the FVM has an obvious advantage of preserving local conservation laws, which is crucial for many physical and engineering applications. Due to its advantages, the FVM has become a popular numerical method for solving partial differential equations (PDEs), see e.g., [6, 19, 20, 22, 29]. The purpose of this paper is to study a quadratic FVM discretization method based on triangular meshes for solving parabolic problems.

The linear FVMs for solving PDEs have been studied a lot and many results have been derived, see e.g., [3, 10, 17, 21]. Even though higher-order FVMs have great challenges in theoretical analysis compared with linear FVMs, they can obtain higher order convergence accuracy and have attracted many scholars' attention. The research on

*Corresponding author.
Emails: yyzhang@ytu.edu.cn (Y. Zhang), xpliu123@126.com (X. Liu)

higher-order FVMs for solving elliptic problems has made great progress in recent years, see e.g., [5,9,25,26,31,33]. For parabolic problems, Gao and Wang in [13] established the super-convergence property of a cubic FVM for one-dimensional parabolic equations. Yu and Li in [28] used optimal stress points to develop a biquadratic FVM on quadrilateral meshes. Yang, Liu and Zou in [27] presented a unified analysis of high order FVMs on quadrilateral meshes and derived their optimal error estimates. To our best knowledge, there is little work about higher-order FVMs based on triangular meshes for solving parabolic problems.

Most of the existing higher-order FVMs based on triangular meshes require that the primary meshes satisfy certain minimal angle conditions to ensure its optimal error estimates, see e.g., [8,26]. However, Zou in [33] first proposed a quadratic FV scheme which possesses the optimal H^1-norm error estimate over any shape regular triangular mesh without any additional minimal angle conditions. What's more, under a novel mapping from the trial space to the test space, its bilinear form can be regarded as a small perturbation of the corresponding quadratic FEM. This fact might greatly simplify its theoretical analysis.

In this paper, we discretize the spacial variables of the parabolic problems adopting the quadratic FVM developed in [33] for elliptic problems. The FVMs for elliptic problems come down to systems of linear equations, so that the inf-sup condition can guarantee the existence and uniqueness of their solutions and optimal error estimates, whereas the semi-discrete FVMs for parabolic equations are converted into ordinary differential equations. Hence, in order to obtain the existence and uniqueness of the solutions of the semi-discrete FVM, we need to prove that the mass matrix is nonsingular. In addition, to derive error estimates of the semi-discrete FVM, we introduce an elliptic projection operator. Then, the error can be written as the sum of two terms. One of them is the error between the exact solution and its projection (denoted by ρ) and the other is the error between the projection and the solution of the FVM (denoted by e). The error ρ can be easily estimated using the results presented in [33]. However, the error e is much more difficulty and we spent a lot of effort to deal with it. Fortunately, we get that its solution can reach optimal error estimate over any shape regular triangular mesh. We further employ the backward Euler method and the Crank-Nicolson method to discretize the time variables to get two full-discrete FVMs. Similar to the error estimate of semi-discrete solution, we mainly focus on estimating the term e and derive that the convergence order of the backward Euler full-discrete FVM reaches 2 in space variable and 1 in time variable, while the Crank-Nicolson full-discrete FVM enjoys the optimal convergence order of 2 in both space and time variables.

The rest of the paper is organized as follows. In Section 2, we present a quadratic FVM for solving parabolic problems and give the semi-discrete scheme. Section 3 is devoted to the theoretical analysis of the semi-discrete FVM, including the existence and uniqueness of the solution and the error estimate. In Section 4, we introduce the backward Euler full-discrete scheme and the Crank-Nicolson full-discrete scheme and give their error estimates. Finally, we present some numerical experiments in Section 5.

In this paper, the notations of Sobolev spaces and associated norms are the same as those in [11] and C will denote a generic positive constant independent of meshes and may be different at different occurrences.

2 A model problem and its quadratic semi-discrete FV scheme

Let $\Omega\subset R^2$ be a polygonal region with boundary $\partial\Omega$. We consider the initial-boundary value problem for the parabolic equations:

$$\begin{cases} u_t-\nabla\cdot(\alpha(x)\nabla u)=f(x,t) & \text{in } \Omega\times(0,T], \\ u=0 & \text{on } \partial\Omega\times(0,T], \\ u(x,0)=u_0(x) & \text{in } \Omega, \end{cases} \tag{2.1}$$

where $u=u(x,t)$ and $u_t=\frac{\partial u}{\partial t}$. We assume that there exist positive constants L, α_* and α^* such that

$$|\alpha'(x)|\le L \quad \text{and} \quad \alpha_*\le\alpha(x)\le\alpha^* \quad \text{for } \forall x\in\Omega. \tag{2.2}$$

Let $\mathcal{T}_h:=\{K\}$ be a triangulation of $\bar{\Omega}:=\Omega\cup\partial\Omega$, where the intersection of any two adjacent elements is either a common side or a common vertex. We use $|K|$ to denote the area of the triangle K, h_K to denote the diameter of K and ρ_K to denote the diameter of the largest ball contained in K, and let $h:=\max\{h_K|K\in\mathcal{T}_h\}$. We assume that $\mathcal{T}_h$ is shape regular (cf. [2]), that is, there exists a constant $\sigma>1$ such that

$$h_K\le\sigma\rho_K, \quad \forall K\in\mathcal{T}_h.$$

The trial space $\mathcal{U}_h$ is chosen as the Lagrange quadratic element space related to the triangulation $\mathcal{T}_h$

$$\mathcal{U}_h:=\left\{u\in C(\bar{\Omega}):u|_K\in\mathbf{P}_2,\ \forall K\in\mathcal{T}_h \text{ and } u|_{\partial\Omega}=0\right\},$$

where $\mathbf{P}_2$ denotes the set of polynomials of degree less than or equal to 2. For a triangle $K=\Delta p_1p_2p_3$ as plotted in Fig. 1, let m_i, $i=1,2,3$ be the midpoint of the edge $\overline{p_{i+1}p_{i+2}}$ with $p_4:=p_1$ and $p_5:=p_2$. We use λ_1, λ_2, λ_3 to denote its barycenter coordinates. It is known that the nodal basis restricted on K are given by

$$\phi_{p_i}=2\lambda_i^2-\lambda_i \quad \text{and} \quad \phi_{m_i}=4\lambda_{i+1}\lambda_{i+2}, \quad i=1,2,3, \tag{2.3}$$

where $\lambda_{i+3}=\lambda_i$, $i=1,2,3$.

We introduce the dual partition $\mathcal{T}_h^*:=\{K^*\}$ of $\mathcal{T}_h$, whose elements are called control volumes. Let $\mathcal{N}_h$ and $\mathcal{M}_h$ be the set of interior vertices and the set of mid-points of the internal edges of the elements in $\mathcal{T}_h$ respectively. Generally speaking, a control volume is a polygon K_p^* surrounding a vertex $p\in\mathcal{N}_h$ or a polygon K_m^* surrounding $m\in\mathcal{M}_h$. To get the control volumes, we need two parameters a and b. For a triangle $K=\triangle p_1p_2p_3$ as

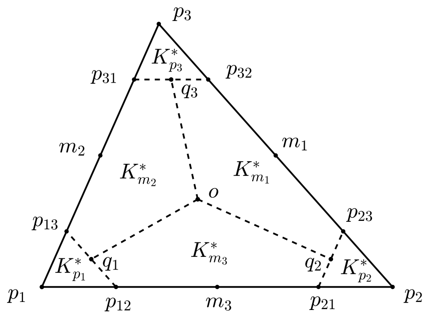
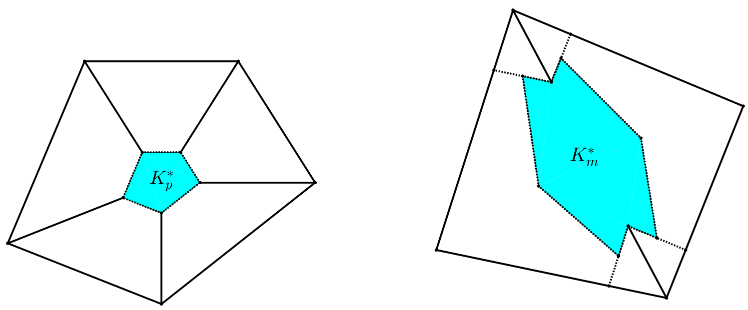


Figure 1: The chosen dual partition restricted on a triangle.



(a) (b)

Figure 2: (a): Control volume around a vertex, (b): Control volume around a midpoint.

plotted in Fig. 1, let o be its barycenter. We choose the points p_{ij}, $i,j=1,2,3$, $i\neq j$ and q_i, $i=1,2,3$ such that

$$\frac{|\overline{p_ip_{ij}}|}{|\overline{p_ip_j}|}=a,\quad \frac{|\overline{p_iq_i}|}{|\overline{p_im_i}|}=b.$$

Then, we get the dual partition $K^*_{p_i}$ and $K^*_{m_i}$, $i=1,2,3$ restricted on K. It is clear that the dual partition depends on the parameters a and b, and there are various choices of them in the existing literatures (cf. [8,15,26,33]). Different choices of a and b lead to different quadratic FVMs. In this paper, we employ that

$$a=b=\frac{1}{2}\left(1-\frac{1}{\sqrt{3}}\right).$$

Under this choice, we plot the dual partition restricted on K in Fig. 1 and the control volumes around a vertex and a midpoint respectively in Fig. 2.

The test space $\mathcal{V}_h$ is taken as the piecewise constant function space related to the dual decomposition $\mathcal{T}_h^*$

$$\mathcal{V}_h=\text{span}\left\{\psi_p,\ \psi_m : p\in\mathcal{N}_h,\ m\in\mathcal{M}_h\right\},$$

where ψ_p and ψ_m are the characteristic functions of K_p^* and K_m^* respectively.

Multiplying (2.1) by $v_h \in \mathcal{V}_h$ and using the Green's formula on the control volumes, we can obtain that

$$(u_t, v_h) + a_h(u, v_h) = (f, v_h), \tag{2.4}$$

where

$$a_h(u, v_h) := -\sum_{K^* \in \mathcal{T}_h^*} \int_{\partial K^*} \alpha \frac{\partial u}{\partial \mathbf{n}} v_h ds. \tag{2.5}$$

The semi-discrete FVM for (2.1) is: Find $u_h = u_h(\cdot, t) \in \mathcal{U}_h$ such that

$$\begin{cases} (u_{h,t}, v_h) + a_h(u_h, v_h) = (f, v_h), & \forall v_h \in \mathcal{V}_h, \quad 0 < t \leq T, \\ u_h(x, 0) = u_{0h}, & \forall x \in \Omega, \end{cases} \tag{2.6}$$

where u_{0h} is a certain approximation of u_0 in $\mathcal{U}_h$. For example, u_{0h} may be taken as a Lagrange interpolation of u_0 in $\mathcal{U}_h$.

We introduce an invertible mapping $\Pi_h^* : \mathcal{U}_h \to \mathcal{V}_h$. For $\forall v_h \in \mathcal{U}_h$, let $v_h^* := \Pi_h^* v_h$ satisfy that for each vertex $p \in \mathcal{N}_h$

$$v_h^*(p) = v_h(p),$$

and for each midpoint $m \in \mathcal{M}_h$

$$v_h^*(m) = \frac{v_h(p_1) + v_h(p_2)}{2} \left(1 - \frac{2}{\sqrt{3}}\right) + \frac{2}{\sqrt{3}} v_h(m),$$

where the edge $\overline{p_1 p_2}$ has m as its midpoint.

Using the mapping Π_h^*, the semi-discrete FVM (2.6) can be rewritten as finding $u_h = u_h(\cdot, t) \in \mathcal{U}_h$ such as

$$\begin{cases} (u_{h,t}, v_h^*) + a_h(u_h, v_h^*) = (f, v_h^*), & \forall v_h \in \mathcal{U}_h, \quad 0 < t \leq T, \\ u_h(x, 0) = u_{0h}, & \forall x \in \Omega. \end{cases} \tag{2.7}$$

By virtue of the basis of $\mathcal{U}_h$, we have that

$$u_h = \sum_{j=1}^{n} u_j(t) \phi_j = \sum_{p \in \mathcal{N}_h} u_p(t) \phi_p + \sum_{m \in \mathcal{M}_h} u_m(t) \phi_m.$$

Then, (2.7) is equivalent to

$$\begin{cases} \displaystyle\sum_{j=1}^{n} \Big[\frac{\partial u_j(t)}{\partial t} (\phi_j, \phi_i^*) + u_j(t) a_h(\phi_j, \phi_i^*) \Big] = (f, \phi_i^*), & 0 < t \leq T, \quad i = 1, \cdots, n, \\ u_j(0) = \alpha_j, & j = 1, \cdots, n, \end{cases} \tag{2.8}$$

where α_j are the coefficients in $u_{0h}=\sum_{j=1}^{n}\alpha_j\phi_j$. We introduce the following matrix and vector notations

$$\begin{aligned}&\mathbf{M}=[(\phi_j,\phi_i^*)]_{n\times n}, \quad \mathbf{K}=[a_h(\phi_j,\phi_i^*)]_{n\times n},\\ &\mathbf{u}=[u_1(t),\cdots,u_n(t)]^T, \quad \mathbf{F}=[(f,\phi_1^*),\cdots,(f,\phi_n^*)]^T, \quad \alpha=[\alpha_1,\cdots,\alpha_n]^T.\end{aligned}$$

We rewrite (2.8) as the following ordinary differential equation

$$\begin{cases}\mathbf{M}\dfrac{\partial \mathbf{u}}{\partial t}+\mathbf{K}\mathbf{u}=\mathbf{F},\\ \mathbf{u}(0)=\alpha.\end{cases} \tag{2.9}$$

As in the finite element method, we call $\mathbf{M}$ a mass matrix and $\mathbf{K}$ a stiff matrix.

3 Analysis of the semi-discrete FVM

In this section, we shall establish the existence and uniqueness of the semi-discrete FVM solution, and prove that the solution enjoys an optimal error estimate order of $\mathcal{O}(h^2)$ in the H^1-norm.

We choose a special triangle $\hat{K}$ with vertices $\hat{p}_1:=(0,0)$, $\hat{p}_2:=(1,0)$ and $\hat{p}_3:=(0,1)$ as a reference triangle. Let $K=\Delta p_1p_2p_3$ be an element in $\mathcal{T}_h$, there exists an unique invertible affine mapping $\mathcal{F}_K$ from $\hat{K}$ to K such that $\mathcal{F}_K(\hat{p}_i)=p_i$, $i=1,2,3$ (cf. [2]). For a u defined on K, we denote

$$\hat{u}:=u\circ\mathcal{F}_K.$$

From (2.3), we derive the basis of the trial space on $\hat{K}$:

$$\hat{\phi}_i:=\hat{\phi}_{p_i}=\phi_{p_i}\circ\mathcal{F}_K, \quad \hat{\phi}_{i+3}:=\hat{\phi}_{m_i}=\phi_{m_i}\circ\mathcal{F}_K, \quad i=1,2,3. \tag{3.1}$$

Similarly, we derive the basis of the test space on $\hat{K}$:

$$\hat{\psi}_i:=\hat{\psi}_{p_i}=\psi_{p_i}\circ\mathcal{F}_K, \quad \hat{\psi}_{i+3}:=\hat{\psi}_{m_i}=\psi_{m_i}\circ\mathcal{F}_K, \quad i=1,2,3. \tag{3.2}$$

We give the positive definiteness of $(\cdot,\Pi_h^*\cdot)$ in the next lemma.

Lemma 3.1. *There exists a constant $\beta>0$ independent of the trial space $\mathcal{U}_h$ such that*

$$(u_h,u_h^*)\geq\beta\|u_h\|_0^2, \quad \forall u_h\in\mathcal{U}_h \quad \text{with } u_h^*:=\Pi_h^*u_h. \tag{3.3}$$

Proof. It is sufficient to prove that for each $K\in\mathcal{T}_h$ (cf. Fig. 1)

$$(u_h,u_h^*)_K\geq\beta\|u_h\|_{0,K}^2. \tag{3.4}$$

By changing variables, we have that

$$(u_h,u_h^*)_K=2|K|(\hat{u}_h,\hat{u}_h^*)_{\hat{K}}, \quad \|u_h\|_{0,K}^2=2|K|(\hat{u}_h,\hat{u}_h)_{\hat{K}}. \tag{3.5}$$

We next discuss $(\hat{u}_h,\hat{u}_h^*)_{\hat{K}}$ and $(\hat{u}_h,\hat{u}_h)_{\hat{K}}$. Let

$$u_i:=u_h(p_i),\quad u_{i+3}:=u_h(m_i),\quad i=1,2,3,\quad \mathbf{w}:=[u_1,u_2,\cdots,u_6]^T,$$

and

$$\mathbf{A}_{\hat{K}}:=[(\hat{\phi}_j,\hat{\phi}_i^*)]_{6\times 6},\quad \mathbf{B}_{\hat{K}}:=[(\hat{\phi}_j,\hat{\phi}_i)]_{6\times 6},$$

where $\hat{\phi}_j$, $j=1,\cdots,6$ are defined as in (3.1) and $\hat{\phi}_i^*:=\Pi_h^*\hat{\phi}_i$. Then

$$(\hat{u}_h,\hat{u}_h^*)_{\hat{K}}=\Big(\sum_{j=1}^{6}u_j\hat{\phi}_j,\sum_{i=1}^{6}u_i\hat{\phi}_i^*\Big)=\mathbf{w}^T\mathbf{A}_{\hat{K}}\mathbf{w}=\mathbf{w}^T\frac{\mathbf{A}_{\hat{K}}^T+\mathbf{A}_{\hat{K}}}{2}\mathbf{w}, \tag{3.6a}$$

$$(\hat{u}_h,\hat{u}_h)_{\hat{K}}=\Big(\sum_{j=1}^{6}u_j\hat{\phi}_j,\sum_{i=1}^{6}u_i\hat{\phi}_i\Big)=\mathbf{w}^T\mathbf{B}_{\hat{K}}\mathbf{w}. \tag{3.6b}$$

Let

$$\bar{\mathbf{A}}_{\hat{K}}=[(\hat{\phi}_j,\hat{\psi}_i)]_{6\times 6},$$

where $\hat{\psi}_i$, $i=1,\ldots,6$ are defined as in (3.2). Using the definition of Π_h^*, the matrices $\mathbf{A}_{\hat{K}}$ and $\bar{\mathbf{A}}_{\hat{K}}$ have the relationship (cf. [33])

$$\mathbf{A}_{\hat{K}}=\bar{\mathbf{A}}_{\hat{K}}\mathbf{C}_K \tag{3.7}$$

with the invertible matrix

$$\mathbf{C}_K=\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0\\ 0 & 1 & 0 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 0 & 0\\ 0 & \frac{1}{2}-\frac{1}{\sqrt{3}} & \frac{1}{2}-\frac{1}{\sqrt{3}} & \frac{2}{\sqrt{3}} & 0 & 0\\ \frac{1}{2}-\frac{1}{\sqrt{3}} & 0 & \frac{1}{2}-\frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{3}} & 0\\ \frac{1}{2}-\frac{1}{\sqrt{3}} & \frac{1}{2}-\frac{1}{\sqrt{3}} & 0 & 0 & 0 & \frac{2}{\sqrt{3}} \end{pmatrix}.$$

Using the bases defined in (3.1) and (3.2), it is easy to calculate the matrices $\bar{\mathbf{A}}_{\hat{K}}$ and $\mathbf{B}_{\hat{K}}$. Then, we get the minimum eigenvalue of $\frac{\mathbf{A}_{\hat{K}}^T+\mathbf{A}_{\hat{K}}}{2}$ and the maximum eigenvalue of $\mathbf{B}_{\hat{K}}$ as follows

$$\lambda_{\min}\left(\frac{\mathbf{A}_{\hat{K}}^T+\mathbf{A}_{\hat{K}}}{2}\right)=0.0121,\quad \lambda_{\max}(\mathbf{B}_{\hat{K}})=0.1785. \tag{3.8}$$

Finally, combing (3.5), (3.6) and (3.8), we get the desired result (3.4) with $\beta=0.0677$. $\square$

With the aid of the above preparation, we are ready to prove the existence and uniqueness of the solution of the semi-discrete FVM.

Proposition 3.1. *The semi-discrete FVM* (2.6) *has a unique solution.*

Proof. According to the ordinary differential equation theory, we only need to prove that the matrix $\mathbf{M}$ in (2.9) is nonsingular. Namely, we have to show that the equation $\mathbf{M}x=0$ has only zero solution. Definition of $\mathbf{M}$ implies that

$$\sum_{j=1}^{n}(\phi_j,\phi_i^*)x_j=0, \quad i=1,\cdots,n,$$

which leads to

$$\Big(\sum_{j=1}^{n}x_j\phi_j,\sum_{i=1}^{n}x_i\phi_i^*\Big)=0. \tag{3.9}$$

Let $u_h:=\sum_{j=1}^{n}x_j\phi_j$. From (3.9) and the linearity of the mapping Π_h^*, we get that $(u_h,u_h^*)=0$. This combining with Lemma 3.1 yields that $\|u_h\|_0=0$, which implies that $x=0$. $\square$

We introduce a discrete L^2-norm on $\mathcal{U}_h$. For each $K\in\mathcal{T}_h$ as plotted in Fig. 1, let

$$\|u_h\|_{0,h,K}^2=|K|\left(\sum_{i=1}^{3}\left(u_{p_i}^2+u_{m_i}^2\right)\right), \quad \|u_h\|_{0,h}=\left(\sum_{K\in\mathcal{T}_h}\|u_h\|_{0,h,K}^2\right)^{\frac{1}{2}}. \tag{3.10}$$

Lemma 3.2. *There exist constants $c_i>0$, $i=1,2,3,4$ such that for all $u_h\in\mathcal{U}_h$ with $u_h^*:=\Pi_h^*u_h$*

$$c_1\|u_h\|_{0,h}\le\|u_h\|_0\le c_2\|u_h\|_{0,h}, \tag{3.11a}$$
$$c_3\|u_h\|_{0,h}\le\|u_h^*\|_0\le c_4\|u_h\|_{0,h}. \tag{3.11b}$$

Proof. The first norm equivalence relationship (3.11a) can be found in [14]. We next prove (3.11b). It is sufficient to prove that for each $u_h\in\mathcal{U}_h$ and each $K\in\mathcal{T}_h$ (cf. Fig. 1)

$$c_3\|u_h\|_{0,h,K}\le\|u_h^*\|_{0,K}\le c_4\|u_h\|_{0,h,K}. \tag{3.12}$$

Let

$$\mathbf{w}:=[u_{p_1},u_{p_2},u_{p_3},u_{m_1},u_{m_2},u_{m_3}]^T. \tag{3.13}$$

From the definition in (3.10), we have that

$$\|u_h\|_{0,h,K}^2=|K|\mathbf{w}^T\mathbf{w}. \tag{3.14}$$

By direct calculation, we get that

$$\|u_h^*\|_{0,K}^2=|K|(\mathbf{w}^*)^T\mathbf{A}\mathbf{w}^*, \tag{3.15}$$

where

$$\begin{aligned}&\mathbf{w}^*:=[u_{p_1}^*,u_{p_2}^*,u_{p_3}^*,u_{m_1}^*,u_{m_2}^*,u_{m_3}^*]^T,\\ &\mathbf{A}=\mathrm{diag}\Big(\frac{1}{3}-\frac{1}{2\sqrt{3}},\frac{1}{3}-\frac{1}{2\sqrt{3}},\frac{1}{3}-\frac{1}{2\sqrt{3}},\frac{1}{2\sqrt{3}},\frac{1}{2\sqrt{3}},\frac{1}{2\sqrt{3}}\Big).\end{aligned}$$

By the definition of the mapping Π_h^*, we get that

$$(\mathbf{w}^*)^T \mathbf{A} \mathbf{w}^* = \mathbf{w}^T \mathbf{C}_K \mathbf{A} \mathbf{C}_K^T \mathbf{w}, \tag{3.16}$$

where $\mathbf{C}_K$ is the same as that in Lemma 3.1. It is easy to see that $\mathbf{C}_K \mathbf{A} \mathbf{C}_K^T$ is a positive definite matrix. Then there exist positive constants c_3^2 and c_4^2 such that

$$c_3^2 \mathbf{w}^T \mathbf{w} \le \mathbf{w}^T \mathbf{C}_K \mathbf{A} \mathbf{C}_K^T \mathbf{w} \le c_4^2 \mathbf{w}^T \mathbf{w}. \tag{3.17}$$

Finally, from (3.14), (3.15), (3.16) and (3.17) we get (3.12). □

From [33], we have the following coercivity and boundedness of the bilinear form $a_h(\cdot,\Pi_h^* \cdot)$.

Lemma 3.3. *There exist constants γ and $M>0$ such that*

$$a_h(u_h,\Pi_h^* u_h) \ge \gamma \|u_h\|_1^2, \quad \forall u_h \in \mathcal{U}_h, \tag{3.18}$$

and

$$a_h(u_h,\Pi_h^* v_h) \le M \|u_h\|_1 \|v_h\|_1, \quad \forall u_h, v_h \in \mathcal{U}_h. \tag{3.19}$$

We introduce the finite element bilinear form

$$a(v,u) := \int_\Omega \alpha \nabla v \cdot \nabla u dx, \quad \forall v,u \in H_0^1(\Omega). \tag{3.20}$$

From [33], we know that the FV bilinear form $a_h(\cdot,\Pi_h^* \cdot)$ and the finite element bilinear form $a(\cdot,\cdot)$ have the following relationship.

Lemma 3.4. *If the coefficient α is piecewise constant with respect to $\mathcal{T}_h$, then there holds*

$$a_h(u_h,v_h^*) = a(u_h,v_h), \quad \forall u_h, v_h \in \mathcal{U}_h \quad \text{with } v_h^* := \Pi_h^* v_h.$$

The next lemma tells us that the bilinear form $a_h(\cdot,\Pi_h^* \cdot)$ is nearly symmetric for the coefficient α satisfying (2.2).

Lemma 3.5. *Under the condition* (2.2), *there holds*

$$|a_h(u_h,v_h^*) - a_h(v_h,u_h^*)| \le Ch \|u_h\|_1 \|v_h\|_1, \quad \forall u_h, v_h \in \mathcal{U}_h,$$

where $u_h^* := \Pi_h^* u_h$ and $v_h^* := \Pi_h^* v_h$.

Proof. For each $K \in \mathcal{T}_h$, we define the average of α on K

$$\alpha_K := \frac{1}{|K|} \int_K \alpha(x) dx,$$

and let

$$\bar{a}_h(u_h,v_h^*)=-\sum_{K\in\mathcal{T}_h}\left(\sum_{p\in\mathcal{N}_h}\int_{\partial K_p^*\cap K}\alpha_K\frac{\partial u_h}{\partial n}v_h^*(p)ds+\sum_{m\in\mathcal{M}_h}\int_{\partial K_m^*\cap K}\alpha_K\frac{\partial u_h}{\partial n}v_h^*(m)ds\right).$$

By Lemma 3.4, we have that

$$\bar{a}_h(u_h,v_h^*)=\bar{a}_h(v_h,u_h^*).$$

Then

$$\begin{aligned}&|a_h(u_h,v_h^*)-a_h(v_h,u_h^*)|\\ \le&|a_h(u_h,v_h^*)-\bar{a}_h(u_h,v_h^*)|+|\bar{a}_h(v_h,u_h^*)-a_h(v_h,u_h^*)|.\end{aligned}\tag{3.21}$$

The proof of Theorem 3.3 in [33] indicates that

$$|a_h(u_h,v_h^*)-\bar{a}_h(u_h,v_h^*)|\le Ch\|u_h\|_1\|v_h\|_1,\tag{3.22a}$$
$$|\bar{a}_h(v_h,u_h^*)-a_h(v_h,u_h^*)|\le Ch\|v_h\|_1\|u_h\|_1.\tag{3.22b}$$

Substituting (3.22a) and (3.22b) into (3.21) yields the result of this lemma. □

Let us introduce an elliptic projection operator

$$P_h:H^2(\Omega)\cap H_0^1(\Omega)\to\mathcal{U}_h,$$

defined by the following equation

$$a_h(P_hu,v_h)=a_h(u,v_h),\quad\forall v_h\in\mathcal{V}_h.\tag{3.23}$$

Lemma 3.3 guarantees that the projection operator is well defined and from [33], we have the following error estimate.

Lemma 3.6. *Let P_hu be the elliptic projection of u defined by* (3.23)*, then*

$$\|u-P_hu\|_1\le Ch^2\|u\|_3.\tag{3.24}$$

We are ready to present the error estimate of the semi-discrete FVM.

Theorem 3.1. *Let u and u_h be the solutions of* (2.1) *and* (2.6) *respectively. Then for all $t\in[0,T]$, there holds*

$$\|u-u_h\|_1\le C\left\{\|u_0-u_{0h}\|_1+h^2\left[\|u_0\|_3+\int_0^t\|u_t\|_3dt+\left(\int_0^t\|u_t\|_3^2dt\right)^{\frac{1}{2}}\right]\right\}.$$

Proof. Let

$$\rho := u - P_h u, \quad e := P_h u - u_h.$$

Then, we have

$$u - u_h = \rho + e \quad \text{and} \quad \|u - u_h\|_1 \le \|\rho\|_1 + \|e\|_1. \tag{3.25}$$

It follows from Lemma 3.6 that

$$\|\rho\|_1 \le Ch^2 \|u\|_3 = Ch^2 \|u_0 + \int_0^t u_t dt\|_3 \le Ch^2 \left[\|u_0\|_3 + \int_0^t \|u_t\|_3 dt \right]. \tag{3.26}$$

We next focus on estimating e. Since u and u_h satisfy (2.4) and (2.6) respectively, we get that

$$(u_t - u_{h,t}, v_h) + a_h(u - u_h, v_h) = 0, \quad \forall v_h \in \mathcal{V}_h. \tag{3.27}$$

From (3.23) and (3.27), we have that

$$\begin{aligned}(e_t, v_h) + a_h(e, v_h) &= (P_h u_t - u_{h,t}, v_h) + a_h(P_h u - u_h, v_h) \\ &= (P_h u_t, v_h) - (u_{h,t}, v_h) + a_h(u, v_h) - a_h(u_h, v_h) \\ &= (P_h u_t, v_h) - (u_t, v_h) = -(\rho_t, v_h), \quad \forall v_h \in \mathcal{V}_h. \end{aligned} \tag{3.28}$$

Choosing $v_h = \Pi_h^* e_t$ in (3.28), we get that

$$(e_t, \Pi_h^* e_t) + a_h(e, \Pi_h^* e_t) = -(\rho_t, \Pi_h^* e_t). \tag{3.29}$$

We note that

$$a_h(e, \Pi_h^* e_t) = \frac{1}{2} \frac{d}{dt} a_h(e, \Pi_h^* e) + \frac{1}{2} [a_h(e, \Pi_h^* e_t) - a_h(e_t, \Pi_h^* e)]. \tag{3.30}$$

From Lemma 3.5 and inverse inequality, we have

$$|a_h(e, \Pi_h^* e_t) - a_h(e_t, \Pi_h^* e)| \le Ch \|e\|_1 \|e_t\|_1 \le C \|e\|_1 \|e_t\|_0. \tag{3.31}$$

Lemma 3.2 leads to that

$$|-(\rho_t, \Pi_h^* e_t)| \le \|\rho_t\|_0 \|\Pi_h^* e_t\|_0 \le C \|\rho_t\|_0 \|e_t\|_0. \tag{3.32}$$

From (3.29), (3.30), and (3.32), we get that

$$(e_t, \Pi_h^* e_t) + \frac{1}{2} \frac{d}{dt} a_h(e, \Pi_h^* e) \le C \|\rho_t\|_0 \|e_t\|_0 + C \|e\|_1 \|e_t\|_0.$$

The above inequality combined with Lemma 3.1 and the Hölder inequality yields that

$$\beta \|e_t\|_0^2 + \frac{1}{2} \frac{d}{dt} a_h(e, \Pi_h^* e) \le C \|\rho_t\|_0^2 + \beta \|e_t\|_0^2 + C \|e\|_1^2,$$

that is

$$\frac{d}{dt}a_h(e,\Pi_h^* e)\le C\|\rho_t\|_0^2+C\|e\|_1^2.$$

Integrating it with respect to t and using Lemma 3.3, we get that

$$\begin{aligned}\gamma\|e\|_1^2\le & a_h(e,\Pi_h^* e)\\ \le & a_h(e(0),\Pi_h^* e(0))+C\int_0^t\|\rho_t\|_0^2dt+C\int_0^t\|e\|_1^2dt\\ \le & M\|e(0)\|_1^2+C\int_0^t\|\rho_t\|_0^2dt+C\int_0^t\|e\|_1^2dt. \end{aligned}\tag{3.33}$$

Note that

$$\begin{aligned}\|e(0)\|_1=&\|P_hu_0-u_{0h}\|_1\\ &\le\|P_hu_0-u_0\|_1+\|u_0-u_{0h}\|_1\\ &\le Ch^2\|u_0\|_3+\|u_0-u_{0h}\|_1, \end{aligned}\tag{3.34a}$$

$$\|\rho_t\|_0=\|u_t-P_hu_t\|_0\le Ch^2\|u_t\|_3. \tag{3.34b}$$

Substituting (3.34a) and (3.34b) into (3.33), we derive that

$$\|e\|_1^2\le C\int_0^t\|e\|_1^2dt+C\left\{\|u_0-u_{0h}\|_1^2+h^4\left[\|u_0\|_3^2+\int_0^t\|u_t\|_3^2dt\right]\right\}.$$

Using the Gronwall lemma to the above inequality, we arrive at

$$\|e\|_1\le C\left\{\|u_0-u_{0h}\|_1+h^2\left[\|u_0\|_3+\left(\int_0^t\|u_t\|_3^2dt\right)^{\frac{1}{2}}\right]\right\}. \tag{3.35}$$

Finally, substituting (3.26) and (3.35) into (3.25) leads to the desired result of the theorem. □

4 The fully discrete FVMs and their error estimates

In the last section, the semi-discrete FVM is obtained by discretizing the spatial variable only. In this section, we further discretize the time variable to get fully discrete FVMs. We shall introduce two methods for time discretization: the backward Euler method and the Crank-Nicolson method.

Let τ denote a time step size and set $t_n:=n\tau$ and $u_h^n:=u_h(t_n)$, $(n=0,1,\cdots)$. At time $t=t_n$, we apply the backward Euler method to approximate $u_{h,t}$ by the difference quotient

$$\partial_t u_h^n=\frac{u_h^n-u_h^{n-1}}{\tau}.$$

Then, we get the backward Euler full-discrete FVM: Finding $u_h^n \in \mathcal{U}_h$ such that

$$\begin{cases} (\partial_t u_h^n, v_h) + a_h(u_h^n, v_h) = (f(t_n), v_h), \quad \forall v_h \in \mathcal{V}_h, \quad n = 1, \cdots, \\ u_h^0 = u_{0h}. \end{cases} \tag{4.1}$$

Or we can equivalently write it as

$$\begin{cases} (u_h^n, v_h) + \tau a_h(u_h^n, v_h) = (u_h^{n-1} + \tau f(t_n), v_h), \quad \forall v_h \in \mathcal{V}_h, \quad n = 1, \cdots, \\ u_h^0 = u_{0h}. \end{cases} \tag{4.2}$$

Using the mapping Π_h^*, we rewrite (4.2) as

$$(u_h^n, v_h^*) + \tau a_h(u_h^n, v_h^*) = (u_h^{n-1} + \tau f(t_n), v_h^*), \quad \forall v_h \in \mathcal{U}_h \quad \text{with } v_h^* := \Pi_h^* v_h. \tag{4.3}$$

We give the existence and uniqueness of the solution of (4.1).

Proposition 4.1. *The backward Euler FVM* (4.1) *has a unique solution.*

Proof. By Lemma 3.1 and Lemma 3.3, we obtain

$$(u_h^n, \Pi_h^* u_h^n) + \tau a_h(u_h^n, \Pi_h^* u_h^n) \geq \beta \|u_h^n\|_0^2 + \tau\gamma \|u_h^n\|_1^2, \quad \forall u_h^n \in \mathcal{U}_h.$$

This guarantees that the coefficient matrix of the unknowns in (4.3) is nonsingular. That is, there exists a unique solution u_h^n for a given u_h^{n-1}. □

We shall prove the following error estimate for the backward Euler full-discrete FVM.

Theorem 4.1. *Let u and u_h^n be the solutions of* (2.1) *and* (4.1), *respectively. Then there holds*

$$\begin{aligned} &\|u(t_n) - u_h^n\|_1 \\ \leq & C\Big\{ \|u_0 - u_{0h}\|_1 + h^2 \Big[\|u_0\|_3 + \int_0^{t_n} \|u_t\|_3 dt + \Big(\int_0^{t_n} \|u_t\|_3^2 dt \Big)^{\frac{1}{2}} \Big] \\ & + \tau \Big(\int_0^{t_n} \|u_{tt}\|_0^2 dt \Big)^{\frac{1}{2}} \Big\}, \quad n = 0, 1, \cdots. \end{aligned} \tag{4.4}$$

Proof. Let

$$\rho^n := u(t_n) - P_h u(t_n), \quad e^n := P_h u(t_n) - u_h^n,$$

where the projection operator P_h is defined as in (3.23). Then

$$\|u(t_n) - u_h^n\|_1 = \|\rho^n + e^n\|_1 \leq \|\rho^n\|_1 + \|e^n\|_1. \tag{4.5}$$

It follows from Lemma 3.6 that

$$\|\rho^n\|_1 \leq Ch^2 \|u(t_n)\|_3 \leq Ch^2 \Big[\|u_0\|_3 + \int_0^{t_n} \|u_t\|_3 dt \Big]. \tag{4.6}$$

We next estimate $\|e^n\|_1$. Set $t=t_n$ in (2.4), we obtain that

$$(u_t(t_n),v_h)+a_h(u(t_n),v_h)=(f(t_n),v_h), \quad \forall v_h\in\mathcal{V}_h. \tag{4.7}$$

From (4.7) and (4.1), we have that

$$(u_t(t_n)-\partial_t u_h^n,v_h)+a_h(u(t_n)-u_h^n,v_h)=0, \quad \forall v_h\in\mathcal{V}_h. \tag{4.8}$$

By virtue of (3.23) and (4.8), we derive that

$$\begin{aligned}(\partial_t e^n,v_h)+a_h(e^n,v_h)&=(\partial_t P_h u(t_n)-\partial_t u_h^n,v_h)+a_h(P_h u(t_n)-u_h^n,v_h)\\&=(\partial_t P_h u(t_n),v_h)-(\partial_t u_h^n,v_h)+a_h(u(t_n),v_h)-a_h(u_h^n,v_h)\\&=(\partial_t P_h u(t_n),v_h)-(u_t(t_n),v_h)\\&=(\omega^n,v_h), \quad \forall v_h\in\mathcal{V}_h,\end{aligned} \tag{4.9}$$

where $\omega^n:=\partial_t P_h u(t_n)-u_t(t_n)$. Setting $v_h=\Pi_h^*\partial_t e^n$ in (4.9), and using Lemmas 3.1 and 3.2 leads to that

$$\beta\|\partial_t e^n\|_0^2+a_h(e^n,\Pi_h^*\partial_t e^n)\le\|\omega^n\|_0\|\partial_t e^n\|_0. \tag{4.10}$$

It follows from Lemma 3.3 that

$$\begin{aligned}&a_h(e^n,\Pi_h^*\partial_t e^n)=a_h\Big(e^n,\Pi_h^*\frac{e^n-e^{n-1}}{\tau}\Big)\\=&\frac{1}{2\tau}[a_h(e^n+e^{n-1},\Pi_h^*(e^n-e^{n-1}))+a_h(e^n-e^{n-1},\Pi_h^*(e^n-e^{n-1}))]\\\ge&\frac{1}{2\tau}a_h(e^n+e^{n-1},\Pi_h^*(e^n-e^{n-1}))\\=&\frac{1}{2\tau}[a_h(e^n,\Pi_h^*e^n)-a_h(e^{n-1},\Pi_h^*e^{n-1})]+\frac{1}{2}[a_h(e^n,\Pi_h^*\partial_t e^n)-a_h(\partial_t e^n,\Pi_h^*e^n)].\end{aligned} \tag{4.11}$$

Substituting (4.11) into (4.10) and applying Lemma 3.5, the Hölder inequality and the inverse inequality produces that

$$a_h(e^n,\Pi_h^*e^n)\le a_h(e^{n-1},\Pi_h^*e^{n-1})+C\tau\|\omega^n\|_0^2+C\tau\|e^n\|_1^2.$$

The above recursion relation leads to that

$$a_h(e^n,\Pi_h^*e^n)\le a_h(e^0,\Pi_h^*e^0)+C\tau\sum_{j=1}^n\|\omega^j\|_0^2+C\tau\sum_{j=1}^n\|e^j\|_1^2.$$

This together with Lemma 3.3 leads to that

$$\|e^n\|_1^2\le C\Big(\|e^0\|_1^2+\tau\sum_{j=1}^n\|\omega^j\|_0^2+\tau\sum_{j=1}^n\|e^j\|_1^2\Big). \tag{4.12}$$

It remains to estimate the right terms of (4.12).

By Lemma 3.6, we have that

$$\|e^0\|_1^2 \le \|P_h u_0 - u_0\|_1^2 + \|u_0 - u_{0h}\|_1^2 \le Ch^4 \|u_0\|_3^2 + \|u_0 - u_{0h}\|_1^2. \tag{4.13}$$

We note that

$$\omega^j = \partial_t P_h u(t_j) - u_t(t_j) = \omega_1^j + \omega_2^j, \tag{4.14}$$

where

$$\omega_1^j := \partial_t P_h u(t_j) - \partial_t u(t_j) = \frac{1}{\tau} \int_{t_{j-1}}^{t_j} (P_h u_t - u_t) dt, \tag{4.15a}$$

$$\omega_2^j := \partial_t u(t_j) - u_t(t_j) = -\frac{1}{\tau} \int_{t_{j-1}}^{t_j} (t - t_{j-1}) u_{tt} dt. \tag{4.15b}$$

By the Cauchy-Schwartz inequality and Lemma 3.6, we have that

$$\tau \sum_{j=1}^n \|\omega_1^j\|_0^2 \le \sum_{j=1}^n \int_{t_{j-1}}^{t_j} \|P_h u_t - u_t\|_0^2 dt \le Ch^4 \int_0^{t_n} \|u_t\|_3^2 dt, \tag{4.16a}$$

$$\tau \sum_{j=1}^n \|\omega_2^j\|_0^2 \le \frac{\tau^2}{3} \int_0^{t_n} \|u_{tt}\|_0^2 dt. \tag{4.16b}$$

From (4.14), (4.16a) and (4.16b), we obtain that

$$\tau \sum_{j=1}^n \|\omega^j\|_0^2 \le C\Big(h^4 \int_0^{t_n} \|u_t\|_3^2 dt + \tau^2 \int_0^{t_n} \|u_{tt}\|_0^2 dt\Big).$$

This together with (4.13) and (4.12) leads to that

$$\begin{aligned} \|e^n\|_1^2 \le C\tau \sum_{j=1}^n \|e^j\|_1^2 + C\Big\{ &\|u_0 - u_{0h}\|_1^2 + h^4 [\|u_0\|_3^2 \\ &+ \int_0^{t_n} \|u_t\|_3^2 dt] + \tau^2 \int_0^{t_n} \|u_{tt}\|_0^2 dt \Big\}. \end{aligned} \tag{4.17}$$

By applying the Gronwall inequality to the above inequality, we derive that

$$\begin{aligned} \|e^n\|_1 \le C\Big\{ &\|u_0 - u_{0h}\|_1 + h^2 \Big[\|u_0\|_3 + \Big(\int_0^{t_n} \|u_t\|_3^2 dt)^{\frac{1}{2}}\Big] \\ &+ \tau \Big(\int_0^{t_n} \|u_{tt}\|_0^2 dt\Big)^{\frac{1}{2}} \Big\}. \end{aligned} \tag{4.18}$$

Finally, the desired result (4.4) comes from (4.5), (4.6) and (4.18). □

Note that the backward Euler full-discrete FVM has only first order convergence rate in the time step size τ. We now present another full-discrete FV scheme: the Crank-Nicolson FVM, which enjoys the second order convergence rate in τ. It reads: find $u_h^n \in \mathcal{U}_h$ such that

$$\begin{cases} (\partial_t u_h^n, v_h) + a_h\Big(\dfrac{u_h^n + u_h^{n-1}}{2}, v_h\Big) = \Big(\dfrac{f(t_n)+f(t_{n-1})}{2}, v_h\Big), & \forall v_h \in \mathcal{V}_h, \quad n=1,\cdots, \\ u_h^0 = u_{0h}. \end{cases} \tag{4.19}$$

Similar as Proposition 4.1, we can get the existence and uniqueness of the solution of (4.19) for a given u_h^{n-1}.

The error estimate of the Crank-Nicolson FVM is established in the next theorem.

Theorem 4.2. *Let u and u_h^n be the solutions of* (2.1) *and* (4.19) *respectively. Then there holds*

$$\begin{aligned} &\|u(t_n) - u_h^n\|_1 \\ \le & C\Big\{ \|u_0 - u_{0h}\|_1 + h^2\Big[\|u_0\|_3 + \int_0^{t_n} \|u_t\|_3 dt + \Big(\int_0^{t_n} \|u_t\|_3^2 dt\Big)^{\frac{1}{2}} \Big] \\ & + \tau^2 \Big(\int_0^{t_n} \|u_{ttt}\|_0^2 dt \Big)^{\frac{1}{2}} \Big\}, \quad n = 0,1,\cdots. \end{aligned} \tag{4.20}$$

Proof. As the proof of Theorem 4.1, we write the error as follows

$$\|u(t_n) - u_h^n\|_1 = \|\rho^n + e^n\|_1 \le \|\rho^n\|_1 + \|e^n\|_1. \tag{4.21}$$

We next estimate $\|e^n\|_1$. By (3.23), (4.7) and (4.19), we have that

$$(\partial_t e^n, v_h) + a_h\Big(\frac{e^n + e^{n-1}}{2}, v_h\Big) = (r^n, v_h), \quad \forall v_h \in \mathcal{V}_h, \tag{4.22}$$

where

$$r^n := \partial_t P_h u(t_n) - \frac{u_t(t_n) + u_t(t_{n-1})}{2}.$$

Taking $v_h = \Pi_h^* \partial_t e^n$ in (4.22), and using Lemmas 3.1 and 3.2, we get tat

$$\beta \|\partial_t e^n\|_0^2 + a_h\Big(\frac{e^n + e^{n-1}}{2}, \Pi_h^* \partial_t e^n\Big) \le C\|r^n\|_0 \|\partial_t e^n\|_0. \tag{4.23}$$

By simply identical transformation in (4.23), we derive that

$$\begin{aligned} &\beta \|\partial_t e^n\|_0^2 + \frac{1}{2\tau}[a_h(e^n, \Pi_h^* e^n) - a_h(e^{n-1}, \Pi_h^* e^{n-1})] \\ \le & C\|r^n\|_0 \|\partial_t e^n\|_0 + \frac{1}{2}[a_h(e^n, \Pi_h^* \partial_t e^n) - a_h(\partial_t e^n, \Pi_h^* e^n)]. \end{aligned}$$

The above inequality combined with Lemma 3.5 yields that

$$a_h(e^n,\Pi_h^* e^n) \le a_h(e^{n-1},\Pi_h^* e^{n-1}) + C\tau \|r^n\|_0^2 + C\tau \|e^n\|_1^2.$$

Using the above recursion relation, we have that

$$a_h(e^n,\Pi_h^* e^n) \le a_h(e^0,\Pi_h^* e^0) + C\tau \sum_{j=1}^n \|r^j\|_0^2 + C\tau \sum_{j=1}^n \|e^j\|_1^2.$$

This together with Lemma 3.3 leads to that

$$\|e^n\|_1^2 \le C\Big(\|e^0\|_1^2 + \tau \sum_{j=1}^n \|r^j\|_0^2 + \tau \sum_{j=1}^n \|e^j\|_1^2 \Big). \tag{4.24}$$

We write

$$r^j = w_1^j + r_2^j,$$

where w_1^j is the same as that in (4.15a) and

$$r_2^j := \partial_t u(t_j) - \frac{u_t(t_j)+u_t(t_{j-1})}{2} = \frac{u(t_j)-u(t_{j-1})}{\tau} - \frac{u_t(t_j)+u_t(t_{j-1})}{2}.$$

The Taylor expansion tells us that

$$u(t_j) = u(t_{j-1}) + \tau u_t(t_{j-1}) + \frac{\tau^2}{2} u_{tt}(t_{j-1}) + \int_{t_{j-1}}^{t_j} \frac{(t_j-t)^2}{2} u_{ttt}(t) dt, \tag{4.25a}$$

$$u_t(t_j) = u_t(t_{j-1}) + \tau u_{tt}(t_{j-1}) + \int_{t_{j-1}}^{t_j} (t_j-t) u_{ttt}(t) dt. \tag{4.25b}$$

From (4.25a) and (4.25b), we have that

$$r_2^j = -\frac{1}{2}\int_{t_{j-1}}^{t_j} (t_j-t) u_{ttt}(t) dt + \frac{1}{\tau}\int_{t_{j-1}}^{t_j} \frac{(t_j-t)^2}{2} u_{ttt}(t) dt.$$

This combined with the Cauchy-Schwartz inequality leads to that

$$|r_2^j| \le \frac{4}{15} \tau^{\frac{3}{2}} \Big(\int_{t_{j-1}}^{t_j} |u_{ttt}|^2 dt \Big)^{\frac{1}{2}}.$$

Hence

$$\tau \sum_{j=1}^n \|r_2^j\|_0^2 \le C\tau^4 \int_0^{t_n} \|u_{ttt}\|_0^2 dt.$$

This together with (4.16a), (4.13) and (4.24) yields that

$$\|e^n\|_1^2 \le C\tau \sum_{j=1}^n \|e^j\|_1^2 + C\Big\{ \|u_0 - u_{0h}\|_1^2 + h^4 \Big[\|u_0\|_3^2 + \int_0^{t_n} \|u_t\|_3^2 dt \Big] + \tau^4 \int_0^{t_n} \|u_{ttt}\|_0^2 dt \Big\}.$$

By applying the discrete form of the Gronwall inequality, we have that

$$\begin{aligned}\|e^n\|_1 \le C\Big\{ &\|u_0-u_{0h}\|_1+h^2\Big[\|u_0\|_3+\Big(\int_0^{t_n}\|u_t\|_3^2dt\Big)^{\frac{1}{2}}\Big]\\ &+\tau^2\Big(\int_0^{t_n}\|u_{ttt}\|_0^2dt\Big)^{\frac{1}{2}}\Big\}.\end{aligned} \tag{4.26}$$

Finally, the desired result of this theorem comes from (4.21), (4.6) and (4.26). □

5 Numerical examples

In order to confirm the theoretical results numerically, we shall give three numerical examples in this section. In all examples, we choose the domain $\bar{\Omega}\times[0,T]=[0,1]^2\times[0,1]$. We subdivide the region $\bar{\Omega}$ into $N\times N$ equal rectangles. The triangle mesh of $\bar{\Omega}$ is then obtained by connecting the diagonal lines of the resulting rectangles.

We consider solving the model problem (2.1). In the first example, we take the analytical solution

$$u(x,y,t)=e^{-t}(x-x^2)(y-y^2),$$

and use the constant coefficient

$$a(x,y)=1.$$

Then the right side function $f(x,y,t)$ and initial value function $u_0(x,y)$ are determined. This solution is sufficiently smooth in the spatial variables. In the second example, we set the analytical solution

$$u(x,y,t)=e^t x^4y^4\ln(x)\ln(y),$$

and use the constant coefficient

$$a(x,y)=1.$$

Obviously, this solution is an element in $H^3(\Omega)$, but not in $H^4(\Omega)$. In the third example, we set the analytical solution

$$u(x,y,t)=e^{-t}(x-x^2)(y-y^2),$$

and use the variable coefficient

$$a(x,y)=x^2+y^2.$$

We compute the H^1-norm errors and the convergence orders of the backward Euler and Crank-Nicolson FVMs at final time $T=1$. The numerical results for the three examples are presented respectively in Table 1, Table 2 and Table 3. We can see that their convergence rates can reach the optimal order of 2, which is consistent with the theoretical findings in previous section.

Table 1: Error estimates and convergence rate for the first example.

N	backward Euler FVM:$\tau=h^2$		Crank-Nicolson FVM:$\tau=h$	
	$\lvert u-u_h\rvert_1$	Rate	$\lvert u-u_h\rvert_1$	Rate
4	3.0523e-3		3.0417e-3	
8	7.7714e-4	1.9737	7.7649e-4	1.9698
16	1.9523e-4	1.9930	1.9520e-4	1.9920
32	4.8868e-5	1.9982	4.8868e-5	1.9980
64	1.2221e-5	1.9995	1.2221e-5	1.9995

Table 2: Error estimates and convergence rate for the second example.

N	backward Euler FVM:$\tau=h^2$		Crank-Nicolson FVM:$\tau=h$	
	$\lvert u-u_h\rvert_1$	Rate	$\lvert u-u_h\rvert_1$	Rate
4	1.1995e-2		1.1963e-2	
8	3.5913e-3	1.7399	3.5817e-3	1.7399
16	9.4369e-4	1.9281	9.4278e-4	1.9257
32	2.3912e-4	1.9806	2.3908e-4	1.9794
64	5.9993e-5	1.9949	5.9995e-5	1.9946

Table 3: Error estimates and convergence rate for the third example.

N	backward Euler FVM:$\tau=h^2$		Crank-Nicolson FVM:$\tau=h$	
	$\lvert u-u_h\rvert_1$	Rate	$\lvert u-u_h\rvert_1$	Rate
4	3.0744e-3		3.1520e-3	
8	7.7865e-4	1.9813	7.8620e-4	2.0033
16	1.9533e-4	1.9951	1.9603e-4	2.0038
32	4.8875e-5	1.9988	4.8936e-4	2.0021
64	1.2221e-5	1.9997	1.2226e-5	2.0009

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under grant No. 11901506, and by the Shandong Province Natural Science Foundation under grant No. ZR2018QA003 and ZR2021MA010.

References

[1] S. ADJERID, M. AIFFA, AND J. E. FLAHERTY, *High-order finite element methods for singularly perturbed elliptic and parabolic problems,* SIAM J. Appl. Math., 55 (1995), pp. 520–543.

[2] S. C. BRENNER AND L. R. SCOTT, The Mathematical Theory of Finite Element Methods, Springer-Verlag, New York, 1994.

[3] P. CHATZIPANTELIDIS, R. D. LAZAROV AND V. THOMEE, *Parabolic finite volume element equations in non-convex polygonal domains,* Numer. Methods PDEs, 25 (2008), pp. 507–525.

[4] P. CHATZIPANTELIDIS, R. D. LAZAROV AND V. THOMEE, *Some error estimates for the lumped mass finite element method for a parabolic problem,* Math. Comput., 81 (2012), pp. 1–20.

[5] L. CHEN, *A new class of high order finite volume methods for second order elliptic equations,* SIAM J. Numer. Anal., 47 (2010), pp. 4021–4043.

[6] P. CHEN, J. SUN, C. LIN AND W. ZHOU, *Application of the finite volume method for geomechanics calculation and analysis on temperature dependent poromechanical stress and displacement fields in enhanced geothermal system,* Geothermics, 95 (2021), p. 102138.

[7] Z. CHEN AND J. FENG, *An adaptive finite element algorithm with reliable and efficient error control for linear parabolic problems,* Math. Comput., 73 (2004), pp. 1167–1193.

[8] Z. CHEN, J. WU AND Y. XU, *Higher-order finite volume methods for elliptic boundary value problem,* Adv. Comput. Math., 37 (2012), pp. 191–253.

[9] Z. CHEN, Y. XU, AND Y. ZHANG, *A construction of higher-order finite volume methods,* Math. Comput., 84 (2015), pp. 599–628.

[10] S. CHOU AND Q. LI, *Error estimates in L^2, H^1, and L^∞ in covolume methods for elliptic and parabolic problems: A unified approach,* Math. Comput., 69 (2000), pp. 103–120.

[11] P. G. CIARLET, The Finite Element Method for Elliptic Problems, North-Holland, Amsterdam, 1978.

[12] F. GAO AND D. LIANG, *A new weighted upwind finite volume element method based on non-standard covolume for time-dependent convection-diffusion problems,* Int. J. Numer. Meth. Fl., 73 (2013), pp. 953–975.

[13] G. GAO AND T. WANG, *Cubic superconvergent finite volume element method for one-dimensional elliptic and parabolic equations,* J. Comput. Appl. Math., 233 (2010), pp. 2285–2301.

[14] R. LI, Z. CHEN AND W. WU, Generalized Difference Methods for Differential Equations: Numerical Analysis of Finite Volume Methods, Marcel Dekker, New York, 2000.

[15] F. LIEBAU, *The finite volume element method with quadratic basis finctions,* Computing, 57 (1996), pp. 281–299.

[16] Y. LIU AND Y. NIE, *A priori and a posteriori error estimates of the weak Galerkin finite element method for parabolic problems,* Comput. Math. Appl., 99 (2021), pp. 73–83.

[17] X. MA, S. SHI AND A. ZHOU, *Symmetric finite volume discretizations for parabolic problems,* Comput. Methods Appl. Mech. Eng., 192 (2003), pp. 4467–4485.

[18] X. MAKRIDAKII AND I. BABUSKA, *On the stability of the discontinuous Galerkin method for the heat equation,* SIAM J. Numer. Anal., 34 (1997), pp. 389–401.

[19] S. C. MISHRA AND H. K. ROY, *Solving transient conduction and radiation heat transfer problems using the lattice Boltzmann method and the finite volume method,* J. Comput. Phys., 223 (2007), pp. 89–107.

[20] R. A. NICOLAIDES, T. A. PORSHING AND C. A. HALL, *Covolume methods in computational fluid dynamics,* Comput. Fluid D. Rev., (1995), pp. 279–299.

[21] R. K. SINHA AND J. GEISER, *Error estimates for finite volume element methods for convection-diffusion-reaction equations,* Appl. Numer. Math., 57 (2007), pp. 59–72.

[22] I. SOKOLOVA, M. G. BASTISYA AND H. HAJIBEYGI, *Multiscale finite volume method for finite-volume-based simulation of poroelasticity,* J. Comput. Phys., 379 (2019), pp. 309–324.

[23] O. J. SUTTON, *Long-time $L^\infty(L^2)$ a posteriori error estimates for fully discrete parabolic problems,* IMA J. Numer. Anal., 40 (2020), pp. 498–529.

[24] V. THOMEE, Galerkin Finite Element Methods for Parabolic Problems, Springer-Verlag, New York, 1997.

[25] X. WANG AND Y. LI, *L^2 error estimates for high order finite volume methods on triangular meshes,* SIAM J. Numer. Anal., 54 (2016), pp. 2729–2749.

[26] J. XU AND Q. ZOU, *Analysis of linear and quadratic simplicial finite volume methods for elliptic equation,* Numer. Math., 111 (2009), pp. 469–492.

[27] M. YANG, J. LIU AND Q. ZOU, *Unified analysis of higher-order finite volume methods for parabolic problems on quadrilateral meshes,* IMA J. Numer. Anal., 36 (2016), pp. 872–896.

[28] C. YU AND Y. LI, *Biquadratic finite volume element methods based on optimal stress points for parabolic problems,* J. Comput. Appl. Math., 236 (2011), pp. 1055–1068.

[29] R. ZHANG, L. ZHANG, J. LUO, Z. YANG AND M. XU, *Numerical simulation of water flooding in natural fractured reservoirs based on control volume finite element method,* J. Petrol. Sci. Eng., 146 (2016), pp. 1211–1225.

[30] Y. ZHANG AND D. SHI, *Superconvergence of an H^1-Galerkin nonconforming mixed finite element method for a parabolic equation,* Comput. Math. Appl., 66 (2013), pp. 2362–2375.

[31] Y. ZHANG AND M. YANG, *A posteriori error analysis of any order finite volume methods for elliptic problems,* Adv. Appl. Math. Mech., 12 (2020), pp. 564–578.

[32] S. ZHOU, F. GAO, B. LI AND Z. SUN, *Weak Galerkin finite element method with second-order accuracy in time for parabolic problems,* Appl. Math. Lett., 90 (2019), pp. 118–123.

[33] Q. ZOU, *An unconditionally stable quadratic finite volume scheme over triangular meshes for elliptic equations,* J. Sci. Comput., 70 (2017), pp. 112–124.