A POSITIVE AND BOUNDED FINITE ELEMENT APPROXIMATION OF THE GENERALIZED BURGERS-HUXLEY EQUATION

V.J. ERVIN *, J.E. MACÍAS-DÍAZ †, AND J. RUIZ-RAMÍREZ ‡

Abstract. We present a finite element scheme capable of preserving the nonnegative and bounded solutions of the generalized Burgers-Huxley equation. Proofs of existence and uniqueness of a solution to the continuous problem together with some results concerning the boundedness and the nonnegativity of the solution are given. Under appropriate conditions on the mesh and the initial and boundary data, boundedness and nonnegativity of the finite element approximation are established. An *a priori* error estimate for the approximation is also derived. Numerical experiments are presented which support the derived theoretical results.

Key words. Finite element method; convection-diffusion-reaction; nonnegativity; boundedness

AMS subject classifications. 65M60, 35B09, 35K57

1. Introduction.

The class of advection-diffusion-reaction problems has been used extensively to model different physical processes such as atmospheric air quality [20], the mobility of fish populations [27], nuclear waste disposal [14], pattern formation [23] and the reactive transport of contaminants [16]. An important member of this class and the object of study in this paper is the generalized Burgers-Huxley equation, which has found application in biology [28], electrodynamics [29] and transport phenomena [1]. For $\Omega \subset \mathbb{R}^2$, the nonlinear partial differential equation under consideration is:

$$\frac{\partial u}{\partial t} + \alpha u^p \left(\frac{\partial u}{\partial x} + \frac{\partial u}{\partial y} \right) - \Delta u - u g(u) = 0, \quad (\mathbf{x}, t) \in \Omega \times (0, T), \tag{1.1}$$

where $\alpha \in \mathbb{R}^+$ is the advection coefficient, $\gamma \in (0, 1)$, $p \geq 1$ are parameters and

$$g(u) = \beta(1 - u^p)(u^p - \gamma), \quad \beta > 0. \tag{1.2}$$

The numerical approximation in 1D of (1.1) using a variety of techniques has received considerable attention. One approach is the Adomian decomposition [15] which needs no discretization, linearization or perturbation, and provides an analytical solution in the form of a power series using Adomian polynomials. A second approach, which can be classified as a Lagrange multiplier method, is the variational iteration method (VIM) [2]. This technique uses a linerization assumption as an initial approximation and later, through a correction functional, the approximation is made more precise. Results have shown that in some cases one iteration of VIM is of comparable accuracy to a 5-term Adomian solution. A third approach which belongs to the class of interpolation techniques is the spectral collocation method (SC) [7]. Using Chebyshev-Gauss-Lobatto collocation points, an interpolant polynomial is constructed and a differential operator in terms of the grid point values is computed. The matrices that appear in

*Department of Mathematical Sciences, Clemson University, Clemson, SC 29634, USA. (vjervin@clemson.edu)

†Departamento de Matemáticas y Física, Universidad Autónoma de Aguascalientes, Aguascalientes, Ags. 20131, México (jemacias@correo.uaa.mx)

‡Department of Mathematical Sciences, Clemson University, Clemson, SC 29634, USA. (javier@clemson.edu)

SC are generally ill-conditioned and some preconditioning is requierd. A fourth approach is the use of Haar wavelets [4]. The procedure relies on the decomposition of a $L^2([a,b])$ function in terms of the orthogonal Haar basis. The matrices that arise in the computations are sparse and the accuracy is in general high even with few collocation points. A fifth approach is finite differences [26], where a Runge-Kutta method of order 4 is used in time and coupled with a 10-th order finite difference scheme in space. A final example of approximation methods investigated in 1D is the finite element method [19], where a three-step Taylor-Galerkin finite element scheme is implemeneted to approximate a problem with the diffusion term being multiplied by a perturbation parameter ε.
Of interest is the design of numerical methods capable of preserving the nonnegative and bounded solutions of (1.1). Among all the aforementioned methods, only finite differences and finite element methods have been successful in being able to supply proofs for the conservation of those properties. Works in this direction in the case of finite differences include [22, 21, 25]. If we restrict ourselves to the smaller class of linear parabolic problems of the form $u_t - \nabla \cdot (\kappa \nabla u) = f$, where κ satisfies $0 < \kappa_{\min} \leq \kappa(\mathbf{x}) \leq \kappa_{\max}$ and f is bounded, for finite differences and finite element the (discrete) nonnegativity preservation property is equivalent to the (discrete) maximum principle, a bridge that was established in [10]. Some articles on finite elements concerning maximum principles for advection-reaction-diffusion problems are [6, 18, 9, 11].
In this document we propose a finite element scheme capable of preserving the nonnegative and bounded solutions of (1.1) under suitable conditions for the computational parameters $\Delta t, h$, and on the triangulation $\mathcal{T}_h$ of the region Ω. In Section 2 we establish a result concerning the existence and uniqueness of a solution to the continuous problem (1.1). Moreover, we prove that for certain initial and boundary conditions, any classical solution to (1.1) satisfies certain boundedness conditions. Then, in Section 3, we state the weak formulation and prove the boundedness of the solution in this setting. The discrete weak formulation is introduced in Section 4 and therein we show the computability of the scheme and state parallel results to the ones derived for the continuous problem. In Section 5 we provide an *a priori* error estimate which guarantees that the method is of first order accurate, with respect to the space and time discretizations. Finally, in Section 6, numerical experiments are presented which support our theoretical results.

2. Continuous problem. In this section we establish the existence of a solution to (1.1) for some time interval of positive length $[0, T_0]$. Then, we show that under suitable boundary and initial conditions, the solution u is nonnegative and bounded.

Before we state our first result, we introduce the following definition.

DEFINITION 2.1. *The real valued function $F(\mathbf{x}, t, z_1, \ldots, z_r)$, $F : \Omega \times \mathbb{R}^+ \cup \{0\} \times \mathbb{R}^r \to \mathbb{R}$, with $\Omega \subset \mathbb{R}^2$, is locally Hölder continuous with respect to $(\mathbf{x}, t)$ if for all $B \subset \overline{\Omega} \times [0, T]$, B compact, there exist a constant $C > 0$ and $0 < \kappa < 1$ such that*

$$|F(\mathbf{x}_1, t_1, z_1, \ldots, z_r) - F(\mathbf{x}_2, t_2, z_1, \ldots, z_r)| \leq C \|(\mathbf{x}_1, t_1) - (\mathbf{x}_2, t_2)\|_\ell^\kappa$$

for all $(\mathbf{x}, t) \in B$, where $\| \cdot \|_\ell$ is any vector norm on $\mathbb{R}^3$.

In order to prove the existence and uniqueness of a solution for (1.1), we use the following result from the theory of quasi-linear parabolic equations.

THEOREM 2.2 ([12], pg. 206). *Consider the following parabolic problem*

$$\begin{aligned} u_t &= Au + F(\mathbf{x}, t, u, \nabla u), \text{ subject to} \\ u(x,t) &= \widehat{u}(x,t) \text{ for } (\mathbf{x}, t) \in \overline{\Omega} \times \{0\} \cup \partial\Omega \times [0, T], \end{aligned} \tag{2.1}$$

where A is an elliptic operator of order 2. If the following assumptions hold:

1. *$\widehat{u}$ is smooth,*
2. *$F(\mathbf{x}, t, u, \nabla u)$ is locally Hölder continuous with respect to $(\mathbf{x}, t)$,*
3. *$F(\mathbf{x}, t, u, \nabla u)$ is Lipschitz continuous in u, uniformly for bounded subsets of $\overline{\Omega} \times [0, T] \times \mathbb{R} \times \mathbb{R}^2$,*
4. *$\widehat{u}_t = A\widehat{u} + F(\mathbf{x}, 0, \widehat{u}, \nabla\widehat{u})$ holds for $(x, t) \in \partial\Omega \times \{0\}$,*

then there exist a unique solution to (2.1) in $\Omega \times [0, T_0]$ for some $0 < T_0 \leq T$.

COROLLARY 2.3. *If the boundary and initial conditions of (1.1) satisfy the assumptions of Theorem 2.2, then there exists $T_0 > 0$ such that a unique solution to (1.1) exists in $\overline{\Omega} \times [0, T_0]$.*

Proof. Using (1.1), let B be a compact subset of $\overline{\Omega} \times [0, T]$, and $F(\mathbf{x}, t, u, \nabla u) = -\alpha u^p \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot \nabla u + ug(u)$. Observe that for $(\mathbf{x}_1, t_1)$, $(\mathbf{x}_2, t_2) \in B$, $|F(\mathbf{x}_1, t_1, u, \nabla u) - F(\mathbf{x}_2, t_2, u, \nabla u)| = 0$, i.e. property 2 is satisfied. Now for B a bounded subset of $\overline{\Omega} \times [0, T] \times \mathbb{R} \times \mathbb{R}^2$, and $(\mathbf{x}, t, u_1, \nabla u)$, $(\mathbf{x}, t, u_2, \nabla u) \in B$,

$$\begin{aligned} |F(\mathbf{x}, t, u_1, \nabla u) - F(\mathbf{x}, t, u_2, \nabla u)| &= |r(u_1) - r(u_2)| \\ &\leq \sup_{(\mathbf{x}, t, u, \nabla u) \in B} |r'(u)||u_1 - u_2|, \end{aligned}$$

where $r(x)$ is a polynomial of degree $2p+1$ whose coefficients may depend on $\mathbf{x}, t$ and ∇u. Thus, property 3 holds and consequently F satisfies all assumptions of Theorem 2.2. □

The next result establishes the boundedness of the solution u to (1.1).

THEOREM 2.4.

1. *If $u(x, t)$ is a classical solution of (1.1) satisfying $0 \leq u(\mathbf{x}, 0) < \gamma^{1/p}$ for $\mathbf{x} \in \overline{\Omega}$ and $0 \leq u(\mathbf{x}, t) < \gamma^{1/p}$ for $\mathbf{x} \in \partial\Omega$ and $t > 0$, then $0 \leq u(\mathbf{x}, t) < \gamma^{1/p}$ for $\mathbf{x} \in \overline{\Omega}$ and $t \geq 0$.*
2. *If $u(x, t)$ is a classical solution of (1.1) satisfying $\gamma^{1/p} < u(\mathbf{x}, 0) \leq 1$ for $\mathbf{x} \in \overline{\Omega}$ and $\gamma^{1/p} < u(\mathbf{x}, t) \leq 1$ for $\mathbf{x} \in \partial\Omega$ and $t > 0$, then $\gamma^{1/p} < u(\mathbf{x}, t) \leq 1$ for $\mathbf{x} \in \overline{\Omega}$ and $t \geq 0$.*

Proof.

1. First we show that $u(\mathbf{x}, t) \geq 0$ for $\mathbf{x} \in \overline{\Omega}$ and $t \geq 0$. Assume that there exists $\mathbf{x}_0$ and t_0 such that $u(\mathbf{x}_0, t_0) < 0$. This implies the existence of δ, t_δ and $\mathbf{x}_\delta$ such that $0 < \delta < \gamma$, and $u(\mathbf{x}_\delta, t_\delta) = -\delta$ with the property that $u(\mathbf{x}_\delta, t_\delta)$ is a local minimum for $u(\mathbf{x}, t_\delta)$. Thus, using (1.1) and noting that $g(u(\mathbf{x}_\delta, t_\delta)) < 0$, we obtain

$$\begin{aligned} u_t(\mathbf{x}_\delta, t_\delta) &= \Delta u - \alpha u^p(u_x + u_y) + ug(u)\Big|_{(\mathbf{x}_\delta, t_\delta)} \\ &\geq 0 - \alpha u^p(\mathbf{x}_\delta, t_\delta)(0 + 0) + u(\mathbf{x}_\delta, t_\delta)g(u(\mathbf{x}_\delta, t_\delta)) \\ &> 0. \end{aligned} \tag{2.2}$$

Now, as $u(\mathbf{x},0) \geq 0$ for $\mathbf{x} \in \overline{\Omega}$, we may assume further that t_δ is the first time when this occurs. Then, for some $\varepsilon > 0$ we must have that $u(\mathbf{x}_\delta, t)$ is strictly decreasing for $t \in (t_\delta - \varepsilon, t_\delta)$. However, this contradicts (2.2). Hence we must conclude that $u(\mathbf{x},t) \geq 0$ for $\mathbf{x} \in \overline{\Omega}$ and $t > 0$.

To establish the second part, note that if $u(\mathbf{x},t)$ satisfying $0 < u(\mathbf{x},t) < \gamma^{1/p}$ and (1.1) has an interior local maximum at $\mathbf{x}_m \in \Omega$ at $t = t_m$, then

$$\begin{aligned} u_t(\mathbf{x}_m, t_m) &= \Delta u - \alpha u^p(u_x + u_y) + u g(u)\Big|_{(\mathbf{x}_m, t_m)} \\ &\leq 0 - \alpha u^p(\mathbf{x}_m, t_m)(0+0) + u(\mathbf{x}_m, t_m) g(u(\mathbf{x}_m, t_m)) \\ &< 0. \end{aligned}$$

Thus $u(\mathbf{x},t)$ must be strictly decreasing at any such interior maximum point. As $u(\mathbf{x},0) < \gamma^{1/p}$ and $u(\mathbf{x},t) < \gamma^{1/p}$ for $\mathbf{x} \in \partial\Omega$ and $t > 0$, then it immediatly follows that $u(\mathbf{x},t) < \gamma^{1/p}$ for $\mathbf{x} \in \overline{\Omega}$, $t \geq 0$.

2. An analogous argument establishes this case.

□

A simple consequence of the previous Theorem is the following.

COROLLARY 2.5. *Under the assumptions of Theorem 2.4, together with the constraint that* $0 \leq u(\mathbf{x},0) \leq 1$ *for* $\mathbf{x} \in \overline{\Omega}$ *and* $0 \leq u(\mathbf{x},t) \leq 1$ *for* $\mathbf{x} \in \partial\Omega$ *and* $t > 0$*, then* $0 \leq u(\mathbf{x},t) \leq 1$ *for* $\mathbf{x} \in \overline{\Omega}$ *and* $t \geq 0$.

Proof. The proof follows from the nonnegativity and the upper bound arguments discussed in Theorem 2.4. □

3. Variational formulation. In this section we introduce the weak formulation of (1.1) and the notation that will be used for the rest of the document. We consider a bounded convex polygonal domain $\Omega \subset \mathbb{R}^2$ and a regular triangulation of it $\mathcal{T}_h$, with mesh parameter h. For $T \in T_h$, $P_1(T)$ will represent the vector space of affine functions on T. As usual, $\|\cdot\|_{L^p(\Omega)}$, $\|\cdot\|_p$ will be used to denote the norms in $L^p(\Omega)$ and $H^p(\Omega)$, respectively. The inner product in L^2 will be denoted by $(\cdot,\cdot)$ and we let $\|\cdot\| = \|\cdot\|_{L^2(\Omega)}$, and $\|\cdot\|_\infty = \|\cdot\|_{L^\infty(\Omega)}$.

The spaces that we will consider in the analysis are

$$X = H_0^1(\Omega) = \left\{v \in H^1(\Omega) : v|_{\partial\Omega} = 0\right\}, \text{ and} \tag{3.1}$$

$$X_h = \left\{v \in C^0(\Omega) : v|_T \in P_1(T), \forall T \in \mathcal{T}_h,\, v|_{\partial\Omega} = 0\right\}. \tag{3.2}$$

Taking a test function $v \in X$, multiplying (1.1) by v, and integrating, we obtain

$$\begin{aligned} (u_t, v) + a(u,v) + b(u^p, u, v) - (ug(u), v) = 0,\ \forall v \in X,\ 0 < t \leq T \\ \text{subject to } u(\mathbf{x},0) = u_0(\mathbf{x}),\ \mathbf{x} \in \Omega, \end{aligned} \tag{3.3}$$

where

$$a(u,v) = \int_\Omega \nabla u \cdot \nabla v \, d\Omega \quad \text{and} \quad b(u,v,w) = \alpha \int_\Omega u \left(\frac{\partial v}{\partial x} + \frac{\partial v}{\partial y}\right) w \, d\Omega. \tag{3.4}$$

The following lemmas will be useful to prove the boundedness of the solution of (3.3).

LEMMA 3.1. *For* $p, s \in \mathbb{R}^+$, $u \in H^1(\Omega) \cap L^{p+s}(\Omega)$*, and* $u|_{\partial\Omega} = 0$*, we have that* $b(u^p, u, u^s) = 0$.

Proof. We have

$$b(u^p, u, u^s) = \int_\Omega u^p \left(\frac{\partial u}{\partial x} + \frac{\partial u}{\partial y}\right) u^s \, d\Omega = \int_\Omega u^{p+s} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot \nabla u \, d\Omega$$
$$= -\int_\Omega u \nabla \cdot \left(u^{p+s} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right) d\Omega = -(p+s) \int_\Omega u^{p+s} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot \nabla u \, d\Omega.$$

Since $p+s$ is nonnegative, this last equality implies that

$$(p+s+1)\, b(u^p, u, u^s) = 0.$$

□

LEMMA 3.2. *For $s \geq 2$ and $u^{s-1} \in H^1(\Omega)$, we have that*

$$a(u, u^{s-1}) = \frac{4(s-1)}{s^2} a(u^{s/2}, u^{s/2}).$$

Proof. Note that

$$\nabla u^{s/2} = \frac{s}{2} u^{s/2-1} \nabla u \iff \nabla u = \frac{2}{s} u^{1-s/2} \nabla u^{s/2}. \tag{3.5}$$

From (3.4) and using (3.5)

$$a(u, u^{s-1}) = \int_\Omega \nabla u \cdot \nabla u^{s-1} \, d\Omega = \int_\Omega \nabla u \cdot (s-1) u^{s-2} \nabla u \, d\Omega$$
$$= \int_\Omega \frac{2}{s} u^{1-s/2} \nabla u^{s/2} \cdot (s-1) u^{s-2} \frac{2}{s} u^{1-s/2} \nabla u^{s/2} \, d\Omega$$
$$= \frac{4(s-1)}{s^2} a(u^{s/2}, u^{s/2}).$$

□

LEMMA 3.3. *Let $m = 2n$ with $n \in \mathbb{N}$ be given, and define $g_b = \max_{s \in [-1,1]} |g(s)|$. If $u_0 \in L^m(\Omega)$, then any solution u to (3.3) that belongs to $L^m(\Omega)$ satisfies the bound*

$$\|u(t)\|^m_{L^m(\Omega)} + \frac{4(m-1)}{m^3} \int_0^t \left(\exp\Big((t-s) g_b / m\Big) \|\nabla u^{m/2}\|^2 \right) ds$$
$$\leq \|u_0\|^m_{L^m(\Omega)} \exp\Big(t\, g_b / m\Big).$$

Proof. With the choice $v = \dfrac{1}{m} u^{m-1}$, and using Lemmas 3.1 and 3.2, we obtain from (3.3)

$$\frac{d}{dt} \|u^{m/2}\|^2 + \frac{4(m-1)}{m^3} \|\nabla u^{m/2}\| - \frac{1}{m} (u g(u), u^{m-1}) = 0.$$

Noting that m is even, g is a polynomial in u, and if $|u| > 1$, then $g(u) < 0$, we obtain the following bound

$$(u g(u), u^{m-1}) = \int_{|u| \leq 1} u^m g(u) \, d\Omega + \int_{|u| > 1} u^m g(u) \, d\Omega$$
$$\leq g_b \int_{|u| \leq 1} u^m \, d\Omega \;\leq\; g_b \|u^{m/2}\|^2.$$

Thus,

$$\frac{d}{dt}\|u^{m/2}\|^2 + \frac{4(m-1)}{m^3}\|\nabla u^{m/2}\| - \frac{g_b}{m}\|u^{m/2}\|^2 \le 0. \tag{3.6}$$

Letting $\mu(t) = -t\, g_b/m$, multiplying (3.6) by $\exp(\mu(t))$ and integrating, we obtain

$$\int_0^t \frac{d}{ds}\left(\exp(\mu(s))\|u^{m/2}\|^2\right) ds + \frac{4(m-1)}{m^3}\int_0^t \exp(\mu(s))\|\nabla u^{m/2}\|^2\, ds \le 0. \tag{3.7}$$

From (3.7) we can conclude that

$$\|u^{m/2}(t)\|^2 + \frac{4(m-1)}{m^3}\int_0^t \exp(\mu(s-t))\|\nabla u^{m/2}\|^2\, ds \le \|u_0^{m/2}\|^2 \exp(\mu(-t)).$$

□

In order to establish that solutions of (3.3) are also bounded in $L^p(\Omega)$, for p odd, we recall the following interpolation result.

Lemma 3.4. *Let $0 < p_1 < p_2 < \infty$ and $f \in L^{p_1}(\Omega) \cap L^{p_2}(\Omega)$. Then $f \in L^p(\Omega)$ for all $p_1 \le p \le p_2$. Moreover, we have that*

$$\|f\|_{L^{p_\lambda}(\Omega)} \le \|f\|^{1-\lambda}_{L^{p_1}(\Omega)}\|f\|^{\lambda}_{L^{p_2}(\Omega)} \text{ for all } 0 \le \lambda \le 1,$$

where $1/p_\lambda = (1-\lambda)/p_1 + \lambda/p_2$.

Proof. Using Hölder's inequality with $p = p_1/((1-\lambda)p_\lambda)$ and $q = p_2/(\lambda p_\lambda)$, we obtain

$$\|f\|^{p_\lambda}_{L^{p_\lambda}(\Omega)} = \int_\Omega |f|^{(1-\lambda)p_\lambda}|f|^{\lambda p_\lambda} \le \|f^{(1-\lambda)p_\lambda}\|_{L^p(\Omega)}\|f^{\lambda p_\lambda}\|_{L^q(\Omega)}. \tag{3.8}$$

Taking the p_λ-th root of (3.8), the result follows. □

Theorem 3.5. *For $u_0 \in L^p(\Omega)$ for all $t \in [0,T]$, we have that any solution $u(\mathbf{x},t)$ of (3.3) is bounded in the L^p norm, $p \ge 2$ on the interval $[0,T]$. Moreover, $\|u(t)\|_\infty \le \|u_0\|_\infty$ for all $t \in [0,T]$.*

Proof. We consider two cases. The case when $p < \infty$ follows from the interpolation property of Lemma 3.4 applied to Lemma 3.3. The case $p = \infty$ follows by taking the m-th root of the bound established in Lemma 3.3, and then taking the limit as $m \to \infty$. □

4. Discrete problem. In this section we introduce the fully discrete finite element scheme and derive some properties of it. For the temporal discretization, for M given, let $\Delta t = T/M$, $t_k = k\Delta t$, and $f^k = f(t_k)$. Replacing the time derivative with a backward Euler discretization and lagging all nonlinearities in (3.3) results in the following problem: Compute $u_h^k \in X_h$ for $k = 1, \dots, M$, such that for all $v \in X_h$

$$(\frac{u_h^k - u_h^{k-1}}{\Delta t}, v) + (\nabla u_h^k, \nabla v) + b((u_h^{k-1})^p, u_h^k, v) - (u_h^k g(u_h^{k-1}), v) = 0, \tag{4.1}$$

$$\text{subject to } u_h^0(\mathbf{x}_i) = u^0(\mathbf{x}_i) \text{ for } i = 1, \dots, N, \tag{4.2}$$

where the $\mathbf{x}_i$ represent the nodes of the triangulation $\mathcal{T}_h$ of Ω.

Let $\{\phi_j(\mathbf{x})\}_{j=1}^N$ be a Lagrangian basis for X_h, where $\phi_j(\mathbf{x}) \in X_h$ is defined by

$$\phi_j(\mathbf{x}) = \begin{cases} 1, & \mathbf{x} = \mathbf{x}_j \\ 0, & \mathbf{x} \ne \mathbf{x}_j \end{cases}. \tag{4.3}$$

With $\mathbf{b}^k = \alpha(u_h^{k-1})^p \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $g^k = g(u_h^{k-1})$, $u_h^k(\mathbf{x}) = \sum_{j=1}^N c_j \phi_j(\mathbf{x})$, and $v = \phi_i$, $i = 1, \ldots, N$, equation (4.1) is equivalent to

$$A\mathbf{c} = \mathbf{r}, \tag{4.4}$$

where

$$\begin{aligned} A_{ij} &= \frac{1}{\Delta t}\int_\Omega \phi_j \phi_i \, d\Omega + \int_\Omega \nabla\phi_j \cdot \nabla\phi_i \, d\Omega + \int_\Omega \mathbf{b}^k \cdot \nabla\phi_j \ \phi_i \, d\Omega \\ &\quad - \int_\Omega g^k \phi_j \, \phi_i \, d\Omega, \text{ and} \qquad (4.5) \\ \mathbf{r}_i &= \frac{1}{\Delta t}\int_\Omega u_h^{k-1}\phi_i \, d\Omega. \qquad (4.6) \end{aligned}$$

In [24], conditions to ensure that A is an M-matrix [13] were investigated. These conditions involved the following assumptions on the mesh: For every triangle T not adjacent to $\partial\Omega$, there are positive constants c_h, C_h, c_θ, c_J and C_J such that:

A1. For e_i the edge opposite to the i-th vertex, $c_h h \le \ell(e_i) \le C_h h$, where ℓ denotes the length.

A2. The angle φ subtended at any vertex satisfies $0 < c_\theta \le \cos\varphi < 1$.

A3. $c_J h^2 \le |J| \le C_J h^2$,

where $|J|$ is the determinant of the matrix that appears in the mapping from the reference triangle $\widehat{T}$ to the triangle $T \in \mathcal{T}_h$ (see [24] (4.6)).
For the approximation of u_h^k in (4.1) we make the following assumption for u_h^{k-1} which we later confirm in Theorem 4.3.

A4. For $k \ge 1$, $0 \le u_h^{k-1} \le 1$.

Note that under assumption ***A4*** we have that $\|\mathbf{b}^k\|_\infty \le \alpha$ and $\|g^k\|_\infty \le \beta(\frac{1-\gamma}{2})^2 \le \beta$. The computability of the numerical scheme can now be established.

THEOREM 4.1. *Under assumptions **A1**-**A4**, for h and Δt satisfying*

$$0 < \frac{C_J h^2}{24}\left(\frac{1}{6C_J}c_h^2 c_\theta - \frac{\sqrt{2}}{6}\alpha\, h\, C_h - \frac{\beta}{24}\left(\frac{1-\gamma}{2}\right)^2 C_J h^2\right)^{-1} \le \Delta t \le \frac{1}{(1-\gamma)\beta p} \tag{4.7}$$

there exists a unique solution u_h^k, $k = 1, 2, \ldots, M$, satisfying (4.1).

Proof. From [24], (4.7) guarantees that the square coeffcent matrix A in (4.5) is an M-matrix. Hence A is invertible, which implies a unique solution u_h^k to (4.1). ☐

We now address the question of boundedness for the numerical approximation of (4.1). Note that with the (nonnegative) Lagrangian basis functions ϕ_j (see (4.3)), $u_h^k(\mathbf{x}) \ge 0$ is equivalent to $\mathbf{c} \ge 0$ (see (4.4)).

LEMMA 4.2. *Let the assumptions of Theorem 4.1 hold.*

1. *If u_h^{k-1} in* (4.1) *satisfies $0 \le u_h^{k-1}(\mathbf{x}) < \gamma^{1/p}$ for $\mathbf{x} \in \overline{\Omega}$, then $0 \le u_h^k < \gamma^{1/p}$ for $\mathbf{x} \in \overline{\Omega}$.*
2. *If u_h^{k-1} in* (4.1) *satisfies $\gamma^{1/p} < u_h^{k-1}(\mathbf{x}) \le 1$ for $\mathbf{x} \in \overline{\Omega}$, then $\gamma^{1/p} < u_h^k \le 1$ for $\mathbf{x} \in \overline{\Omega}$.*

Proof.

1. The nonnegativity follows from the fact that A is an M-matrix (A^{-1} is nonnegative) and that the right hand side of (4.4) is nonnegative.

In order to establish the upper bound, we let $\mathbf{w} = \mathbf{e}\tau - \mathbf{u}^k$ where $\mathbf{e}$ is a vector with all components equal to 1, and τ a nonnegative number to be determined below. Replacing $\mathbf{c}$ in (4.4) with $\mathbf{u}^k$, we observe that $A\mathbf{w} = A\mathbf{e}\tau - A\mathbf{u}^k$ implies the following chain of equalities

$$\begin{aligned}
(A\mathbf{w})_i &= \tau(A\mathbf{e})_i - (\mathbf{r})_i \\
&= \tau\sum_{j=1}^{n} A_{ij} - (u_h^{k-1}, \phi_i) \\
&\quad \text{(see the proof of Theorem 2.3 in [24])} \\
&= \tau(1, v_i) - \tau\Delta t\,(g(u_h^{k-1}), \phi_i) - (u_h^{k-1}, \phi_i) \\
&= (\tau - \tau\Delta t\, g(u_h^{k-1}) - u_h^{k-1}, \phi_i). \qquad (4.8)
\end{aligned}$$

Motivated by (4.8), we investigate the behavior of $s(x) = \tau(1 - \beta\Delta t(1 - x^p)(x^p - \gamma)) - x$. Of interest is $s(x) > 0$. Observe that for $\tau = \gamma^{1/p}$ and $x < \gamma^{1/p}$, the coefficient of τ, $(1 - \beta\Delta t(1 - x^p)(x^p - \gamma)) > 1$. Thus, $s(x) > 0$ and therefore $A\mathbf{e}\tau - A\mathbf{u}^k > 0$, implying that $\mathbf{u}^k < \gamma^{1/p}\mathbf{e}$, i.e. $u_h^k(\mathbf{x}) < \gamma^{1/p}$ for $\mathbf{x} \in \overline{\Omega}$.

2. We start with the lower bound. Let $\mathbf{w} = \mathbf{u}^k - \tau\mathbf{e}$. Proceeding as before, we note that

$$\begin{aligned}
(A\mathbf{w})_i &= (\mathbf{r})_i - \tau(A\mathbf{e})_i \\
&= -(\tau - \tau\Delta t\, g(u_h^{k-1}) - u_h^{k-1}, \phi_i). \qquad (4.9)
\end{aligned}$$

   In this case, $s(x) = -\tau(1 - \beta\Delta t(1 - x^p)(x^p - \gamma)) + x$. For $\tau = \gamma^{1/p}$ and $x > \gamma^{1/p}$, the coefficient of τ in $s(x)$ is negative and consequently $s(x) > 0$. The bound now follows.
   For the upper bound we consider $\tau = 1$ in (4.8) and note that $s(1) = 0$. Therefore, it is sufficient to show that $s'(x) \leq 0$ for $0 \leq x \leq 1$. Computing $s'(x) \leq 0$ we obtain

$$-\beta\Delta t(px^{p-1} - 2px^{2p-1} + p\gamma x^{p-1}) - 1 \leq 0,$$
$$\text{which is equivalent to } -\beta\Delta t p x^{p-1}((1+\gamma) - 2x^p) \leq 1.$$

   Noting that the left hand side is maximized when $x = 1$, it is sufficient to require

$$\Delta t \leq \frac{1}{(1-\gamma)\beta p} \text{ (see (4.7)).}$$

   Finally, since A is an M-matrix and $A\mathbf{e}\tau - A\mathbf{u}^k$ is nonnegative, we obtain that $\mathbf{w}$ is also nonnegative and consequently that $\mathbf{u}^k \leq 1$, i.e. $u_h^k(\mathbf{x}) \leq 1$ for $\mathbf{x} \in \overline{\Omega}$.

□

The previous discussion yields the following Theorem.

THEOREM 4.3. *Under the assumptions of Lemma 4.2, if* $0 \leq u_h^{k-1}(\mathbf{x}) \leq 1$ *for* $\mathbf{x} \in \overline{\Omega}$ *then* $0 \leq u_h^k \leq 1$.

Proof. The inequality follows from the positivity of $\mathbf{u}^k$ and the second part of Lemma 4.2. □

Using the results derived for the continuous and discrete problems, we are ready to derive an *a priori* error estimate.

5. Error analysis. Throughout this section we assume that both the initial and boundary conditions are bounded in the interval $[0,1]$, so that in view of Theorem 3.5 and Theorem 4.3, the estimates $\|u^k\|_\infty \le 1$ and $\|u_h^k\|_\infty \le 1$ hold for the continuous solution and the discrete approximation, respectively. Moreover, from the previous discussion, we have that the functions $g(\cdot)$ as defined in (1.2) and $f(x) = x^p$, which are evaluated at either u_h^k or u^k, are bounded in $L^\infty(\Omega)$ with bounds β and 1, respectively. Using these facts, we note that f and g are Lipschitz continuous. We denote their corresponding Lipschitz constant by p_{L} and g_{L}, respectively. The following Lemmas will be helpful in the analysis.

LEMMA 5.1 ([24]). *For u, u_t, u_{tt} in $L^2((0,T], L^2(\Omega))$,*

$$\|u_t^k - \frac{u^k - u^{k-1}}{\Delta t}\|^2 \le \frac{\Delta t}{3} \int_{t_{k-1}}^{t_k} \|u_{tt}\|^2. \tag{5.1}$$

LEMMA 5.2. *For a function $f(\mathbf{x}, t)$ such that f and f_t are $L^2((0,T], L^2(\Omega))$, we have*

$$\|f^k - f^{k-1}\|^2 \le (\Delta t)^2 \int_{t_{k-1}}^{t_k} \|f_t\|^2 \, dt.$$

We will also use the discrete version of Gronwall's lemma [17].

LEMMA 5.3. *Let Δt, H, and a_n, b_n, c_n, γ_n (for integers $n \ge 0$), be nonnegative numbers such that*

$$a_\ell + \Delta t \sum_{n=0}^{\ell} b_n \le \Delta t \sum_{n=0}^{\ell} \gamma_n a_n + \Delta t \sum_{n=0}^{\ell} c_n + H \quad \text{for } \ell \ge 0.$$

Suppose that $\Delta t \gamma_n < 1$, for all n, and set $\sigma_n = (1 - \Delta t \gamma_n)^{-1}$. Then

$$a_\ell + \Delta t \sum_{n=0}^{\ell} b_n \le \exp\left(\Delta t \sum_{n=0}^{\ell} \sigma_n \gamma_n\right) \left(\Delta t \sum_{n=0}^{\ell} c_n + H\right) \quad \text{for } \ell \ge 0.$$

Additional norms that we will use in the analysis are the discrete norms

$$|||v|||_s = \left(\sum_{k=0}^{M} \|v(\mathbf{x}, t_k)\|^2_{H^s(\Omega)} \Delta t\right)^{1/2}, \quad |||v|||_\infty = \sup_{0 \le k \le M} \|v^k\|_\infty. \tag{5.2}$$

We are now in position to state the main result of this section.

THEOREM 5.4. *Let $u(\mathbf{x}, t) \in L^\infty((0,T], H^2(\Omega))$, $u_t(\mathbf{x}, t) \in L^2((0,T], H^2(\Omega))$, $u_{tt}(\mathbf{x}, t) \in L^2((0,T], L^2(\Omega))$ be the solution of (3.3) on the interval $(0,T]$. Assume further that the initial condition u_0 and the boundary conditions are bounded in $[0,1]$. Then, the finite element approximation u_h^k converges to u as $\Delta t, h \to 0$, provided that*

Δt and h satisfy Theorem 4.1. In addition, there exists a constant $C > 0$, such that the approximation u_h^k satisfies the following error estimate:

$$\begin{aligned}
\|u^k - u_h^k\|^2 + \sum_{n=1}^{k} \|\nabla(u^k - u_h^k)\|^2 \Delta t &\leq 2C^2 h^2 |||u|||_2^2 + 2C^2 h^4 \|u^k\|_{H^2(\Omega)}^2 \\
&+ 2K\Big((\Delta t)^2\, \mathbf{F}_3 \int_0^T \|u_t\|^2\, dt + C(h^4(\mathbf{F}_4 + \beta) + 4h^2)|||u|||_2^2 \\
&+ \frac{(\Delta t)^2}{3} \int_0^T \|u_{tt}\|^2\, dt + C\frac{h^4}{2} \int_0^T \|u_t\|_2^2\, dt\Big),
\end{aligned} \tag{5.3}$$

where $\mathbf{F}_i = |||\mathbf{F}_i|||_\infty$,

$$\begin{aligned}
\mathbf{F}_2^k &= 3 + 6\, p_{\mathrm{L}} \|\nabla u^k\|_\infty + 3\beta + (\sqrt{2} + \frac{5\sqrt{2}}{2}) + 3g_{\mathrm{L}}, \\
\mathbf{F}_3^k &= 2\, p_{\mathrm{L}} \|\nabla u^k\|_\infty + g_{\mathrm{L}}, \\
\mathbf{F}_4^k &= g_{\mathrm{L}} + 2\, p_{\mathrm{L}} \|\nabla u^k\|_\infty, \text{ and}
\end{aligned}$$

$$K = \exp\left(\frac{T(\mathbf{F}_2 + \mathbf{F}_4)}{1 - \Delta t(\mathbf{F}_2 + \mathbf{F}_4)}\right).$$

Proof. Recalling the weak formulation

$$(u_t^k, v) + (\nabla u^k, \nabla v) + b((u^k)^p, u^k, v) - (u^k g(u^k), v) = (f^k, v) \tag{5.4}$$

and its discretization

$$(\frac{u_h^k - u_h^{k-1}}{\Delta t}, v) + (\nabla u_h^k, \nabla v) + b((u_h^{k-1})^p, u_h^k, v) - (u_h^k g(u_h^{k-1}), v) = (f^k, v), \tag{5.5}$$

we subtract (5.5) from (5.4) to obtain

$$\begin{aligned}
(\frac{\partial u^k}{\partial t} - \frac{u_h^k - u_h^{k-1}}{\Delta t}, v) + (\nabla(u^k - u_h^k), \nabla v) + b((u^k)^p, u^k, v) - b((u_h^{k-1})^p, u_h^k, v) \\
-(u^k g(u^k), v) + (u_h^k g(u_h^{k-1}), v) = 0.
\end{aligned} \tag{5.6}$$

Rewriting (5.6)

$$\begin{aligned}
(\frac{\partial u^k}{\partial t} - \frac{u_h^k - u_h^{k-1}}{\Delta t}, v) &+ (\nabla(u^k - u_h^k), \nabla v) + b((u^k)^p - (u^{k-1})^p, u^k, v) \\
&+ b((u^{k-1})^p - (u_h^{k-1})^p, u^k, v) + b((u_h^{k-1})^p, u^k - u_h^k, v) \\
&- (u^k(g(u^k) - g(u^{k-1})), v) - ((u^k - u_h^k) g(u^{k-1}), v) \\
&- (u_h^k(g(u^{k-1}) - g(u_h^{k-1})), v) = 0.
\end{aligned} \tag{5.7}$$

For $U^k \in X_h$, let $e^k = u^k - u_h^{k-1} = \Lambda^k + E^k$ where $\Lambda^k = u^k - U^k$ and $E^k = U^k - u_h^k$. Substituting into (5.7) we obtain

$$\begin{aligned}
(\frac{\partial u^k}{\partial t} - \frac{u^k - u^{k-1}}{\Delta t}, v) &+ (\frac{\Lambda^k - \Lambda^{k-1} + E^k - E^{k-1}}{\Delta t}, v) + (\nabla(\Lambda^k + E^k), \nabla v) \\
&+ b((u^k)^p - (u^{k-1})^p, u^k, v) + b((u^{k-1})^p - (u_h^{k-1})^p, u^k, v) \\
&+ b((u_h^{k-1})^p, \Lambda^k + E^k, v) - (u^k(g(u^k) - g(u^{k-1})), v) \\
&- ((\Lambda^k + E^k) g(u^{k-1}), v) - (u_h^k(g(u^{k-1}) - g(u_h^{k-1})), v) = 0.
\end{aligned} \tag{5.8}$$

Choosing $v = E^k$, multiplying (5.8) by Δt and rearranging, yields

$$\begin{aligned}
(E^k - E^{k-1}, E^k) + \Delta t(\nabla E^k, \nabla E^k) &= -\Delta t\,(\frac{\partial u^k}{\partial t} - \frac{u^k - u^{k-1}}{\Delta t}, E^k) \\
&- \Delta t\, b((u^k)^p - (u^{k-1})^p, u^k, E^k) - \Delta t\, b((u^{k-1})^p - (u_h^{k-1})^p, u^k, E^k) \\
&- \Delta t\, b((u_h^{k-1})^p, \Lambda^k + E^k, E^k) + \Delta t\,(u^k(g(u^k) - g(u^{k-1})), E^k) \\
&+ \Delta t\,((\Lambda^k + E^k) g(u^{k-1}), E^k) + \Delta t\,(u_h^k(g(u^{k-1}) - g(u_h^{k-1})), E^k) \\
&- (\Lambda^k - \Lambda^{k-1}, E^k) - \Delta t\,(\nabla \Lambda^k, \nabla E^k).
\end{aligned} \tag{5.9}$$

Using Cauchy-Schwarz and Young's inequality results in the following bounds for the terms appearing in (5.9).

$$\begin{aligned}
(E^k - E^{k-1}, E^k) = \|E^k\|^2 - \int_\Omega E^{k-1}E^k &\geq \|E^k\|^2 - \frac{1}{2}\|E^k\|^2 - \frac{1}{2}\|E^{k-1}\|^2 \\
&= \frac{1}{2}\|E^k\|^2 - \frac{1}{2}\|E^{k-1}\|^2.
\end{aligned}$$

A direct application of Lemma 5.1 yields

$$\begin{aligned}
(\frac{\partial u^k}{\partial t} - \frac{u^k - u^{k-1}}{\Delta t}, E^k) &\leq \frac{1}{2}\|u_t^k - \frac{u^k - u^{k-1}}{\Delta t}\|^2 + \frac{1}{2}\|E^k\|^2 \\
&\leq \frac{\Delta t}{6}\int_{t_{k-1}}^{t_k} \|u_{tt}\|^2 + \frac{1}{2}\|E^k\|^2.
\end{aligned}$$

Using the Lipschitz continuity of g and $f(x) = x^p$ and Lemma 5.2, we obtain

$$\begin{aligned}
|b((u^k)^p - (u^{k-1})^p, u^k, E^k)| &\leq \int_\Omega |(u^k)^p - (u^{k-1})^p|\,|\begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot \nabla u^k|\,|E^k|\, d\Omega \\
&\leq 2\,p_{\mathrm{L}}\|\nabla u^k\|_\infty \int_\Omega |u^k - u^{k-1}|\,|E^k|\, d\Omega \\
&\leq p_{\mathrm{L}}\|\nabla u^k\|_\infty(\|u^k - u^{k-1}\|^2 + \|E^k\|^2) \\
&\leq p_{\mathrm{L}}\|\nabla u^k\|_\infty(\Delta t \int_{t_{k-1}}^{t_k} \|u_t\|^2\, dt + \|E^k\|^2).
\end{aligned}$$

$$\begin{aligned}
|b((u^{k-1})^p - (u_h^{k-1})^p, u^k, E^k)| &\leq \int_\Omega |(u^{k-1})^p - (u_h^{k-1})^p|\,|\begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot \nabla u^k|\,|E^k|\, d\Omega \\
&\leq 2\,p_{\mathrm{L}}\|\nabla u^k\|_\infty \int_\Omega |u^{k-1} - u_h^{k-1}|\,|E^k|\, d\Omega \\
&= 2\,p_{\mathrm{L}}\|\nabla u^k\|_\infty \int_\Omega |E^{k-1} + \Lambda^{k-1}|\,|E^k|\, d\Omega \\
&\leq p_{\mathrm{L}}\|\nabla u^k\|_\infty(\|E^{k-1}\|^2 + \|\Lambda^{k-1}\|^2 + 2\|E^k\|^2).
\end{aligned}$$

For the next term, using Young's inequality yields

$$\begin{aligned}
|b((u_h^{k-1})^p, \Lambda^k + E^k, E^k)| &\leq \int_\Omega |(u_h^{k-1})^p|\,|\begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot \nabla(\Lambda^k + E^k)|\,|E^k|\, d\Omega \\
&\leq \|(u_h^{k-1})^p\|_\infty \int_\Omega |\begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot \nabla(\Lambda^k + E^k)|\,|E^k|\, d\Omega \\
&\leq \sqrt{2}(\frac{1}{2}\|\nabla \Lambda^k\|^2 + (\frac{1}{2} + \frac{5}{4})\|E^k\|^2 + \frac{1}{5}\|\nabla E^k\|^2).
\end{aligned}$$

Using Lemma 5.2,

$$\begin{aligned}
|(u^k(g(u^k)-g(u^{k-1})),E^k)| &\le \int_\Omega |u^k|\,|g(u^k)-g(u^{k-1})|\,|E^k|\,d\Omega \\
&\le \|u^k\|_\infty g_{\mathrm{L}} \int_\Omega |u^k-u^{k-1}|\,|E^k|\,d\Omega \\
&\le \frac{g_{\mathrm{L}}}{2}(\Delta t\int_{t_{k-1}}^{t_k} \|u_t\|^2\,dt + \|E^k\|^2).
\end{aligned}$$

$$\begin{aligned}
|((\Lambda^k+E^k)g(u^{k-1}),E^k)| &\le \int_\Omega |g(u^{k-1})|\,|\Lambda^k+E^k|\,|E^k|\,d\Omega \\
&\le \beta(\frac{3}{2}\|E^k\|^2+\frac{1}{2}\|\Lambda^k\|^2).
\end{aligned}$$

$$\begin{aligned}
|(u_h^k(g(u^{k-1})-g(u_h^{k-1})),E^k)| &\le \int_\Omega |u_h^k|\,|g(u^{k-1})-g(u_h^{k-1})|\,|E^k|\,d\Omega \\
&\le \|u_h^k\|_\infty g_{\mathrm{L}} \int_\Omega |u^{k-1}-u_h^{k-1}|\,|E^k|\,d\Omega \\
&\le g_{\mathrm{L}} \int_\Omega |E^{k-1}+\Lambda^{k-1}|\,|E^k|\,d\Omega \\
&\le g_{\mathrm{L}}(\frac{1}{2}\|E^{k-1}\|^2+\|E^k\|^2+\frac{1}{2}\|\Lambda^{k-1}\|^2).
\end{aligned}$$

Again, applying Lemma 5.2 and Young's inequality we bound the following two terms by

$$\begin{aligned}
|(\Lambda^k-\Lambda^{k-1},E^k)| &\le \Delta t\|E^k\|^2+\frac{1}{4\Delta t}\|\Lambda^k-\Lambda^{k-1}\|^2 \\
&\le \Delta t\|E^k\|^2+\frac{1}{4}\int_{t_{k-1}}^{t_k}\|\Lambda_t\|^2\,dt.
\end{aligned}$$

$$|(\nabla\Lambda^k,\nabla E^k)| \le \frac{1}{5}\|\nabla E^k\|^2+\frac{5}{4}\|\nabla\Lambda^k\|^2.$$

Substituting the above estimates into (5.9) and rearranging, we obtain

$$\begin{aligned}
&\|E^k\|^2-\|E^{k-1}\|^2+\Delta t\,\mathbf{F}_1^k\|\nabla E^k\|^2 \le \\
&\quad \Delta t\,\mathbf{F}_2^k\|E^k\|^2+(\Delta t)^2\,\mathbf{F}_3^k\int_{t_{k-1}}^{t_k}\|u_t\|^2\,dt+\Delta t\,\mathbf{F}_4^k\|E^{k-1}\|^2 \\
&\quad +\Delta t\,\mathbf{F}_4^k\|\Lambda^{k-1}\|^2+\Delta t\,\mathbf{F}_5^k\|\nabla\Lambda^k\|^2+\Delta t\,\beta\|\Lambda^k\|^2 \\
&\quad +\frac{(\Delta t)^2}{3}\int_{t_{k-1}}^{t_k}\|u_{tt}\|^2\,dt+\frac{1}{2}\int_{t_{k-1}}^{t_k}\|\Lambda_t\|^2\,dt,
\end{aligned} \tag{5.10}$$

where

$$
\begin{aligned}
\mathbf{F}_1^k &= 2 - 2(1+\sqrt{2})/5 \geq 1, \\
\mathbf{F}_2^k &= 3 + 6\, p_{\mathrm{L}} \|\nabla u^k\|_\infty + 3\beta + (\sqrt{2} + \frac{5\sqrt{2}}{2}) + 3 g_{\mathrm{L}}, \\
\mathbf{F}_3^k &= 2\, p_{\mathrm{L}} \|\nabla u^k\|_\infty + g_{\mathrm{L}}, \\
\mathbf{F}_4^k &= g_{\mathrm{L}} + 2\, p_{\mathrm{L}} \|\nabla u^k\|_\infty, \\
\mathbf{F}_5^k &= \frac{5}{2} + \sqrt{2} \leq 4.
\end{aligned} \tag{5.11}
$$

Summing (5.10) from $k = 1$ to $k = \ell$ and defining $\mathbf{F}_i = \|\!|\mathbf{F}_i|\!\|_\infty$, yields

$$
\begin{aligned}
&\|E^\ell\|^2 + \sum_{k=1}^{\ell} \|\nabla E^k\|^2 \Delta t \leq \\
&\quad \Delta t\, (\mathbf{F}_2 + \mathbf{F}_4) \sum_{k=0}^{\ell} \|E^k\|^2 + (\Delta t)^2\, \mathbf{F}_3 \int_0^T \|u_t\|^2\, dt \\
&\quad + \Delta t \sum_{k=0}^{\ell} ((\mathbf{F}_4 + \beta)\|\Lambda^k\|^2 + 4\|\nabla \Lambda^k\|^2) \\
&\quad + \frac{(\Delta t)^2}{3} \int_0^T \|u_{tt}\|^2\, dt + \frac{1}{2} \int_0^T \|\Lambda_t\|^2\, dt.
\end{aligned} \tag{5.12}
$$

Choosing Δt so that $\Delta t(\mathbf{F}_2 + \mathbf{F}_4) < 1$ and applying Lemma 5.3 results in the following inequality

$$
\begin{aligned}
\|E^\ell\|^2 + \sum_{k=1}^{\ell} \|\nabla E^k\|^2 \Delta t \leq K\Big(&(\Delta t)^2\, \mathbf{F}_3 \int_0^T \|u_t\|^2\, dt \\
&+ \Delta t \sum_{k=0}^{\ell} ((\mathbf{F}_4 + \beta)\|\Lambda^k\|^2 + 4\|\nabla \Lambda^k\|^2) \\
&+ \frac{(\Delta t)^2}{3} \int_0^T \|u_{tt}\|^2\, dt + \frac{1}{2} \int_0^T \|\Lambda_t\|^2\, dt\Big),
\end{aligned} \tag{5.13}
$$

where

$$
K = \exp\left(\frac{T(\mathbf{F}_2 + \mathbf{F}_4)}{1 - \Delta t(\mathbf{F}_2 + \mathbf{F}_4)} \right).
$$

From the theory of finite element interpolation [3, 5], we have that for $\mathcal{I}_h$, the interpolant of the exact solution u in the space of piecewise linear continuous polynomials, and $\Lambda^k = u^k - \mathcal{I}_h u^k$, there exist a constant $C \geq 0$ such that

$$
\|\Lambda^k\| + h\|\nabla \Lambda^k\| \leq C h^2 \|u^k\|_{H^2(\Omega)}. \tag{5.14}
$$

Owing to (5.14) and (5.2), (5.13) becomes

$$
\begin{aligned}
\|E^\ell\|^2 + \sum_{k=1}^{\ell} \|\nabla E^k\|^2 \Delta t \leq K\Big(&(\Delta t)^2\, \mathbf{F}_3 \int_0^T \|u_t\|^2\, dt \\
&+ C(h^4(\mathbf{F}_4 + \beta) + 4h^2)|\!|\!|u|\!|\!|_2^2 \\
&+ \frac{(\Delta t)^2}{3} \int_0^T \|u_{tt}\|^2\, dt + C\frac{h^4}{2} \int_0^T |\!|\!|u_t|\!|\!|_2^2\, dt\Big).
\end{aligned} \tag{5.15}
$$

Finally, noting that

$$\|u^\ell - u_h^\ell\| \leq \|E^\ell\| + \|\Lambda^\ell\|,$$

we get the bound given in (5.3). □

6. Numerical results. In this section three numerical experiments are performed to investigate the theoretical results presented in Lemma 4.2 and Theorem 5.4. Numerical experiments 1 and 2 investigate the nonnegativity and boundedness of the approximation. Numerical experiment 3 investigates the predicted *a priori* error estimate. A traveling wave solution of (1.1) for the case when $\beta = 1$ is given [8]

$$u(\mathbf{x}, t) = \Big(\frac{\gamma}{2} + \frac{\gamma}{2}\tanh\left(\sigma\gamma\left(z - ct\right)\right)\Big)^{1/p}, \tag{6.1}$$

where

$$z = \frac{1}{\sqrt{2}}(x+y),\ c = \frac{(\overline{\alpha} - \rho)\gamma + (\overline{\alpha} + \rho)(p+1)}{2(p+1)},\ \sigma = \frac{p(\rho - \overline{\alpha})}{4(p+1)},$$
$$\rho = \sqrt{\overline{\alpha}^2 + 4(1+p)},\ \overline{\alpha} = \alpha\sqrt{2}.$$

We note that (6.1) lies in the interval $(0, \gamma^{1/p})$. A second traveling-wave solution of (1.1) for $\beta = 1$ and $p = 1$ is the following [22]

$$u(\mathbf{x}, t) = \frac{1}{2} - \frac{1}{2}\tanh\left(\frac{z - \nu t}{r - \overline{\alpha}}\right), \tag{6.2}$$

where $r = \sqrt{\overline{\alpha}^2 + 8}$, $\nu = (2\overline{\alpha} + (\overline{\alpha} - r)(2\gamma - 1))/4$. In this case the particular solution (6.2) lies in the interval $(0, 1)$.

Experiment 1. We let $\Omega = (-50, 50) \times (-50, 50)$ and approximate (1.1) using (3.3) for $t \in [0, 2]$. The physical parameters under consideration are $\alpha = 0.7$, $\gamma = 0.4$ and $p = 2$. The computational parameters are $\Delta t = 0.1$ and $h = 3/5$. In order to validate our results concerning the boundedness of the approximations, we use solution (6.1) for the initial and boundary conditions and compute the minimum and maximum of u_h^k for every discrete time t_k.
Note that the initial data and boundary conditions are bounded between 0 and $\gamma^{1/p}$. Figure 6.1 illustrates the results and exhibits the nonnegativity of u_h^k and $\gamma^{1/p} - u_h^k$, implying that $0 \leq u_h^k < \gamma^{1/p}$, consistent with Lemma 4.2.

Experiment 2. For this experiment we consider $\Omega = (-20, 30) \times (-20, 30)$ and $t \in [0, 2]$. The physical parameters are $\alpha = 0.4$, $\gamma = 0.3$ and $p = 1$. The computational parameters are $\Delta t = 0.1$ and $h = 7/9$. This time we imposed initial and boundary conditions with numeric values lying in $(0, 1)$ using solution (6.2). As before, we compute the minimum and maximum of the approximations u_h^k for every discrete time t_k. The results are shown in Figure 6.2. We note that the approximation $u_h^k \in (0, 1)$.

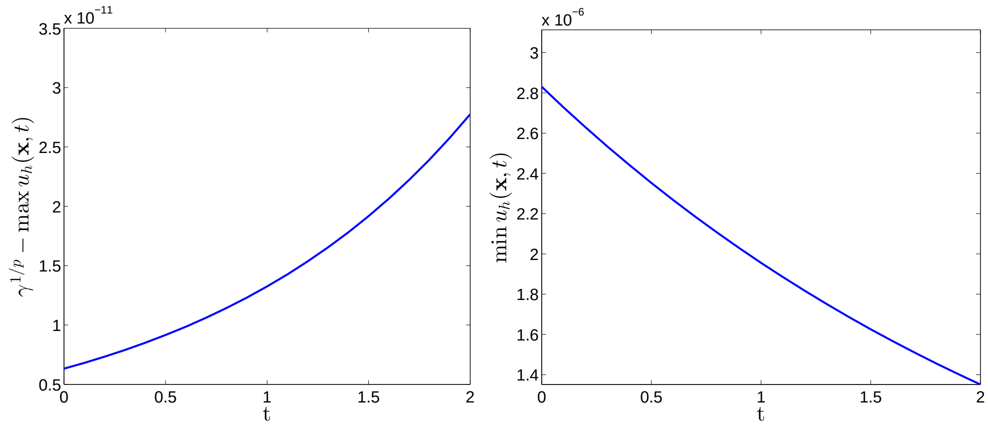

FIG. 6.1. *(Left) Evolution of the maximum of the approximation $u_h(\mathbf{x},t)$ through time with respect to the theoretical maximum $\gamma^{1/p}$. (Right) Evolution of the minimum of $u_h(\mathbf{x},t)$ through time.*

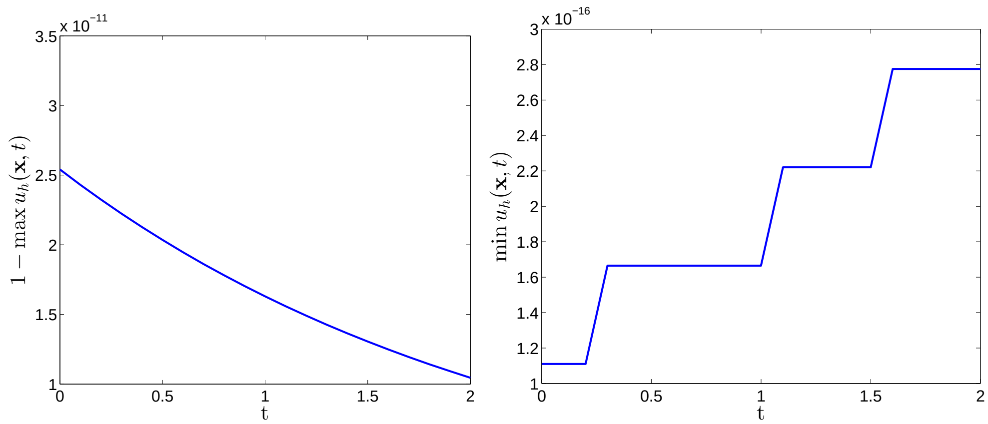

FIG. 6.2. *(Left) Evolution of the quantity $1 - \max u_h(\mathbf{x},t)$ through time. (Right) Evolution of the minimum of $u_h(\mathbf{x},t)$ through time.*

Experiment 3. As a means of verifying our *a priori* error estimate, we approximated the solution of (1.1) on $\Omega = (-40, 80) \times (-40, 80)$ for $t \in [0, 2]$ using the physical parameters $\alpha = 1$, $p = 1$ and $\gamma = 0.5$. We considered the quantity $u - u_h$ under the norms $\|\cdot\|$ and $|||\cdot|||$ for different values of Δt and h.
As we chose $\Delta t \propto h$, Theorem 5.4 predicts that $\|u(T) - u_h^M\|$ and $|||u - u_h||| \sim Ch^r$, with $r = 1$. We compute the experimental convergence rate $\widetilde{r}_{\|\cdot\|}$ for $\|\cdot\|$ using $\widetilde{r}_{\|\cdot\|} = \log(\|u - u_{h_1}\|/\|u - u_{h_2}\|)/\log(h_1/h_2)$. The quantity $\widetilde{r}_{|||\cdot|||}$ is similarly computed. The results are shown in Table 6.1 and illustrate the first order convergence in space and time of the scheme.

7. Conclusion. In the first part of this paper, we established the existence, nonnegativity and boundedness of the solution to (1.1). The second part of the paper was dedicated to the study of the proposed finite element scheme, capable of preserving the nonnegativity and boundedness of the solution under relatively mild conditions on the computational parameters Δt, h and the mesh. An *a priori* error estimate

Δt	h	$\|u - u_h\|$	$\widetilde{r}_{\|\cdot\|}$	$\|\|\|u - u_h\|\|\|$	$\widetilde{r}_{\|\|\|\cdot\|\|\|}$
0.800	15	1.38	-	1.72	-
0.400	15/2	0.41	1.74	0.56	1.62
0.200	15/4	0.15	1.49	0.17	1.71
0.100	15/8	0.07	1.07	0.07	1.35
0.050	15/16	0.04	0.95	0.03	1.08
0.025	15/32	0.02	0.95	0.02	1.00

TABLE 6.1

Convergence rates of the numerical approximation computed using scheme (3.3).

was derived, which shows that the numerical approximation converges to the solution as $\Delta t, h \to 0$. This same analysis also demonstrates that our numerical scheme is of first order in time and space. The numerical experiments, designed to investigate the properties of the approximation scheme, were in agreement with the theory.

REFERENCES

[1] M. Ablowitz and B. Fuchssteiner. *Topics in soliton theory and exactly solvable nonlinear equations*. World Scientific Publishing Co., Singapore, 1987.

[2] B. Batiha, M.S.M. Noorani, and I. Hashim. Application of variational iteration method to the generalized Burgers-Huxley equation. *Chaos, Solitons & Fractals*, 36(3):660 – 663, 2008.

[3] D. Braess. *Finite Elements*. Cambridge University Press, third edition, 2007. Cambridge Books.

[4] İ. Çelik. Haar wavelet method for solving generalized Burgers-Huxley equation. *Arab Journal of Mathematical Sciences*, 18(1):25 – 37, 2012.

[5] P.G. Ciarlet. *Finite Element Method for Elliptic Problems*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2002.

[6] R. Codina. Comparison of some finite element methods for solving the diffusion-convection-reaction equation. *Computer Methods in Applied Mechanics and Engineering*, 156(14):185 – 210, 1998.

[7] M.T. Darvishi, S. Kheybari, and F. Khani. Spectral collocation method and Darvishi's preconditionings to solve the generalized Burgers-Huxley equation. *Communications in Nonlinear Science and Numerical Simulation*, 13(10):2091 – 2103, 2008.

[8] X. Deng. Travelling wave solutions for the generalized Burgers-Huxley equation. *Applied Mathematics & Computation*, 204(2):733 – 737, 2008. Special Issue on New Approaches in Dynamic Optimization to Assessment of Economic and Environmental Systems.

[9] J. Douglas, Jr. and T.F. Russell. Numerical methods for convection-dominated diffusion problems based on combining the method of characteristics with finite element or finite difference procedures. *SIAM Journal on Numerical Analysis*, 19(5):871–885, 1982.

[10] I. Faragó and R. Horváth. Discrete maximum principle and adequate discretizations of linear parabolic problems. *SIAM J. Sci. Comput.*, 28(6):2313–2336, 2006.

[11] L.P. Franca, A. Nesliturk, and M. Stynes. On the stability of residual-free bubbles for convection-diffusion problems and their approximation by a two-level finite element method. *Computer Methods in Applied Mechanics and Engineering*, 166(12):35 – 49, 1998. Advances in Stabilized Methods in Computational Mechanics.

[12] A. Friedman. *Partial Differential Equations of Parabolic Type*. Prentice-Hall, Inc., Englewood Cliffs, NJ, USA, 1964.

[13] T. Fujimoto and R.R. Ranade. Two characterizations of inverse-positive matrices: the Hawkins-Simon condition and the Le Chatelier-Braun principle. *Electron. J. Linear Algebra*, 11:59–65, 2004.

[14] M.J. Gander, L. Halpern, and M. Kern. A Schwarz waveform relaxation method for advection-diffusion-reaction problems with discontinuous coefficients and non-matching grids. In Olof B. Widlund and David E. Keyes, editors, *Domain Decomposition Methods in Science and Engineering XVI*, volume 55 of *Lecture Notes in Computational Science and Engineering*, pages 283–290. Springer Berlin Heidelberg, 2007.

[15] I. Hashim, M.S.M. Noorani, and M.R. Said Al-Hadidi. Solving the generalized Burgers-Huxley

equation using the adomian decomposition method. *Mathematical and Computer Modelling*, 43(1112):1404 – 1411, 2006.

[16] F. Heße, F.A. Radu, M. Thullner, and S. Attinger. Upscaling of the advection-diffusion-reaction equation with Monod reaction. *Advances in Water Resources*, 32(8):1336 – 1351, 2009.

[17] J.G. Heywood and R. Rannacher. Finite-element approximation of the nonstationary Navier-Stokes problem. IV. Error analysis for second-order time discretization. *SIAM J. Numer. Anal.*, 27(2):353–384, 1990.

[18] S. Idelsohn, N. Nigro, M. Storti, and G. Buscaglia. A Petrov-Galerkin formulation for advection-reaction-diffusion problems. *Computer Methods in Applied Mechanics and Engineering*, 136(12):27 – 46, 1996.

[19] B.V.R. Kumar, V. Sangwan, S. Murthy, and M. Nigam. A numerical study of singularly perturbed generalized Burgers-Huxley equation using three-step Taylor-Galerkin method. *Computers & Mathematics with Applications*, 62(2):776 – 786, 2011.

[20] D. Lanser and J.G. Verwer. Analysis of operator splitting for advection-diffusion-reaction problems from air pollution modelling. *Journal of Computational and Applied Mathematics*, 111(12):201 – 216, 1999.

[21] J.E. Macías-Díaz. Sufficient conditions for the preservation of the boundedness in a numerical method for a physical model with transport memory and nonlinear damping. *Computer Physics Communications*, 182(12):2471 – 2478, 2011.

[22] J.E. Macías-Díaz, J. Ruiz-Ramírez, and J. Villa. The numerical solution of a generalized Burgers-Huxley equation through a conditionally bounded and symmetry-preserving method. *Computers & Mathematics with Applications*, 61(11):3330 – 3342, 2011.

[23] C.R. Nugent, W.M. Quarles, and T.H. Solomon. Experimental studies of pattern formation in a Reaction-Advection-Diffusion system. *Phys. Rev. Lett.*, 93:218301, Nov 2004.

[24] J. Ruiz-Ramírez. Preserving nonnegativity of an affine finite element approximation for a convection-diffusion-reaction problem. Technical report, Dept. Math. Sci., Clemson University, 2014. Available at: `http://www.clemson.edu/ces/math/technical_reports/TR2014_8_jr.linear_paper.pdf`.

[25] J. Ruiz-Ramírez and J.E. Macías-Díaz. A skew symmetry-preserving computational technique for obtaining the positive and the bounded solutions of a time-delayed advection-diffusion-reaction equation. *Journal of Computational and Applied Mathematics*, 250(0):256 – 269, 2013.

[26] M. Sari, G. Gürarslan, and A. Zeytinoğlu. High-order finite difference schemes for numerical solutions of the generalized Burgers-Huxley equation. *Numerical Methods for Partial Differential Equations*, 27(5):1313–1326, 2011.

[27] J.R. Sibert, J. Hampton, D.A. Fournier, and P.J. Bills. An advection-diffusion-reaction model for the estimation of fish movement parameters from tagging data, with application to skipjack tuna (katsuwonus pelamis). *Canadian Journal of Fisheries and Aquatic Sciences*, 56(6):925–938, 1999.

[28] X.Y. Wang, Z.S. Zhu, and Y.K. Lu. Solitary wave solutions of the generalised Burgers-Huxley equation. *J. Phys. A*, 23(3):271–274, 1990.

[29] S. Zhou and X. Cheng. A linearly semi-implicit compact scheme for the Burgers-Huxley equation. *International Journal of Computer Mathematics*, 88(4):795–804, 2011.