Journal of Computational and Applied Mathematics 280 (2015) 347–366

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

An energy-conserving second order numerical scheme for nonlinear hyperbolic equation with an exponential nonlinear term

Lingdi Wang [a], Wenbin Chen [a], Cheng Wang [b,*]

[a] School of Mathematical Sciences, Fudan University, China

[b] Mathematics Department, The University of Massachusetts, North Dartmouth, MA, USA

ARTICLE INFO

Article history:
Received 25 September 2013
Received in revised form 11 August 2014

MSC:
35L70
65M06
65M12

Keywords:
Nonlinear hyperbolic equation
Energy stability
Positivity
Linear iteration
Convergence analysis

ABSTRACT

We present a second order accurate numerical scheme for a nonlinear hyperbolic equation with an exponential nonlinear term. The solution to such an equation is proven to have a conservative nonlinear energy. Due to the special nature of the nonlinear term, the positivity is proven to be preserved under a periodic boundary condition for the solution. For the numerical scheme, a highly nonlinear fractional term is involved, for the theoretical justification of the energy stability. We propose a linear iteration algorithm to solve this nonlinear numerical scheme. A theoretical analysis shows a contraction mapping property of such a linear iteration under a trivial constraint for the time step. We also provide a detailed convergence analysis for the second order scheme, in the $\ell^\infty(0, T; \ell^\infty)$ norm. Such an error estimate in the maximum norm can be obtained by performing a higher order consistency analysis using asymptotic expansions for the numerical solution. As a result, instead of the standard comparison between the exact and numerical solutions, an error estimate between the numerical solution and the constructed approximate solution yields an $O(\Delta t^3 + h^4)$ convergence in $\ell^\infty(0, T; \ell^2)$ norm, which leads to the necessary ℓ^∞ error estimate using the inverse inequality, under a standard constraint $\Delta t \leq Ch$. A numerical accuracy check is given and some numerical simulation results are also presented.

1. Introduction

Consider the following nonlinear hyperbolic equation in the domain $\Omega = [0, L]^2$, with a periodic boundary condition,

$$\begin{cases} \partial_t^2 u - \Delta u = \alpha e^{-u}, & \text{in } \Omega_T, \\ u\,|_{t=0} = C_0, \qquad \partial_t u\,|_{t=0} = 0. \end{cases} \tag{1.1}$$

Here Ω_T is defined as $\Omega \times (0, T], \alpha = \alpha(x, y)$ is a periodic non-negative function with finite upper bound $\bar{\alpha}$ and is differentiable with respect to space up to the rth order, and C_0 is a non-negative constant. An integral equation arises from Johnson–Mehl–Avrami–Kolmogorov theory [1–3] and Cahn's time cone method [4], which characterizes the nucleation and growth phenomenon of nuclei. In the framework of the time cone method, u stands for the number of nuclei and satisfies an integral equation in the time–space region. Recently, Liu and Yamamoto successfully proved that the integral equation

* Corresponding author.
E-mail addresses: wanglingdi888@gmail.com (L. Wang), wbchen@fudan.edu.cn (W. Chen), cwang1@umassd.edu (C. Wang).

http://dx.doi.org/10.1016/j.cam.2014.11.043
0377-0427/

could be reduced to a hyperbolic system [5], which is exactly the same as (1.1) in one dimensional case. In addition, similar nonlinear hyperbolic systems can be derived in high dimensional cases. The righthand term αe^{-u} describes the nucleation rate. The introduction of the nonlinear term e^{-u} is an available assumption based on the physical fact that nucleation rate decreases with the increase of the number of nuclei.

Similar to the linear problem, the nonlinear equation (1.1) also admits an energy law. Indeed, we can define the nonlinear energy as follows,

$$\begin{aligned} E(t) &= \int_\Omega \left(\frac{1}{2}|\partial_t u|^2 + \frac{1}{2}|\nabla u|^2 + \alpha e^{-u}\right) d\mathbf{x} \\ &= \frac{1}{2}\left(\|\partial_t u\|^2_{L^2(\Omega)} + \|\nabla u\|^2_{L^2(\Omega)}\right) + \int_\Omega \alpha e^{-u} d\mathbf{x}. \end{aligned} \tag{1.2}$$

By taking inner product with $\partial_t u$ for the equation of (1.1), it is straightforward to observe that

$$E'(t) \equiv 0, \quad \text{i.e. } E(t) = E(0),\ \forall t > 0, \tag{1.3}$$

which shows a conservation for the energy defined as (1.2).

Although the energy $E(t)$ is always conserved for all time, we should point out that the existence and uniqueness of the solution of (1.1) are not straightforward. Similar energy conservation also holds if the exponential nonlinear term e^{-u} is replaced by a polynomial one $|u|^p$, while the solution may not exist globally in time and blow up once p exceeds a critical value [6]. Up to now, the existence of a global (in time) solution of (1.1) with a general boundary condition and initial value has been an open problem. In this paper we prove there exists a global in time, unique positive solution of (1.1) with periodic boundary conditions and special initial values. The positivity property is important to prove the well-posedness of the problem, and a new metric of the solution space is introduced to establish the global (in time) existence. On the other hand, the positivity property is also required by the physical problem, since the solution corresponds to the expected number of nuclei, which is expected to be positive. The technique used here cannot be directly applied to the equation with Dirichlet boundary condition, since the solution may not always be positive and the well-posedness is not clear in two or three dimensional cases. Meanwhile, the local existence can be established in enough small time interval, and the numerical scheme also works well for the equation with Dirichlet boundary condition.

There are many numerical schemes to solve nonlinear hyperbolic equations, for example, see [7–9]. However, very few have dealt with an exponential nonlinear term. Due to the special nature of the nonlinear term appearing in (1.1), the positivity property is proven to be preserved under a periodic boundary condition. Based on the positivity, we present an energy-conserving second order numerical scheme. For this scheme, a highly nonlinear fractional term is involved, for the theoretical justification of the energy stability. For the implementation part, we propose a linear iteration method to solve this nonlinear numerical scheme. A theoretical analysis will be given to prove a contraction mapping property of such a linear iteration under a trivial constraint for the time step.

We also provide a detailed convergence analysis for the second order scheme, in the $\ell^\infty(0, T; \ell^\infty)$ norm. Such an error estimate in the maximum norm can be obtained by performing a higher order consistency analysis using asymptotic expansions for the numerical solution. As a result, instead of the standard comparison between the exact and numerical solutions, an error estimate between the numerical solution and the constructed approximate solution yields an $O(\Delta t^3 + h^4)$ convergence in $\ell^\infty(0, T; \ell^2)$ norm, which leads to the necessary ℓ^∞ error estimate using the inverse inequality, under a standard constraint $\Delta t \le Ch$, where C stands for some given constant.

The rest of the paper is organized as follows. In Section 2, we discuss the well-posedness and positivity of the solution to problem (1.1). In Section 3, the second order scheme is proposed and the energy-conserving property is proven. In Section 4, to facilitate the numerical implementation, we modify the numerical scheme based on the positivity of the solution and discuss some details in the implementation process. Next we give the corresponding convergence analysis in Section 5. Subsequently, several numerical examples are carried out in Section 6. Finally, some conclusions are made in Section 7.

2. Well-posedness and positivity

In this section, we prove the well-posedness and positivity of the solution to (1.1). The main tool we use here is Banach's fixed point theorem [10]; also see the related discussions in [11].

Let $C(0, T; C_\#(\Omega))$ denote the space of all functions which are continuous with respect to both time and space. Herein and what after, the subscript # means that the functions in the corresponding space are periodic with respect to space, i.e. for any function f, there exists its periodic extension defined in $\mathbb{R}^2$, still denoted as f, which satisfies

$$f(x + mL, y + nL) = f(x, y), \tag{2.1}$$

with any integers m and n. We denote X as the subspace of $C(0, T; C_\#(\Omega))$, which contains all non-negative functions.

Besides, we introduce the following metric. For any functions $f, g \in C(0, T; C_\#(\Omega))$, define

$$\delta(f, g) = \max_{t\in[0,T]} \left(e^{-st} \|f(x, y, t) - g(x, y, t)\|_{C(\Omega)}\right), \tag{2.2}$$

as the metric on $C(0, T; C_\#(\Omega))$, where s is a given weight. Here the norm $\|\cdot\|_{C(\Omega)}$ is defined as

$$\|f(x, y, t)\|_{C(\Omega)} = \max_{(x,y)\in\Omega} |f(x, y, t)| . \tag{2.3}$$

The following lemma defines an operator $A : X \to X$ through a linear problem.

Lemma 1. *Assume that the non-negative function $\alpha \in C_\#(\Omega)$. Consider the following linear hyperbolic equation with a periodic boundary condition,*

$$\begin{cases}\partial_t^2 w - \Delta w = \alpha e^{-u}, & \text{in } \Omega_T, \\ w\,|_{t=0} = C_0, & \partial_t w\,|_{t=0} = 0,\end{cases} \tag{2.4}$$

with $u \in X$. Then there exists a unique solution $w \in X$. Hence, we could define an operator $A : X \to X$ as $A[u] = w$.

Proof. For simplicity, we denote $f = \alpha e^{-u}$. Thanks to the periodic boundary condition, we could transform the initial–boundary problem (2.4) to the corresponding Cauchy problem by periodic extension. According to [12,13], the mild solution to the Cauchy problem in two dimensional case is given by

$$w(x, y, t) = C_0 + \frac{1}{2\pi}\int_0^t d\tau \iint_{\Theta_{t-\tau}} \frac{f(x+\xi, y+\eta, \tau)}{\sqrt{(t-\tau)^2 - \xi^2 - \eta^2}} d\xi d\eta, \tag{2.5}$$

where $\Theta_t = \{(\xi, \eta) \in \mathbb{R}^2 : \xi^2 + \eta^2 < t^2\}$. By the periodicity of α and u, we have

$$f(x + mL, y + nL, t) = f(x, y, t), \quad \forall (x, y) \in \mathbb{R}^2 \text{ and } m, n \in \mathbb{Z}.$$

Thus,

$$w(x + mL, y + nL, t) = w(x, y, t), \quad \forall (x, y) \in \mathbb{R}^2 \text{ and } m, n \in \mathbb{Z},$$

i.e. w is also a periodic function. Since $\alpha \in C_\#(\Omega)$ and $u \in C(0, T; C_\#(\Omega))$, it is obvious that $f \in C(0, T; C_\#(\Omega))$. Meanwhile, by observing that

$$\frac{1}{2\pi}\iint_{\Theta_t} \frac{1}{\sqrt{t^2 - \xi^2 - \eta^2}} d\xi d\eta = t, \tag{2.6}$$

we conclude that w is continuous with respect to space.

For the temporal continuity, let $t^\varepsilon = t + \varepsilon$, we have

$$\begin{aligned}2\pi |w(x, y, t^\varepsilon) - w(x, y, t)| &= \left| \int_0^{t^\varepsilon} d\tau \iint_{\Theta_\tau} \frac{f(x+\xi, y+\eta, t^\varepsilon - \tau)}{\sqrt{\tau^2 - \xi^2 - \eta^2}} d\xi d\eta \right. \\ &\quad \left. - \int_0^t d\tau \iint_{\Theta_\tau} \frac{f(x+\xi, y+\eta, t-\tau)}{\sqrt{\tau^2 - \xi^2 - \eta^2}} d\xi d\eta \right| \\ &\le \left| \int_0^{t^\varepsilon} d\tau \iint_{\Theta_\tau} \frac{f(x+\xi, y+\eta, t^\varepsilon - \tau) - f(x+\xi, y+\eta, t-\tau)}{\sqrt{\tau^2 - \xi^2 - \eta^2}} d\xi d\eta \right| \\ &\quad + \left| \int_0^{t^\varepsilon - t} d\tau \iint_{\Theta_\tau} \frac{f(x+\xi, y+\eta, t-\tau)}{\sqrt{\tau^2 - \xi^2 - \eta^2}} d\xi d\eta \right| .\end{aligned} \tag{2.7}$$

Due to the temporal continuity of f and (2.6), we see that both parts of (2.7) tend to zero as $\varepsilon \to 0$, which yields the continuity of w with respect to time.

Finally, the positivity of w comes directly from (2.5), since both C_0 and f are non-negative. Above all, we get $w|_{\Omega_T} = A[u] \in X$. □

Now we turn to the original nonlinear problem (1.1) and give the following result.

Theorem 1. *For problem (1.1), there exists a unique solution u in X.*

Proof. By Lemma 1, the original problem (1.1) is equivalent to the operator equation $A[u] = u$. Thanks to Banach's fixed point theorem, we only need to show that operator A is a strict contraction in X with metric defined as (2.2). For any u and v in X, due to the positivity, we have

$$\left\|\alpha e^{-u} - \alpha e^{-v}\right\|_{C(\Omega)} \le C \|\alpha\|_{C(\Omega)} \|u - v\|_{C(\Omega)} , \tag{2.8}$$

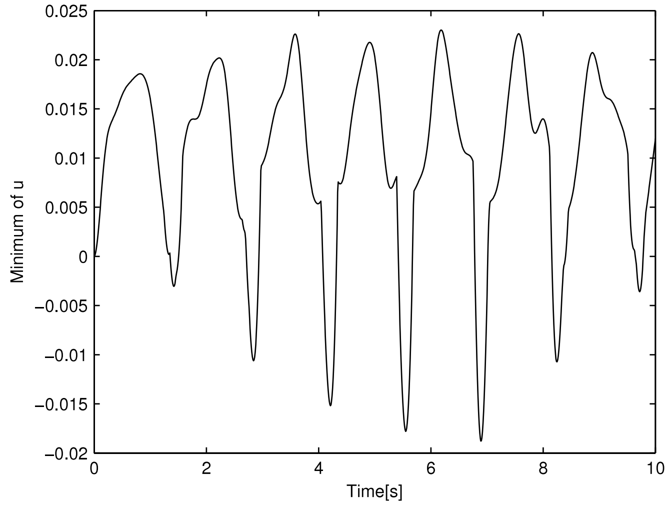

Fig. 2.1. Minimum of the numerical solution to (1.1) with Dirichlet boundary condition.

for some constant $C \geq 0$. On the other hand, by combining with (2.6), we have

$$
\begin{aligned}
|A[u] - A[v]| &= \frac{1}{2\pi} \left| \int_0^t d\tau \iint_{\Theta_{t-\tau}} \frac{(\alpha e^{-u} - \alpha e^{-v})(x+\xi, y+\eta, \tau)}{\sqrt{(t-\tau)^2 - \xi^2 - \eta^2}} d\xi d\eta \right| \\
&\leq \int_0^t (t-\tau) \left\| \alpha e^{-u} - \alpha e^{-v} \right\|_{C(\Omega)} d\tau. \qquad (2.9)
\end{aligned}
$$

Substituting (2.8) into (2.9), we obtain

$$
\begin{aligned}
|A[u] - A[v]| &\leq C \, \|\alpha\|_{C(\Omega)} \int_0^t (t-\tau) \, \|u - v\|_{C(\Omega)} \, d\tau \\
&= C \, \|\alpha\|_{C(\Omega)} \int_0^t e^{-s\tau} \, \|u - v\|_{C(\Omega)} \, (t-\tau) e^{s\tau} d\tau \\
&\leq C \, \|\alpha\|_{C(\Omega)} \, \delta(u, v) \int_0^t (t-\tau) e^{s\tau} d\tau \\
&\leq C \, \|\alpha\|_{C(\Omega)} \, \delta(u, v) \frac{e^{st}}{s^2}.
\end{aligned}
$$

Hence by combining with (2.2) and (2.3), we get

$$
\delta(A[u], A[v]) \leq \frac{C \, \|\alpha\|_{C(\Omega)}}{s^2} \delta(u, v). \qquad (2.10)
$$

Thus, by taking the weight s satisfying $\frac{C\|\alpha\|_{C(\Omega)}}{s^2} < 1$, A becomes a strict contraction. Therefore, the existence, uniqueness and positivity to the solution to (1.1) come directly from Banach's fixed point theorem. □

Remark 1. In the above proof, thanks to the periodic boundary condition, we transform the original initial–boundary problem to the corresponding Cauchy problem by periodic extension and we can use the mild solution formula (2.5) to prove the existence, uniqueness and positivity. However, this method fails if the Dirichlet boundary condition is applied. Specially, the positivity is no longer preserved with the Dirichlet boundary condition. To illustrate this, let us consider the problem (1.1) with a homogeneous Dirichlet boundary condition. Set $\alpha = 1$ and $C_0 = 0$. Fig. 2.1 gives the minimum of the numerical solution. It shows that the numerical solution to problem (1.1) with a homogeneous Dirichlet boundary condition is not always positive. Corresponding analysis would be presented in our future work.

Corollary 1. *The solution to problem* (1.1) *belongs to* $X \bigcap C(0, T; C^r(\Omega))$, *where* $C(0, T; C^r(\Omega))$ *denotes the space containing all the functions that are continuous with respect to time and differentiable with respect to space up to the* r*th order, based on that non-negative function* $\alpha \in C^r_\#(\Omega)$.

Proof. With mild solution formula (2.5), we have

$$u = A[u] = C_0 + \frac{1}{2\pi}\int_0^t d\tau \iint_{\Theta_{t-\tau}} \frac{\alpha e^{-u}(x+\xi, y+\eta, \tau)}{\sqrt{(t-\tau)^2 - \xi^2 - \eta^2}} d\xi d\eta. \tag{2.11}$$

Taking derivative with respect to x on both sides, we get

$$\partial_x u = \frac{1}{2\pi}\int_0^t d\tau \iint_{\Theta_{t-\tau}} \frac{[(\partial_x\alpha)e^{-u} - \alpha e^{-u}\partial_x u](x+\xi, y+\eta, \tau)}{\sqrt{(t-\tau)^2 - \xi^2 - \eta^2}}. \tag{2.12}$$

On the other hand, we consider the following equation:

$$w = \frac{1}{2\pi}\int_0^t d\tau \iint_{\Theta_{t-\tau}} \frac{[(\partial_x\alpha)e^{-u} - \alpha e^{-u}w](x+\xi, y+\eta, \tau)}{\sqrt{(t-\tau)^2 - \xi^2 - \eta^2}} \doteq B[w]. \tag{2.13}$$

With a similar process as in Theorem 1, we could prove that (2.13) admits a solution $w \in C(0, T; C(\Omega))$. Indeed, it is easy to check that operator B maps from $C(0, T; C(\Omega))$ to $C(0, T; C(\Omega))$. Moreover, for any $w, v \in C(0, T; C(\Omega))$, we have

$$\begin{aligned}
|B[w] - B[v]| &= \left| \frac{1}{2\pi}\int_0^t d\tau \iint_{\Theta_{t-\tau}} \frac{\alpha e^{-u}(w-v)(x+\xi, y+\eta, \tau)}{\sqrt{(t-\tau)^2 - \xi^2 - \eta^2}} \right| \\
&\le C\|\alpha\|_{C(\Omega)} \int_0^t (t-\tau)\|w - v\|_{C(\Omega)} \\
&= C\,\|\alpha\|_{C(\Omega)} \int_0^t e^{-s\tau}\,\|w - v\|_{C(\Omega)}\,(t-\tau)e^{s\tau} d\tau \\
&\le C\,\|\alpha\|_{C(\Omega)}\,\delta(w, v) \int_0^t (t-\tau)e^{s\tau} d\tau \\
&\le C\,\|\alpha\|_{C(\Omega)}\,\delta(w, v)\frac{e^{st}}{s^2},
\end{aligned}$$

which yields

$$\delta(B[w], B[v]) \le \frac{C\,\|\alpha\|_{C(\Omega)}}{s^2}\delta(w, v). \tag{2.14}$$

By taking s such that $\frac{C\|\alpha\|_{C(\Omega)}}{s^2} < 1$, B is a strict contraction. Thus the existence and uniqueness of solution w to (2.13) come directly from Banach's fixed point theorem. Hence we conclude that $\partial_x u = w$ belongs to $C(0, T; C(\Omega))$. With the same process, we could prove that all the derivatives of u with respect to space, up to the rth order, belong to $C(0, T; C(\Omega))$. Hence the solution to (1.1) belongs to $C(0, T; C^r(\Omega))$. □

Corollary 2. *The solution to problem (1.1) belongs to* $X \bigcap C^r(\Omega_T)$, *where* $C^r(\Omega_T) = \{\partial_t^{i_1}\partial_x^{i_2}\partial_y^{i_2}u \in C(\Omega_T),\ |(i_1, i_2, i_3)| \le r\}$ *denotes the space containing all the functions that are differentiable up to the* r*th order, based on that non-negative function* $\alpha \in C^r_{\#}(\Omega)$.

Proof. In polar coordinate, (2.11) can be rewritten as

$$u = C_0 + \frac{1}{2\pi}\int_0^t d\tau \int_0^{t-\tau} dr \int_0^{2\pi} d\theta \frac{(\alpha e^{-u})(x + r\cos\theta, y + r\sin\theta, \tau)r}{\sqrt{(t-\tau)^2 - r^2}}. \tag{2.15}$$

Let $r = (t-\tau)\sin\psi$, $\psi \in [0, \frac{\pi}{2}]$, then we get

$$u = C_0 + \frac{1}{2\pi}\int_0^t d\tau \int_0^{\frac{\pi}{2}} d\psi \int_0^{2\pi} d\theta \left\{(t-\tau)\sin\psi(\alpha e^{-u})(x', y', \tau)\right\}, \tag{2.16}$$

where $x' = x + (t-\tau)\sin\psi\cos\theta$, $y' = y + (t-\tau)\sin\psi\sin\theta$. Taking derivative with respect to time on both sides of (2.16), we arrive at

$$\partial_t u = \frac{1}{2\pi}\int_0^t d\tau \int_0^{\frac{\pi}{2}} d\psi \int_0^{2\pi} d\theta \left\{\sin\psi(\alpha e^{-u})(x', y', \tau) + (t-\tau)\sin\psi\,\partial_t(\alpha e^{-u})(x', y', \tau)\right\}. \tag{2.17}$$

From the definitions of x' and y', we have

$$\begin{aligned}
\partial_t(\alpha e^{-u})(x', y', \tau) &= (s_1\partial_x\alpha(x', y') + s_2\partial_y\alpha(x', y'))e^{-u}(x', y', \tau) \\
&\quad - \alpha(x', y')e^{-u}(x', y', \tau)(s_1\partial_x u(x', y', \tau) + s_2\partial_y u(x', y', \tau)),
\end{aligned} \tag{2.18}$$

where $s_1 = \sin\psi\cos\theta$ and $s_2 = \sin\psi\sin\theta$. Therefore,

$$|\partial_t u| \le C(t^2+1)\|u\|_{C(0,T;C^1(\Omega))}.$$

By Corollary 1, we conclude that $u \in C^1(\Omega_T)$. Note that the derivative of u with respect to time is converted into the derivatives of α and u with respect to space with the chain rule of the derivative, so similarly we could show that $u \in C^r(\Omega_T)$ if $u \in C(0,T;C^r(\Omega))$ and $\alpha \in C^r(\Omega)$, which is the result proven by Corollary 1. □

Remark 2. As shown by the above proof, due to the special form of mild solution formula (2.5), the derivative of the solution u respect to time could be converted into the derivatives of α and u with respect to space. Thus α, although independent on time, determines the regularity of u with respect to both time and space. On the other hand, from the above proof, we could conclude that $\|u\|_{C^r(\Omega)} \le Ce^{st}$ and $\|u\|_{C^r(\Omega_T)} \le Cp(T)e^{sT}$ where $p(T)$ is a polynomial bound of T. Moreover, even if another source term $k(x,y,t)$ is added on the righthand side of (1.1), there still exists a unique solution to the modified problem as long as $k(x,y,t)$ satisfies certain regularity. Specially, with the same process as in Theorem 1, Corollaries 1 and 2, we could prove that the solution belongs to $X\bigcap C^r(\Omega_T)$ if $k(x,y,t)$ belongs to $X\bigcap C(0,T;C^r(\Omega))$.

3. The second order scheme with energy conservation

For the 2D domain $\Omega = [0,L]^2$, we take the uniform numerical grid (x_i, y_j) with $\Delta x = \Delta y = h = \frac{L}{N}$. Denote $\Delta_h = D_x^2 + D_y^2$ as the standard second order centered difference operator. A second order scheme is formulated at a pointwise level as

$$\frac{\psi^{n+1}-\psi^n}{\Delta t} - \frac{1}{2}\Delta_h\left(u^{n+1}+u^n\right) = \alpha\frac{-e^{-u^{n+1}}+e^{-u^n}}{u^{n+1}-u^n}, \tag{3.1}$$

$$\frac{u^{n+1}-u^n}{\Delta t} = \frac{\psi^{n+1}+\psi^n}{2}, \tag{3.2}$$

with $u^0 = C_0$, $\psi^0 = 0$.

With a substitution $\psi^{n+1} = \frac{2(u^{n+1}-u^n)}{\Delta t} - \psi^n$, scheme (3.1) can be reformulated as a closed equation for u^{n+1}:

$$\frac{2u^{n+1}}{\Delta t^2} - \frac{1}{2}\Delta_h u^{n+1} = \alpha\frac{-e^{-u^{n+1}}+e^{-u^n}}{u^{n+1}-u^n} + \frac{1}{2}\Delta_h u^n + \frac{\frac{2u^n}{\Delta t}+2\psi^n}{\Delta t}. \tag{3.3}$$

This equation is nonlinear. However, a more detailed derivation shows that it can be efficiently solved using a linear iteration.

3.1. Energy conservation

For any grid functions f and g evaluated at the uniform numerical mesh, we define the discrete L^2 inner product and L^2 norm in the following standard way:

$$\langle f,g\rangle = h^2\sum_{i,j=0}^{N-1} f_{i,j}g_{i,j}, \qquad \|f\|_2 = \sqrt{\langle f,f\rangle}. \tag{3.4}$$

And the following summation by parts formula could be derived in a straightforward way:

$$\langle f, \Delta_h g\rangle = -\langle \nabla_h f, \nabla_h g\rangle, \tag{3.5}$$

where

$$\|\nabla_h f\|_2^2 = \|D_x f\|_2^2 + \|D_y f\|_2^2$$

with

$$\|D_x f\|_2^2 = h^2\sum_{i,j=0}^{N-1}(f_{i+1,j}-f_{i,j})^2/h^2, \qquad \|D_y f\|_2^2 = h^2\sum_{i,j=0}^{N-1}(f_{i,j+1}-f_{i,j})^2/h^2.$$

In addition, we also define the discrete L^∞ norm as follows:

$$\|f\|_\infty = \max_{0\le i,j\le N-1}|f_{i,j}|. \tag{3.6}$$

Define the discrete energy as

$$E_h^n = \frac{1}{2}\left\|\psi^n\right\|_2^2 + \frac{1}{2}\left\|\nabla_h u^n\right\|_2^2 + E_{nl,h}^n, \quad \text{with } E_{nl,h}^n = h^2\sum_{i,j=0}^{N-1}(\alpha e^{-u^n})_{i,j}. \tag{3.7}$$

Taking inner product of (3.1) with $u^{n+1} - u^n = \frac{1}{2}\Delta t \left(\psi^{n+1} + \psi^n\right)$ yields

$$\frac{1}{2}\left\langle\left(\psi^{n+1} - \psi^n\right), \left(\psi^{n+1} + \psi^n\right)\right\rangle - \frac{1}{2}\left\langle\Delta_h\left(u^{n+1} + u^n\right), \left(u^{n+1} - u^n\right)\right\rangle - \left\langle\alpha\frac{-e^{-u^{n+1}} + e^{-u^n}}{u^{n+1} - u^n}, \left(u^{n+1} - u^n\right)\right\rangle = 0. \quad (3.8)$$

The first term becomes

$$\begin{aligned}\frac{1}{2}\left\langle\left(\psi^{n+1} - \psi^n\right), \left(\psi^{n+1} + \psi^n\right)\right\rangle &= \frac{1}{2}h^2\sum_{i,j=0}^{N-1}\left((\psi^{n+1})^2 - (\psi^n)^2\right)_{i,j} \\ &= \frac{1}{2}\left(\left\|\psi^{n+1}\right\|_2^2 - \left\|\psi^n\right\|_2^2\right). \quad (3.9)\end{aligned}$$

The second term can be analyzed as

$$\begin{aligned}-\frac{1}{2}\left\langle\Delta_h\left(u^{n+1} + u^n\right), \left(u^{n+1} - u^n\right)\right\rangle &= \frac{1}{2}\left\langle\nabla_h\left(u^{n+1} + u^n\right), \nabla_h\left(u^{n+1} - u^n\right)\right\rangle \\ &= \frac{1}{2}\left(\left\|\nabla_h u^{n+1}\right\|_2^2 - \left\|\nabla_h u^n\right\|_2^2\right), \quad (3.10)\end{aligned}$$

in which the summation by parts was applied in the first step. The analysis for the nonlinear term is also straightforward:

$$\begin{aligned}-\left\langle\alpha\frac{-e^{-u^{n+1}} + e^{-u^n}}{u^{n+1} - u^n}, \left(u^{n+1} - u^n\right)\right\rangle &= h^2\sum_{i,j=0}^{N-1}\alpha_{i,j}\left(e^{-u^{n+1}} + e^{-u^n}\right)_{i,j} \\ &= E_{nl,h}^{n+1} - E_{nl,h}^n. \quad (3.11)\end{aligned}$$

As a result, a combination of (3.9)–(3.11) shows that

$$E_h^{n+1} - E_h^n = 0, \quad (3.12)$$

which indicates that the scheme is energy conserving.

4. An alternate nonlinear energy; an alternate second order scheme

As stated above, with scheme (3.1)–(3.2), we must solve u^{n+1} from (3.3) at every step. On the other hand, solving such a nonlinear equation is not a trivial work. To facilitate the numerical implementation, we introduce the following functional $F(u)$:

$$F(u) = \begin{cases} e^{-u}, & \text{if } u \geq 0, \\ 1 - u, & \text{if } u < 0. \end{cases} \quad (4.1)$$

Note that $F(u)$ is equivalent to the nonlinear term e^{-u} when $u \geq 0$, and it becomes a linear functional for $u < 0$. Clearly, it is a C^1 functional, since the left and right directional derivatives equal to each other at $u = 0$. In addition, we observe that

$$F'(u) = -1, \quad \text{if } u < 0, \qquad F'(u) = -e^{-u}, \quad \text{if } u \geq 0. \quad (4.2)$$

With this functional, we propose a modified second order scheme as follows,

$$\frac{\psi^{n+1} - \psi^n}{\Delta t} - \frac{1}{2}\Delta_h\left(u^{n+1} + u^n\right) = \alpha\frac{-F(u^{n+1}) + F(u^n)}{u^{n+1} - u^n}, \quad (4.3)$$

$$\frac{u^{n+1} - u^n}{\Delta t} = \frac{\psi^{n+1} + \psi^n}{2}. \quad (4.4)$$

Thanks to the positivity of the exact solution, there would not be any trouble in the consistency of this scheme.

In turn, we introduce a modified discrete energy as follows:

$$\tilde{E}_h^n = \frac{1}{2}\left\|\psi^n\right\|_2^2 + \frac{1}{2}\left\|\nabla_h u^n\right\|_2^2 + \tilde{E}_{nl,h}^n, \quad \text{with } \tilde{E}_{nl,h}^n = h^2\sum_{i,j=0}^{N-1}\left(\alpha F(u^n)\right)_{i,j}. \quad (4.5)$$

With the similar work as in Section 3.1, we could show that the modified discrete energy (4.5) is conserved for this alternate scheme, i.e.

$$\tilde{E}_h^{n+1} = \tilde{E}_h^n. \quad (4.6)$$

4.1. Numerical implementation; linear iteration

Similar to (3.3), the alternate scheme (4.3) can be rewritten as

$$\frac{2u^{n+1}}{\Delta t^2} - \frac{1}{2}\Delta_h u^{n+1} = \alpha \frac{-F(u^{n+1}) + F(u^n)}{u^{n+1} - u^n} + g\left(u^n, \psi^n\right), \tag{4.7}$$

with $g\left(u^n, \psi^n\right) = \frac{1}{2}\Delta_h u^n + \frac{\frac{2u^n}{\Delta t} + 2\psi^n}{\Delta t}$.

To overcome the difficulty associated with the implicit treatment of the nonlinear term, the following linear iteration algorithm is proposed to solve the above scheme:

$$\frac{2u^{n+1,(m+1)}}{\Delta t^2} - \frac{1}{2}\Delta_h u^{n+1,(m+1)} = \alpha \frac{-F(u^{n+1,(m)}) + F(u^n)}{u^{n+1,(m)} - u^n} + g\left(u^n, \psi^n\right), \tag{4.8}$$

in which $u^{n+1,(m)}$ corresponds to the solution at the m-th iteration. The following theorem shows that such an iteration algorithm is convergent.

Theorem 2. *The linear iteration (4.8) is a contraction mapping under a trivial requirement* $\Delta t < \sqrt{\frac{2}{\alpha}}$.

To prove the theorem, a preliminary estimate is needed.

Lemma 2. *For any* $a, x \in R$ *with* $x \neq a$, *we have*

$$\left|h'(x)\right| \leq 1, \quad \text{where } h(x) = \frac{F(x) - F(a)}{x - a}. \tag{4.9}$$

Proof. A direct calculation yields

$$h'(x) = \frac{F'(x) \cdot (x - a) - (F(x) - f(a))}{(x - a)^2} = \frac{F'(x) - \frac{F(x) - F(a)}{x - a}}{x - a}. \tag{4.10}$$

An application of intermediate value theorem gives

$$\frac{F(x) - F(a)}{x - a} = F'(\xi), \quad \text{so that } h'(x) = \frac{F'(x) - F'(\xi)}{x - a}, \tag{4.11}$$

with ξ between x and a. Meanwhile, with the construction of functional F, it is easy to check that

$$\left|F'(x) - F'(\xi)\right| \leq |x - \xi|. \tag{4.12}$$

Going back to (4.11), we have

$$\left|h'(x)\right| = \frac{\left|F'(x) - F'(\xi)\right|}{|x - a|} \leq \frac{|x - \xi|}{|x - a|} \leq 1, \tag{4.13}$$

in which the inequality $|x - \xi| \leq |x - a|$ was used. □

Now we proceed into the proof of Theorem 2.

Proof. We denote the iteration error at every step as

$$e^{(m)} = u^{n+1,(m)} - u^{n+1}, \tag{4.14}$$

with $u^{n+1,(m)}$ generated by the iteration scheme (4.8). Subtracting (4.7) from (4.8) yields

$$\left(\frac{2}{\Delta t^2} - \frac{1}{2}\Delta_h\right) e^{(m+1)} = \alpha \left(\frac{-F(u^{n+1,(m)}) + F(u^n)}{u^{n+1,(m)} - u^n} - \frac{-F(u^{n+1}) + F(u^n)}{u^{n+1} - u^n}\right).$$

Taking inner product with $e^{(m+1)}$ leads to

$$\frac{2}{\Delta t^2}\left\|e^{(m+1)}\right\|_2^2 + \frac{1}{2}\left\|\nabla_h e^{(m+1)}\right\|_2^2 = \left\langle \alpha \left(\frac{-F(u^{n+1,(m)}) + F(u^n)}{u^{n+1,(m)} - u^n} - \frac{-F(u^{n+1}) + F(u^n)}{u^{n+1} - u^n}\right), e^{(m+1)}\right\rangle. \tag{4.15}$$

The nonlinear iteration error term can be bounded with the help of Lemma 2 by setting $a = u^n$:

$$\begin{aligned} \left|\frac{-F(u^{n+1,(m)}) + F(u^n)}{u^{n+1,(m)} - u^n} - \frac{-F(u^{n+1}) + F(u^n)}{u^{n+1} - u^n}\right| &= \left|h(u^{n+1,(m)}) - h(u^{n+1})\right| \\ &= \left|h'(\xi)(u^{n+1,(m)} - u^{n+1})\right| \\ &\le \left|h'(\xi)\right| \cdot |e^{(m)}| \le |e^{(m)}|, \end{aligned} \tag{4.16}$$

at the pointwise level, with ξ between $u^{n+1,(m)}$ and u^{n+1}. This in turn indicates that

$$\begin{aligned} \left\langle \alpha\left(\frac{-F(u^{n+1,(m)}) + F(u^n)}{u^{n+1,(m)} - u^n} - \frac{-F(u^{n+1}) + F(u^n)}{u^{n+1} - u^n}\right), e^{(m+1)}\right\rangle &\le \bar{\alpha}\left|\langle e^{(m)}, e^{(m+1)}\rangle\right| \\ &\le \frac{\bar{\alpha}}{2}\left(\left\|e^{(m)}\right\|_2^2 + \left\|e^{(m+1)}\right\|_2^2\right). \end{aligned} \tag{4.17}$$

Combine with (4.15), we get

$$\left(\frac{2}{\Delta t^2} - \frac{\bar{\alpha}}{2}\right)\left\|e^{(m+1)}\right\|_2^2 + \frac{1}{2}\left\|\nabla_h e^{(m+1)}\right\|_2^2 \le \frac{\bar{\alpha}}{2}\left\|e^{(m)}\right\|_2^2. \tag{4.18}$$

As a direct consequence, if we have

$$\frac{2}{\Delta t^2} - \frac{\bar{\alpha}}{2} > \frac{\bar{\alpha}}{2}, \quad \text{i.e. } \Delta t < \sqrt{\frac{2}{\bar{\alpha}}}. \tag{4.19}$$

the contraction mapping property could be assured. □

Remark 3. A linear iteration can be applied to solve a nonlinear equation. A geometric convergence in certain norm could be established if the nonlinear derivative is bounded by the linear operator in a functional way. The related work for an epitaxial thin film growth model can be found in a recent article [14].

4.2. The convergence result in maximum norm for a fixed final time

For the alternate scheme (4.3)–(4.4), we have the following convergence result at the pointwise level.

Theorem 3. *Suppose u_e, $\psi_e = \partial_t u_e$ as the exact solution to the original PDE (1.1). Denote (u, ψ) as the numerical solution given by the proposed second order scheme (4.3)–(4.4). Then, provided Δt and h are sufficiently small with the linear refinement path constraint $\Delta t \le Ch$, with C any fixed constant, we have*

$$\left\|\psi_e(t^n) - \psi^n\right\|_\infty + \left\|\nabla_h\left(u_e(t^n) - u^n\right)\right\|_\infty \le C\left(\Delta t^2 + h^2\right), \tag{4.20}$$

with the constant C only dependent on the exact solution and the final time T, independent on the numerical grid sizes Δt and h.

The proof of this theorem is not trivial, so we will give it in the next section.

5. Proof for Theorem 3

In this section we provide a detailed convergence analysis for the alternate second order scheme (4.3)–(4.4). To obtain the desired convergence in $\|\cdot\|_\infty$ norm, we perform a higher order consistency analysis, up to $O(\Delta t^4 + h^4)$, in addition to the leading $O(\Delta t^2 + h^2)$ consistency for the exact solution, through a construction of approximate solution which satisfies the numerical scheme with higher order accuracy. In turn, an $O(\Delta t^3 + h^4)$ convergence in $\ell^\infty(0, T; \ell^2)$ norm can be derived, between the numerical solution and the constructed approximate solution, and this convergence result is applied to derive an error estimate in $\|\cdot\|_\infty$ norm, using an inverse inequality under a mild linear refinement constraint, $\Delta t \le Ch$.

5.1. Higher order consistency analysis; asymptotic expansion of the numerical solution

By standard consistency analysis, the exact solution (u_e, ψ_e) solves the discrete equation

$$\frac{\psi_e^{k+1} - \psi_e^k}{\Delta t} - \frac{1}{2}\Delta_h\left(u_e^k + u_e^{k+1}\right) = -\alpha\frac{F(u_e^{k+1}) - F(u_e^k)}{u_e^{k+1} - u_e^k} + \tau_1^k, \tag{5.1}$$

$$\frac{u_e^{k+1} - u_e^k}{\Delta t} = \frac{1}{2}\left(\psi_e^{k+1} + \psi_e^k\right) + \tau_2^k, \tag{5.2}$$

$$\text{with } \left|(\tau_1^k)_{i,j}\right|, \left|(\tau_2^k)_{i,j}\right| \le C\left(\Delta t^2 + h^2\right), \tag{5.3}$$

for some $C \geq 0$ that depends only on T, Ω and the exact solution. Superscript k denotes that the variable is taken at time $t^k = k\Delta t$. We note that the consistency of the nonlinear term appearing in (5.1) comes from the following fact,

$$u_e^{k+1},\ u_e^k > 0, \quad \text{so that } F(u_e^{k+1}) = e^{-u_e^{k+1}}, \qquad F(u_e^k) = e^{-u_e^k}. \tag{5.4}$$

By Taylor expansion, we have,

$$\frac{e^{-\xi} - e^{-\eta}}{\xi - \eta} = -e^{\frac{-\xi-\eta}{2}} \left(1 + \frac{1}{24}(\xi - \eta)^2 + O\left((\xi - \eta)^4\right)\right).$$

Thus

$$-\frac{F(u_e^{k+1}) - F(u_e^k)}{u_e^{k+1} - u_e^k} = e^{-\frac{u_e^{k+1}+u_e^k}{2}} + O\left(\Delta t^2\right) = e^{-u_e^{k+\frac{1}{2}}} + O(\Delta t^2). \tag{5.5}$$

Meanwhile, it is observed that the leading local truncation error in (5.1)–(5.3) will not be enough to recover a full $O(\Delta t^2 + h^2)$ error estimate in the $\|\cdot\|_\infty$ norm. To remedy this, we need to construct supplementary fields so that the proposed numerical scheme (4.3)–(4.4) satisfies a higher $O(\Delta t^4 + h^4)$ consistency.

For simplicity of presentation, given the exact solution (u_e, ψ_e), we denote $U = u_e$ and introduce the following function

$$\Psi(x, y, t) := \partial_t U(x, y, t) - \frac{\Delta t^2}{12}\partial_t^3 U(x, y, t). \tag{5.6}$$

It is clear that Ψ is an $O(\Delta t^2)$ approximation to the exact solution $\psi_e = \partial_t u_e$. A detailed Taylor expansion shows that

$$\frac{U^{k+1} - U^k}{\Delta t} = \frac{1}{2}\left(\Psi^{k+1} + \Psi^k\right) + O(\Delta t^4). \tag{5.7}$$

We note that an $O(\Delta t^2)$ correction term is added in the construction (5.6) for Ψ so that a higher order consistency is obtained in (4.4). Such a correction term is based on an asymptotic expansion of the numerical scheme and the resulting higher order consistency is crucial in the stability and convergence analysis.

Lemma 3. *For numerical scheme (4.3)–(4.4), there exists an approximate solution $(\hat{U}, \hat{\psi})$ such that an $O(\Delta t^4 + h^4)$ consistency is satisfied. The approximate solution could be expressed as*

$$\hat{U} = U + h^2 U_{h,1} + \Delta t^2 U_{\Delta t,1}, \qquad \hat{\Psi} = \Psi + h^2 \Psi_{h,1} + \Delta t^2 \Psi_{\Delta t,1}, \tag{5.8}$$

where Ψ is the $O(\Delta t^2)$ approximation to $\partial_t u_e$ as defined in (5.6), and $U_{h,1}$, $U_{\Delta t,1}$, $\Psi_{h,1}$, $\Psi_{\Delta t,1}$ are constructed fields depending solely on the exact solution (u_e, ψ_e).

Proof. The construction of the approximate solution consists of two steps. In the spatial discretization, the following truncation error analysis can be obtained by using a straightforward Taylor expansion for the exact solution:

$$\partial_t^2 U - \Delta_h U = \alpha e^{-U} + h^2 g^{(0)} + O(h^4). \tag{5.9}$$

Here the spatially discrete function $g^{(0)}$ is smooth enough in the sense that its discrete derivatives are bounded. Also note that there is no $O(h^3)$ truncation error term, due to the fact that the centered difference used in the spatial discretization gives local truncation errors with only even order terms, $O(h^2)$, $O(h^4)$, etc.

The spatial correction function $\left(U_{h,1}, \Psi_{h,1}\right)$ is given by solving the following equation:

$$\partial_t^2 U_{h,1} - \Delta_h U_{h,1} = -\alpha e^{-U} U_{h,1} - g^{(0)}, \tag{5.10}$$

$$\Psi_{h,1} = \partial_t U_{h,1} - \frac{\Delta t^2}{12}\partial_t^3 U_{h,1}. \tag{5.11}$$

Existence of a solution to the above linear system of ODEs is a standard exercise. Note that the solution depends only on the exact solution U. In addition, the divided differences of $\left(U_{h,1}, \Psi_{h,1}\right)$ of various orders are bounded.

Now, we define

$$U_h^* := U + h^2 U_{h,1}, \qquad \Psi_h^* := \Psi + h^2 \Psi_{h,1}. \tag{5.12}$$

A combination of (5.6), (5.9), (5.10) and (5.11) leads to the fourth order local truncation error for $\left(U_h^*, \Psi_h^*\right)$:

$$\partial_t^2 U_h^* - \Delta_h U_h^* = \alpha e^{-U_h^*} + O(h^4), \tag{5.13}$$

$$\Psi_h^* = \partial_t U_h^* - \frac{\Delta t^2}{12}\partial_t^3 U_h^*, \tag{5.14}$$

for which the following estimate is used:

$$e^{-U_h^*} = e^{-U-h^2U_{h,1}} = e^{-U} - e^{-U}h^2U_{h,1} + O(h^4). \tag{5.15}$$

In addition, using the same type of Taylor expansion as in (5.7), we are able to derive the following higher order estimate:

$$\frac{(U_h^*)^{k+1} - (U_h^*)^k}{\Delta t} = \frac{1}{2}\left((\Psi_h^*)^{k+1} + (\Psi_h^*)^k\right) + O(\Delta t^4), \tag{5.16}$$

based on the construction (5.6), (5.11) and (5.12). In other words, the $O(\Delta t^2)$ correction terms added in the construction (5.6), (5.11) for Ψ and $\Psi_{h,1}$ ensure that an $O(\Delta t^4)$ order consistency is satisfied as in (5.16).

For the temporal correction term, we observe that an application of the proposed second order scheme (4.3) for the profile $\left(U_h^*, \Psi_h^*\right)$ gives

$$\begin{aligned}&\frac{\left(\Psi_h^*\right)^{k+1} - \left(\Psi_h^*\right)^k}{\Delta t} - \frac{1}{2}\Delta_h\left((U_h^*)^{k+1} + (U_h^*)^k\right) + \alpha\frac{F\left((U_h^*)^{k+1}\right) - F\left((U_h^*)^k\right)}{(U_h^*)^{k+1} - (U_h^*)^k}\\ &\quad = \Delta t^2 g^{(1)} + O(\Delta t^4) + O(h^4),\end{aligned} \tag{5.17}$$

using a similar Taylor expansion as in (5.5), where $g^{(1)}$ behaves similarly as $g^{(0)}$ in (5.9) and only depends on U_h^*. In turn, the first order temporal correction function $\left(U_{\Delta t,1}, \Psi_{\Delta t,1}\right)$ is given by the solution to the following linear system of ODEs,

$$\partial_t^2 U_{\Delta t,1} - \Delta_h U_{\Delta t,1} = -\alpha e^{-U_h^*}U_{\Delta t,1} - g^{(1)}, \tag{5.18}$$

$$\Psi_{\Delta t,1} = \partial_t U_{\Delta t,1} - \frac{\Delta t^2}{12}\partial_t^3 U_{\Delta t,1}. \tag{5.19}$$

Again, the solution to (5.18)–(5.19), which exists and is unique, depends solely on the profile U_h^*, hence depends solely on the exact solution U, and is smooth enough in the sense that its divided differences of various orders are bounded. Similar to (5.17), an application of the linearized second order scheme to $\left(U_{\Delta t,1}, \Psi_{\Delta t,1}\right)$ reads

$$\begin{aligned}&\frac{\Psi_{\Delta t,1}^{k+1} - \Psi_{\Delta t,1}^k}{\Delta t} - \frac{1}{2}\Delta_h\left(U_{\Delta t,1}^{k+1} + U_{\Delta t,1}^k\right) + \frac{1}{2}\alpha e^{-\frac{(U_h^*)^{k+1}+(U_h^*)^k}{2}}\left(U_{\Delta t,1}^{k+1} + U_{\Delta t,1}^k\right)\\ &\quad = -\mathbf{g}^{(1)} + O(\Delta t^2) + O(\Delta t^2h^2).\end{aligned} \tag{5.20}$$

For the nonlinear term, a careful Taylor expansion indicates that

$$\begin{aligned}&\frac{F\left((U_h^*)^{k+1}\right) - F\left((U_h^*)^k\right)}{(U_h^*)^{k+1} - (U_h^*)^k} + \frac{1}{2}\Delta t^2 e^{-\frac{(U_h^*)^{k+1}+(U_h^*)^k}{2}}\left(U_{\Delta t,1}^{k+1} + U_{\Delta t,1}^k\right)\\ &\quad = \frac{F\left((U_h^* + \Delta t^2U_{\Delta t,1})^{k+1}\right) - F\left((U_h^* + \Delta t^2U_{\Delta t,1})^k\right)}{(U_h^* + \Delta t^2U_{\Delta t,1})^{k+1} - (U_h^* + \Delta t^2U_{\Delta t,1})^k} + O(\Delta t^4)\\ &\quad = \frac{F\left(\hat{U}^{k+1}\right) - F\left(\hat{U}^k\right)}{\hat{U}^{k+1} - \hat{U}^k} + O(\Delta t^4),\end{aligned} \tag{5.21}$$

as long as Δt, h are small enough, with the full expansion for $\hat{U}$ recalled in (5.8).

Therefore, a combination of (5.17), (5.20) and (5.21) shows that

$$\frac{\hat{\Psi}^{k+1} - \hat{\Psi}^k}{\Delta t} - \frac{1}{2}\Delta_h\left(\hat{U}^{k+1} + \hat{U}^k\right) + \alpha\frac{F\left(\hat{U}^{k+1}\right) - F\left(\hat{U}^k\right)}{\hat{U}^{k+1} - \hat{U}^k} = O(\Delta t^4 + \Delta t^4h^2 + h^4), \tag{5.22}$$

$$\frac{\hat{U}^{k+1} - \hat{U}^k}{\Delta t} = \frac{1}{2}\left(\hat{\Psi}^{k+1} + \hat{\Psi}^k\right) + O(\Delta t^4), \tag{5.23}$$

in which a similar estimate as in (5.7) and (5.16), based on a careful Taylor expansion, is performed to derive (5.23). □

Lemma 3 illustrates the construction of the approximate solution which has an $O(\Delta t^4 + h^4)$ consistency. This higher order consistency is crucial in the stability and convergence analysis, as will be shown in the next subsection.

5.2. $O(\Delta t^3 + h^4)$ convergence in $\ell^\infty(0, T; \ell^2)$ norm

As stated earlier, the purpose of the higher order expansion (5.8) is to obtain an ℓ^∞ estimate of the error function via its ℓ^2 norm in higher order accuracy by utilizing an inverse inequality in spatial discretization, which will be shown below. Instead of a direct comparison between the numerical solution (u, ψ) and the exact solution (u_e, ψ_e), we estimate the error between the numerical solution and the constructed solution to obtain a higher order convergence in $\|\cdot\|_2$ norm. Define the error function as follows,

$$\tilde{u}^k := \hat{U}^k - u^k, \qquad \tilde{\psi}^k := \hat{\Psi}^k - \psi^k. \tag{5.24}$$

Subtracting (4.3)–(4.4) from (5.22)–(5.23) yields

$$\frac{\tilde{\psi}^{k+1} - \tilde{\psi}^k}{\Delta t} - \frac{1}{2}\Delta_h\left(\tilde{u}^{k+1} + \tilde{u}^k\right) + \alpha\left(\frac{F\left(\hat{U}^{k+1}\right) - F\left(\hat{U}^k\right)}{\hat{U}^{k+1} - \hat{U}^k} - \frac{F\left(u^{k+1}\right) - F\left(u^k\right)}{u^{k+1} - u^k}\right) = \hat{\tau}_1^k, \tag{5.25}$$

$$\left|(\hat{\tau}_1)_{i,j}^k\right| \le C\left(\Delta t^4 + h^4\right),$$

$$\frac{\tilde{u}^{k+1} - \tilde{u}^k}{\Delta t} = \frac{1}{2}\left(\tilde{\psi}^{k+1} + \tilde{\psi}^k\right) + \Delta t\,\hat{\tau}_2^k, \qquad \left|(\hat{\tau}_2)_{i,j}^k\right| \le C\Delta t^3. \tag{5.26}$$

Taking an inner product with the error difference function $\left(\tilde{u}^{k+1} - \tilde{u}^k\right)$ gives

$$\begin{aligned}
&\frac{1}{\Delta t}\left\langle \tilde{\psi}^{k+1} - \tilde{\psi}^k, \tilde{u}^{k+1} - \tilde{u}^k\right\rangle - \frac{1}{2}\left\langle \Delta_h\left(\tilde{u}^{k+1} + \tilde{u}^k\right), \tilde{u}^{k+1} - \tilde{u}^k\right\rangle \\
&\quad = -\left\langle \alpha\left(\frac{F\left(\hat{U}^{k+1}\right) - F\left(\hat{U}^k\right)}{\hat{U}^{k+1} - \hat{U}^k} - \frac{F\left(u^{k+1}\right) - F\left(u^k\right)}{u^{k+1} - u^k}\right), \tilde{u}^{k+1} - \tilde{u}^k\right\rangle + \left\langle \hat{\tau}_1^k, \tilde{u}^{k+1} - \tilde{u}^k\right\rangle.
\end{aligned} \tag{5.27}$$

By (5.26), the term associated with the second order temporal derivative can be analyzed as

$$\begin{aligned}
\frac{1}{\Delta t}\left\langle \tilde{\psi}^{k+1} - \tilde{\psi}^k, \tilde{u}^{k+1} - \tilde{u}^k\right\rangle &= \frac{1}{2}\left\langle \tilde{\psi}^{k+1} - \tilde{\psi}^k, \tilde{\psi}^{k+1} + \tilde{\psi}^k\right\rangle + \Delta t\left\langle \tilde{\psi}^{k+1} - \tilde{\psi}^k, \hat{\tau}_2^k\right\rangle \\
&= \frac{1}{2}\left(\left\|\tilde{\psi}^{k+1}\right\|_2^2 - \left\|\tilde{\psi}^k\right\|_2^2\right) + \Delta t\left\langle \tilde{\psi}^{k+1} - \tilde{\psi}^k, \hat{\tau}_2^k\right\rangle \\
&\ge \frac{1}{2}\left(\left\|\tilde{\psi}^{k+1}\right\|_2^2 - \left\|\tilde{\psi}^k\right\|_2^2\right) - \frac{1}{2}\Delta t\left(\left\|\tilde{\psi}^{k+1} - \tilde{\psi}^k\right\|_2^2 + \left\|\hat{\tau}_2^k\right\|_2^2\right) \\
&\ge \frac{1}{2}\left(\left\|\tilde{\psi}^{k+1}\right\|_2^2 - \left\|\tilde{\psi}^k\right\|_2^2\right) - \frac{1}{2}\Delta t\left\|\hat{\tau}_2^k\right\|_2^2 - \Delta t\left(\left\|\tilde{\psi}^{k+1}\right\|_2^2 + \left\|\tilde{\psi}^k\right\|_2^2\right).
\end{aligned} \tag{5.28}$$

The reason for the higher order temporal expansion as in (5.6), (5.11) and (5.19) can be observed in the above derivation. In more detail, the second order temporal derivative may introduce a reduction of accuracy in the stability and convergence estimate; therefore, a higher order consistency analysis is needed to avoid such an accuracy reduction.

The diffusion and local truncation error terms in (5.27) can be handled in a standard way, with the help of summation by parts:

$$\begin{aligned}
-\left\langle \Delta_h\left(\tilde{u}^{k+1} + \tilde{u}^k\right), \tilde{u}^{k+1} - \tilde{u}^k\right\rangle &= \left\langle \nabla_h\left(\tilde{u}^{k+1} + \tilde{u}^k\right), \nabla_h\left(\tilde{u}^{k+1} - \tilde{u}^k\right)\right\rangle \\
&= \left\|\nabla_h\tilde{u}^{k+1}\right\|_2^2 - \left\|\nabla_h\tilde{u}^k\right\|_2^2,
\end{aligned} \tag{5.29}$$

and

$$\begin{aligned}
\left\langle \hat{\tau}_1^k, \tilde{u}^{k+1} - \tilde{u}^k\right\rangle &= \Delta t\left\langle \hat{\tau}_1^k, \frac{\tilde{u}^{k+1} - \tilde{u}^k}{\Delta t}\right\rangle = \frac{1}{2}\Delta t\left\langle \hat{\tau}_1^k, \tilde{\psi}^{k+1} + \tilde{\psi}^k\right\rangle + \Delta t^2\left\langle \hat{\tau}_1^k, \hat{\tau}_2^k\right\rangle \\
&\le \frac{1}{2}\Delta t\left(\frac{1}{2}\left\|\hat{\tau}_1^k\right\|_2^2 + \frac{1}{2}\left\|\tilde{\psi}^{k+1} + \tilde{\psi}^k\right\|_2^2\right) + \Delta t\left(\frac{1}{2}\left\|\hat{\tau}_1^k\right\|_2^2 + \frac{1}{2}\Delta t^2\left\|\hat{\tau}_2^k\right\|_2^2\right) \\
&\le \frac{1}{4}\Delta t\left\|\hat{\tau}_1^k\right\|_2^2 + \frac{1}{2}\Delta t\left(\left\|\tilde{\psi}^{k+1}\right\|_2^2 + \left\|\tilde{\psi}^k\right\|_2^2\right) + \frac{1}{2}\Delta t\left\|\hat{\tau}_1^k\right\|_2^2 + \frac{1}{2}\Delta t^3\left\|\hat{\tau}_2^k\right\|_2^2 \\
&= \frac{3}{4}\Delta t\left\|\hat{\tau}_1^k\right\|_2^2 + \frac{1}{2}\Delta t^3\left\|\hat{\tau}_2^k\right\|_2^2 + \frac{1}{2}\Delta t\left(\left\|\tilde{\psi}^{k+1}\right\|_2^2 + \left\|\tilde{\psi}^k\right\|_2^2\right).
\end{aligned} \tag{5.30}$$

Moreover, the following estimate is crucial in the analysis for the nonlinear error term.

Lemma 4. *We have*

$$\left| \frac{F\left(\hat{U}^{k+1}\right) - F\left(\hat{U}^k\right)}{\hat{U}^{k+1} - \hat{U}^k} - \frac{F\left(u^{k+1}\right) - F\left(u^k\right)}{u^{k+1} - u^k} \right| \le \left|\tilde{u}^{k+1}\right| + \left|\tilde{u}^k\right|, \tag{5.31}$$

at a pointwise level.

Proof. We start from the following rewriting:

$$\frac{F\left(\hat{U}^{k+1}\right) - F\left(\hat{U}^k\right)}{\hat{U}^{k+1} - \hat{U}^k} - \frac{F\left(u^{k+1}\right) - F\left(u^k\right)}{u^{k+1} - u^k} = \frac{F\left(\hat{U}^{k+1}\right) - F\left(\hat{U}^k\right)}{\hat{U}^{k+1} - \hat{U}^k} - \frac{F\left(u^{k+1}\right) - F\left(\hat{U}^k\right)}{u^{k+1} - \hat{U}^k}$$
$$+ \frac{F\left(\hat{U}^k\right) - F\left(u^{k+1}\right)}{\hat{U}^k - u^{k+1}} - \frac{F\left(u^k\right) - F\left(u^{k+1}\right)}{u^k - u^{k+1}}. \tag{5.32}$$

Each part can be analyzed with an application of Lemma 2:

$$\left| \frac{F\left(\hat{U}^{k+1}\right) - F\left(\hat{U}^k\right)}{\hat{U}^{k+1} - \hat{U}^k} - \frac{F\left(u^{k+1}\right) - F\left(\hat{U}^k\right)}{u^{k+1} - \hat{U}^k} \right| = \left| h(\hat{U}^{k+1}) - h(u^{k+1}) \right|,$$
$$= \left| h'(\xi_1)\left(\hat{U}^{k+1} - u^{k+1}\right) \right| \le \left| h'(\xi_1) \right| \cdot \left| \hat{U}^{k+1} - u^{k+1} \right| \le |\tilde{u}^{k+1}|, \quad (a = \hat{U}^k), \tag{5.33}$$

$$\left| \frac{F\left(\hat{U}^k\right) - F\left(u^{k+1}\right)}{\hat{U}^k - u^{k+1}} - \frac{F\left(u^k\right) - F\left(u^{k+1}\right)}{u^k - u^{k+1}} \right| = \left| h(\hat{U}^k) - h(u^k) \right|,$$
$$= \left| h'(\xi_2)\left(\hat{U}^k - u^k\right) \right| \le \left| h'(\xi_2) \right| \cdot \left| \hat{U}^k - u^k \right| \le |\tilde{u}^k|, \quad (a = u^{k+1}), \tag{5.34}$$

at the pointwise level, with ξ_1 between u^{k+1} and $\hat{U}^{k+1}$, ξ_2 between u^k and $\hat{U}^k$, respectively. In turn, a combination of (5.32)–(5.34) yields (5.31). □

Consequently, we arrive at the following nonlinear estimate

$$\begin{aligned}
&-\left\langle \alpha \left(\frac{F\left(\hat{U}^{k+1}\right) - F\left(\hat{U}^k\right)}{\hat{U}^{k+1} - \hat{U}^k} - \frac{F\left(u^{k+1}\right) - F\left(u^k\right)}{u^{k+1} - u^k} \right), \tilde{u}^{k+1} - \tilde{u}^k \right\rangle \\
&\le \bar{\alpha} \left\langle \left|\tilde{u}^{k+1}\right| + \left|\tilde{u}^k\right|, \left|\tilde{u}^{k+1} - \tilde{u}^k\right| \right\rangle \\
&\le \frac{\bar{\alpha}}{2}\Delta t \left\langle \left|\tilde{u}^{k+1}\right| + \left|\tilde{u}^k\right|, \left|\tilde{\psi}^{k+1}\right| + \left|\tilde{\psi}^k\right| \right\rangle + \bar{\alpha}\Delta t^2 \left\langle \left|\tilde{u}^{k+1}\right| + \left|\tilde{u}^k\right|, \left|\hat{\tau}_2^k\right| \right\rangle \\
&\le \frac{\bar{\alpha}}{4}\Delta t \left(\left\| |\tilde{u}^{k+1}| + |\tilde{u}^k| \right\|_2^2 + \left\| |\tilde{\psi}^{k+1}| + |\tilde{\psi}^k| \right\|_2^2 \right) + \frac{\bar{\alpha}}{2}\Delta t \left(\left\| |\tilde{u}^{k+1}| + |\tilde{u}^k| \right\|_2^2 + \Delta t^2 \left\| \hat{\tau}_2^k \right\|_2^2 \right) \\
&\le \frac{\bar{\alpha}}{2}\Delta t \left(\left\| \tilde{u}^{k+1} \right\|_2^2 + \left\| \tilde{u}^k \right\|_2^2 + \left\| \tilde{\psi}^{k+1} \right\|_2^2 + \left\| \tilde{\psi}^k \right\|_2^2 \right) + \bar{\alpha}\Delta t \left(\left\| \tilde{u}^{k+1} \right\|_2^2 + \left\| \tilde{u}^k \right\|_2^2 \right) + \frac{\bar{\alpha}}{2}\Delta t^3 \left\| \hat{\tau}_2^k \right\|_2^2 \\
&= \frac{3}{2}\bar{\alpha}\Delta t \left(\left\| \tilde{u}^{k+1} \right\|_2^2 + \left\| \tilde{u}^k \right\|_2^2 \right) + \frac{1}{2}\bar{\alpha}\Delta t \left(\left\| \tilde{\psi}^{k+1} \right\|_2^2 + \left\| \tilde{\psi}^k \right\|_2^2 \right) + \frac{1}{2}\bar{\alpha}\Delta t^3 \left\| \hat{\tau}_2^k \right\|_2^2.
\end{aligned} \tag{5.35}$$

Finally, a combination of (5.27)–(5.30) and (5.35) gives

$$\begin{aligned}
&\frac{1}{2}\left(\left\| \tilde{\psi}^{k+1} \right\|_2^2 - \left\| \tilde{\psi}^k \right\|_2^2 + \left\| \nabla_h \tilde{u}^{k+1} \right\|_2^2 - \left\| \nabla_h \tilde{u}^k \right\|_2^2 \right) \\
&\le \frac{3}{2}\bar{\alpha}\Delta t \left(\left\| \tilde{u}^{k+1} \right\|_2^2 + \left\| \tilde{u}^k \right\|_2^2 \right) + \frac{1}{2}(3 + \bar{\alpha})\Delta t \left(\left\| \tilde{\psi}^{k+1} \right\|_2^2 + \left\| \tilde{\psi}^k \right\|_2^2 \right) \\
&\quad + \frac{3}{4}\Delta t \left\| \hat{\tau}_1^k \right\|_2^2 + \frac{1}{2}\Delta t \left\| \hat{\tau}_2^k \right\|_2^2 + \frac{1}{2}(1 + \bar{\alpha})\Delta t^3 \left\| \hat{\tau}_2^k \right\|_2^2.
\end{aligned} \tag{5.36}$$

Meanwhile, by (5.26), we have the following expansion for $\tilde{u}^k$ in terms of $\tilde{\psi}^l,\ 0 \le l \le k$:

$$\tilde{u}^k = \tilde{u}^0 + \sum_{l=0}^{k-1} \left(\frac{1}{2}\Delta t \left(\tilde{\psi}^{l+1} + \tilde{\psi}^l \right) + \Delta t^2 \hat{\tau}_2^l \right)$$

$$= \frac{1}{2}\Delta t \sum_{l=0}^{k-1} \left(\tilde{\psi}^{l+1} + \tilde{\psi}^l \right) + \Delta t^2 \sum_{l=0}^{k-1} \hat{\tau}_2^l, \quad (\text{since } \tilde{u}^0 \equiv 0). \tag{5.37}$$

Let $P_1 = \frac{1}{2}\Delta t \sum_{l=0}^{k-1} \left(\tilde{\psi}^{l+1} + \tilde{\psi}^l \right), P_2 = \Delta t^2 \sum_{l=0}^{k-1} \hat{\tau}_2^l$. An application of Cauchy inequality implies that

$$\left|\tilde{u}^k\right|^2 = (P_1 + P_2)^2 \le 2(P_1^2 + P_2^2)$$

$$\le \frac{1}{2}\Delta t^2 k \sum_{l=0}^{k-1} (\tilde{\psi}^{l+1} + \tilde{\psi}^l)^2 + 2\Delta t^4 k \sum_{l=0}^{k-1} (\hat{\tau}_2^l)^2$$

$$\le k\Delta t^2 \sum_{l=0}^{k-1} \left(\left|\tilde{\psi}^{l+1}\right|^2 + \left|\tilde{\psi}^l\right|^2 \right) + 2k\Delta t^4 \sum_{l=0}^{k-1} \left|\hat{\tau}_2^l\right|^2$$

$$\le 2T\Delta t \sum_{l=0}^{k} \left|\tilde{\psi}^l\right|^2 + 2T\Delta t^3 \sum_{l=0}^{k-1} \left|\hat{\tau}_2^l\right|^2, \quad \text{since } k\Delta t \le T, \tag{5.38}$$

at a pointwise level. This in turn shows that

$$\|\tilde{u}^k\|_2^2 \le 2T\Delta t \sum_{l=0}^{k} \left\|\tilde{\psi}^l\right\|_2^2 + C\Delta t^8, \qquad \|\tilde{u}^{k+1}\|_2^2 \le 2T\Delta t \sum_{l=0}^{k+1} \left\|\tilde{\psi}^l\right\|_2^2 + C\Delta t^8, \tag{5.39}$$

in which the local truncation error bound $\left|(\hat{\tau}_2)_{i,j}^l\right| \le C\Delta t^3$ is used. Going back to (5.36), we get

$$\frac{1}{2}\left(\left\|\tilde{\psi}^{k+1}\right\|_2^2 - \left\|\tilde{\psi}^k\right\|_2^2 + \|\nabla_h \tilde{u}^{k+1}\|_2^2 - \|\nabla_h \tilde{u}^k\|_2^2 \right)$$

$$\le 6\bar{\alpha}T\Delta t^2 \sum_{l=0}^{k+1} \left\|\tilde{\psi}^l\right\|_2^2 + \frac{1}{2}(3+\bar{\alpha})\Delta t \left(\left\|\tilde{\psi}^{k+1}\right\|_2^2 + \left\|\tilde{\psi}^k\right\|_2^2 \right) + C\Delta t \left(\Delta t^6 + h^8 \right). \tag{5.40}$$

Summing over k and using $\tilde{u}^0 \equiv 0$, $\tilde{\psi}^0 \equiv 0$, we obtain

$$\frac{1}{2}\left(\left\|\tilde{\psi}^n\right\|_2^2 + \|\nabla_h \tilde{u}^n\|_2^2 \right) \le 6\bar{\alpha}T\Delta t^2 \sum_{k=0}^{n-1} \sum_{l=0}^{k+1} \left\|\tilde{\psi}^l\right\|_2^2 + (3+\bar{\alpha})\Delta t \sum_{k=0}^{n} \left\|\tilde{\psi}^k\right\|_2^2 + CT\left(\Delta t^6 + h^8 \right). \tag{5.41}$$

Meanwhile, by the following observation

$$\sum_{l=0}^{k+1} \left\|\tilde{\psi}^l\right\|_2^2 \le \sum_{l=0}^{n} \left\|\tilde{\psi}^l\right\|_2^2, \quad \forall 0 \le k+1 \le n, \tag{5.42}$$

we arrive at

$$\frac{1}{2}\left(\left\|\tilde{\psi}^n\right\|_2^2 + \|\nabla_h \tilde{u}^n\|_2^2 \right) \le \left(6\bar{\alpha}T^2 + 3 + \bar{\alpha} \right) \Delta t \sum_{k=0}^{n} \left\|\tilde{\psi}^k\right\|_2^2 + CT\left(\Delta t^6 + h^8 \right). \tag{5.43}$$

By denoting $\tilde{\hat{E}}^k = \left\|\tilde{\psi}^k\right\|_2^2 + \|\nabla_h \tilde{u}^k\|_2^2$, the above inequality shows that

$$\tilde{\hat{E}}^n \le \left(12\bar{\alpha}T^2 + 6 + 2\bar{\alpha} \right) \Delta t \sum_{k=0}^{n} \tilde{\hat{E}}^k + CT\left(\Delta t^6 + h^8 \right). \tag{5.44}$$

An application of the discrete Gronwall inequality yields the desired result:

$$\left\|\tilde{\psi}^n\right\|_2^2 + \|\nabla_h \tilde{u}^n\|_2^2 = \tilde{\hat{E}}^n \le C\left(\Delta t^6 + h^8 \right),$$

$$\text{so that } \left\|\tilde{\psi}^n\right\|_2 + \|\nabla_h \tilde{u}^n\|_2 \le C\left(\Delta t^3 + h^4 \right), \tag{5.45}$$

with C independent on Δt and h. Subsequently, its substitution to (5.39) gives

$$\left\|\tilde{u}^n\right\|_2^2 \le 2T\Delta t \sum_{l=0}^{n} \left\|\tilde{\psi}^l\right\|_2^2 + C\Delta t^8 \le C\left(\Delta t^6 + h^8\right),$$

$$\text{i.e.,} \quad \left\|\tilde{u}^n\right\|_2 \le C\left(\Delta t^3 + h^4\right). \tag{5.46}$$

In turn, the following estimate of valid:

$$\left\|\tilde{\psi}^n\right\|_2 + \left\|\tilde{u}^n\right\|_2 + \left\|\nabla_h \tilde{u}^n\right\|_2 \le C\left(\Delta t^3 + h^4\right), \tag{5.47}$$

i.e. the $O(\Delta t^3 + h^4)$ convergence in $\ell^\infty(0, T; \ell^2)$ norm, between the numerical solution (u, ψ) and the constructed approximate solution $\left(\hat{U}, \hat{\Psi}\right)$.

5.3. $O(\Delta t^2 + h^2)$ convergence in $\ell^\infty(0, T; \ell^\infty)$ norm

As shown above, we have already had the convergence result (5.47), in ℓ^2 norm. To get the corresponding result in l^∞ norm, the following discrete inverse inequality in two dimensional case is necessary:

$$\|u\|_\infty \le \frac{\|u\|_2}{h}, \tag{5.48}$$

where h is the discrete mesh size.

Thus we get

$$\left\|\tilde{\psi}^n\right\|_\infty + \left\|\tilde{u}^n\right\|_\infty \le \frac{C\left(\left\|\tilde{\psi}^n\right\|_2 + \left\|\tilde{u}^n\right\|_2\right)}{h} \le \frac{C\left(\Delta t^3 + h^4\right)}{h} \le C\left(\Delta t^2 + h^2\right), \tag{5.49}$$

with the linear refinement constraint $\Delta t \le Ch$.

On the other hand, by Lemma 3, we have

$$\left\|\hat{U} - U\right\|_\infty \le C(\Delta t^2 + h^2), \qquad \left\|\hat{\Psi} - \Psi\right\|_\infty \le C(\Delta t^2 + h^2). \tag{5.50}$$

Hence, a direct application of triangle inequality yields

$$\left\|\psi^n - \psi_e^n\right\|_\infty \le \left\|\psi^n - \hat{\Psi}^n\right\|_\infty + \left\|\Psi^n - \hat{\Psi}^n\right\|_\infty + \left\|\psi_e - \Psi^n\right\|_\infty \le C\left(\Delta t^2 + h^2\right), \tag{5.51}$$

$$\left\|u^n - u_e^n\right\|_\infty \le \left\|u^n - \hat{U}^n\right\|_\infty + \left\|U^n - \hat{U}^n\right\|_\infty \le C\left(\Delta t^2 + h^2\right), \tag{5.52}$$

for any $n \cdot \Delta t \le T$, which is just the convergence result stated in Theorem 3.

Remark 4. In Theorem 3, the linear refinement path constraint $\Delta t \le Ch$ is a necessary condition since we use the constraint in the discrete inverse inequality to get the l^∞ norm convergence result from the l^2 norm convergence result. However, due to the special structure of the nonlinear term and our choice of functional $F(u)$, the l^2 norm convergence result is unconditional. Actually, we have

$$\left\|u^n - u_e^n\right\|_2 \le \left\|u^n - \hat{U}^n\right\|_2 + \left\|U^n - \hat{U}^n\right\|_2, \tag{5.53}$$

where $u_e = U$ is the exact solution. The first part is of order $O(\Delta t^3 + h^4)$ due to (5.47) and the second part is of order $O(\Delta t^2 + h^2)$ due to (5.8). Then we have

$$\left\|u^n - u_e^n\right\|_2 \le C(\Delta t^2 + h^2), \tag{5.54}$$

which is independent of the choice of Δt and h.

6. Numerical simulation

In this section, we give several numerical examples to illustrate the theoretical results analyzed above. We take $L = 1$ so that the two dimensional domain becomes $\Omega = [0, 1]^2$.

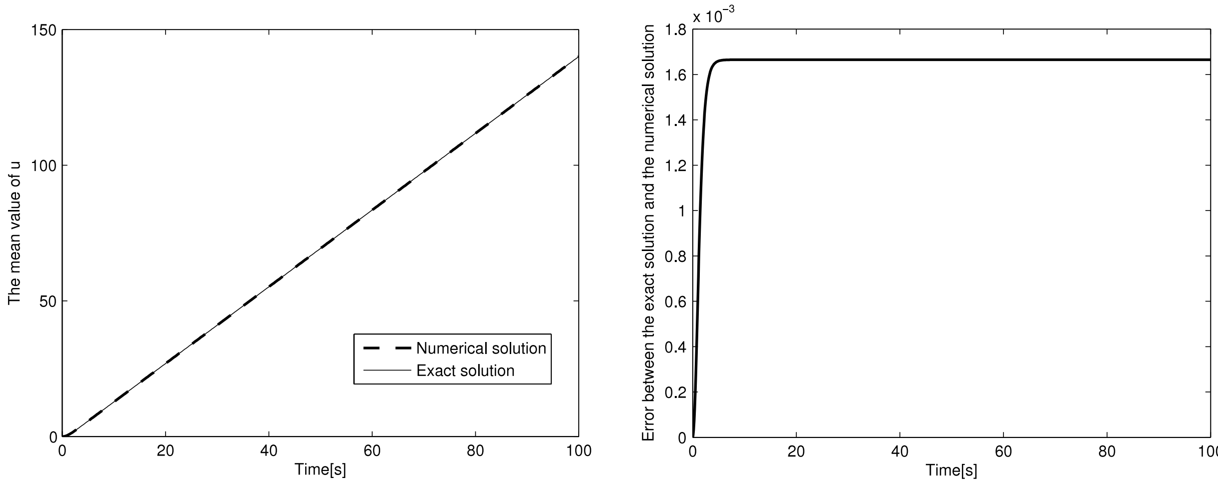

Fig. 6.1. The left figure shows both exact solution and numerical solution. The dotted line corresponds to the mean of the numerical solution to (1.1) using scheme (4.3)–(4.4), and the solid line corresponds to that of the exact solution to (6.1). The right figure illustrates the error.

Table 6.1
Relative error corresponding to different time steps. The order column is calculated by $\log_2(\frac{e_{\Delta t}}{e_{2\Delta t}})$.

Δt	Relative error	Order
0.1	$1.33e^{-3}$	/
0.1×2^{-1}	$3.33e^{-4}$	1.9978
0.1×2^{-2}	$8.34e^{-5}$	1.9974
0.1×2^{-3}	$2.08e^{-5}$	2.0035
0.1×2^{-4}	$5.21e^{-6}$	1.9972
0.1×2^{-5}	$1.30e^{-6}$	2.0028
0.1×2^{-6}	$3.26e^{-7}$	1.9956

6.1. *A special case for constant α*

When α is a constant, it is easy to show that under a periodic boundary condition, the solution to (1.1) in Ω at any time is also a constant, whose value is equal to the solution to the following ODE problem,

$$\begin{cases} \dfrac{d^2u}{dt^2} = \alpha e^{-u}, \\ u\,|_{t=0} = C_0 \geq 0, \qquad \partial_t u\,|_{t=0} = 0. \end{cases} \tag{6.1}$$

And the solution to the ODE problem is given by

$$u(t) = -\ln\left\{\frac{C - C\left(\frac{e^{\sqrt{2C}t}-1}{e^{\sqrt{2C}t}+1}\right)^2}{\alpha}\right\}, \quad \text{with } C = \alpha e^{-C_0}. \tag{6.2}$$

Here we set $C_0 = 0$ and $\alpha = 1$ and use the numerical scheme (4.3)–(4.4) to perform the simulation. To calculate u^{n+1} from u^n and ψ^n, we use the iteration algorithm (4.8) and take the solution calculated from leap-frog scheme as the initial guess at every time step.

Fig. 6.1 gives the results. The two lines, which represent exact and numerical solutions, respectively, match well in the left figure. The right figure illustrates the error.

To illustrate the convergence rate, we take time step as 10^{-1}, $2^{-1}\times10^{-1}$, $2^{-2}\times10^{-1}$, $2^{-3}\times10^{-1}$, $2^{-4}\times10^{-1}$, respectively. Fig. 6.2 shows the semilog plot of the relative error, and Table 6.1 shows the relative error corresponding to different time steps. Both illustrate the second order convergence rate in time, which agrees with the above theoretical analysis.

Moreover, Fig. 6.3 presents the discrete energy corresponding to the simulation. The left figure corresponds to the value of $\tilde{E}_h^n - \tilde{E}_h^0$, and the right figure corresponds to the value of $\tilde{E}_h^n$. We could clearly observe the conservation for the discrete energy E_h^n defined as (4.5), up to a machine error.

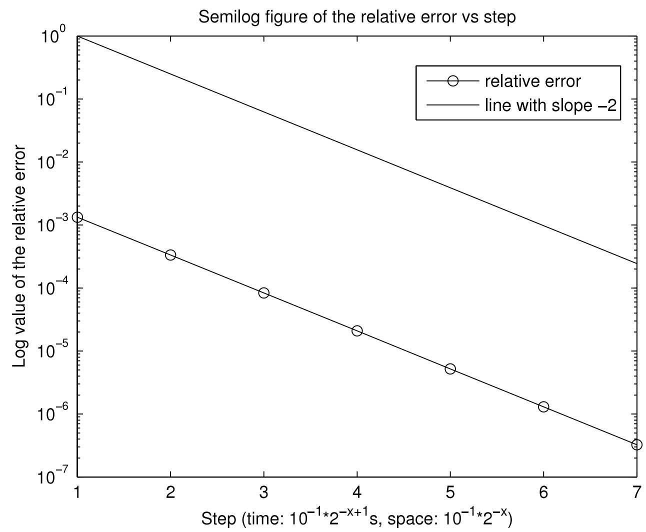

Fig. 6.2. Semilog figure of the relative error corresponding to (1.1) with $\alpha = 1$ and $C_0 = 0$.

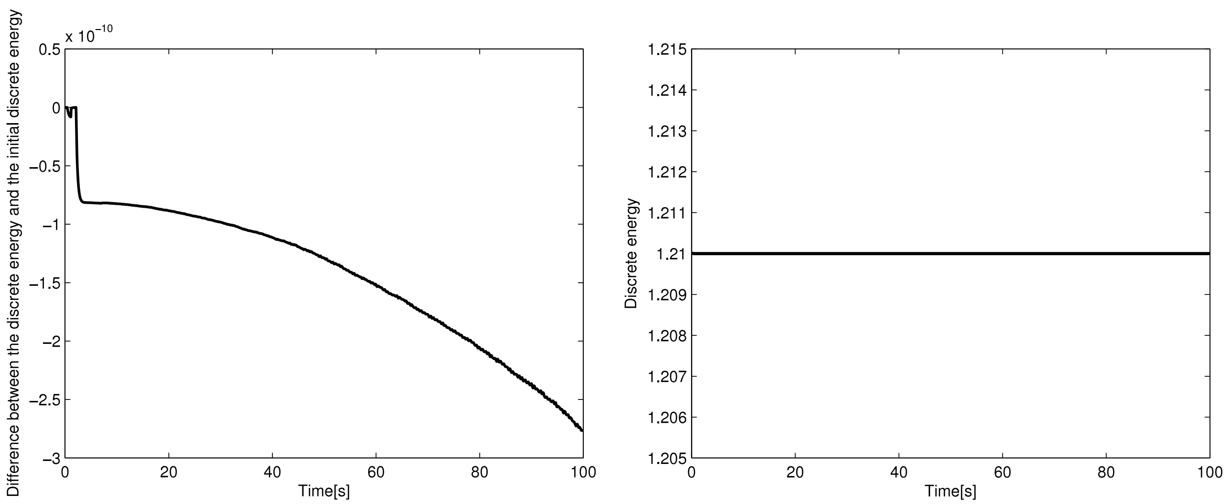

Fig. 6.3. The discrete energy of the numerical solution.

6.2. The case for nonconstant α

Now consider the following modified nonlinear equation with a periodic boundary condition,

$$\begin{cases} \partial_t^2 u - \Delta u = \alpha e^{-u} + k, & \text{in } \Omega_T, \\ u\,|_{t=0} \equiv C_0 \geq 0, & \partial_t u\,|_{t=0} \equiv 0. \end{cases} \tag{6.3}$$

Set

$$u(x, y, t) = t^2(1 + \cos t)(2 + \sin 2\pi x + \sin 2\pi y) \tag{6.4}$$

as the solution to (6.3). And set $\alpha = \alpha(x, y) = 1 + \sin 2\pi x \sin 2\pi y$ and $C_0 = 0$. Then we have

$$\begin{aligned} k(x, y, t) = {} & (2 + 2\cos t - 4t \sin t - t^2 \cos t)(2 + \sin 2\pi x + \sin 2\pi y) + 4\pi^2 t^2(1 + \cos t)(\sin 2\pi x + \sin 2\pi y) \\ & - \alpha(x, y, t)e^{-t^2(1+\cos t)(2+\sin 2\pi x+\sin 2\pi y)}. \end{aligned} \tag{6.5}$$

For this example, we also use the numerical scheme (4.3)–(4.4) to perform the simulation, where we take k as $k^{n+\frac{1}{2}}$ to keep the accuracy. For the iteration part, again we use the numerical solution obtained from leap-frog scheme as our initial guess for u^{n+1} at every time step.

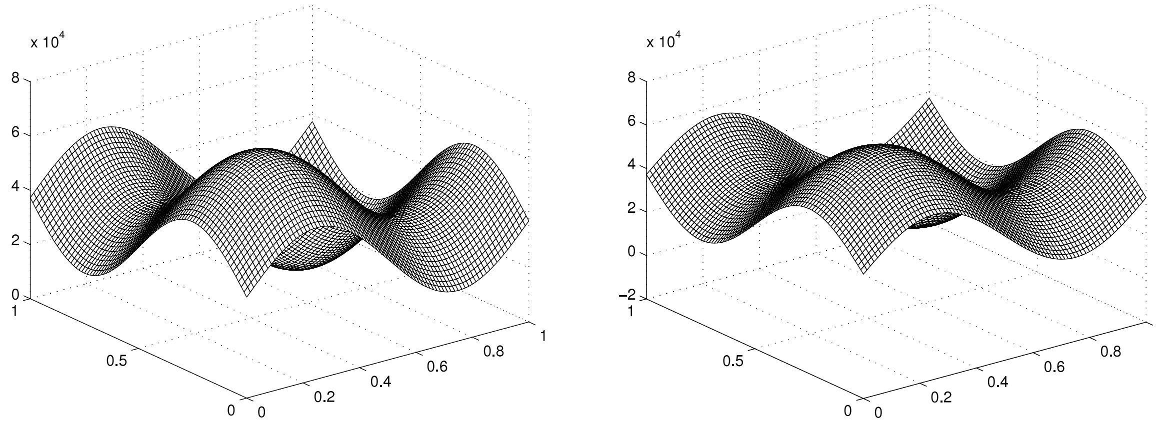

Fig. 6.4. The exact and numerical solution to (6.3) at $T = 100$.

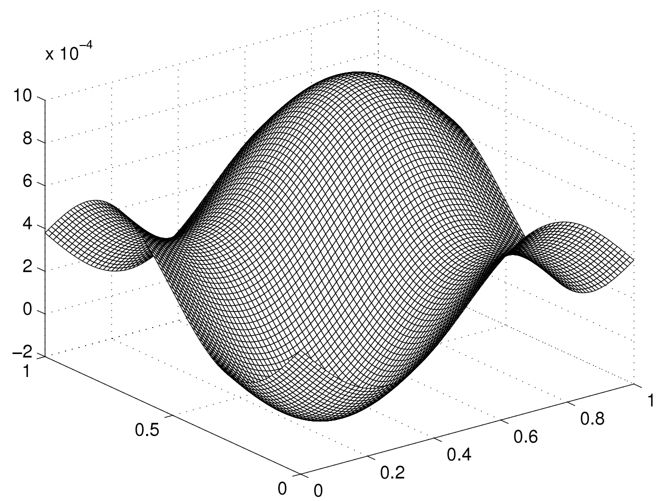

Fig. 6.5. The relative error between exact solution and numerical solution to (6.3) at $T = 100$.

Table 6.2
Relative errors corresponding to different time steps and space steps. The order column is calculated by $\log_2(\frac{e_h}{e_{2h}})$.

$\Delta t = h$	Relative error	Order
0.1	$1.31e^{-2}$	/
0.1×2^{-1}	$3.29e^{-3}$	1.9916
0.1×2^{-2}	$8.25e^{-4}$	1.9941
0.1×2^{-3}	$2.07e^{-4}$	1.9953
0.1×2^{-4}	$5.18e^{-5}$	1.9972
0.1×2^{-5}	$1.30e^{-5}$	1.9985
0.1×2^{-6}	$3.24e^{-6}$	1.9992

Fig. 6.4 presents the results at $T = 100$, in which the left figure corresponds to the exact solution and the right figure corresponds to the numerical solution, respectively. Fig. 6.5 gives the relative error. From both figures we could see that two results are very close.

To illustrate the convergence rate, initially, we take both time step and space step as 0.1. Then we halve both in the same time for the later simulation. Fig. 6.6 shows the semilog plot of the relative error. Table 6.2 shows the relative error corresponding to different time steps and space steps. Both illustrate that the convergence rate is $O(\Delta t^2 + h^2)$, which agrees with the theoretical analysis above.

Moreover, although we could not prove that scheme (4.3)–(4.4) preserves positivity at a theoretical level, from the convergence of the scheme and the positivity of the exact solution, we still could believe that the numerical solution keeps

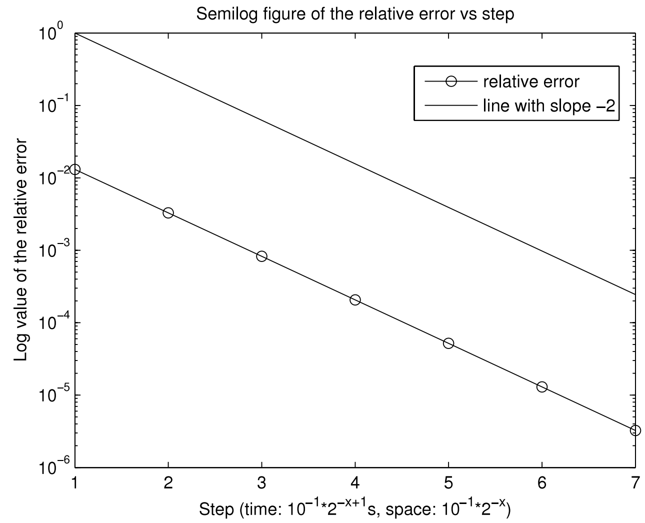

Fig. 6.6. Semilog figure of relative error corresponding to (6.3).

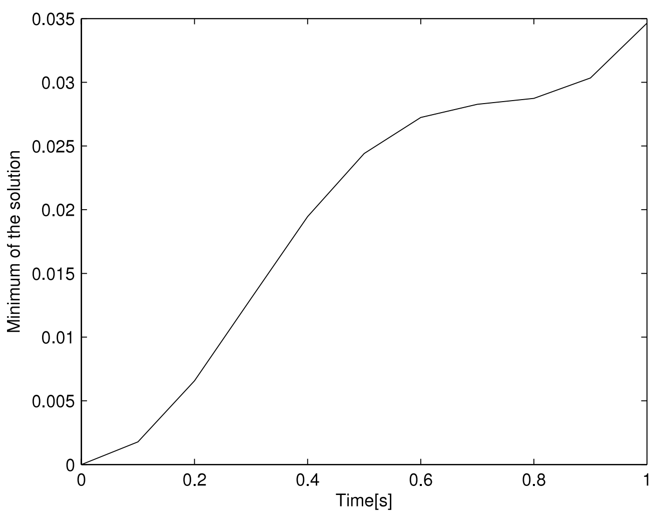

Fig. 6.7. Minimum of the numerical solution to (6.3).

non-negative. Fig. 6.7 gives the minimum of the numerical solution, which shows its positivity preserving property at a numerical level.

7. Conclusions

In this work, we propose an energy-conserving second order scheme to solve a nonlinear hyperbolic equation with an exponential nonlinear term, i.e. (1.1). Based on the positivity of the solution, a modification of the nonlinear term is introduced. With this modification, we apply a linear iteration algorithm to solve the numerical scheme and prove that the iteration is a contraction mapping. We also analyze the numerical scheme in detail and give an $O(\Delta t^2 + h^2)$ convergence analysis in $\ell^\infty(0, T; \ell^\infty)$ norm, under a standard constraint $\Delta t \leq Ch$. Furthermore, several numerical examples are presented, which justify our theoretical results.

Acknowledgments

The authors greatly appreciate many helpful discussions with Masahiro Yamamoto, Xiaoming Wang and Jin Cheng, in particular for their insightful suggestion and comments. This work is supported in part by the NSF DMS-1115420 (C. Wang), NSF DMS-1418689 (C. Wang), NSFC 11271281 (C. Wang), the Ministry of Education of China and the State Administration of Foreign Experts Affairs of China under 111 project grant B08018 (W. Chen), and National Science Foundation of China

11171077, 91130004 and 11331004 (W. Chen). L. Wang is partially supported by Shanghai excellent academic leaders plan (13XD1400900). C. Wang also thanks the Key Laboratory of Mathematics for Nonlinear Sciences (EZH1411104), Fudan University, for support during his visit.

References

[1] M. Avrami, Kinetics of phase change. I General theory, J. Chem. Phys. 7 (12) (1939) 1103–1112.
[2] W. Johnson, R. Mehl, Reaction kinetics in processes of nucleation and growth, Trans. AIME 135 (1939) 416–442.
[3] A.N. Kolmogorov, On the statistical theory of the crystalization of metals, Bull. Acad. Sci. USSR Math. Ser. 1 (1937) 355–359.
[4] J.W. Cahn, The time cone method for nucleartion and growth kinetics on a finite domain, Mater. Res. Soc. Symp. Proc. 398 (1995) 425–438.
[5] Y. Liu, M. Yamamoto, On the multiple hyperbolic systems modeling phase transformation kinetics, Appl. Anal. 93 (2014) 1297–1318.
[6] W.A. Strauss, Nonlinear Wave Equations, AMS, 1989.
[7] A. Baskaran, J.S. Lowengrub, C. Wang, S.M. Wise, Convergence analysis of a second order convex splitting scheme for the modified phase field crystal equation, SIAM J. Numer. Anal. 51 (2013) 2851–2873.
[8] C. Wang, S.M. Wise, An energy stable and convergent finite-difference scheme for the modified phase field crystal equation, SIAM J. Numer. Anal. 49 (3) (2011) 945–969.
[9] A. Baskaran, Z. Hu, J.S. Lowengrub, C. Wang, S.M. Wise, P. Zhou, Energy stable and efficient finite-difference nonlinear multigrid schemes for the modified phase field crystal equation, J. Comput. Phys. 250 (2013) 270–292.
[10] F.E. Browder, Nonlinear Operators and Nonlinear Equations of Evolution in Banach Spaces, AMS, Providence, 1976.
[11] L.C. Evans, Partial Differential Equation, AMS, Providence, Rohode Island, 1998.
[12] R. Courant, D. Hilbert, Methods of Mathematical Physics, Interscience Publishers, New York, 1962.
[13] M.H. Protter, H.F. Weinberger, Maximum Principles in Differential Equations, Springer-Verlag, New York, 1984.
[14] W. Chen, C. Wang, X. Wang, S.M. Wise, A linear iteration algorithm for energy stable second-order scheme for a thin film model without slope selection, J. Sci. Comput. 59 (2014) 574–601.