Research Article

Mathematical Methods in the Applied Sciences

Received 9 April 2016

Published online 15 February 2017 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/mma.4317
MOS subject classification: 2010 MSC: 70K75; 65M70; 65M12

Numerical approximation of the conservative Allen–Cahn equation by operator splitting method

Zhifeng Weng[*†] **and Qingqu Zhuang**

Communicated by T. Monovasilis

In this paper, a second-order fast explicit operator splitting method is proposed to solve the mass-conserving Allen–Cahn equation with a space–time-dependent Lagrange multiplier. The space–time-dependent Lagrange multiplier can preserve the volume of the system and keep small features. Moreover, we analyze the discrete maximum principle and the convergence rate of the fast explicit operator splitting method. The proposed numerical scheme is of spectral accuracy in space and of second-order accuracy in time, which greatly improves the computational efficiency. Numerical experiments are presented to confirm the accuracy, efficiency, mass conservation, and stability of the proposed method. Copyright © 2017 John Wiley & Sons, Ltd.

Keywords: mass-conserving Allen–Cahn equation; operator splitting; spectral method; the discrete maximum principle; error estimate

1. Introduction

The Allen–Cahn equation was introduced originally as a phenomenological model for antiphase domain coarsening in a binary alloy [1]. The Allen–Cahn equation and its various modified forms have been applied in addressing a range of problems, for example, crystal growth [2], image analysis [3, 4], and phase separation [5]. Because the well-known classical Allen–Cahn equation [6–9] does not have mass conservation property, Rubinstein and Sternberg [10] introduced a nonlocal Allen–Cahn equation with a time-dependent Lagrange multiplier to enforce conservation of mass.

In this paper, we focus on the following mass conserving Allen–Cahn equation [11] with space–time-dependent Lagrange multiplier $\beta(u)$

$$\begin{cases} \frac{\partial u}{\partial t}(\mathbf{x},t) = \mathcal{M}\Big(\Delta u(\mathbf{x},t) - \frac{1}{\epsilon^2}[F'(u) - \beta(u)]\Big), & (\mathbf{x},t) \in \Omega \times (0,T], \\ u(\mathbf{x},0) = u_0(\mathbf{x}), & \mathbf{x} \in \Omega. \end{cases} \tag{1}$$

Then, we consider two kinds of boundary conditions,which are periodic boundary conditions and homogeneous Neumann boundary conditions, for the aforementioned equation. Here, $\mathcal{M}$ is a positive kinetic coefficient. $\Omega = [a,b]^n (n = 1,2,3)$ and T denote a bounded domain in $\mathbb{R}^n$ and finite time, respectively. The function u is defined as the difference between the concentrations of the two components in a mixture, and ϵ is a small parameter that describes the thickness of the phase boundary. Meanwhile, we assume the initial data $u_0 \in [0,1]$; it follows from the maximum principe that $u \in [0,1]$ in [12]. The potential $F(u) = \frac{1}{2}u^2(1-u)^2$ is a double-well potential with equal well depth, taking its global minimum value at $u = 0$ and $u = 1$. In other words, the values 0 and 1 of the function u prevail in Ω, whereas the transition between them forms a thin interface layer. And we write $F'(u)$ to indicate the derivative of $F(u)$ with respect to u. Without lose of generality, we assume $\mathcal{M} = 1$.

The term

$$\beta(u,t) = \frac{\sqrt{2F(u)}}{\int_\Omega \sqrt{2F(u)}d\mathbf{x}} \int_\Omega F'(u)d\mathbf{x}, \tag{2}$$

School of Mathematics Science, Huaqiao University, Quanzhou 362021, China
** Correspondence to: Zhifeng Weng, School of Mathematics Science, Huaqiao University, Quanzhou 362021, China.*
† E-mail: zfwmath@163.com.

can be understood as a nonlocal and local Lagrange multiplier for the mass constraint

$$\frac{d}{dt}M(u) = 0$$

with

$$M(u) = \int_\Omega u(\mathbf{x}, t)d\mathbf{x}. \tag{3}$$

Another important feature of the nonlocal Allen–Cahn equation (1) is that it can be viewed as the L^2-gradient flow of the Lyapunov energy functional [11]

$$E(u) = \int_\Omega \left(\frac{1}{2}|\nabla u|^2 + \frac{1}{\epsilon^2}\left[F'(u) - \int_0^u \beta(s)ds\right]\right)d\mathbf{x} \tag{4}$$

and the total energy $E(u)$ is nonincreasing in time because

$$\frac{d}{dt}E(u) = -\int_\Omega \left(\frac{\partial u}{\partial t}\right)^2 d\mathbf{x} \le 0.$$

Compared with large amount of studies for the classical Allen–Cahn equation, there are few numerical results on non-local Allen–Cahn equation. The nonlocal Allen–Cahn equation [10] was first proposed the by Rubinsteinand Sternberg in 1992. In their paper, the nonlocal term is as follows:

$$\beta(u, t) = \frac{1}{|\Omega|}\int_\Omega F'(u)d\mathbf{x}, \tag{5}$$

which is related to the mean-curvature flow with the constraint of constant volume enclosed by the evolving curve. In [13], the axisymmetric case was studied to provide information on the singular-limit behavior of the selected level set. As described, for example, in [14], problem (1) with (5) is closely related to the diffuse interface models of phase transitions. Then this mass-conserving Allen–Cahn equation with nonlocal Lagrange multiplier was widely studied in [15–18]. However, this model is difficult to keep small features (nonzero regions whose area/volume is small compared with the total area/volume of the domain) because they dissolve into the bulk region because mass conservation is realized by a global correction using the time-dependent Lagrange multiplier (5). Then Brassel and Bretin [11] proposed the model (1) with the Lagrange multiplier (2); it seems that Eq. (2) has better volume preservation properties than Eq. (5), where the volume $|\Omega(t)|$ is defined as follows:

$$\Omega(t) = \left\{\mathbf{x} \in \Omega | u(\mathbf{x}, t) \ge \frac{1}{2}\right\}.$$

In this paper, we use the same definition.

Recently, for the mass-conserving Allen–Cahn equation with Lagrange multiplier (2), Alfaro and Alifrangis [12] performed formal asymptotic expansions of the solution and proved the convergence to a weak volume-preserving motion by mean curvature via energy estimates. Kim *et al*. [18] proposed a first-order operator splitting method, but they did not give convergence analysis of the operator splitting scheme. Very recently, for solving the similar problem, Zhai *et al*. [19] investigated some numerical methods, including the Crank–Nicolson finite difference method and the finite difference/spectral operator splitting method. Based on the aforementioned studies, we will extend the second-order operator splitting scheme and Fourier spectral method to the non-local Allen–Cahn equation. Furthermore, we will strictly prove the maximum principle of the method with the nonlocal term (2) and analyze the convergence rate of the fast explicit operator splitting (FEOS) method.

In our work, we concentrate mainly on the maximum principle and the convergence analysis for the mass-conserving Allen–Cahn equation (1) with the nonlocal term (2) based on a second-order operator splitting method and Fourier spectral method. First, we use Fourier spectral method in spatial discretization for the linear homogeneous heat equation and use the central difference scheme in time for the Lagrange multiplier equation. The nonlinear part was solved analytically. Second, we analyze the discrete maximum principle of the FEOS method, which guarantees the stability of the second-order operator splitting scheme. And we study the error estimate of the FEOS method. Our method is a generalization of the second-order operator splitting schemes for three subequations.

The rest of the paper is organized as follows. In Section 2, we propose a second-order FEOS method for problem (1). In Section 3, the discrete maximum principle and convergence analysis are given for this method. Then the numerical results confirming the accuracy and efficiency of the proposed method are presented in Section 4. Finally, conclusions are drawn in Section 5.

2. A fast explicit operator splitting method

In this section, we present an FEOS method to simulate the asymptotic behavior of solution for Eq. (1). The proposed method is based on Fourier spectral method for the space and second-order Strang splitting scheme for the time. For simplicity, we only consider two-dimensional (2D) space. The 1D/3D case is defined analogously. In order to illustrate the main idea of the proposed method, we adopted the following definition.

Definition 2.1 ([20])
The 2D Laplacian $(-\Delta)$ has a complete set of orthonormal eigenfunctions φ_{pq} corresponding to eigenvalues λ_{pq} on the bounded region $\Omega = [a, b]^2$, that is, $(-\Delta)\varphi_{pq} = \lambda_{pq}\varphi_{pq}$. Let

$$\mathcal{U} = \left\{ u = \sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} \hat{u}_{pq}\varphi_{pq}, \quad \hat{u}_{pq} = \langle u, \varphi_{pq} \rangle, \quad \sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} (|\hat{u}_{pq}|^2 |\lambda_{pq}|) < \infty, \right\},$$

Then for any $u \in \mathcal{U}$, it is easy to have

$$(-\Delta)u = \sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} \lambda_{pq}\hat{u}_{pq}\varphi_{pq}, \tag{6}$$

where λ_{pq} and φ_{pq} will depend on the specified boundary conditions.

For periodic condition,

$$\begin{cases} \lambda_{pq} = \left(\frac{2\pi p}{b-a}\right)^2 + \left(\frac{2\pi q}{b-a}\right)^2, & \forall p, q. \\ \varphi_{pq} = \exp\left(\sqrt{-1}\frac{2\pi p(x-a)+2\pi q(y-a)}{b-a}\right), & \forall p, q. \\ \hat{u}_{pq}(t) = \frac{1}{(b-a)^2}\int_{\Omega} u\varphi_{pq} d\mathbf{x}, & \forall p, q. \end{cases} \tag{7}$$

Here, the imaginary unit is denoted by $\sqrt{-1}$.
For homogeneous Neumann condition,

$$\begin{cases} \lambda_{pq} = 0, \quad \varphi_{pq} = 0, & \forall p, q < 0. \\ \lambda_{pq} = \left(\frac{\pi p}{b-a}\right)^2 + \left(\frac{\pi q}{b-a}\right)^2, & \forall p, q \geq 0. \\ \varphi_{pq} = \frac{2}{b-a}\cos\left(\frac{\pi p(x-a)}{b-a}\right)\cos\left(\frac{\pi q(y-a)}{b-a}\right), & \forall p, q \geq 0. \\ \hat{u}_{pq}(t) = \frac{1}{C_p^{neu}C_q^{neu}}\int_{\Omega} u\varphi_{pq} d\mathbf{x}, & \forall p, q \geq 0. \end{cases} \tag{8}$$

where C_p^{neu} and C_q^{neu} are defined as

$$C_r^{neu} = \begin{cases} 2, & r = 0, \\ 1, & r > 0. \end{cases}$$

Now, we will introduce two spatial grids

$$\Omega_h^{per} = \left\{ (x_i, y_j) = \left(a + ih, a + jh\right), \quad 0 \leq i, j \leq N - 1) \right\},$$

$$\Omega_h^{neu} = \left\{ (x_i, y_j) = \left(a + (i + 0.5)h, a + (j + 0.5)h\right), \quad 0 \leq i, j \leq N - 1) \right\},$$

for periodic condition and homogeneous Neumann condition, respectively. Here, $h = (b-a)/N$. Let $u_{ij}^m = u(x_i, y_j, m\tau)$, where $\tau = T/M$ is the time step and M is the total number of time steps.

Based on the theory of spectral method [21], we use the FFT algorithm to compute the discrete Fourier coefficients $\{\hat{u}_{pq}(t)\}$ from the point values $\{u_{ij}(t)\}$ and approximate $\{u_{ij}(t)\}$ by its Fourier expansion:

$$\begin{cases} \hat{u}_{pq}(t) = \frac{h^2}{C_p^{per}C_q^{per}(b-a)^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} u_{ij}(t)\varphi_{pq}(x_i, y_j), \\ u_{ij}(t) = \sum_{p=-N/2}^{N/2} \sum_{q=-N/2}^{N/2} \hat{u}_{pq}(t)\varphi_{pq}(x_i, y_j), \end{cases} \quad \text{on} \quad \Omega_h^{per} \tag{9}$$

and

$$\begin{cases} \hat{u}_{pq}(t) = \frac{h^2}{C_p^{neu}C_q^{neu}} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} u_{ij}(t)\varphi_{pq}(x_i, y_j), \\ u_{ij}(t) = \sum_{p=0}^{N-1} \sum_{q=0}^{N-1} \hat{u}_{pq}(t)\varphi_{pq}(x_i, y_j), \end{cases} \quad \text{on} \quad \Omega_h^{neu}, \tag{10}$$

where C_p^{per}, C_q^{per} and C_p^{neu}, C_q^{neu} are respectively defined as

$$C_r^{per} = \begin{cases} 2, & |r| = \frac{N}{2}, \\ 1, & |r| < \frac{N}{2}, \end{cases}$$

and

$$C_r^{neu} = \begin{cases} 2, & \text{mod } (r, N) = 0, \\ 1, & \text{otherwise.} \end{cases}$$

2.1. *The fast explicit operator splitting method*

Now, we use an operator splitting method to split the original problem (1), which can be briefly described as follows. We first denote by $\mathcal{S}^A$, $\mathcal{S}^B$, and $\mathcal{S}^C$ the exact solution operators associated with the nonlinear equation

$$\mathcal{S}^A: \quad u_t = -\frac{1}{\epsilon^2} F'(u), \tag{11}$$

the heat equation

$$\mathcal{S}^B: \quad u_t = \Delta u, \tag{12}$$

and the Lagrange multiplier equation

$$\mathcal{S}^C: \quad u_t = \frac{1}{\epsilon^2} \beta(u, t). \tag{13}$$

Thus, we can evolve the solution from t to $t + \tau$ using the following second-order Strang splitting method [22]:

$$u(x, y, t + \tau) = \mathcal{S}^A\left(\frac{\tau}{2}\right) \mathcal{S}^B\left(\frac{\tau}{2}\right) \mathcal{S}^C(\tau) \mathcal{S}^B\left(\frac{\tau}{2}\right) \mathcal{S}^A\left(\frac{\tau}{2}\right) u(x, y, t) + \mathcal{O}(\tau^3). \tag{14}$$

Higher-order splitting schemes can be found in [23, 24]. Next, the exact solution operators $\mathcal{S}^A$, $\mathcal{S}^B$, and $\mathcal{S}^C$ are replaced by their numerical approximations $\mathcal{S}_h^A$, $\mathcal{S}_h^B$, and $\mathcal{S}_h^C$.

Equation (11) is solved analytically, whose solution $u^{m+1,(1)}$ is given as follows [18]:

$$\mathcal{S}_h^A(\tau): \quad u^{m+1,(1)} = \frac{1}{2} - \frac{1 - 2u^m}{2\sqrt{(1 - 2u^m)^2 + 4u^m(1 - u^m)e^{-2\tau/\epsilon^2}}}. \tag{15}$$

Then, by combining with Eq. (6), Eq. (12) for the (pq)th Fourier mode simply becomes

$$\frac{\partial \hat{u}_{pq}}{\partial t} = -\lambda_{pq} \hat{u}_{pq}. \tag{16}$$

By the method of separation of variables [25], the aforementioned equation can be solved analytically with an initial condition $\hat{u}_{pq}^{m+1,(1)}$, and the solution $\hat{u}_{pq}^{m+1,(2)}$ is given as follows:

$$\hat{u}_{pq}^{m+1,(2)} = e^{-\lambda_{pq}\tau} \hat{u}_{pq}^{m+1,(1)}. \tag{17}$$

Hence, we have

$$\mathcal{S}_h^B(\tau): \quad u^{m+1,(2)} = \mathcal{F}^{-1}\left[e^{-\lambda_{pq}\tau} \mathcal{F}(u^{m+1,(1)})(p, q)\right],$$

where $\mathcal{F}$ denotes the discrete cosine transform and $\mathcal{F}^{-1}$ its inverse.

Finally, we discuss the Lagrange multiplier equation (13), which can be written as

$$u_t = \frac{1}{\epsilon^2} \frac{\int_\Omega F'(u) d\mathbf{x}}{\int_\Omega \sqrt{2F(u)} d\mathbf{x}} \sqrt{2F(u)} = S_F(u) \sqrt{2F(u)},$$

where $S_F(u) = \frac{1}{\epsilon^2} \frac{\int_\Omega F'(u) d\mathbf{x}}{\int_\Omega \sqrt{2F(u)} d\mathbf{x}}$. Because $u \in [0, 1]$, the aforementioned equation can be recast as

$$u_t = S_F(u)\, u(1 - u). \tag{18}$$

Then based on the central difference scheme, a stable and novel second-order discretization of Eq. (18) is presented as follows:

$$\frac{u^{m+1} - u^{m+1,(2)}}{\tau} = S_F\left(\frac{u^{m+1,(2)} + u^{m+1}}{2}\right)\left[\frac{u^{m+1,(2)} + u^{m+1} - 2u^{m+1,(2)}u^{m+1}}{2}\right]. \tag{19}$$

Summing both sides of the aforementioned formula over i and j, we obtain

$$S_F\left(\frac{u^{m+1,(2)} + u^{m+1}}{2}\right)$$
$$= \frac{1}{\tau} \sum_{i,j=0}^{N-1} \left(u_{ij}^{m+1} - u_{ij}^{m+1,(2)}\right) \Big/ \sum_{i,j=0}^{N-1} \left[\frac{u_{ij}^{m+1,(2)} + u_{ij}^{m+1} - 2u_{ij}^{m+1,(2)}u_{ij}^{m+1}}{2}\right].$$

By the property of mass conservation $\sum_{i,j=0}^{N-1} u_{ij}^0 = \sum_{i,j=0}^{N-1} u_{ij}^{m+1}$, we have

$$\begin{aligned} &S_F\left(\frac{u^{m+1,(2)}+u^{m+1}}{2}\right) \\ &=\frac{1}{\tau}\sum_{i,j=0}^{N-1}\left(u_{ij}^0-u_{ij}^{m+1,(2)}\right)\Big/\sum_{i,j=0}^{N-1}\left[\frac{u_{ij}^{m+1,(2)}+u_{ij}^0-2u_{ij}^{m+1,(2)}u_{ij}^{m+1}}{2}\right]. \end{aligned} \tag{20}$$

The similar idea has also been mentioned in [18].

However, the aforementioned equation could not provide explicit scheme to handle $S_F(u)$. Thus, a modified form $S_F^{m+1,(2)}$ of Eq. (20) is proposed:

$$S_F^{m+1,(2)} = \frac{1}{\tau}\sum_{i,j=0}^{N-1}\left(u_{ij}^0-u_{ij}^{m+1,(2)}\right)\Big/\sum_{i,j=0}^{N-1}\left[\frac{u_{ij}^{m+1,(2)}+u_{ij}^0-2\left(u_{ij}^{m+1,(2)}\right)^2}{2}\right], \tag{21}$$

with

$$S_F\left(\frac{u^{m+1,(2)}+u^{m+1}}{2}\right)-S_F^{m+1,(2)}=\mathcal{O}(\tau^2). \tag{22}$$

In fact, combining Eq. (20) with Eq. (21), we obtain immediately

$$\begin{aligned} &S_F\left(\frac{u^{m+1,(2)}+u^{m+1}}{2}\right)-S_F^{m+1,(2)} \\ &=\frac{1}{\tau}\sum_{i,j=0}^{N-1}\left(u_{ij}^0-u_{ij}^{m+1,(2)}\right)\frac{\sum_{i,j=0}^{N-1}u_{ij}^{m+1,(2)}\left(u_{ij}^{m+1,(2)}-u_{ij}^{m+1}\right)}{\sum_{i,j=0}^{N-1}\left[\frac{u_{ij}^{m+1,(2)}+u_{ij}^0-2u_{ij}^{m+1,(2)}u_{ij}^{m+1}}{2}\right]\sum_{i,j=0}^{N-1}\left[\frac{u_{ij}^{m+1,(2)}+u_{ij}^0-2\left(u_{ij}^{m+1,(2)}\right)^2}{2}\right]} \\ &=\frac{\sum_{i,j=0}^{N-1}\left(u_{ij}^0-u_{ij}^{m+1,(2)}\right)}{\sum_{i,j=0}^{N-1}\left[\frac{u_{ij}^{m+1,(2)}+u_{ij}^0-2u_{ij}^{m+1,(2)}u_{ij}^{m+1}}{2}\right]\sum_{i,j=0}^{N-1}\left[\frac{u_{ij}^{m+1,(2)}+u_{ij}^0-2\left(u_{ij}^{m+1,(2)}\right)^2}{2}\right]}\sum_{i,j=0}^{N-1}u_{ij}^{m+1,(2)}\frac{u_{ij}^{m+1,(2)}-u_{ij}^{m+1}}{\tau} \\ &=\mathcal{O}(\tau^2). \end{aligned}$$

Thus the estimate (22) is proved.

Substituting $S_F^{m+1,(2)}$ for $S_F\left(\frac{u^{m+1,(2)}+u^{m+1}}{2}\right)$, after rearrangement of terms, then the scheme (19) simply becomes

$$\mathcal{S}_h^C(\tau):\quad u^{m+1}=\frac{\left(1+\frac{\tau S_F^{m+1,(2)}}{2}\right)u^{m+1,(2)}}{1+\frac{\tau S_F^{m+1,(2)}}{2}\left(2u^{m+1,(2)}-1\right)}. \tag{23}$$

Now, an FEOS method for Eq. (1) is proposed as follows:

$$u^{m+1,(1)}=\frac{1}{2}-\frac{1-2u^m}{2\sqrt{(1-2u^m)^2+4u^m(1-u^m)e^{-\tau/\epsilon^2}}},$$

$$u^{m+1,(2)}=\mathcal{F}^{-1}\left[e^{-\lambda_{pq}\tau/2}\mathcal{F}[u^{m+1,(1)}](p,q)\right],$$

$$u^{m+1,(3)}=\frac{\left(1+\frac{\tau S_F^{m+1,(2)}}{2}\right)u^{m+1,(2)}}{1+\frac{\tau S_F^{m+1,(2)}}{2}(2u^{m+1,(2)}-1)},$$

$$u^{m+1,(4)}=\mathcal{F}^{-1}\left[e^{-\lambda_{pq}\tau/2}\mathcal{F}[u^{m+1,(3)}](p,q)\right],$$

$$u^{m+1}=\frac{1}{2}-\frac{1-2u^{m+1,(4)}}{2\sqrt{(1-2u^{m+1,(4)})^2+4u^m(1-u^{m+1,(4)})e^{-\tau/\epsilon^2}}}.$$

Our proposed scheme combines with a half-time nonlinear equation solver, a half-time linear homogeneous heat equation solver, a full-time midpoint solver for the Lagrange multiplier equation, and a half-time linear homogeneous heat equation solver followed by a final half-time nonlinear equation solver.

The integral term

$$S_F(u) = \frac{1}{\epsilon^2}\frac{\int_\Omega F'(u)dx}{\int_\Omega \sqrt{2F(u)}dx}$$

is used in the aforementioned derivation. From the expression of $S_F(u)$, we know that it is only dependent on time t, for a fixed time, such as $t = t_m$, $S_F(u^m)$ is a constant C, which maybe different at different time. Furthermore, combining the expression of $F(u)$, for any $u \in [0,1]$, we have

$$\begin{aligned}|S_F(u)| &= \frac{1}{\epsilon^2}\left|\frac{\int_\Omega F'(u)d\mathbf{x}}{\int_\Omega \sqrt{2F(u)}d\mathbf{x}}\right| = \frac{1}{\epsilon^2}\left|\frac{\int_\Omega u(1-u)(1-2u)d\mathbf{x}}{\int_\Omega u(1-u)d\mathbf{x}}\right| \\ &\le \frac{1}{\epsilon^2}\max_{u\in[0,1]}|1-2u|\frac{\int_\Omega u(1-u)d\mathbf{x}}{\int_\Omega u(1-u)d\mathbf{x}} = \frac{1}{\epsilon^2},\end{aligned}$$

which means that $|S_F(u)|$ is bounded.

3. The discrete maximum principle and convergence analysis

In this section, the discrete maximum principle and convergence analysis of the method FEOS are discussed. First, we will prove that if the initial value $u_0 \in [0,1]$, then the entire solution is also bounded by 0 and 1, that is,

$$u(x,y,t) \in [0,1], \quad \forall t > 0.$$

This is also called maximum principle for Eq. (1). Now, we deduce the following discrete maximum principle, and the corresponding proof is also given.

Theorem 3.1

If the initial value satisfies $u_0 \in [0,1]$, then numerical solution of the problem (1) obtained by method FEOS is uniformly bounded for any $\tau > 0$, that is,

$$u^{m+1} \in [0,1], \qquad \forall m \ge 0.$$

Proof

For convenience, we only give the stability of the method FEOS with homogeneous Neumann boundary condition; the case of periodic ones can be discussed similarly.

Using the monotonicity of function and assuming $u^m \in [0,1]$, it is easy to see that

$$0 \le \frac{1}{2} - \frac{1-2u^m}{2\sqrt{(1-2u^m)^2 + 4u^m(1-u^m)e^{-\tau/\epsilon^2}}} \le 1, \tag{24}$$

that is,

$$0 \le u^{m+1,(1)} \le 1. \tag{25}$$

Combining Definition 2.1 and the Cauchy–Schwarz inequality, it follows from the second formula of method FEOS that

$$\begin{aligned}|u^{m+1,(2)}| &= \left|\sum_{p,q=0}^{N-1} e^{-\lambda_{pq}\frac{\tau}{2}}\hat{u}_{pq}^{m+1,(1)}\varphi_{pq}\right| \\ &\le \left|\sum_{p,q=0}^{N-1} e^{-\lambda_{ij}\frac{\tau}{2}}\right|\left|\sum_{p,q=0}^{N-1}\hat{u}_{pq}^{m+1,(1)}\varphi_{pq}\right| = |u^{m+1,(1)}|\sum_{p,q=0}^{N-1} e^{-\lambda_{pq}\frac{\tau}{2}}.\end{aligned} \tag{26}$$

Suppose that $b-a = \pi$; then $\lambda_{pq} = p^2 + q^2$. Thus, we have

$$\sum_{p,q=0}^{N-1} e^{-\lambda_{pq}\frac{\tau}{2}} = \left(\sum_{p=0}^{N-1} e^{-p^2\frac{\tau}{2}}\right)^2 \le \left(\int_0^1 \omega^{\frac{\tau}{2}}d\omega\right)^2 \le 1,$$

where $\omega = e^{-x^2}$ with $x \in [0,+\infty)$.

Combining the aforementioned equation, it follows from Eq. (25) that

$$|u^{m+1,(2)}| \le |u^{m+1,(1)}| \le 1. \tag{27}$$

Next, we will further verify $u^{m+1,(2)} \geq 0$. Because

$$\begin{aligned} u^{m+1,(2)} &= \sum_{p,\,q=0}^{N-1} e^{-\lambda_{pq}\frac{\tau}{2}} \hat{u}_{pq}^{m+1,(1)} \varphi_{pq} \\ &= \sum_{p,\,q=0}^{N-1} \left(1 - \lambda_{pq}\frac{\tau}{2} + \frac{1}{2!}\left(\lambda_{pq}\frac{\tau}{2}\right)^2 - \frac{1}{3!}\left(\lambda_{pq}\frac{\tau}{2}\right)^3 + \cdots\right) \hat{u}_{pq}^{m+1,(1)} \varphi_{pq} \\ &= \left[I + \frac{\tau}{2}\Delta + \frac{1}{2!}\left(\frac{\tau}{2}\right)^2 \Delta^2 + \frac{1}{3!}\left(\frac{\tau}{2}\right)^3 \Delta^3 + \cdots\right] u^{m+1,(1)}, \end{aligned} \tag{28}$$

where I denotes the unit operator. Then we have

$$\begin{aligned} &\left(u^{m+1,(1)}, u^{m+1,(2)}\right) = \left(u^{m+1,(1)}, u^{m+1,(1)}\right) + \frac{\tau}{2}\left(u^{m+1,(1)}, \Delta u^{m+1,(1)}\right) + \\ &\frac{1}{2!}\left(\frac{\tau}{2}\right)^2 \left(u^{m+1,(1)}, \Delta^2 u^{m+1,(1)}\right) + \frac{1}{3!}\left(\frac{\tau}{2}\right)^3 \left(u^{m+1,(1)}, \Delta^3 u^{m+1,(1)}\right) + \cdots. \end{aligned} \tag{29}$$

It is well known that operator Δ is a negative definite operator in L^2-norm in the sense $(u^{m+1,(1)}, \Delta u^{m+1,(1)}) = \gamma \parallel u^{m+1,(1)} \parallel_0^2$ with $\gamma < 0$. So it follows from Eq. (29)

$$\begin{aligned} (u^{m+1,(1)}, u^{m+1,(2)}) &= \left(1 + \frac{\tau}{2}\gamma + \frac{1}{2!}\left(\frac{\tau}{2}\gamma\right)^2 + \frac{1}{3!}\left(\frac{\tau}{2}\gamma\right)^3 + \cdots\right) \parallel u^{m+1,(1)} \parallel_0^2 \\ &= e^{\frac{\tau}{2}\gamma} \parallel u^{m+1,(1)} \parallel_0^2 \geq 0. \end{aligned} \tag{30}$$

that is,

$$\left(u^{m+1,(1)}, u^{m+1,(2)}\right) \geq 0, \quad \text{for any} \quad u^{m+1,(1)} \geq 0.$$

Thus, it is easy to obtain

$$u^{m+1,(2)} \geq 0.$$

Above all,

$$u^{m+1,(2)} \in [0, 1]. \tag{31}$$

Now, we discuss the third formula of the method FEOS. From Remark 2.1, we know that $S_F^{m+1,(2)}$ is a constant at t_{m+1}. Then combining the monotonicity of function $g(x) = \left[(1 + \frac{\tau S_F^{m+1,(2)}}{2})x\right] \Big/ \left[1 + \frac{\tau S_F^{m+1,(2)}}{2}(2x-1)\right]$ with $x \in [0, 1]$, we find that it easily leads to

$$0 \leq \frac{\left(1 + \frac{\tau S_F^{m+1,(2)}}{2}\right) u^{m+1,(2)}}{1 + \frac{\tau S_F^{m+1,(2)}}{2}(2u^{m+1,(2)} - 1)} \leq 1,$$

that is,

$$0 \leq u^{m+1,(3)} \leq 1. \tag{32}$$

Finally, using the similar idea in Eqs (31) and (24), we have

$$u^{m+1,(4)} \in [0, 1], \quad \text{and then} \quad u^{m+1} \in [0, 1].$$

This completes the proof. □

Thus, the stability of second-order operator splitting schemes is guaranteed by the discrete maximum principle.

In the following, we present the main convergence result. Let us start by introducing some notations and some useful results. In the following, we use C to denote a generic positive constant whose value may change from place to place but remains independent of the mesh parameter. Meanwhile, we will restrict ourselves to the homogeneous Neumann condition. For the periodic condition, we can use the similar idea to finish the proof.

Dropping the truncated term $\mathcal{O}(\tau^3)$ in scheme (14), then let

$$\underline{u}(x, y, t_m) \quad \text{and} \quad U^m$$

denote the exact solution of scheme (14) and the numerical solution of the method FEOS at t_m. Next we define a grid function space on Ω_h^{per}

$$\mathcal{W}^h = \left\{U \middle| U = \{u_{ij} \mid 0 \leq i, j \leq N-1\}\right\}.$$

and a projection mapping $I^h : H^2_{neu}(\Omega) \longrightarrow \mathcal{W}_h$ by

$$I^h(u) = U,$$

where $H^2_{neu}(\Omega) = \{u \in H^2(\Omega) \mid \nabla u(\mathbf{x}) \cdot n_\Omega = 0\}$.

Lemma 3.1
For any grid function $U \in \mathcal{W}^h$ with $U = \{u_{ij} \in [0,1]\}$, it holds that

$$\| \mathcal{S}^A_h U \| \leq e^{\tau/\epsilon^2} \| U \| .$$

Proof
By (15), we have

$$\begin{aligned} \| \mathcal{S}^A_h U \| &= \left\| \frac{1}{2} - \frac{1-2U}{2\sqrt{(1-2U)^2 + 4U(1-U)e^{-2\tau/\epsilon^2}}} \right\| \\ &= \left\| \frac{\sqrt{(1-2U)^2 + 4U(1-U)e^{-2\tau/\epsilon^2}} - (1-2U)}{2\sqrt{(1-2U)^2 + 4U(1-U)e^{-2\tau/\epsilon^2}}} \right\|. \end{aligned}$$

For $x \in [0,1]$, using the fact that

$$0 \leq \sqrt{(1-2x)^2 + 4x(1-x)e^{-2\tau/\epsilon^2}} - (1-2x) \leq 2x$$

and the monotonicity of function

$$g(x) = (1-2x)^2 + 4x(1-x)e^{-2\tau/\epsilon^2} \Longrightarrow g(x) \geq g\left(\frac{1}{2}\right) = e^{-2\tau/\epsilon^2},$$

we have

$$\| \mathcal{S}^A_h U \| \leq \left\| \frac{2U}{2e^{-\tau/\epsilon^2}} \right\| = e^{\tau/\epsilon^2} \| U \| .$$

This end the proof. □

Lemma 3.2
For any grid function $U \in \mathcal{W}^h$, it holds that

$$\| \mathcal{S}^B_h(t) U \| \leq \| U \| .$$

Proof
Using the Parseval's formula and the fact that $|e^{-\lambda_{pq}\tau}| \leq 1 (\forall p, q)$, we can obtained easily the required inequality. The proof is completed. □

Lemma 3.3
For any functions $u_0 \in H^s_{Neu}(\Omega)$ with $s \in \mathbb{N}$ and $s > 1$, it holds that

$$\| I^h \mathcal{S}^B(t) u_0 - \mathcal{S}^B_h(t) I^h u_0 \| \leq C h^s |u_0|_s,$$

where $C > 0$ is a constant independent of τ and h.

Proof
For convenience of exposition, we introduce some notations as follows:

$$\begin{cases} u(x,y,t) = \sum\limits_{p=0}^{\infty} \sum\limits_{q=0}^{\infty} \hat{u}_{pq}(t)\, \varphi_{pq}(x,y), \\ \bar{u}(x,y,t) = \sum\limits_{p=0}^{\infty} \sum\limits_{q=0}^{N-1} \hat{u}_{pq}(t)\, \varphi_{pq}(x,y), \\ \bar{\bar{u}}(x,y,t) = \sum\limits_{p=0}^{N-1} \sum\limits_{q=0}^{N-1} \hat{u}_{pq}(t)\, \varphi_{pq}(x,y), \\ \tilde{u}(x,y,t) = \sum\limits_{p=0}^{N-1} \sum\limits_{q=0}^{N-1} \tilde{u}_{pq}(t)\, \varphi_{pq}(x,y), \end{cases} \tag{33}$$

where $\hat{u}_{pq}(t)$ and $\tilde{u}_{pq}(t)$ are given by Eqs (8) and (10), respectively.

Assuming $u(x,y,0) = \bar{u}(x,y,0) = \bar{\bar{u}}(x,y,0) = \tilde{u}(x,y,0) = u_0(x,y)$, then we have

$$\begin{aligned} \| I^h \mathcal{S}^B u_0 - \mathcal{S}^B_h I^h u_0 \|^2 &= h^2 \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} |u(x_i, y_j, \tau) - \tilde{u}(x_i, y_j, \tau)|^2 \\ &\leq 4A_1 + 4A_2 + 2B, \end{aligned} \tag{34}$$

where

$$\begin{cases} A_1 = h^2 \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} |u(x_i,y_j,\tau) - \bar{u}(x_i,y_j,\tau)|^2, \\ A_2 = h^2 \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} |\bar{u}(x_i,y_j,\tau) - \bar{\bar{u}}(x_i,y_j,\tau)|^2, \\ B = h^2 \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} |\bar{\bar{u}}(x_i,y_j,\tau) - \tilde{u}(x_i,y_j,\tau)|^2. \end{cases}$$

We first consider the term A_1. From Eq. (33), we have

$$\begin{aligned} A_1 &= h^2 \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} \left| \sum_{p=0}^{\infty}\sum_{q\geq N}^{\infty} \hat{u}_{pq}(\tau)\, \varphi_{pq}(x_i,y_j) \right|^2 \\ &= \sum_{p=0}^{\infty}\sum_{q\geq N}^{\infty} C_p^{neu} C_q^{neu} |\hat{u}_{pq}(\tau)|^2 \\ &= \sum_{p=0}^{\infty}\sum_{q\geq N}^{\infty} C_p^{neu} C_q^{neu} \left| \exp\left(-\lambda_{pq} t\right) \hat{u}_{pq}(0) \right|^2 \\ &\leq 4N^{-2s} \sum_{p=0}^{\infty}\sum_{q\geq N}^{\infty} q^{2s} |\hat{u}_{pq}(0)|^2. \end{aligned}$$

Similarly, we have

$$\begin{aligned} A_2 &= h^2 \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} \left| \sum_{p\geq N}^{\infty}\sum_{q=0}^{N-1} \hat{u}_{pq}(\tau)\, \varphi_{pq}(x_i,y_j) \right|^2 \\ &\leq 4N^{-2s} \sum_{p\geq N}^{\infty}\sum_{q=0}^{N-1} p^{2s} |\hat{u}_{pq}(0)|^2. \end{aligned}$$

Thus, a direct calculation leads to

$$\begin{aligned} A_1 + A_2 &\leq 4N^{-2s} \left\{ \sum_{p=0}^{\infty}\sum_{q\geq N}^{\infty} q^{2s} |\hat{u}_{pq}(0)|^2 + \sum_{p\geq N}^{\infty}\sum_{q=0}^{N-1} p^{2s} |\hat{u}_{pq}(0)|^2 \right\} \\ &< 4N^{-2s} \sum_{p=0}^{\infty}\sum_{q=0}^{\infty} (p^{2s} + q^{2s}) |\hat{u}_{pq}(0)|^2. \end{aligned}$$

Because the semi-norm $|\cdot|_s$ of $H^s(\Omega)$ can be characterized in the frequency space, we have

$$A_1 + A_2 \leq 4N^{-2s} |u_0|_s^2.$$

It is time to estimate the term B. From Eq. (33), we have

$$\begin{aligned} B &= h^2 \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} \left| \sum_{p=0}^{N-1}\sum_{q=0}^{N-1} \left(\hat{u}_{pq}(\tau) - \tilde{u}_{pq}(\tau) \right) \varphi_{pq}(x_i,y_j) \right|^2 \\ &= \sum_{p=0}^{N-1}\sum_{q=0}^{N-1} C_p^{neu} C_q^{neu} |\hat{u}_{pq}(\tau) - \tilde{u}_{pq}(\tau)|^2 \\ &= \sum_{p=0}^{N-1}\sum_{q=0}^{N-1} C_p^{neu} C_q^{neu} \left| \exp\left(-\lambda_{pq} t\right) \left(\hat{u}_{pq}(0) - \tilde{u}_{pq}(0) \right) \right|^2 \\ &\leq 4 \sum_{p=0}^{N-1}\sum_{q=0}^{N-1} \left| \hat{u}_{pq}(0) - \tilde{u}_{pq}(0) \right|^2. \end{aligned}$$

Before proceeding further, we first prove that

$$\tilde{u}_{pq}(t) = \hat{u}_{pq}(t) + \sum_{l_1=1}^{\infty}\sum_{l_2=1}^{\infty} (-1)^{l_1+l_2} \hat{u}_{2l_1N,\, 2l_2N}(t).$$

In fact, from Eq. (10), we have

$$\begin{aligned}\tilde{u}_{pq}(t) &= \frac{h^2}{c_p c_q}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} u_{ij}(t)\varphi_{pq}(x_i,y_j)\\ &= \frac{h^2}{c_p c_q}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left(\sum_{r_1=0}^{\infty}\sum_{r_2=0}^{\infty}\hat{u}_{r_1,r_2}(t)\varphi_{r_1,r_2}(x_i,y_j)\right)\varphi_{pq}(x_i,y_j)\\ &= \frac{h^2}{c_p c_q}\sum_{r_1=0}^{\infty}\sum_{r_2=0}^{\infty}\hat{u}_{r_1,r_2}(t)\left(\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\varphi_{r_1,r_2}(x_i,y_j)\varphi_{pq}(x_i,y_j)\right)\\ &= \frac{h^2}{c_p c_q}\sum_{l_1=0}^{\infty}\sum_{l_2=0}^{\infty}\hat{u}_{2l_1N,2l_2N}(t)\left(c_p c_q(-1)^{l_1+l_2}\left(\frac{2}{b-a}\right)^2\left(\frac{N}{2}\right)^2\right)\\ &= \hat{u}_{pq}(t)+\sum_{l_1=1}^{\infty}\sum_{l_2=1}^{\infty}(-1)^{l_1+l_2}\hat{u}_{2l_1N,\,2l_2N}(t).\end{aligned}$$

Hence, a combination of the aforementioned estimates leads to

$$\begin{aligned}B &\le 4\sum_{p=0}^{N-1}\sum_{q=0}^{N-1}\left|\sum_{l_1=1}^{\infty}\sum_{l_2=1}^{\infty}\hat{u}_{2l_1N,\,2l_2N}(t)\right|^2\\ &\le 4\sum_{p=0}^{N-1}\sum_{q=0}^{N-1}\left\{\left(\sum_{l_1=1}^{\infty}\sum_{l_2=1}^{\infty}\left[(p+2l_1N)^2+(q+2l_2N)^2\right]^{-s}\right)\right.\\ &\quad\left.\left(\sum_{l_1=1}^{\infty}\sum_{l_2=1}^{\infty}\left[(p+2l_1N)^2+(q+2l_2N)^2\right]^{s}|\hat{u}_{2l_1N,\,2l_2N}(0)|^2\right)\right\}\\ &\le 4\max_{0\le p,q\le N-1}\left(\sum_{l_1=1}^{\infty}\sum_{l_2=1}^{\infty}\left[(p+2l_1N)^2+(q+2l_2N)^2\right]^{-s}\right)\\ &\quad\left(\sum_{p=0}^{N-1}\sum_{q=0}^{N-1}\sum_{l_1=1}^{\infty}\sum_{l_2=1}^{\infty}\left[(p+2l_1N)^2+(q+2l_2N)^2\right]^{s}|\hat{u}_{2l_1N,\,2l_2N}(0)|^2\right).\end{aligned}$$

It is clear that

$$\begin{aligned}&\max_{0\le p,q\le N-1}\left(\sum_{l_1=1}^{\infty}\sum_{l_2=1}^{\infty}\left[(p+2l_1N)^2+(q+2l_2N)^2\right]^{-s}\right)\\ \le& N^{-2s}\sum_{l_1=1}^{\infty}\sum_{l_2=1}^{\infty}\frac{1}{\left[(2l_1)^2+(2l_2)^2\right]^s}\\ \le&\frac{1}{2}N^{-2s}\left(\sum_{l_1=1}^{\infty}\frac{1}{(2l_1)^s}\right)\left(\sum_{l_2=1}^{\infty}\frac{1}{(2l_2)^s}\right)\le\frac{1}{2^7}N^{-2s},\end{aligned}$$

and

$$\sum_{p=0}^{N-1}\sum_{q=0}^{N-1}\sum_{l_1=1}^{\infty}\sum_{l_2=1}^{\infty}\left[(p+2l_1N)^2+(q+2l_2N)^2\right]^{s}|\hat{u}_{2l_1N,\,2l_2N}(0)|^2\le 2|u|_s^2.$$

Thus, we have

$$B\le 2^{-4}N^{-2s}|u_0|_s^2.$$

Based on the aforementioned arguments, we obtain

$$\|\, I^h\mathcal{S}_{\mathcal{L}}u-\mathcal{S}^h_{\mathcal{L}}I^h u\,\|^2\le CN^{-2s}|u_0|_s^2.$$

The proof is completed. □

Lemma 3.4

For any functions $u_1(x,y), u_2(x,y)\in H^2_{neu}(\Omega)$, and $u_1(x,y), u_2(x,y)\in[0,1]$, it holds that

$$\|\,\mathcal{S}^C(t)u_1-\mathcal{S}^C(t)u_2\,\|_{L^2(\Omega)}\le\ e^{t/\epsilon^2}\ \|\,u_1-u_2\,\|_{L^2(\Omega)},$$

where $\|\cdot\|_{L^2(\Omega)}$ represents the L^2-norm.

Proof

Set $u(x,y,t)$ and $v(x,y,t)$ to be the solutions of (13) with the initial data $u(x,y,0) = u_1(x,y)$ and $v(x,y,0) = u_2(x,y)$, respectively. From Remark 2.1, we know $S_F(u)$ is only dependent on time t; thus, we denote $S_F(u)$ by $S_F(t)$. Then let $\omega = u - v$; by Eq. (13), we have

$$\omega_t = S_F(t)(1-u-v)\omega,$$

which means that

$$|\omega(t)| = |\omega(0)| \exp\left(\int_0^t S_F(s)(1-u-v)ds\right).$$

Furthermore, because $u \in [0,1]$, $v \in [0,1]$, and $|S_F(t)| \leq \frac{1}{\epsilon^2}$, it follows from the aforementioned equation that

$$|\omega(t)| \leq |\omega(0)|\, e^{t/\epsilon^2}.$$

The proof is completed. □

Lemma 3.5

For any functions $u(x,y) \in H^2_{neu}(\Omega)$, it holds that

$$\| I^h \mathcal{S}^C(\tau)u - \mathcal{S}_h^C(\tau) I^h u \| \leq C\tau^3.$$

Proof

Scheme (19) shows that this is a second-order central difference formula; thus, the truncation error of (23) is $\mathcal{O}(\tau^3)$. Then the validity of the aforementioned conclusion can be directly derived. The proof is completed. □

Lemma 3.6

For any grid function $U_1, U_2 \in \mathcal{W}^h$, it holds that

$$\| \mathcal{S}_h^C(\tau)U_1 - \mathcal{S}_h^C(\tau)U_2 \| \leq C(\tau^3 + h^s) + e^{\tau/\epsilon^2} \| U_1 - U_2 \|,$$

where $\| \cdot \|$ represents the discrete L^2-norm.

Proof

Assuming $\bar{u}, \bar{v} \in H^s_{neu}(\Omega)$ with $s > 1$ such that $I^h\bar{u} = U_1$ and $I^h\bar{v} = U_2$, where $H^s_{Neu}(\Omega) = \{u \in H^s(\Omega) \mid \nabla u(\mathbf{x}) \cdot n_\Omega = 0\}$. Then we have

$$\begin{aligned} &\| \mathcal{S}_h^C U_1 - \mathcal{S}_h^C U_2 \| \\ \leq &\| \mathcal{S}_h^C U_1 - I^h\mathcal{S}^C\bar{u} \| + \| I^h\mathcal{S}^C\bar{u} - I^h\mathcal{S}^C\bar{v} \| + \| I^h\mathcal{S}^C\bar{v} - \mathcal{S}_h^C U_2 \| \\ \leq &\, C_1\tau^3 + \| I^h\mathcal{S}^C\bar{u} - I^h\mathcal{S}^C\bar{v} \| + C_1\tau^3. \end{aligned}$$

Because the L^2-norm can be approximated by the discrete L^2-norm with spectral accuracy [21], by Lemma 3.1 it follows

$$\begin{aligned} \| I^h\mathcal{S}^C\bar{u} - I^h\mathcal{S}^C\bar{v} \| &\leq \| \mathcal{S}^C\bar{u} - \mathcal{S}^C\bar{v} \|_{L^2(\Omega)} + C_2h^s \\ &\leq e^{t/\epsilon^2} \| \bar{u} - \bar{v} \|_{L^2(\Omega)} + C_2h^s \\ &\leq e^{t/\epsilon^2}\left(\| U_1 - U_2 \|_{L^2(\Omega)} + C_3h^s \right) + C_2h^s. \end{aligned}$$

Combining the aforementioned equations, the conclusion of Lemma 3.2 would be obtained immediately. The proof is completed. □

Next,we will present the main convergence result.

Theorem 3.2

Suppose that $u^{m+1} \in H^s_{Neu}(\Omega)$ with $s > 1$ and U^{m+1} are the exact solution of Eq. (1) and the numerical solution of the proposed method, respectively. Then there exists a positive constant C independent of τ and h, such that

$$\| I^h u^{m+1} - U^{m+1} \| \leq C\left(\tau^2 + \frac{h^s}{\tau}\right).$$

Proof

For $m \geq 0$, we have

$$\| I^h u^{m+1} - U^{m+1} \| \leq \| I^h u^{m+1} - I^h\underline{u}^{m+1} \| + \| I^h\underline{u}^{m+1} - U^{m+1} \|.$$

Because the Strang splitting method is second order, the first term on the left-hand side of the aforementioned inequality is bounded by

$$\| I^h\underline{u}^{m+1} - I^h u^{m+1} \| \leq C_1\tau^2,$$

where C_1 is a positive constant independent of τ and h.

Then we consider the second term

$$\begin{aligned}
&\| I^h \underline{u}^{m+1} - U^{m+1} \| \\
=& \| I^h \mathcal{S}^A \mathcal{S}^B \mathcal{S}^C \mathcal{S}^B \mathcal{S}^A \underline{u}^m - \mathcal{S}_h^A \mathcal{S}_h^B \mathcal{S}_h^C \mathcal{S}_h^B \mathcal{S}_h^A U^m \| \\
\le& \| I^h \mathcal{S}^A \mathcal{S}^B \mathcal{S}^C \mathcal{S}^B \mathcal{S}^A \underline{u}^m - \mathcal{S}_h^A I^h \mathcal{S}^B \mathcal{S}^C \mathcal{S}^B \mathcal{S}^A \underline{u}^m \| \\
&+ \| \mathcal{S}_h^A I^h \mathcal{S}^B \mathcal{S}^C \mathcal{S}^B \mathcal{S}^A \underline{u}^m - \mathcal{S}_h^A \mathcal{S}_h^B \mathcal{S}_h^C \mathcal{S}_h^B \mathcal{S}_h^A U^m \| \\
=& \| \mathcal{S}_h^A I^h \mathcal{S}^B \mathcal{S}^C \mathcal{S}^B \mathcal{S}^A \underline{u}^m - \mathcal{S}_h^A \mathcal{S}_h^B \mathcal{S}_h^C \mathcal{S}_h^B \mathcal{S}_h^A U^m \| .
\end{aligned}$$

The last step is because $\mathcal{S}^A$ is solved analytically.

Using Lemmas 3.1–3.3 yields

$$\begin{aligned}
&\| \mathcal{S}_h^A I^h \mathcal{S}^B \mathcal{S}^C \mathcal{S}^B \mathcal{S}^A \underline{u}^m - \mathcal{S}_h^A \mathcal{S}_h^B \mathcal{S}_h^C \mathcal{S}_h^B \mathcal{S}_h^A U^m \| \\
\le& e^{\tau/\epsilon^2} \| I^h \mathcal{S}^B \mathcal{S}^C \mathcal{S}^B \mathcal{S}^A \tilde{u}^m - \mathcal{S}_h^B \mathcal{S}_h^C \mathcal{S}_h^B \mathcal{S}_h^A U^m \| \\
\le& e^{\tau/\epsilon^2} \Big[\| (I^h \mathcal{S}^B - \mathcal{S}_h^B I^h) \mathcal{S}^C \mathcal{S}^B \mathcal{S}^A \underline{u}^m \| + \| \mathcal{S}_h^B (I^h \mathcal{S}^C \mathcal{S}^B \mathcal{S}^A \underline{u}^m - \mathcal{S}_h^C \mathcal{S}_h^B \mathcal{S}_h^A U^m) \| \Big] \\
\le& e^{\tau/\epsilon^2} C_2 h^s |\mathcal{S}^C \mathcal{S}^B \mathcal{S}^A \underline{u}^m|_s + e^{\tau/\epsilon^2} \| I^h \mathcal{S}^C \mathcal{S}^B \mathcal{S}^A \underline{u}^m - \mathcal{S}_h^C \mathcal{S}_h^B \mathcal{S}_h^A U^m \| .
\end{aligned}$$

By Lemmas 3.5 and 3.6, it follows

$$\begin{aligned}
&\| I^h \mathcal{S}^C \mathcal{S}^B \mathcal{S}^A \underline{u}^m - \mathcal{S}_h^C \mathcal{S}_h^B \mathcal{S}_h^A U^m \| \\
\le& \| \left(I^h \mathcal{S}^C - \mathcal{S}_h^C I^h \right) \mathcal{S}^B \mathcal{S}^A \underline{u}^m \| + \| \mathcal{S}_h^C \left(I^h \mathcal{S}^B \mathcal{S}^A \underline{u}^m - \mathcal{S}_h^B \mathcal{S}_h^A U^m \right) \| \\
\le& C_3 \tau^3 + C_4 (\tau^3 + h^s) + e^{\tau/\epsilon^2} \| I^h \mathcal{S}^B \mathcal{S}^A \underline{u}^m - \mathcal{S}_h^B \mathcal{S}_h^A U^m \| .
\end{aligned}$$

By Lemmas 3.1–3.3 again, we know

$$\begin{aligned}
&\| I^h \mathcal{S}^B \mathcal{S}^A \underline{u}^m - \mathcal{S}_h^B \mathcal{S}_h^A U^m \| \\
\le& \| \left(I^h \mathcal{S}^B - \mathcal{S}_h^B I^h \right) \mathcal{S}^A \underline{u}^m \| + \| \mathcal{S}_h^B \left(I^h \mathcal{S}^A \underline{u}^m - \mathcal{S}_h^A U^m \right) \| \\
\le& C_2 h^s |\mathcal{S}^A \underline{u}^m|_s + \| I^h \mathcal{S}^A \underline{u}^m - \mathcal{S}_h^A U^m \| \\
=& C_2 h^s |\mathcal{S}^A \underline{u}^m|_s + \| \mathcal{S}_h^A I^h \underline{u}^m - \mathcal{S}_h^A U^m \| \\
\le& C_2 h^s |\mathcal{S}^A \underline{u}^m|_s + e^{\tau/\epsilon^2} \| I^h \underline{u}^m - U^m \| .
\end{aligned}$$

Hence, a combination of the aforementioned estimates leads to

$$\begin{aligned}
&\| I^h \underline{u}^{m+1} - U^{m+1} \| \\
\le& e^{\tau/\epsilon^2} \Big(C_2 |\mathcal{S}^C \mathcal{S}^B \mathcal{S}^A \underline{u}^m|_s + C_4 + C_2 e^{\tau/\epsilon^2} \Big) h^s + e^{\tau/\epsilon^2} \Big(C_3 + C_4 \Big) \tau^3 + e^{2\tau/\epsilon^2} \| I^h \underline{u}^m - U^m \| \\
\le& e^{2\tau/\epsilon^2} \| I^h \underline{u}^m - U^m \| + C_5 (h^s + \tau^3), \\
\le& e^{2T/\epsilon^2} \| I^h \underline{u}^0 - U^0 \| + \frac{e^{2T/\epsilon^2} - 1}{e^{2\tau/\epsilon^2} - 1} C_5 (h^s + \tau^3) \\
<& \frac{\epsilon^2}{2} (e^{2T/\epsilon^2} - 1) C_5 (h^s/\tau + \tau^2),
\end{aligned}$$

where $C_5 = \max \left\{ e^{\tau/\epsilon^2} \left(C_2 |\mathcal{S}^C \mathcal{S}^B \mathcal{S}^A \underline{u}|_s + C_4 + C_2 e^{\tau/\epsilon^2} \right), e^{\tau/\epsilon^2} \left(C_3 + C_4 \right) \right\}$. And the last step is because $I^h u^0 = I^h \underline{u}^0 = U^0$ and $e^x - 1 > x$ $(x > 0)$.

This end the proof. □

Remark 3.1

In the estimate of Theorem 3.2, the error contribution from the spatial approximation is affected by the inverse of the time step. However, for regular enough (in space) solution, h^s can be much smaller than τ; therefore, this affection generally would not reduce the global accuracy. Moreover, from the aforementioned proof, we know that the constant C in Theorem 3.2 depends on $e^{\frac{1}{\varepsilon^2}}$, which is not very useful estimate for small ε. For standard Allen–Cahn equation [1], there are some works improving this type of results, for example, [26]. Our future work will further improve the convergence result.

4. Numerical experiments

In this section, four numerical experiments in 1D, 2D, and 3D are carried out to study the performance of the FEOS method. We compare some of the results with the theory from previous sections, and all of them are consistent with the theories.

4.1. *Convergence test*

In order to verify the convergence rate of the proposed method, we first consider the 1D problem on $\Omega = (-\pi, \pi) \times (0, 1]$ with the following initial condition:

$$u_0(x) = \frac{2 + \cos(x)}{4}.$$

In order to verify the temporal numerical accuracy, we choose $h = \pi/50$ and different temporal step sizes τ. The ratio for the estimated convergence rate can be written as

$$\text{Rate} = \log 2\left(\frac{Err_{\tau/2}^{\tau/4}}{Err_{\tau}^{\tau/2}}\right),$$

where $Err_{\tau}^{\tau/2} = \| U^{\tau/2} - U^{\tau} \|_0$ and U^{τ} denotes the numerical result at T with time step τ.

The computational results for both boundary conditions at $T = 1$ are tabulated in Tables I and II corresponding to $\epsilon^2 = 0.1$ and $\epsilon^2 = 0.01$, respectively. It is observed that the temporal approximation orders are close to 2. Moreover, these results in both tables indicate that the different value of ϵ influences the numerical precision.

4.2. *Mass conserving, energy decreasing, and volume preserving*

In order to show that the conservative Allen–Cahn equation is mass conserving, energy decreasing, and volume preserving, we consider the second problem with $\epsilon^2 = 0.1$ on $(-\pi, \pi)^2 \times (0, 1]$ with the initial condition

$$u_0(x, y) = \begin{cases} 1, & \text{if } x^{0.8} + y^{0.8} < 2, \\ 0, & \text{else,} \end{cases}$$

and the periodic boundary condition.

The spatial step $h = \pi/20$ and temporal step $\tau = 1/50$ are used. This experiment may well illustrate the basic mechanism of Eq. (1), which also has been verified in [18]. Figure 1 shows the graph of the initial data u_0 and the numerical result for $u(x, y, T)$. Compared with the shape u_0, we find that the results at $t = 1$ could keep volume conservation. In order to better illustrate this property, we also made a comparison of the dynamics of the level set at different time. As is shown in Figure 2(c), the result at time $t = 0$, $t = 0.1$, and $t = 1$ further verifies the preserving of volume. Meanwhile, Figure 2(a) and (b) respectively shows that problem (1) is almost mass conserving and energy decreasing.

Remark 4.1

Note that the mass $M(u)$ in Figure 2(a) is not exactly conserved. Thus, we further reduce the time step to $\tau = 1/500$; the results have been presented in Figure 2(d), from which one can find that the mass conservation is greatly improved, although the curve is not a horizontal line. In future work, we will further discuss the conservation of the proposed scheme by theoretical analysis.

Table I. Convergence rate in time at $T = 1$ with $\epsilon^2 = 0.1$, $h = \frac{\pi}{50}$, and different τ for Problem 1.

Boundary conditions	τ	$\frac{1}{20}$	$\frac{1}{40}$	$\frac{1}{80}$	$\frac{1}{160}$	$\frac{1}{320}$
Periodic	$Err_{\tau}^{\tau/2}$	2.14e-4	5.71e-5	1.46e-5	3.66e-6	9.17e-7
	Rate	—	1.91	1.97	1.99	2.00
Neumann	$Err_{\tau}^{\tau/2}$	2.14e-4	5.71e-5	1.46e-5	3.66e-6	9.17e-7
	Rate	—	1.91	1.97	1.99	2.00

Table II. Convergence rate in time at $T = 1$ with $\epsilon^2 = 0.01$, $h = \frac{\pi}{50}$, and different τ for Problem 1.

Boundary conditions	τ	$\frac{1}{40}$	$\frac{1}{80}$	$\frac{1}{160}$	$\frac{1}{320}$	$\frac{1}{640}$
Periodic	$Err_{\tau}^{\tau/2}$	1.67e-3	5.96e-4	1.81e-4	4.94e-5	1.27e-5
	Rate	–	1.48	1.72	1.87	1.96
Neumann	$Err_{\tau}^{\tau/2}$	1.71e-3	6.02e-4	1.82e-4	4.95e-5	1.28e-5
	Rate	–	1.51	1.73	1.88	1.96

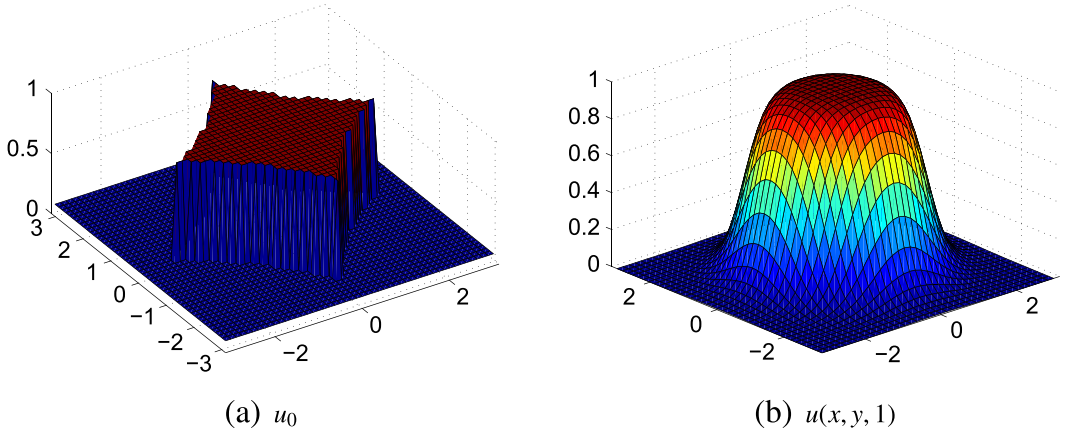


Figure 1. The initial data u_0 and the numerical result for u at $t = 1$ for Problem 2. (a) u_0 and (b) $u(x, y, 1)$. [Colour figure can be viewed at wileyonlinelibrary.com]

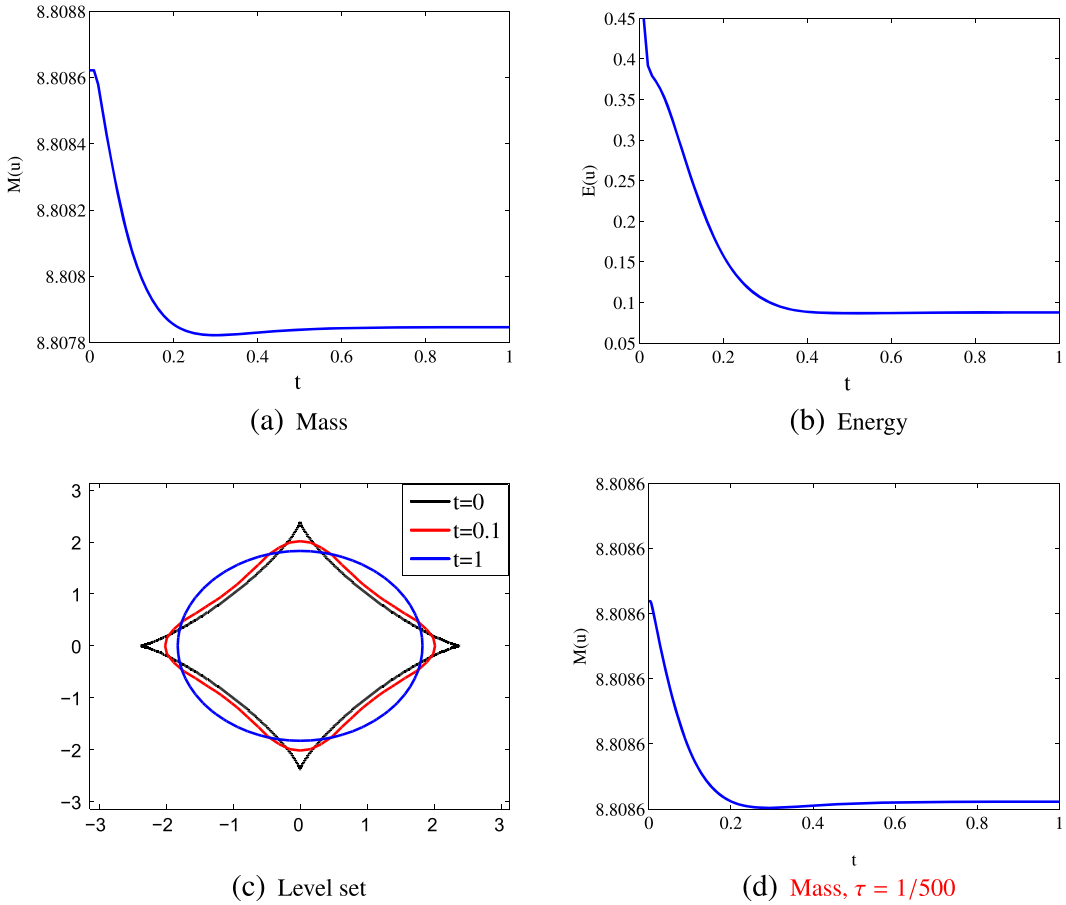


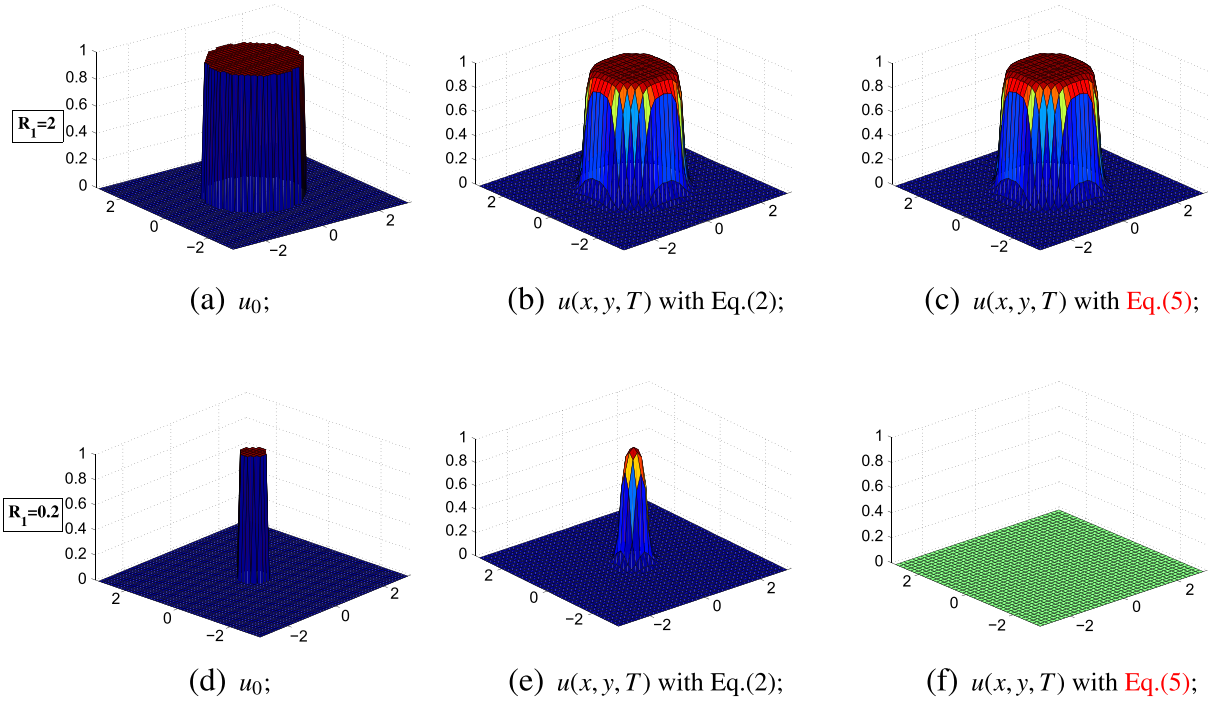
Figure 2. Dynamics of the mass, the energy, and the level set for Problem 2. (a) Mass, (b) energy, (c) level set, and (d) mass, $\tau = 1/500$. [Colour figure can be viewed at wileyonlinelibrary.com]

4.3. *Small geometry test*

In order to compare the difference between two Lagrange multipliers β in Eqs (2) and (5), both 2D and 3D problems with periodic boundary condition are considered on $(-\pi, \pi)^2$ and $(-\pi, \pi)^3$, respectively. The parameter ϵ^2 is chosen as 0.01.

The correspondingly initial conditions are respectively given by

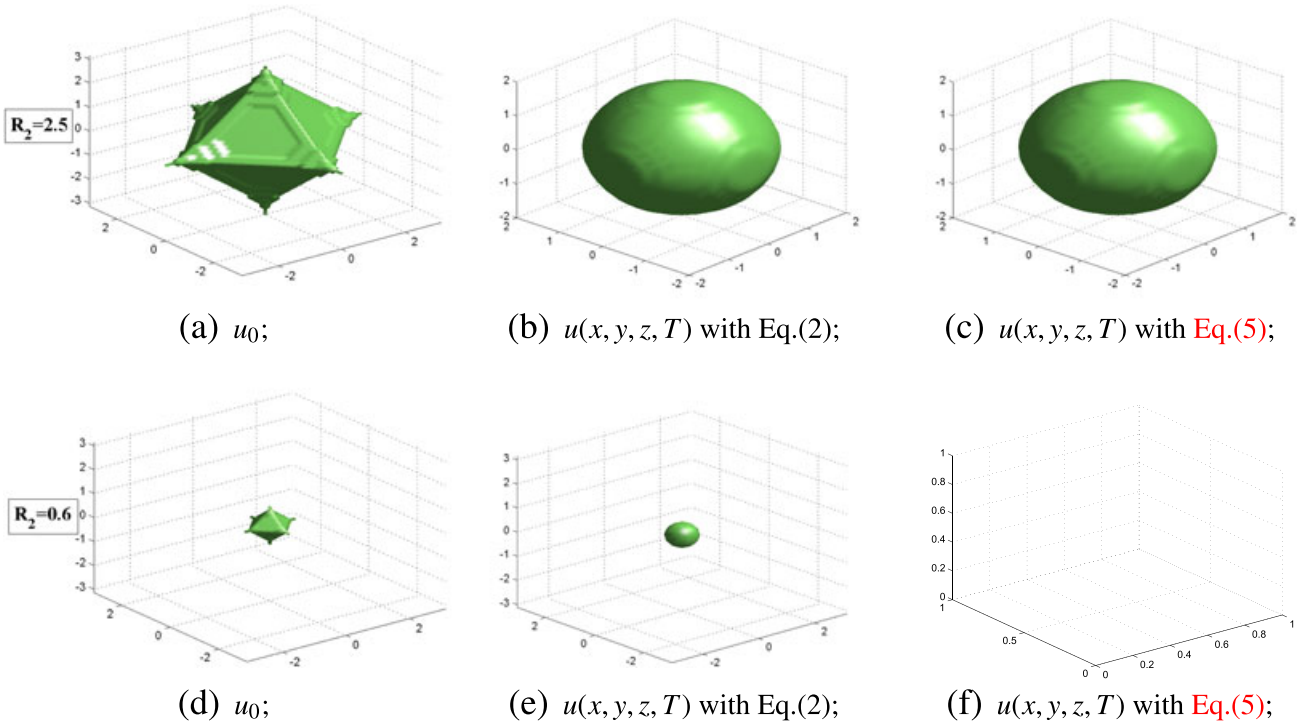
$$u_0(x, y) = \begin{cases} 1, & \text{if } \ x^2 + y^2 < R_1, \\ 0, & \text{else,} \end{cases}$$



(a) u_0; (b) $u(x, y, T)$ with Eq.(2); (c) $u(x, y, T)$ with Eq.(5);

(d) u_0; (e) $u(x, y, T)$ with Eq.(2); (f) $u(x, y, T)$ with Eq.(5);

Figure 3. The initial data u_0 and the numerical results for u in 2D case at $T = 1$ with different R_1 for Problem 3. (a) u_0, (b) $u(x, y, T)$ with Eq. (2), (c) $u(x, y, T)$ with Eq. (5), (d) u_0, (e) $u(x, y, T)$ with Eq. (2), and (f) $u(x, y, T)$ with Eq. (5). [Colour figure can be viewed at wileyonlinelibrary.com]



(a) u_0; (b) $u(x, y, z, T)$ with Eq.(2); (c) $u(x, y, z, T)$ with Eq.(5);

(d) u_0; (e) $u(x, y, z, T)$ with Eq.(2); (f) $u(x, y, z, T)$ with Eq.(5);

Figure 4. The initial data u_0 and the numerical results for u in 3D case at $T = 1$ with different R_2 for Problem 3. (a) u_0, (b) $u(x, y, z, T)$ with Eq. (2), (c) $u(x, y, z, T)$ with Eq. (5), (d) u_0, (e) $u(x, y, z, T)$ with Eq. (2), and (f) $u(x, y, z, T)$ with Eq. (5). [Colour figure can be viewed at wileyonlinelibrary.com]

and

$$u_0(x, y, z) = \begin{cases} 1, & \text{if } \ x^{0.8} + y^{0.8} + z^{0.8} < R_2, \\ 0, & \text{else.} \end{cases}$$

This experiment also may well indicate that Eq. (2) can preserve small features, which has been observed in [18].

We choose $T = 1$, $h = \pi/20$, and $\tau = 1/50$. The results with different R_1 and R_2 are given in Figures 3 and 4. For the 2D case, the comparison of different R_1 for both Lagrange multiplier terms (2) and (5) is shown in Figure 3. When $R_1 = 2$, we can see that both models work well when the initial feature is large enough, which can be observed in Figure 3(a–c). However, if the initial shape is small (Figure 3(d)), model (2) still keeps the small feature , while the geometry has disappeared with Eq. (5). Figure 4 shows the numerical results of 3D case; similar comparison can be made for this figure to reach similar conclusions.

Remark 4.2

We would like to comment to the fact that 'the small feature' refers to nonzero regions whose area/volume is small compared with the total area/volume of the domain. However, there is no specific standard to measure. In this paper, the proportion of nonzero regions in the total area is less than 5%; it will be treated as 'the small feature'. Moreover, it is worth pointing out that Eq. (1) with the Lagrange

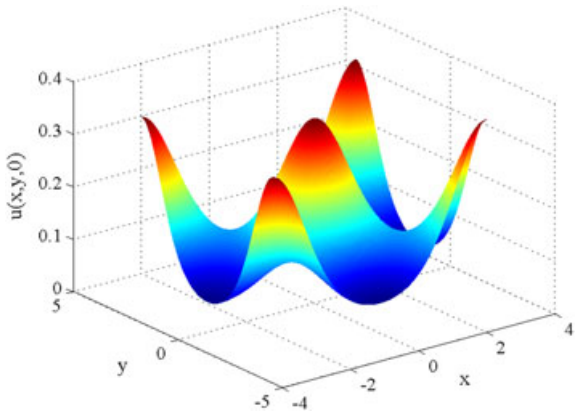


Figure 5. The initial data u_0. [Colour figure can be viewed at wileyonlinelibrary.com]

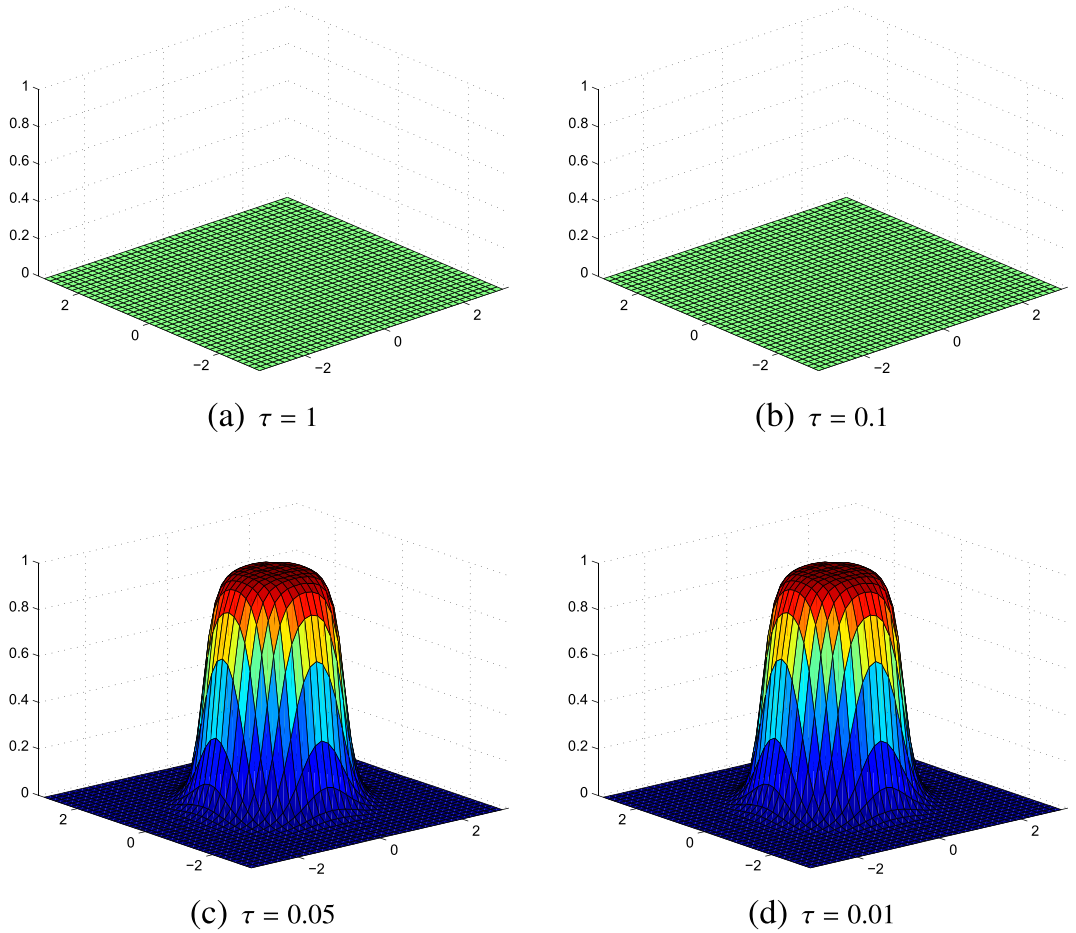


Figure 6. $u(x, y, 50)$ for the periodic boundary condition for Problem 4. (a) $\tau = 1$, (b) $\tau = 0.1$, (c) $\tau = 0.05$, and (d) $\tau = 0.01$. [Colour figure can be viewed at wileyonlinelibrary.com]

multiplier terms (2) can keep any small feature as long as the time step τ is moderate, but Eq. (1) with the Lagrange multiplier terms (5) does not work even for enough small time step.

4.4. *Stability test*

In order to verify that the FEOS method in this paper is stable and analysis the influence of different boundary conditions, we considered in $(-\pi,\pi)^2 \times (0,50]$ the fourth problem with $\epsilon^2 = 0.04$ and periodic/Neumann boundary condition. The initial condition is given by

$$u_0(x,y) = \frac{1}{8}\exp\big(\cos(x)\cos(y)\big) + 0.01.$$

We fix $h = \pi/20$ and then choose $\tau = 1, 1/10, 1/20$, and $1/100$ respectively. Figure 5 shows the graph of the initial data u_0 . Figures 6 and 7 show the the numerical solution plots at the same time $T = 50$ for periodic and Neumann boundary conditions, respectively.

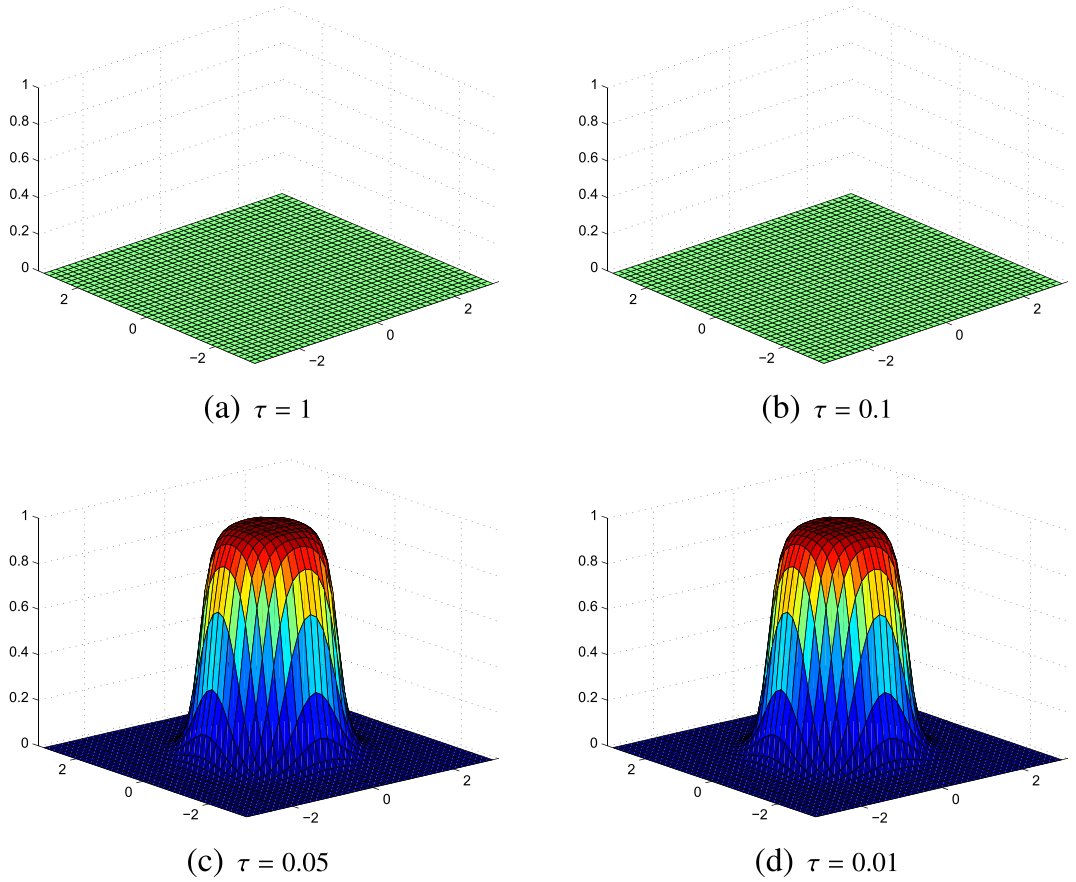


Figure 7. $u(x,y,50)$ for the Neumann boundary condition for Problem 4. (a) $\tau = 1$, (b) $\tau = 0.1$, (c) $\tau = 0.05$, and (d) $\tau = 0.01$. [Colour figure can be viewed at wileyonlinelibrary.com]

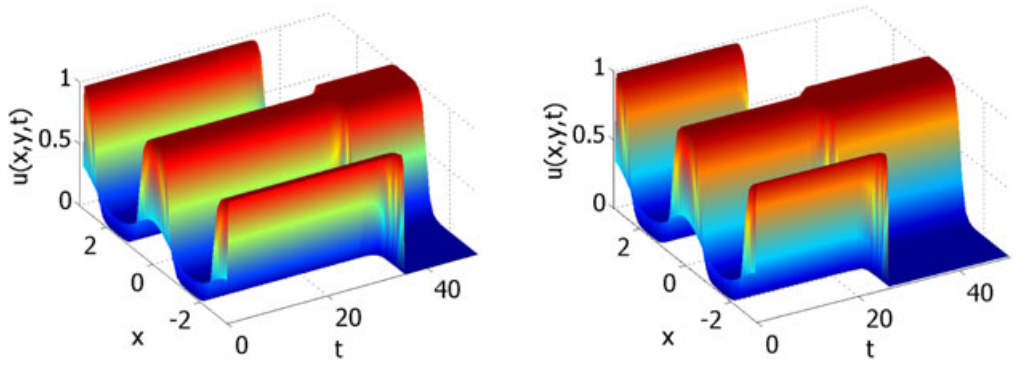


Figure 8. $u(x,y,50)$ with $x = y$ for both kinds of boundary conditions for Problem 4. (a) Periodic boundary condition and (b) Neumann boundary condition. [Colour figure can be viewed at wileyonlinelibrary.com]

These results suggest that the FEOS method cannot obtain the right results (Figures 6(a) and (b) and 7(a) and (b)) when $\tau = 1$ or $\tau = 1/10$ because the time step is not small enough. Meanwhile, note that the solutions computed with $\tau = 1/20$ and $\tau = 1/100$ are almost indistinguishable, which means that the solution has already converged with $\tau = 1/20$.

Figure 8 shows the evolution of the solutions with both different boundary conditions as we form the transition layer and metastable state and then reach the stable state, respectively. We see the fast dynamics from the initial condition to the metastable state, where two transition layers are formed. Furthermore, as Mello *et al*. mentioned in [8], we find that, concerning phase separation, the Neumann boundary condition is less stiff and faster achieved. For instance, in Figure 8(b), the solution with Neumann boundary condition arrives at the steady state around at $t = 25$, whereas the solutions with periodic boundary conditions are stable around at $t = 35$ (Figure 8(a)).

5. Conclusions

In this work, throught combination of the high-accurate Fourier spectral method for the space with the second-order operator splitting technique in time, an FEOS method for the mass-conserving Allen–Cahn equation was developed. A variety of numerical experiments were presented to confirm the accuracy, volume preserving, and stability of the proposed method. The numerical results also indicate that this method could preserve small features. In particular, it is applicable to various problems with a mass conservation constraint. In future work, we plan to extend this method to 3D Cahn–Hilliard equation [27]. Moreover, the design of adaptive time-stepping algorithms [28, 29] suitable to be used in the phase-field framework is also our future work.

Acknowledgements

The authors would like to thank the editor and referees for their valuable comments and suggestions that helped us improve the results of this paper. This work is in part supported by the Scientific Research Foundation of Huaqiao University (no. 15BS307), the Natural Science Foundation of Fujian Province (no. 2016J05007), the Promotion Program for Young and Middle-aged Teacher in Science and Technology Research of Huaqiao University(no. ZQN-PY201), and the NSF of China (nos 11526094 and 11501224).

References

1. Allen SM, Cahn JW. A microscopic theory for antiphase boundary motion and its application to antiphase domain coarsening. *Acta Metallurgica* 1979; **27**:1085–1095.
2. Wheeler AA, Boettinger WJ, McFadden GB. Phase-field model for isothermal phase transitions in binary alloys. *Physical Review A* 1992; **45**:7424–7439.
3. Benes M, Chalupecky V, Mikula K. Geometrical image segmentation by the Allen–Cahn equation. *Applied Numerische Mathematik* 2004; **51**:187–205.
4. Dobrosotskaya JA, Bertozzi AL. A wavelet-Laplace variational technique for image deconvolution and inpainting. *IEEE Transactions on Image Processing* 2008; **17**:657–663.
5. Chen LQ. Phase-field models for microstructure evolution. *Annual Review of Materials Research* 2002; **32**:113–140.
6. Zhang J, Du Q. Numerical studies of discrete approximations to the Allen–Cahn equation in the sharp interface limit. *SIAM Journal of Scientific Computing* 2009; **31**:3042–3063.
7. Choi JW, Lee HG, Jeong D, Kim J. An unconditionally gradient stable numerical method for solving the Allen–Cahn equation. *Physics A* 2009; **388**:1791–1803.
8. Zhai SY, Feng XL, He YN. Numerical simulation of three dimensional Allen–Cahn equation by high-order compact ADI method. *Computer Physics Communications* 2014; **10**:2449–2455.
9. Jeong D, Lee S, Lee D, Shin J, Kim J. Comparison study of numerical methods for solving the Allen–Cahn equation. *Computational Materials Science* 2016; **111**:131–136.
10. Rubinstein J, Sternberg P. Nonlocal reaction diffusion equation and nucleation. *IMA Journal of Applied Mathematics* 1992; **48**:249–264.
11. Brassel M, Bretin E. A modified phase field approximation for mean curvature flow with conservation of the volume. *Mathematical Methods in the Applied Sciences* 2011; **34**:1157–1180.
12. Alfaro M, Alifrangis P. Convergence of a mass conserving Allen–Cahn equation whose Lagrange multiplier is nonlocal and local. *Interfaces and Free Boundaries* 2014; **16**:243–268.
13. Bronsard L, Stoth B. Volume-preserving mean curvature flow as a limit of a nonlocal Ginzburg-Landau equation. *SIAM Journal on Mathematical Analysis* 1997; **28**:769–807.
14. Beneš M. Mathematical and computational aspects of solidification of pure substances. *Acta Mathematica Universitatis Comenianae* 2001; **70**: 123–152.
15. Chen X, Hilhorst D, Logak E. Mass conserving Allen–Cahn equation and volume preserving mean curvature flow. *Interfaces and Free Boundaries* 2010; **12**:527–549.
16. Bates PW, Jin J. Global dynamics of boundary droplets. *Discrete and Continuous Dynamical Systems* 2013; **34**:1–7.
17. Yue P, Zhou C, Feng JJ. Spontaneous shrinkage of drops and mass conservation in phase-field simulations. *Journal of Computational Physics* 2007; **223**:1–9.
18. Kim J, Lee S, Choi Y. A conservative Allen–Cahn equation with a spacetime dependent Lagrange multiplier. *International Journal of Engineering Science* 2014; **84**:11–17.
19. Zhai SY, Weng ZF, Feng XL. Investigations on several numerical methods for the non-local Allen–Cahn equation. *International Journal of Heat and Mass Transfer* 2015; **87**:111–118.
20. Bueno-Orovio A, Kay D, Burrage K. Fourier spectral methods for fractional in-space reaction diffusion equations. *BIT Numerical Mathematics* 2014; **54**:937–954.
21. Shen J, Tang T, Wang LL. *Spectral Methods Algorithms: Analyses and Applications* 1St Ed. Springer: Berlin, 2010.

22. Strang G. On the construction and comparison of difference schemes. *SIAM Journal on Numerical Analysis* 1968; **5**:506–517.
23. Blanes S, Moan PC. Practical symplectic partitioned Runge–Kutta and Runge–Kutta–Nyström methods. *Journal Computational and Applied Mathematics* 2002; **142**:313–330.
24. Csomós P, Faragó I, Havasi Á. Weight sequential splitting and their analysis. *Mathematical and Computational Application* 2005; **50**:1017–1031.
25. Stuart A, Humphries AR. *Dynamical System and Numerical Analysis*. Cambridge University Press: Cambridge, 1998.
26. Feng XB, Prohl A. Numerical analysis of the Allen–Cahn equation and approximation for mean curvature flows. *Numerische Mathematik* 2003; **94**: 33–65.
27. He YN, Liu YX, Tang T. On large time-stepping methods for the Cahn–Hilliard equation. *Applied Numerical Mathematics* 2007; **57**:616–628.
28. Wodo O, Ganapathysubramanian B. Computationally efficient solution to the Cahn–Hilliard equation: adaptive implicit time schemes, mesh sensitivity analysis and the 3D isoperimetric problem. *Journal of Computational Physics* 2011; **230**:6037–6060.
29. González FG, Tierra G. Second order schemes and time-step adaptivity for Allen–Cahn and Cahn–Hilliard models. *Computers & Mathematics with Applications* 2014; **68**:821–846.