Available online at www.sciencedirect.com
ScienceDirect
Comput. Methods Appl. Mech. Engrg. 381 (2021) 113746
Computer methods in applied mechanics and engineering
www.elsevier.com/locate/cma

Supplementary variable method for thermodynamically consistent partial differential equations

Yuezheng Gong[a,c,e], Qi Hong[b,c], Qi Wang[d,*]

[a] College of Science, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
[b] Beijing Computational Science Research Center, Beijing 100193, China
[c] Jiangsu Key Laboratory for Numerical Simulation of Large Scale Complex Systems, Nanjing Normal University, Nanjing 210023, China
[d] Department of Mathematics, University of South Carolina, Columbia, SC 29208, USA
[e] MIIT Key Laboratory of Mathematical Modelling and High Performance Computing of Air Vehicles, Nanjing 210016, China

Received 11 October 2020; accepted 20 February 2021
Available online xxxx

Abstract

We present a paradigm for developing thermodynamical consistent numerical algorithms for thermodynamically consistent partial differential equation (TCPDE) systems, called the supplementary variable method (SVM). We add a proper number of supplementary variables to the TCPDE system coupled with its energy equation and other deduced equations through perturbations to arrive at a consistent, well-determined, solvable and structurally stable system. The extended system not only reduces to the TCPDE system at specific values of the supplementary variables, but also allows one to retain consistency and solvability after a consistent numerical approximation. Among virtually infinite many possibilities to add the supplementary variables, we present two that maintain thermodynamical consistency in the extended system before and after the approximation. A pseudo-spectral method is used in space to arrive at fully discrete schemes. The new schemes are compared with the energy stable SAV scheme and the fully implicit Crank–Nicolson scheme. The numerical results favor the new schemes in the overall performance.

Keywords: Supplementary variable method; Thermodynamically consistent models; Gradient flows; Energy-production-rate preserving schemes; Finite difference methods; Pseudo-spectral methods

1. Introduction

Nonequilibrium phenomena are usually described by dynamical models that follow nonequilibrium thermodynamical principles. The second law of thermodynamics or the equivalent generalized Onsager principle, are essential for one to develop such models for nonequilibrium phenomena not far from equilibria [1–3]. Assuming the variables of a thermodynamic system are denoted as $\Phi \in \mathbf{R}^m$, where m is the number of the variables, and the free energy of the system is described by a functional of Φ: $F[\Phi]$ in isothermal cases (for nonisothermal cases, we use an entropy functional instead). The Onsager linear response theory yields the dynamical equation for the system as follows

$$\partial_t \Phi = -\mathbf{M}\frac{\delta F}{\delta \Phi}, \quad \mathbf{x} \in \Omega, \tag{1.1}$$

* Corresponding author.
E-mail address: qwang@math.sc.edu (Q. Wang).

https://doi.org/10.1016/j.cma.2021.113746
0045-7825/

where $\mathbf{M}$ is the mobility operator, its inverse if exists $\mathbf{R} = \mathbf{M}^{-1}$ is the friction operator, $\frac{\delta F}{\delta \Phi}$ is the variation of F known as the chemical potential, and Ω is the domain in which the dynamical system is defined. This equation system provides relaxation dynamics for the nonequilibrium state to return to equilibrium in dissipative systems or to oscillate in nondissipative systems.

In general, system (1.1) with proper adiabatical boundary conditions possesses an intrinsic energy law

$$\frac{\mathrm{d}F}{\mathrm{d}t} = -\int_{\Omega} \left(\frac{\delta F}{\delta \Phi}\right)^{T} \mathbf{M} \frac{\delta F}{\delta \Phi} \mathrm{d}\mathbf{x}. \tag{1.2}$$

If $\mathbf{M}$ is skew-adjoint, the system is called a conservative system where $\mathrm{d}F/\mathrm{d}t = 0$, which corresponds to a reversible process physically. While $\mathbf{M}$ is positive semi-definite, it leads to a dissipative system where $\mathrm{d}F/\mathrm{d}t \leq 0$, corresponding to an irreversible process physically. Apparently, energy equation (1.2) is deduced from (1.1). It is an identity in the solution manifold of (1.1). There can be other deduced equations of relevant physical implications which also serve as identities in the solution manifold of (1.1). For example, in the KdV equation, there are many other invariants which are conserved quantities in the solution manifold besides the energy and momentum. We emphasize that these are identities instead of constraints on the solution manifold of (1.1) so that they have no impact on the solutions of (1.1).

When approximating (1.1) numerically, one would like to retain the consistent relationship between the discretized solution and the discretized deduced equations including the energy equation to produce the so-called structure-preserving algorithm [4,5]. In order to achieve this goal, discrete gradient methods originally proposed by Gonzalez [6] and McLachlan et al. [7] were generalized to PDE systems (1.1) to develop discrete variational derivative methods that inherit energy conservation or dissipation properties [8,9]. The averaged vector field method was originally proposed for canonical Hamiltonian systems [10] and was later extended to general conservative or dissipative systems as well [11]. However, there are many other numerical algorithms developed for thermodynamically consistent models that are unable to satisfy the discrete consistency requirement, so they are not structure-preserving in this sense.

For dissipative systems, if one is more interested in the steady state than actual transient dynamics, a compromise has been widely adopted in the computational mathematics, engineering and science community, in which one settles with a numerical algorithm satisfying an energy dissipation inequality upon discretization instead of the rigorous equation. In these algorithms, energy decays in time, but may not follow the actual energy dissipation rate. All numerical approaches with this property are called energy stable schemes. In recent years, there have been a great deal of works contributed to the investigation of such numerical algorithms, for instance, the convex splitting method [12–15] and the stabilizing method [16–18]. Although these methods satisfy the property of energy decay, they do not preserve the actual energy dissipation rate, which may cause non-physical dissipation, either enhanced or weakened and thereby lose accuracy in tracking dynamics of the system.

Recently, inspired by the seminal work in [19,20], Yang, Zhao and Wang proposed the energy quadratization (EQ) framework to obtain linear energy stable schemes for general thermodynamically consistent models, which preserve the system's energy dissipation law with a transformed, quadratic free energy density [21,22]. Subsequently, the scalar auxiliary variable (SAV) method following the same idea with a slightly different formulation of the quadratic free energy was proposed by Shen et al. [23]. It yields a class of efficient energy stable schemes in which one only need to solve linear equations with constant coefficients at each time step. Due to the generality of EQ and SAV approaches, they have been applied successfully for many existing gradient flow models for both dissipative and conservative processes, please refer to review articles [24,25] and references therein for the latest development. Several other extensions of EQ and SAV approaches have been added recently [26–29]. In addition, several high-order energy stable schemes have developed for gradient flow models, including linear and nonlinear EQ Runge–Kutta methods [30–32], linear and nonlinear SAV Runge–Kutta schemes [33,34].

Although the convex splitting scheme and the stabilizing method yield an energy dissipation property, they often result in a loss of numerical accuracy because of splitting errors [35]. The EQ and SAV approaches only preserve an energy dissipation law for a reformulated free energy, not the original energy. In fact, implicit energy-dissipation-rate preserving methods had been derived for the Cahn–Hilliard equation sometime ago in [36,37], where the original energy dissipation law is rigorously conserved. These implicit schemes can be viewed as special cases of the discrete variational derivative method and require nonlinear iterations to solve them numerically. Recently, Cheng, Liu and Shen proposed a new Lagrange multiplier method for gradient flows inspired by the SAV method,

whose computational cost is only the cost of solving one nonlinear algebraic equation more than that for the SAV method [38]. Different from the existing projection methods [39], the Lagrange multiplier method is based on the extended PDE system consisting of the original PDE and its energy equation, and the discrete energy of each time step is obtained implicitly.

In this paper, we will look at the issue on preserving thermodynamical consistency of the discretized system from a new perspective. In the extended model, if we view F as one of the thermodynamical variables, (1.1), (1.2) and the definition of the free energy functional constitute an *over-determined, extended system* with $m + 2$ equations and $m + 1$ unknowns (Φ, F), where m is the dimension of system (1.1). This system of equations is consistent and solvable should gradient flow model (1.1) is solvable with the given free energy $F[\Phi]$ and proper initial and boundary conditions. We reiterate again that Eq. (1.2) as a deduced equation is in fact an identity in the solution manifold of (1.1). However, the consistency among the equations can be easily broken under small perturbations in the extended model system leading to a system that is not consistent nor solvable. We describe this scenario in the over-determined, extended system as structurally unstable. How can we make it structurally stable so that consistency would be retained after a small perturbation. One way to achieve this is to augment the system by a proper number of supplementary variables to make up the deficiency in the number of unknowns so that the number of unknowns matches the number of equations in the augmented system or perturbed system. In fact, there are virtually unlimited number of ways for one to accomplish this with supplementary variables. To be consistent with the original problem, we require that the perturbed, well-determined system yields solutions that include the solution of the original thermodynamically consistent model at selected values of the supplementary variables. If such an augmented/perturbed model can be developed, its numerical approximation would retain its structural stability readily so that the discrete equations remain consistent. Then, one can develop numerical methods for the perturbed system to arrive at the so-called energy dissipation rate preserving schemes. We name this numerical strategy the supplementary variable method (SVM), which originates from the idea on how to enforce structural stability in the over-determined PDE system. Clearly, the Lagrange multiplier method and several others can be effectively identified as specific implementations of this idea. But, their development originated from different perspectives.

When the thermodynamically consistent model has additional deduced equation or conserved quantities (invariants) in conservative systems that are of important physical significance, these equations or invariants can be preserved as well by SVM with additional supplementary variables. It thus makes SVM an effective paradigm for developing structural-preserving numerical algorithms for not only thermodynamically consistent models, but also any mathematical models with deduced equations or invariants. In the following, we enumerate the advantages of SVM.

1. The numerical algorithms constructed by SVM preserve the thermodynamical consistency of the original model upon discretization.
2. SVM provides a straightforward way to preserve as many deduced equations and invariants as one wishes for the thermodynamical model.
3. Each supplementary variable corresponds to an algebraic equation in the discretized system. So, the additional computational cost incurred in an SVM method compared with other energy stable methods is exactly the cost for solving the corresponding nonlinear algebraic equations.
4. SVM provides a new paradigm to develop structure-preserving algorithms with a great deal of flexibilities.

In this paper, we will present the basic framework for designing structure-preserving algorithms for thermodynamically consistent models with additional deduced equations using SVM and compare the efficiency and accuracy of the new algorithms with an SAV approach and an implicit method, respectively.

The rest of the paper is organized as follows. In Section 2, we motivate readers on the issue of structural stability in the TCPDE with deduced equations and the need for introducing supplementary variables. In Section 3, we discuss a thermodynamically consistent SVM reformulation of the extended model system. In Section 4, we discuss an SVM algorithm and a couple of efficient implementation strategies. In Section 5, the SVM is generalized for a system with multiple deduced equations (invariants). In Section 6, the spatial discretization is discussed. In Section 7, we present some numerical results and comparisons with the existing energy dissipation rate preserving numerical schemes. The conclusion is given in the end.

2. Motivation for supplementary variable methods (SVM)

We first examine a simple example to bring out the underlying issue and the idea. Let us consider an over-determined and yet consistent linear equation system given below

$$\begin{cases} x + y = 1, \\ x - y = 1, \\ 3x + y = 3. \end{cases} \tag{2.1}$$

The third equation in the 3×2 system is deduced from the first two equations and is therefore consistent with the two equations. It is uniquely solvable with a solution $(1, 0)$. However, the system is structurally unstable because if we perturb the linear system with a small perturbation, the perturbed system would become inconsistent and thus unsolvable. This is what we call structurally unstable for an equation system.

An easy fix to the issue of structural instability is to introduce a third, known as the supplementary variable, z, as follows

$$\begin{cases} x + y + az = 1, \\ x - y + bz = 1, \\ 3x + y + cz = 3, \end{cases} \tag{2.2}$$

where a, b, c are arbitrary constants so long as $2a + b - c \neq 0$. At $z = 0$, this system reduces to the original one. The supplementary variable z is introduced consistently without creating a new equation for it. This perturbed system is well-defined and uniquely solvable with a unique solution $(1, 0, 0)$ and moreover, it is structurally stable in that any small perturbation to the system will render a solvable system with a unique solution which is close to the original one so long as the perturbation is small enough. This problem is a trivial version of the issue associated with designing structure preserving numerical approximations to partial differential equation systems with deduced equations. We now return to the thermodynamically consistent model to discuss the issue of adding supplementary variables to make it structurally stable and designing structure-preserving numerical schemes.

We consider a thermodynamically consistent partial differential equation (TCPDE) model coupled with the deduced energy equation and other equations for invariants as follows

$$\begin{cases} \dfrac{\partial \Phi}{\partial t} = M[\Phi], \\ \dfrac{\mathrm{d}F[\Phi]}{\mathrm{d}t} = G[\Phi], \\ H[\Phi] = 0, \end{cases} \tag{2.3}$$

where F, G, H are given functionals of the unknown function $\Phi(\mathbf{x}, t)$ and M is a functional of Φ and its spatial derivatives. Note that F, G and H are functions of time only and may be a vector. We assume the solution at t_n is already given by Φ^n. We want to solve (2.3) numerically to get an approximate solution at $t_{n+1} > t_n$. Let us assume the three equations in (2.3) are approximated respectively by kth order methods

$$\begin{cases} D_t \Phi = M_1[\Phi, \Phi^n, \ldots], \\ D_t F_1 = G_1[\Phi, \Phi^n, \ldots], \\ H_1[\Phi, \Phi^n, \ldots] = 0, \end{cases} \tag{2.4}$$

where $F_1 = F_1[\Phi, \Phi^n, \ldots]$ is an approximate free energy at $t > t_n$ with accuracy up to the kth order, M_1, G_1 and H_1 are approximations to M, G and H up to the kth order and $D_t = \frac{\partial}{\partial t}$. Then, the truncation error of the numerical scheme is $\mathcal{O}(\Delta t_n^k)$, where $\Delta t_n = t - t_n$.

Note that (2.3) is over-determined and yet consistent. (2.4) inherits the over-determinedness, but may no longer be consistent anymore due to the approximations. In fact, special cares must be taken to ensure consistency in (2.4). In the past, the EQ and SAV method, the Lagrange multiplier method and some implicit methods have been developed to address the issue. To partially address the issue, sometimes one replaces the energy equation by an inequality to meet a weaker consistent condition. The methods that can lead to this class of energy stable approximations include the convex splitting method, stabilization method and some implicit methods. The key that may prevent one from arriving at structure-preserving numerical approximations is rooted in the structure instability of the original coupled TCPDE system. Adding necessary number of supplementary variables to the coupled system via perturbations is a

viable and systematic approach to solve the issue of inconsistency and solvability after the perturbation. The issue on consistency and solvability in fact warrants the structure-preservation of the system after discretization. This motivates us to develop the supplementary variable method.

We note that the deduced equations (energy and invariants in the current case) are not constraints because they do not impose additional constraints in addition to the domain of the TCPDE system. They are identities on the solution manifold of the TCPDE model. Therefore the supplementary variables are not Lagrange multipliers at the PDE level. A closer examination of (2.4) reveals that the equation's inconsistency occurs at the truncated equation level after discretization. Namely, if we add the terms compatible with the truncation error back to the approximate equation system, the inconsistency issue can be addressed satisfactorily. This indicates that some supplementary variables in fact play the role of Lagrange multipliers after the system is discretized (at least semi-discretization) where inconsistency issues arisen. In the SVM approach, we address the inconsistency issue systematically in a more versatile and practical manner.

In general, we extend the coupled TCPDE model via perturbations as follows

$$\begin{cases} \dfrac{\partial \Phi}{\partial t} = M_2[\Phi, \alpha(t)], \\ \dfrac{\mathrm{d}F_2[\Phi, \alpha(t)]}{\mathrm{d}t} = G_2[\Phi, \alpha(t)], \\ H_2[\Phi, \alpha(t)] = 0, \end{cases} \tag{2.5}$$

where M_2, F_2, G_2, H_2 are perturbed functions of M, F, G, H with $M_2[\Phi, 0] = M[\Phi]$, $F_2[\Phi, 0] = F[\Phi]$, $G_2[\Phi, 0] = G[\Phi]$, $H_2[\Phi, 0] = H[\Phi]$. Here we note that α is a vector whose dimension is equal to the sum of the dimensions of F and H. The extended model admits a solution $(\Phi, \alpha(t) = 0)$, where Φ is the solution of the original TCPDE. There could be more solutions for $\alpha \neq 0$. But those solutions would have nothing to do with the original system and are therefore not of interest. In the following, we focus on a special class of perturbations to the extended TCPDE to implement the supplementary variable method. We name it the thermodynamically consistent supplementary variable method (SVM) reformulation, in which we enforce thermodynamically consistency and the other deduced relations in the perturbed system. In principle, this reformulation is an option rather than a necessity.

Next, we detail on how to derive an SVM algorithm for TCPDE models to preserve the thermodynamical consistency in the energy dissipation rate. Then, we discuss how to preserve additional deduced equations or invariants of the TCPDE model.

3. Thermodynamically consistent SVM reformulation

To enforce the original energy dissipation law, we first perturb the extended system by introducing a supplementary variable $\alpha(t)$ such that

$$\begin{cases} \partial_t \Phi = -\mathbf{M}\frac{\delta F}{\delta \Phi} + \alpha g_1[\Phi], \\ \frac{\mathrm{d}F}{\mathrm{d}t} = -\Big(\frac{\delta F}{\delta \Phi} + \alpha g_2[\Phi], \mathbf{M}\big(\frac{\delta F}{\delta \Phi} + \alpha g_2[\Phi]\big)\Big), \\ F = \frac{1}{2}(\Phi, \mathcal{L}\Phi) + \big(f(\Phi, \nabla\Phi), 1\big) + \alpha g_3[\Phi], \end{cases} \tag{3.1}$$

where $g_1[\Phi]$, $g_2[\Phi]$, $g_3[\Phi]$ are user supplied functions or functionals of Φ. They may not be completely independent though. When $\alpha = 0$, this perturbed system reduces to the original extended system. If Φ is a solution of the original extended system, $(\Phi, \alpha = 0)$ is a solution of the perturbed system. If one of the $g_1[\Phi]$, $g_2[\Phi]$, $g_3[\Phi]$ is not zero, the number of unknowns is equal to the number of equations in the perturbed system, so that the perturbed system is structurally stable. Apparently, there is an infinite number of ways for one to obtain the perturbed system.

Unlike the EQ/SAV schemes that preserve a modified energy dissipation law, our target is to construct a numerical scheme that warrants the original energy dissipative structure. To this end, we consider only the perturbation of (1.1) within the thermodynamically consistent models. Namely, the energy dissipation rate in the perturbed system is a quadratic functional of the chemical potential or time rate of change of the thermodynamical variable. Without loss of generality, we set $g_3 = 0$.

Denoting

$$\mu = \frac{\delta F}{\delta \Phi} + \alpha g_2[\Phi], \quad g[\Phi] = \mathbf{M}g_2[\Phi] + g_1[\Phi], \tag{3.2}$$

we can rewrite system (3.1) into the following form

$$\begin{cases} \partial_t \Phi = -\mathbf{M}\mu + \alpha g[\Phi], \\ \dfrac{\mathrm{d}F}{\mathrm{d}t} = -(\mu, \mathbf{M}\mu), \\ F = \frac{1}{2}(\Phi, \mathcal{L}\Phi) + \big(f(\Phi, \nabla\Phi), 1\big), \\ \mu = \mathcal{L}\Phi + \frac{\delta f}{\delta \Phi} + \alpha g_2[\Phi]. \end{cases} \tag{3.3}$$

The question with respect to this perturbed system now is how to choose the user supplied functions $g[\Phi]$ and $g_2[\Phi]$ to make the solution procedure accurate and stable and the energy dissipation rate preserved after discretization. If we choose $g[\Phi] = 0,\ g_2[\Phi] = \frac{\delta f}{\delta \Phi}$ and let $\eta = 1 + \alpha$, then the SVM reformulation reduces to the Lagrange multiplier system proposed in [38]. However, in the SVM, there're an infinite number of possibilities for choosing $g[\Phi]$ and $g_2[\Phi]$.

It is worth noting that in theory we are looking for solutions $(\Phi, \alpha = 0)$ of the SVM reformulated system. Although it is highly unlikely, but not impossible that the discretized extended system admits multiple solutions near 0. With this in mind, we can also implement the SVM idea in the form of optimization subject to PDE constraints. Assuming we have already obtained the solution of the original system up to $t = t_n \geq 0$, we would like to find the solution up to $t_{n+1} > t_n$. In order to do it, we introduce a time-dependent supplementary variable $\alpha(t)$ and reformulate the extended system in $t \in (t_n, t_{n+1}]$ into the following constrained optimization problem:

$$\begin{cases} \min\limits_{\alpha(t)} & \int_{t_n}^{t_{n+1}} |\alpha(t)|^2 \mathrm{d}t, \\ \text{s.t.} & \partial_t \Phi = -\mathbf{M}\mu + \alpha g[\Phi], \\ & \frac{\mathrm{d}F}{\mathrm{d}t} = -(\mu, \mathbf{M}\mu), \\ & F = \frac{1}{2}(\Phi, \mathcal{L}\Phi) + \big(f(\Phi, \nabla\Phi), 1\big), \\ & \mu = \mathcal{L}\Phi + \frac{\delta f}{\delta \Phi} + \alpha g_2[\Phi]. \end{cases} \tag{3.4}$$

This seems to be a trivial mathematical statement given that $\alpha = 0$ is a trivial solution to the problem. However, upon numerical discretization, it provides a platform to design structure-preserving, approximate numerical schemes for the original TCPDE system.

Remark 1. By supplementing the over-determined coupled TCPDE system with a new variable without introducing an equation, we effectively remove the over-determinedness by making the extended system a well-determined one in a higher dimensional phase space! We reiterate that there is a great deal of flexibility to place the supplementary variable in the thermodynamically consistent model, which generalizes it beyond the Lagrange multiplier method and the generalized SAV method, where a specific form of the perturbation is followed. How to add the supplementary variable to make the extended system easier or more efficient to solve would be an interesting and open question to explore.

Next we will show how to develop efficient energy dissipation rate preserving schemes systematically from the SVM reformulated system.

4. Supplementary variable method (SVM)

In this section, we apply the prediction–correction strategy to obtain energy dissipation rate preserving schemes for the TCPDE system via the SVM reformulation given by (3.3), where the discrete, original energy dissipation law is inherited naturally. The proposed method is devised with the expectation that it is second order accurate in time and can be solved efficiently. For simplicity, we denote

$$\delta_t^+(\bullet)^n = \frac{(\bullet)^{n+1} - (\bullet)^n}{\tau}, \quad (\bullet)^{n+\frac{1}{2}} = \frac{(\bullet)^{n+1} + (\bullet)^n}{2}, \quad \overline{(\bullet)}^{n+\frac{1}{2}} = \frac{3(\bullet)^n - (\bullet)^{n-1}}{2}, \quad \forall\, n \geq 0, \tag{4.1}$$

where we set $(\bullet)^{-1} = (\bullet)^0$ and τ is the time step.

Combining the prediction–correction strategy and the semi-implicit Crank–Nicolson scheme for system (3.3), we obtain the following energy dissipation rate preserving scheme.

Scheme 4.1. *Given Φ^{n-1} and Φ^n, we compute Φ^{n+1} through the following two steps:*

1. *Prediction: predict $\Phi_*^{n+\frac{1}{2}}$ via*

$$\frac{\Phi_*^{n+\frac{1}{2}} - \Phi^n}{\tau/2} = -\mathbf{M}\left(\mathcal{L}\Phi_*^{n+\frac{1}{2}} + \frac{\delta f}{\delta \Phi}[\overline{\Phi}^{n+\frac{1}{2}}]\right). \tag{4.2}$$

2. *SVM correction:*

$$\begin{cases} \delta_t^+ \Phi^n = -\mathbf{M}\left(\mathcal{L}\Phi^{n+\frac{1}{2}} + \frac{\delta f}{\delta \Phi}[\Phi_*^{n+\frac{1}{2}}] + \alpha g_2[\Phi_*^{n+\frac{1}{2}}]\right) + \alpha g[\Phi_*^{n+\frac{1}{2}}], \\ \delta_t^+ F[\Phi^n] = -(\mu_*^{n+\frac{1}{2}}, \mathbf{M}\mu_*^{n+\frac{1}{2}}), \\ F[\Phi] = \frac{1}{2}(\Phi, \mathcal{L}\Phi) + \big(f(\Phi, \nabla\Phi), 1\big), \\ \mu_*^{n+\frac{1}{2}} = \mathcal{L}\Phi_*^{n+\frac{1}{2}} + \frac{\delta f}{\delta \Phi}[\Phi_*^{n+\frac{1}{2}}] + \alpha g_2[\Phi_*^{n+\frac{1}{2}}]. \end{cases} \tag{4.3}$$

In this paper, we focus on $g[\Phi] \neq 0$ and $g_2[\Phi] = 0$. It is obvious that Scheme 4.1 preserves the discrete original energy dissipation law

$$F[\Phi^{n+1}] - F[\Phi^n] = -\tau\left(\mu_*^{n+\frac{1}{2}}, \mathbf{M}\mu_*^{n+\frac{1}{2}}\right) \leq 0. \tag{4.4}$$

Note that Scheme 4.1 is a three time-level scheme. If we replace $\overline{\Phi}^{n+\frac{1}{2}}$ with Φ^n, we obtain a two time-level scheme, which provides the initial datum for the second level value of Scheme 4.1. According to the local truncation error analysis, we deduce from (4.2) that

$$\Phi_*^{n+\frac{1}{2}} = \Phi(t_{n+\frac{1}{2}}) + \mathcal{O}(\tau^2). \tag{4.5}$$

Next we discuss how to solve Scheme 4.1 efficiently. Eq. (4.2) implies

$$\Phi_*^{n+\frac{1}{2}} = \left(1 + \frac{\tau}{2}\mathbf{M}\mathcal{L}\right)^{-1}\left(\Phi^n - \frac{\tau}{2}\mathbf{M}\frac{\delta f}{\delta \Phi}[\overline{\Phi}^{n+\frac{1}{2}}]\right). \tag{4.6}$$

Let

$$\widehat{\Phi}^{n+1} = (1 + \frac{\tau}{2}\mathbf{M}\mathcal{L})^{-1}\left((1 - \frac{\tau}{2}\mathbf{M}\mathcal{L})\Phi^n - \tau\mathbf{M}\frac{\delta f}{\delta \Phi}[\Phi_*^{n+\frac{1}{2}}]\right), \tag{4.7}$$

$$w^n = (1 + \frac{\tau}{2}\mathbf{M}\mathcal{L})^{-1} g[\Phi_*^{n+\frac{1}{2}}], \tag{4.8}$$

$$\beta = \tau\alpha. \tag{4.9}$$

It follows from the first equation of (4.3) that

$$\Phi^{n+1} = \widehat{\Phi}^{n+1} + \beta w^n. \tag{4.10}$$

Then we plug Φ^{n+1} into the second equation in (4.3) to obtain

$$F[\widehat{\Phi}^{n+1} + \beta w^n] = F[\Phi^n] - \tau(\mu_*^{n+\frac{1}{2}}, \mathbf{M}\mu_*^{n+\frac{1}{2}}), \tag{4.11}$$

which is an algebraic equation for β. In general, it can have multiple solutions, but one of them must be close to 0 and it approaches to zero as $\tau \to 0$. So, we solve for this solution using an iterative method such as the Newton iteration with 0 as the initial condition, it generally converges to a solution close to 0 when τ is not too large. After obtaining β, we update Φ^{n+1} using (4.10). Following the work of [39,40], the existence of solution β is guaranteed by the conditions of the following theorem.

Theorem 4.1. *If $\left(\frac{\delta F}{\delta \Phi}[\Phi^n], g[\Phi^n]\right) \neq 0$, there exists a $\tau^* > 0$ such that (4.11) defines a unique function $\beta = \beta(\tau)$ for all $\tau \in [0, \tau^*]$ and Scheme 4.1 is second order.*

Proof. For τ, β in a neighborhood of $(0, 0)$, we define the real function

$$u(\tau, \beta) = F[\widehat{\Phi}^{n+1} + \beta w^n] - F[\Phi^n] + \tau(\mu_*^{n+\frac{1}{2}}, \mathbf{M}\mu_*^{n+\frac{1}{2}}). \tag{4.12}$$

It follows from (4.6)–(4.8) that

$$u(0,0)=0,\quad \frac{\partial u}{\partial \beta}(0,0)=\left(\frac{\delta F}{\delta \Phi}[\Phi^n],g[\Phi^n]\right)\neq 0. \tag{4.13}$$

According to the implicit function theorem, there exists a $\tau^*>0$ such that the equation $u(\tau,\beta)=0$ defines a unique smooth function $\beta=\beta(\tau)$ satisfying $\beta(0)=0$ and $u\big(\tau,\beta(\tau)\big)=0$ for all $\tau\in[0,\tau^*]$.

Since $\widehat{\Phi}^{n+1}$ satisfies the following scheme

$$\frac{\widehat{\Phi}^{n+1}-\Phi^n}{\tau}=-\mathbf{M}\left(\mathcal{L}\frac{\widehat{\Phi}^{n+1}+\Phi^n}{2}+\frac{\delta f}{\delta \Phi}[\Phi_*^{n+\frac{1}{2}}]\right), \tag{4.14}$$

through local truncation error analysis we have

$$\widehat{\Phi}^{n+1}=\Phi(t_{n+1})+\mathcal{O}(\tau^3),\quad F[\widehat{\Phi}^{n+1}]=F[\Phi(t_{n+1})]+\mathcal{O}(\tau^3), \tag{4.15}$$

where (4.5) was used. In addition, we expand

$$u(\tau,\beta)=u(\tau,0)+\beta\frac{\partial u}{\partial \beta}(\tau,0)+\mathcal{O}(\beta^2), \tag{4.16}$$

with

$$\begin{aligned} &u(\tau,0)=F[\widehat{\Phi}^{n+1}]-F[\Phi^n]+\tau(\mu_*^{n+\frac{1}{2}},\mathbf{M}\mu_*^{n+\frac{1}{2}})=\mathcal{O}(\tau^3),\\ &\frac{\partial u}{\partial \beta}(\tau,0)=\frac{\partial u}{\partial \beta}(0,0)+\mathcal{O}(\tau)=\left(\frac{\delta F}{\delta \Phi}[\Phi^n],g[\Phi^n]\right)+\mathcal{O}(\tau). \end{aligned} \tag{4.17}$$

Then $\beta=\beta(\tau)=\mathcal{O}(\tau^3)$ and the proposed scheme is of order 2. □

Remark 2. If we choose $g[\Phi]=-\mathbf{M}\frac{\delta f}{\delta \Phi}[\Phi]$, we have

$$\partial_t\Phi=-\mathbf{M}\left(\mathcal{L}\Phi+(1+\alpha)\frac{\delta f}{\delta \Phi}[\Phi]\right), \tag{4.18}$$

which can be regarded as a perturbation to the free energy or chemical potential of (1.1). In this particular implementation, the SVM reduces to the Lagrange multiplier approach proposed in [38]. If we choose $g[\Phi]=-\mathbf{M}_1\mu$, we have

$$\partial_t\Phi=-\widehat{\mathbf{M}}\mu, \tag{4.19}$$

where $\widehat{\mathbf{M}}=\mathbf{M}+\alpha\mathbf{M}_1$. In this case, the perturbation is done to the mobility. For convenience, we name the resulting schemes **SVM-I** and **SVM-II** obtained by the two choices of $g[\Phi]$, respectively. If not explicitly specified, we always set $\mathbf{M}_1=\mathbf{M}$ in the scheme **SVM-II** to conduct numerical experiments in the following section.

As we noted previously, the addition of the supplementary variables to the TCPDE system through perturbations can be fairly general. For instance, we could use simpler user supplied function $g[\Phi]=-\mathbf{M}\Phi$ to simplify the computation. What user specified perturbation can bring the most benefit to the scheme is an open problem worthy of a further investigation. It is clear that the SVM approach is more general and versatile in terms of enforcing an energy dissipation law and other invariants of the TCPDE system.

Remark 3. When the numerical discretization mentioned above is applied to the constrained optimization system (3.4), we can also obtain a numerical approximation similar to Scheme 4.1. The only difference is that we are going to solve for an optimization problem given below

$$\begin{cases} \min\limits_{\beta} \quad \beta^2,\\ \text{s.t.} \quad F[\widehat{\Phi}^{n+1}+\beta w^n]-F[\Phi^n]+\tau(\mu_*^{n+\frac{1}{2}},\mathbf{M}\mu_*^{n+\frac{1}{2}})=0, \end{cases} \tag{4.20}$$

where $\mu_*^{n+\frac{1}{2}}$, $\widehat{\Phi}^{n+1}$ and w^n are defined by (4.3), (4.7) and (4.8), respectively. According to Theorem 4.1, for $\tau<\tau^*$, the algebraic equation (4.11) is uniquely solvable, which implies system (4.20) is equivalent to (4.11). It is worth noting that the constrained optimization system (3.4) can provide an elegant platform for developing arbitrarily high-order energy dissipation rate preserving schemes, which will be discussed further in our future work.

Remark 4. In fact, there are a variety of numerical methods can be used on the SVM reformulated system to construct energy dissipation rate preserving schemes. In order to solve the problem efficiently, we note that the perturbed thermodynamically consistent model equation in the SVM reformulated system should be discretized implicitly in the linear terms and explicitly in nonlinear terms. In addition, some numerical tricks can be implemented in the partition of the linear vs nonlinear free energy components such as adding and subtracting certain quadratic terms to stabilize the numerical schemes etc.

For any given kth order linear algorithm of the SVM reformulated system, a same order linear algorithm can be devised so that $\Phi^{n+1} = \widehat{\Phi}^{n+1} + \beta w^n$, where $\widehat{\Phi}^{n+1}$ and w^n are solutions of two linear systems. This is the consequence of the fact $\beta \sim O(\tau^{k+1})$.

5. Thermodynamically consistent models with multiple deduced equations

For a given TCPDE system, there could be multiple pairs of mobility and free energy corresponding to it. Thus, there can exist multiple deduced energy laws corresponding to a given TCPDE system and even other deduced equations or conserved quantities. For conservative systems, it is well-known that they can have multiple conservation laws. For example, the KdV equation has infinitely many conservation laws. SVM provides a straightforward solution to preserve the multiple deduced equations in either dissipative laws or conservative laws at user's discretion. In this section, we use the KdV equation as an example to illustrate how to design SVM schemes to preserve one conserved quantity, two conserved quantities or three conserved quantities.

We consider the KdV equation with a periodic boundary condition

$$u_t + \eta u u_x + \mu^2 u_{xxx} = 0, \tag{5.1}$$

where η and μ are two real parameters. When $\eta = 1$, KdV equation (5.1) has infinite numbers of conservation laws. For convenience, we mainly focus on the following three invariants [41]

$$\mathcal{H}_1 = \int_\Omega \left(-\frac{1}{6}u^3 + \frac{1}{2}\mu^2 u_x^2\right) \mathrm{d}x, \tag{5.2}$$

$$\mathcal{H}_2 = \frac{1}{2}\int_\Omega u^2 \mathrm{d}x, \tag{5.3}$$

$$\mathcal{H}_3 = \int_\Omega \left(\frac{\mu^4}{2}|u_{xx}|^2 - \frac{5}{6}\mu^2 u|u_x|^2 + \frac{5}{72}|u|^4\right) \mathrm{d}x, \tag{5.4}$$

where $\mathcal{H}_1$ and $\mathcal{H}_2$ are known as the energy and momentum functional of the KdV equation, respectively.

To enforce the above conservation laws, we perturb the governing system of KdV by choosing three supplementary variables and user supplied functions g_i $(i = 1, 2, 3)$ following the SVM as follows

$$\begin{cases} u_t + \eta u u_x + \mu^2 u_{xxx} - \alpha_i g_i[u] = 0, \\ \frac{\mathrm{d}\mathcal{H}_i}{\mathrm{d}t} = 0, \end{cases} \tag{5.5}$$

where $i = 1, 2, 3$, respectively, and

$$\begin{cases} u_t + \eta u u_x + \mu^2 u_{xxx} - \alpha_1 g_1[u] - \alpha_2 g_2[u] - \alpha_3 g_3[u] = 0, \\ \frac{\mathrm{d}\mathcal{H}_1}{\mathrm{d}t} = 0, \\ \frac{\mathrm{d}\mathcal{H}_2}{\mathrm{d}t} = 0, \\ \frac{\mathrm{d}\mathcal{H}_3}{\mathrm{d}t} = 0. \end{cases} \tag{5.6}$$

In the following, we present four numerical schemes that preserve $\mathcal{H}_1$, $\mathcal{H}_2$ and $\mathcal{H}_3$, respectively or all simultaneously.

Scheme 5.1. *Given u^{n-1} and u^n, we obtain u^{n+1} via the following system:*

$$\begin{cases} \frac{u_*^{n+\frac{1}{2}} - u^n}{\tau/2} + \eta \overline{u}^{n+\frac{1}{2}} \partial_x \overline{u}^{n+\frac{1}{2}} + \mu^2 \partial_{xxx} u_*^{n+\frac{1}{2}} = 0, \\ \frac{u^{n+1} - u^n}{\tau} + \eta u_*^{n+\frac{1}{2}} \partial_x u_*^{n+\frac{1}{2}} + \mu^2 \partial_{xxx} u^{n+\frac{1}{2}} - \alpha_i^{n+\frac{1}{2}} g_i[u_*^{n+\frac{1}{2}}] = 0, \\ \mathcal{H}_i[u^{n+1}] = \mathcal{H}_i[u^n], \end{cases} \tag{5.7}$$

where $i = 1, 2, 3$ correspond to the scheme **SVM-$\mathcal{H}_1(\alpha_1)$**, **SVM-$\mathcal{H}_2(\alpha_2)$** *and* **SVM-$\mathcal{H}_3(\alpha_3)$** *that preserve $\mathcal{H}_1$, $\mathcal{H}_2$ and $\mathcal{H}_3$, respectively.*

The following scheme preserves all the three invariants simultaneously.

Scheme 5.2 (*SVM-all(α_1, α_2, α_3)*)**.** *Given u^{n-1} and u^n, we obtain u^{n+1} via the following system:*

$$\begin{cases} \frac{u_*^{n+\frac{1}{2}}-u^n}{\tau/2} + \eta\overline{u}^{n+\frac{1}{2}}\partial_x\overline{u}^{n+\frac{1}{2}} + \mu^2\partial_{xxx}u_*^{n+\frac{1}{2}} = 0, \\ \frac{u^{n+1}-u^n}{\tau} + \eta u_*^{n+\frac{1}{2}}\partial_x u_*^{n+\frac{1}{2}} + \mu^2\partial_{xxx}u^{n+\frac{1}{2}} - \alpha_1^{n+\frac{1}{2}}g_1[u_*^{n+\frac{1}{2}}] - \alpha_2^{n+\frac{1}{2}}g_2[u_*^{n+\frac{1}{2}}] - \alpha_3^{n+\frac{1}{2}}g_3[u_*^{n+\frac{1}{2}}] = 0, \\ \mathcal{H}_1[u^{n+1}] = \mathcal{H}_1[u^n], \\ \mathcal{H}_2[u^{n+1}] = \mathcal{H}_2[u^n], \\ \mathcal{H}_3[u^{n+1}] = \mathcal{H}_3[u^n]. \end{cases}$$

These schemes are second order in time. A theorem analogous to the one given previously on the order of α_i, $i = 1, 2, 3$ in terms of the time step can be proved. But, we will not rephrase it here. We next discuss the spatial discretization of the semidiscrete schemes.

6. Spatial discretization

We use the Fourier pseudo-spectral method to discretize the semidiscrete systems in 2 dimensional space. For simplicity, we use the periodic boundary condition in the implementations. Let N_x and N_y be positive integers. We consider domain $\Omega = [0, L_x] \times [0, L_y]$ with the spatial step size $h_x = L_x/N_x$ and $h_y = L_y/N_y$. We denote the collocation points as $(x_j, y_k) = (jh_x, kh_y)$, where $0 \le j \le N_x - 1$ and $0 \le k \le N_y - 1$. Let $\Phi_{j,k}^n$ be an approximation of $\Phi(x_j, y_k, t_n)$, where $t_n = n\tau$.

Given $\{\Phi_{j,k}^n | j = 0, \ldots, N_x - 1, k = 0, \ldots, N_y - 1\}$, we denote

$$S_N = \{X_j(x)Y_k(y),\ j = 0, \ldots, N_x - 1;\ k = 0, \ldots, N_y - 1\} \tag{6.1}$$

as the interpolation space with $X_j(x)$ and $Y_k(y)$ given by

$$X_j(x) = \frac{1}{N_x}\sum_{p=-N_x/2}^{N_x}\frac{1}{a_p}e^{\mathrm{i}p\mu_x(x-x_j)}, \quad Y_k(y) = \frac{1}{N_y}\sum_{q=-N_y/2}^{N_y}\frac{1}{b_q}e^{\mathrm{i}q\mu_y(y-y_k)}, \tag{6.2}$$

where $\mu_x = 2\pi/L_x$, $\mu_y = 2\pi/L_y$, $a_p = \begin{cases} 1, & |p| < N_x/2, \\ 2, & |p| = N_x/2, \end{cases}$ and $b_q = \begin{cases} 1, & |q| < N_y/2, \\ 2, & |q| = N_y/2. \end{cases}$ We define interpolation operator $I_N : C(\Omega) \to S_N$ as follows:

$$I_N\Phi(x, y) = \sum_{j=0}^{N_x-1}\sum_{k=0}^{N_y-1}\Phi_{j,k}X_j(x)Y_k(y). \tag{6.3}$$

Then, we differentiate (6.3) and evaluate the resulting expression at collocation point (x_j, y_k) as follows

$$\partial_x^{s_1}\partial_y^{s_2}I_N\Phi(x_j, y_k) = \sum_{p=0}^{N_x-1}\sum_{q=0}^{N_y-1}\Phi_{p,q}(\mathbf{D}_{s_1}^x)_{j,p}(\mathbf{D}_{s_2}^y)_{k,q}, \tag{6.4}$$

where $\mathbf{D}_{s_1}^x$ and $\mathbf{D}_{s_2}^y$ are $N_x \times N_x$ and $N_y \times N_y$ matrices, respectively, with entries given by

$$(\mathbf{D}_{s_1}^x)_{j,p} = \frac{\mathrm{d}^{s_1}X_p(x_j)}{\mathrm{d}x^{s_1}}, \quad (\mathbf{D}_{s_2}^y)_{k,q} = \frac{\mathrm{d}^{s_2}Y_q(y_k)}{\mathrm{d}y^{s_2}}. \tag{6.5}$$

Thus, the continuous operator Δ is approximated by the discrete operator Δ_h as follows:

$$\Delta_h\Phi = \mathbf{D}_2^x\Phi + \Phi(\mathbf{D}_2^y)^T. \tag{6.6}$$

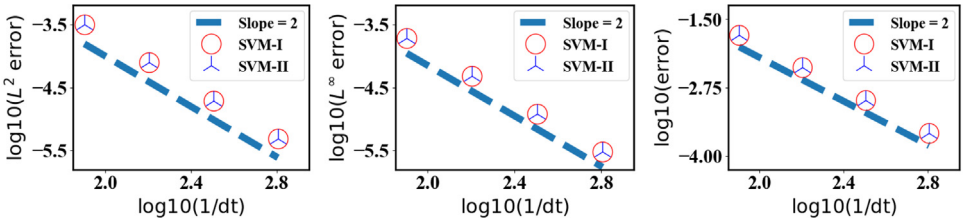


Fig. 1. Mesh refinement test in time for the two schemes. Left: the L^2 error for ϕ; Middle: the L^∞ error for ϕ; Right: the error plots of supplementary variable $\alpha(t)$. Here, we fix the number of spatial meshes at $N_x = N_y = 256$. Second order convergence rates are confirmed.

It is worth noting that the $\mathbf{D}_2^a$ ($a = x$ or y) can be calculated by the discrete fast Fourier transform (FFT) as follows

$$\mathbf{D}_2^a = F_{N_a}^{-1} \mathbf{\Lambda}_a^2 F_{N_a}, \text{ with } \mathbf{\Lambda}_a^2 = \left[\mathrm{i}\mu_a \mathrm{diag}\left(0, \ldots, \frac{N_a}{2} - 1, \frac{N_a}{2}, -\frac{N_a}{2} + 1, \ldots, -1\right)\right]^2. \tag{6.7}$$

Applying the pseudo-spectral discretization, we obtain the fully discrete numerical scheme of (4.2) and (4.3), where $\mathcal{L}$ is set as $-\Delta$ for the sake of presentation.

$$\frac{\Phi_*^{n+\frac{1}{2}} - \Phi^n}{\tau/2} = -\mathbf{M}\left(-\Delta_h \Phi_*^{n+\frac{1}{2}} + \frac{\delta f}{\delta \Phi}[\overline{\Phi}^{n+\frac{1}{2}}]\right), \tag{6.8}$$

and

$$\delta_t^+ \Phi^n = -\mathbf{M}\left(-\Delta_h \Phi^{n+\frac{1}{2}} + \frac{\delta f}{\delta \Phi}[\Phi_*^{n+\frac{1}{2}}] + \alpha g_2[\Phi_*^{n+\frac{1}{2}}]\right) + \alpha g[\Phi_*^{n+\frac{1}{2}}]. \tag{6.9}$$

Remark 5. For the KdV equation, we discretize ∂_x, ∂_{xx} and ∂_{xxx} by spectral differential matrices D_1^x, D_2^x and D_3^x, where $D_3^x = (D_1^x)^3$. D_k^x (when k odd) can be computed efficiently using FFT,

$$\mathbf{D}_k^x = F_{N_x}^{-1} \mathbf{\Lambda}_x^k F_{N_x}, \text{ with } \mathbf{\Lambda}_x^k = \left[\mathrm{i}\mu_x \mathrm{diag}\left(0, \ldots, \frac{N_a}{2} - 1, 0, -\frac{N_a}{2} + 1, \ldots, -1\right)\right]^k. \tag{6.10}$$

7. Numerical results

We implement the SVM on some selected phase field models with periodic boundary conditions for testing its accuracy and efficiency on some benchmark examples. The choice of periodic boundary condition is for convenience and the use of fast solvers. Physical boundary conditions can be handled without problems. We begin with the Cahn–Hilliard equation for a binary material system.

7.1. Cahn-Hilliard Equation

Example 1 (*Cahn–Hilliard Model*)**.** We consider the following Cahn–Hilliard model with phase variable ϕ

$$\partial_t \phi = \lambda \Delta \left(-\varepsilon^2 \Delta \phi + \phi^3 - \phi\right),$$

where the free energy is given by

$$F = \frac{\varepsilon^2}{2} \|\nabla \phi\|^2 + \frac{1}{4}(\phi^2 - 1)^2.$$

First of all, we present the mesh refinement test in time to confirm the order of accuracy of the schemes. To this end, we make the following function an exact solution of the system modified by an appropriate forcing function: $\phi(x, y, t) = \sin(x)\sin(y)\cos(t)$. We set the computational domain as $\Omega = [0, 2\pi]^2$ and the parameter values as $\varepsilon = 10^{-2}$ and $\lambda = 10^{-3}$, respectively. This model is discretized spatially using a Fourier pseudo-spectral method with 256×256 spatial meshes. Both the discrete L^2 and L^∞ errors of numerical solution ϕ and the plot of the supplementary variable $\alpha(t)$ at $t = 1$ are shown in Fig. 1, where we observe that the proposed schemes yield second order convergence rates in time. Moreover, the third-order accuracy for $\beta(t) = \tau\alpha(t)$ is reached as well.

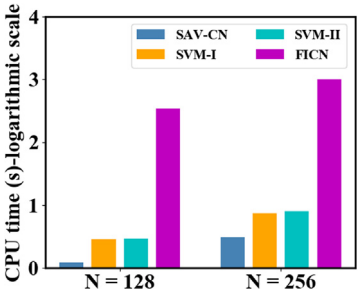


Fig. 2. Comparison of CPU time in logarithmic scale using the four different numerical methods with two sets of spatial meshes up to $t = 1$, where the time step is chosen as $\tau = 1.0e{-}2$. The charts show that the two proposed schemes perform equally well, more slowly than **SAV-CN**, but much faster than **FICN**.

Next, we examine the computational efficiency of the two schemes: **SVM-I** and **SVM-II**. We compare the two proposed schemes with the SAV scheme using the Crank–Nicolson method **SAV-CN** [23] and the fully implicit Crank–Nicolson scheme **FICN** [36]. The result in the total CPU time to solve for the solution using each scheme is summarized in Fig. 2. We observe that **SVM-I** and **SVM-II** for this model are less efficient than scheme **SAV-CN**, but much more efficient than scheme **FICN**. The price we pay using the proposed schemes in CPU time is that we have to solve a scalar nonlinear equation at each time step. In contrast, **SAV** scheme solves a linear system while **FICN** solves a nonlinear system.

Finally, we compare **SVM-I**, **SVM-II** and **SAV-CN** in their accuracy in resolving the energy-dissipation-rate. We do it using a spatial discretization on 128^2 meshes in $\Omega = [0, 1]^2$. The parameter values are $\lambda = 1$ and $\varepsilon = 1.0e{-}2$. We use the following initial condition [32]

$$\phi(x, y, 0) = 0.05\left(\cos(6\pi x)\cos(8\pi y) + (\cos(8\pi x)\cos(6\pi y))^2 + \cos(2\pi x - 10\pi y)\cos(4\pi x - 2\pi y)\right).$$

The initial profile undergoes a fast coarsening dynamics such that the algorithms must use small time steps in order to capture the correct coarsening dynamics. The simulation results are depicted in Fig. 3, where the snapshots of $\phi(x, y, t)$ at $t = 0.1$ are shown using different schemes and time steps. We observe that scheme **SAV-CN** predicts the correct solution at time step $\tau = 1.5625e{-}6$ but fails at time step $\tau = 3.125e{-}6$. For **SVM-I**, it predicts correct numerical result at time step $\tau = 5.0e{-}5$, and **SVM-II** performs even better with time step $\tau = 2.0e{-}4$. The errors in the total volume and the energy are plotted in Fig. 4. The numerical results show that both schemes conserve the total volume and capture the energy-dissipation-rate correctly using relatively larger time steps compared to the **SAV** scheme. In addition, the supplementary variable $\alpha(t)$ remains close to zero except at a few initial time spots. This numerical experiment demonstrates that among the overall numerical performance of the schemes, the SVM schemes tend to be more efficient allowing larger time steps.

7.2. Allen-Cahn equation

Example 2 (*Allen–Cahn Equation*)**.** Another type of gradient flow models is the Allen-Cahn equation. Here we consider the Swift–Hohenberg equation for crystal growth [42] with the free energy given by

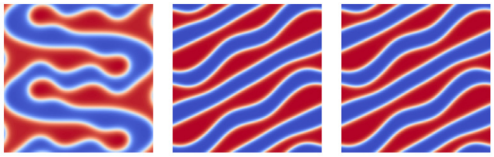
$$F = \int_\Omega \left(\frac{1}{4}\phi^4 - \frac{g}{3}\phi^3 + \frac{1-\epsilon}{2}\phi^2 - |\nabla\phi|^2 + \frac{1}{2}|\Delta\phi|^2\right) d\mathbf{x}, \tag{7.1}$$

where ϵ and g are constant parameters. The dynamical equation in the Swift–Hohenberg model is given by an Allen-Cahn equation:

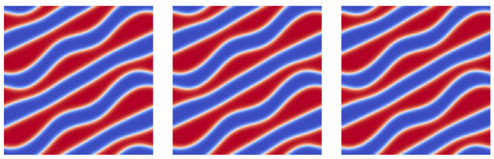
$$\phi_t(\mathbf{x}, t) = -\lambda\left(\phi^3 - g\phi^2 + (1-\epsilon)\phi + 2\Delta\phi + \Delta^2\phi\right), \quad \mathbf{x} \in \Omega. \tag{7.2}$$

where λ is the constant mobility coefficient.

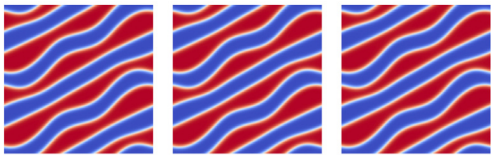
Here, we first show the accuracy test for the numerical solution. We take all parameter values and the exact solution the same as the one in the previous example and $g = 0$. In Fig. 5, we compare the numerical solution with the exact one at $t = 1$, and compute the discrete L^2 and L^∞ errors by choosing time steps at $\tau = 1.25e{-}2$, $6.25e{-}3$, $3.125e{-}3$, $1.5625e{-}3$, respectively, where we fix $N_x = N_y = 256$.



(a) ϕ at $t = 0.1$ using **SAV-CN** with time steps: $\tau = 3.125e-6$, $\tau = 1.5625e-6$ and $\tau = 1.0e-6$ (reference) (from left to right).



(b) ϕ at $t = 0.1$ using **SVM-I** with time steps: $\tau = 5.0e-5$, $\tau = 4.0e-5$ and $\tau = 1.0e-6$ (reference) (from left to right).



(c) ϕ at $t = 0.1$ using **SVM-II** with time steps: $\tau = 2.0e-4$, $\tau = 1.5625e-4$ and $\tau = 1.0e-6$ (reference) (from left to right).

Fig. 3. Comparison of three schemes with respect to different time steps. The first sub-figure in each row indicates the "maximum possible" time step to predict correct dynamics. These snapshots show that both schemes **SVM-I** and **SVM-II** allow larger time steps than **SAV-CN** and **SVM-II** allows the largest one.

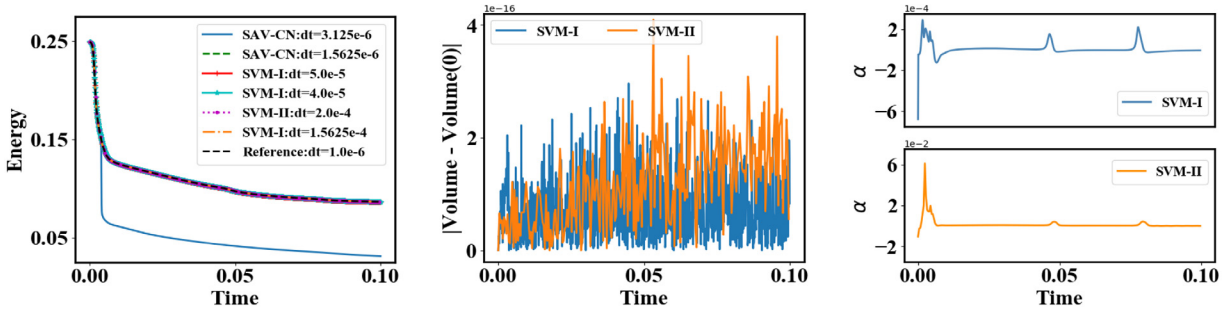


Fig. 4. (a) Time evolution of the free energy computed using the three schemes (**SVM-I, SVM-II, SAV-CN**) with various time steps. The subfigure shows that **SVM-I** and **SVM-II** can work well with much larger time steps than **SAC-CN** does, while **SVM-II** performs the best. (b) Time evolution of the error in the total volume using **SVM-I** and **SVM-II** with $\tau = 5.0e-5$ and $\tau = 2.0e-4$, respectively. The results show that the proposed schemes preserve the total volume very well. (c) The evolution of supplementary variable α. This subfigure indicates α may fluctuate near zero initially, but eventually, settles down close to zero.

To study energy dissipation rate preserving property of the Swift–Hohenberg model, we consider the following initial condition

$$\phi(x, y, 0) = 0.07 - 0.02\cos\left(\frac{2\pi(x-12)}{32}\right)\sin\left(\frac{2\pi(y-1)}{32}\right) + 0.02\cos^2\left(\frac{\pi(x+10)}{32}\right)\sin^2\left(\frac{\pi(y+3)}{32}\right) - 0.01\sin^2\left(\frac{4\pi x}{32}\right)\sin^2\left(\frac{4\pi(y-6)}{32}\right).$$

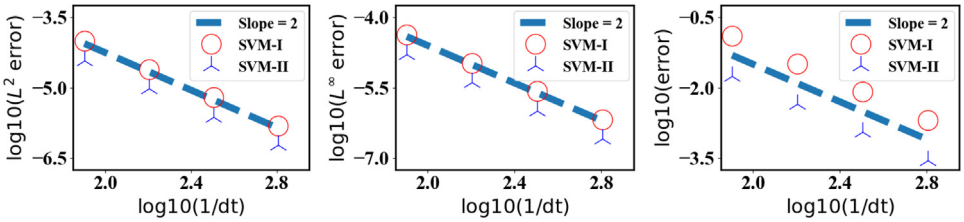
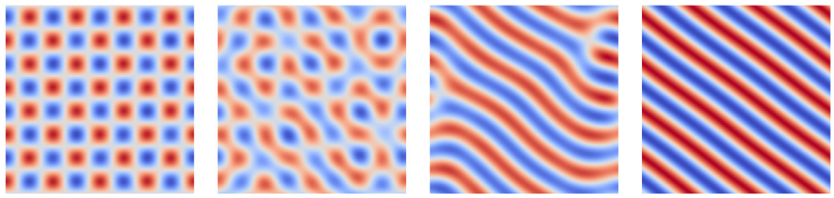
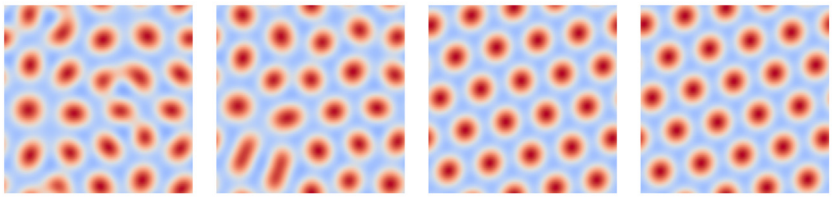


Fig. 5. Example 2: The discrete L^2 (left) and L^∞ (middle) errors, as well as the error plots of supplementary variable $\alpha(t)$ (right) of the temporal refinement test. The results show a convergence rate of 2 as expected.



(a) Evolution of ϕ at $t = 60$, 120, 240 and 480 for $g = 0$. (from left to right).



(b) Evolution of ϕ at $t = 60$, 120, 240 and 480 for $g = 1$. (from left to right).

Fig. 6. Example 2: The snapshots of phase variable ϕ computed using **SVM-I** at different times with $\epsilon = 0.25$, $\tau = 1.0e{-}2$ and $N_x = N_y = 256$.

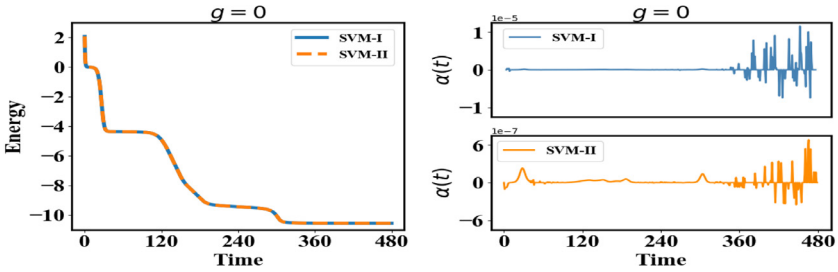
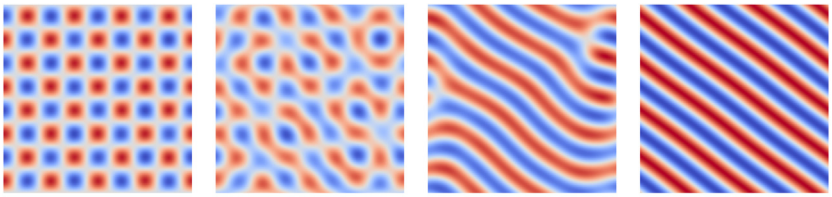


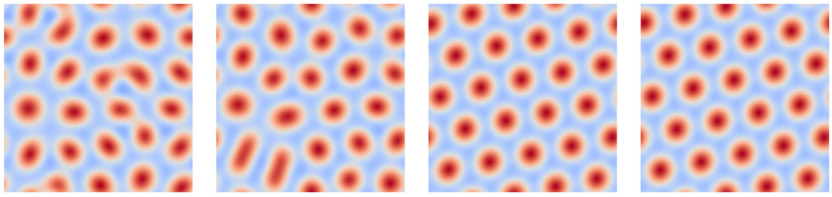
Fig. 7. Example 2: Time evolution of the total free energy computed using **SVM-I/II** for $g = 0$.

This time, we consider domain $\Omega = [0, 32]^2$ with 256×256 spatial meshes. We choose the parameter values as $\epsilon = 0.25$ and $\lambda = 1$. The profile of phase variable ϕ at different times for $g = 0$ and $g = 1$ are plotted in Figs. 6 and 8, respectively, where values of g are shown to influence the coarsening dynamics. In Figs. 7 and 9, for $g = 0$ and $g = 1$, respectively, we plot the free energy and the supplementary variable $\alpha(t)$ as functions of time with $\tau = 0.01$ up to $t = 480$. We observe clearly that two different values of g affect the energy dissipation of the free energy. In addition, $\alpha(t)$ fluctuates within a certain range during the course of the simulations without demonstrating any apparent correlation with the change of the free energy.

Finally, we investigate pattern formation in the model with respect to model parameters. As we have shown, the phase-field crystal growth model can produce different patterns such as the striped and hexagonal pattern, depending on parameter values ϵ and $\overline{\phi}$ even at $g = 0$. Here, we examine pattern formation using the model with a random



(a) Evolution of ϕ at $t = 60$, 120, 240 and 480 for $g = 0$. (from left to right).



(b) Evolution of ϕ at $t = 60$, 120, 240 and 480 for $g = 1$. (from left to right).

Fig. 8. Example 2: The profiles of phase variable ϕ using **SVM-II** at different times with $\epsilon = 0.25$, $\tau = 1.0e{-}2$ and $N_x = N_y = 256$.

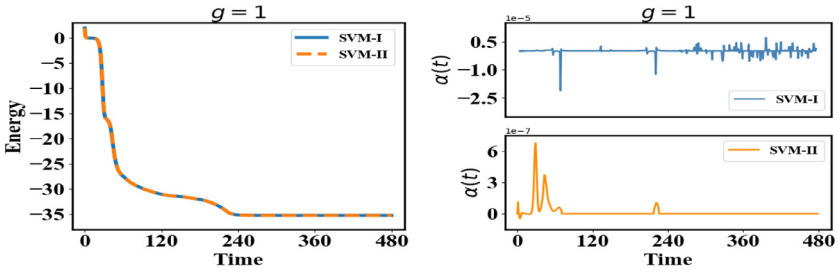


Fig. 9. Example 2: Time evolution of the free energy calculated using **SVM-I/II** for $g = 1$.

initial state. We generate the initial condition $\phi(x, y, 0) = \overline{\phi} + 0.1\text{rand}(x, y)$, where $\text{rand}(x, y)$ denotes a random number in the range of $[-1, 1]$. We choose the domain as $\Omega = [0, 32] \times [0, 32]$ with 128^2 spatial meshes. In Fig. 10, numerical solutions at $t = 1000$ are shown with $\tau = 0.1$ for different values of ϵ, $\overline{\phi}$ and g. By comparison, we observe that pattern formation in the model is mainly affected by g rather than by $\overline{\phi}$ once ϵ is given.

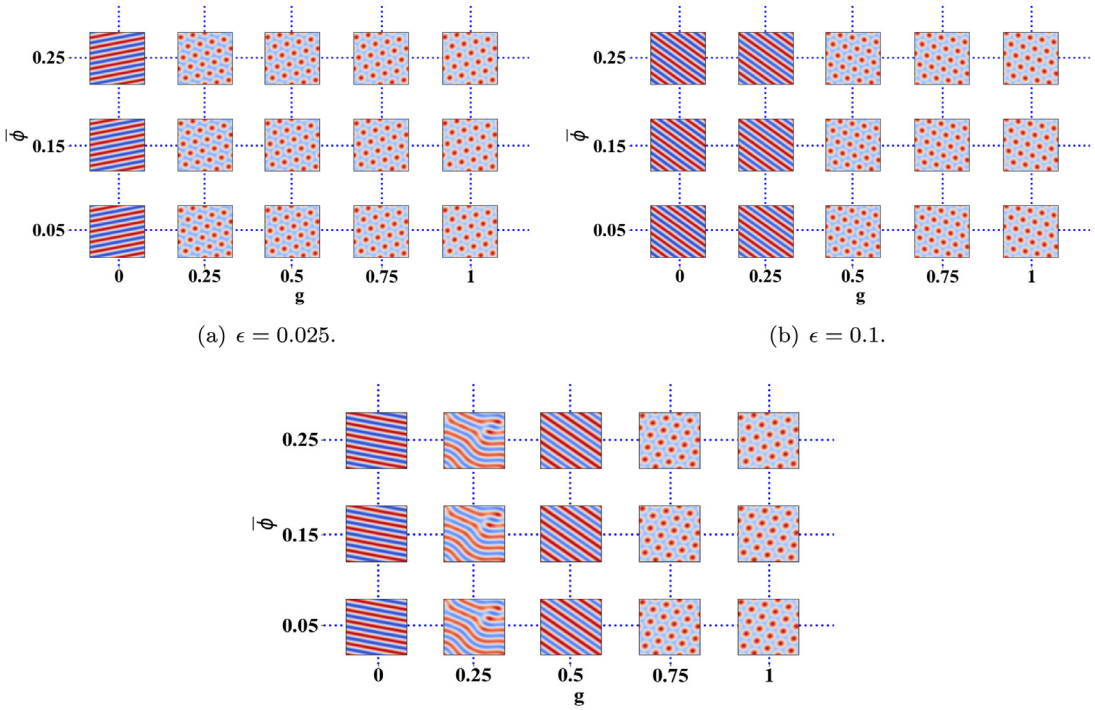
7.3. Preserving multiple invariants in a conservative system

The implementation of the SVM to thermodynamically consistent models with multiple invariants can be very simple, whereas most existing structure-preserving numerical algorithms can preserve one invariant but fail to preserve the others. In principle, the SVM can preserve as many invariants or deduced equations as one wishes. Here, we illustrate the capability of the SVM's using a conservation KdV equation by developing multiple invariant preserving schemes.

Example 3 (*Multiple Solitary Waves in KdV Equation*)**.** In this example, we consider the KdV equation (5.1) with the following initial condition

$$u(x, 0) = \cos(\pi x). \tag{7.3}$$

The model parameter values are taken as $\eta = 1$ and $\mu = 0.022$. We solve the KdV model in a periodic domain $\Omega = [0, 2]$ using a pseudo-spectral method in space with $N_x = 256$. For convenience, we choose $g_1 = \delta\mathcal{H}_1/\delta u$, $g_2 = \delta\mathcal{H}_2/\delta u$ and $g_3 = \delta\mathcal{H}_3/\delta u$ to conduct the simulations. The numerical results obtained using the three schemes are summarized in Figs. 11 and 12. We remark that these numerical phenomena are consistent with those reported



(a) $\epsilon = 0.025$. (b) $\epsilon = 0.1$.

(c) $\epsilon = 0.25$.

Fig. 10. Example 2: Phase diagram of near steady state profiles of ϕ at $t = 1000$ with three different values of ϵ in phase space $(g, \bar{\phi})$.

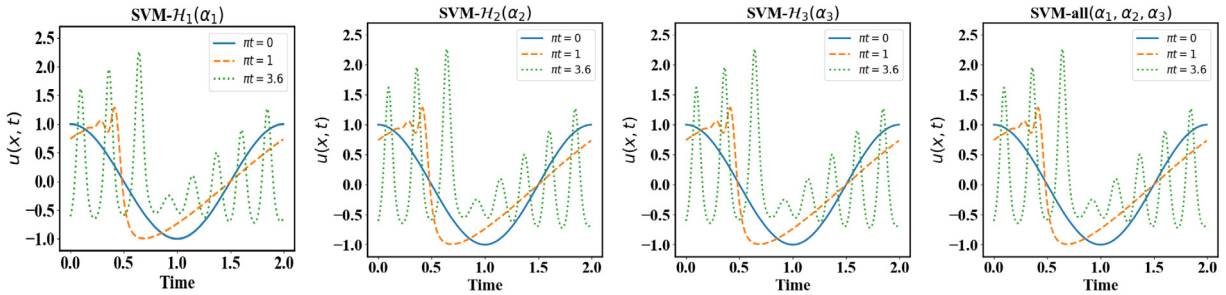
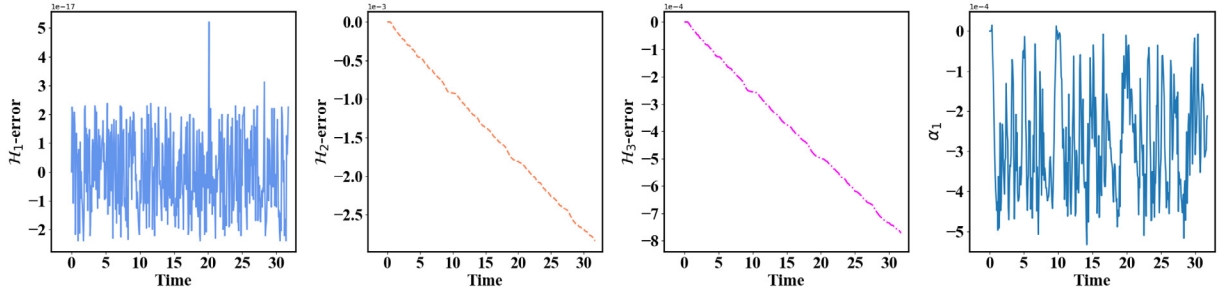


Fig. 11. Example 3: The propagation and interaction of multiple solitary waves computed using four schemes, where $\tau = 0.005/\pi$ and $h_x = 2/400$. We denote the waveform for $\pi t = 0$ by the solid curve, for $\pi t = 1$ by the dashdot curve, and for $\pi t = 3.6$ by the dotted curve.

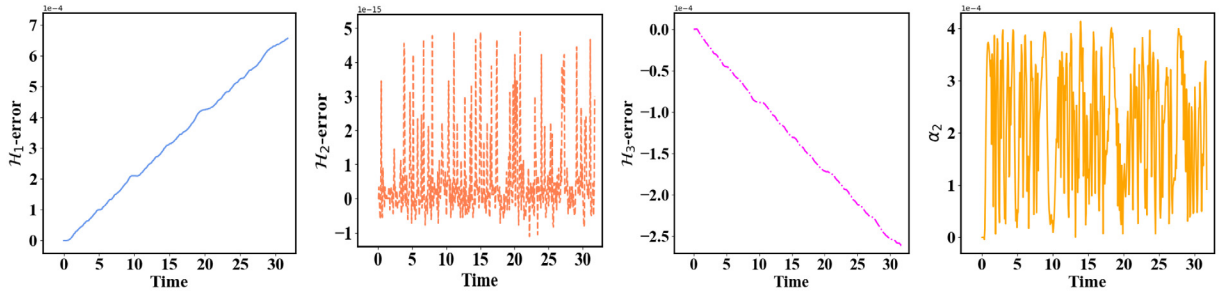
in the literature [43]. We also observe that **SVM-**$\mathcal{H}_1(\alpha_1)$, **SVM-**$\mathcal{H}_2(\alpha_2)$ and **SVM-**$\mathcal{H}_3(\alpha_3)$ preserve the targeted invariant very well, while **SVM-all**(α_1, α_2, α_3) simultaneously preserves the three invariants very accurately. This example demonstrates that the SVM as a structure preserving paradigm for models with deduced equations is very effective and yet simple to implement.

8. Conclusions

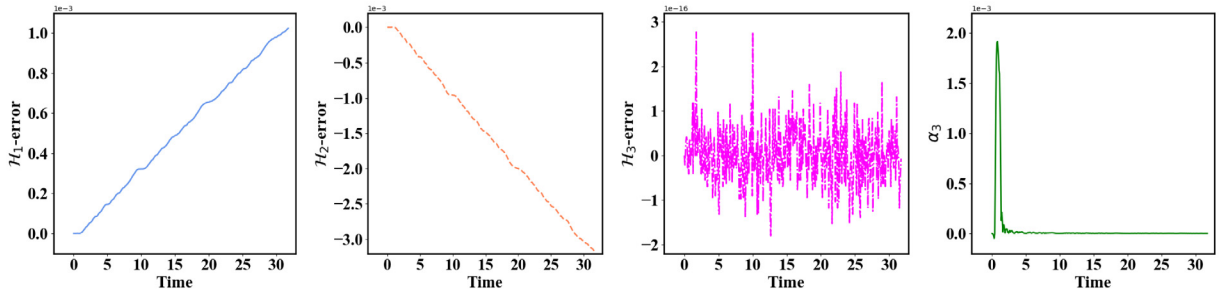
We present a new perspective on developing structure-preserving algorithms for thermodynamically consistent models based on the supplementary variable method. By introducing some supplementary variables, the original model coupled with deduced equations (physical laws, invariants etc.) is first transformed into a structural stable, extended system, which provides a robust mathematical platform for developing structure-preserving algorithms. The development of the extended system can be very flexible so long as the resulting system is well-determined,



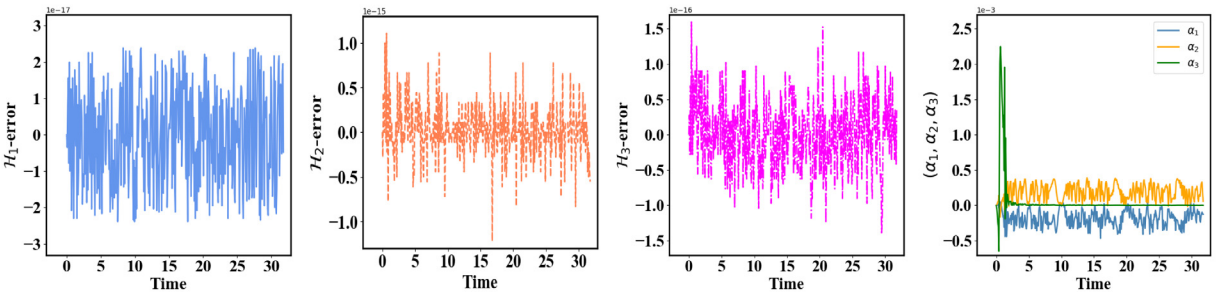
(a) **SVM-$\mathcal{H}_1(\alpha_1)$.**



(b) **SVM-$\mathcal{H}_2(\alpha_2)$.**



(c) **SVM-$\mathcal{H}_3(\alpha_3)$.**



(d) **SVM-all(α_1, α_2, α_3).**

Fig. 12. Example 3: (a) The errors in the $\mathcal{H}_1$, $\mathcal{H}_2$, $\mathcal{H}_3$ and supplementary variable α_1 using **SVM-$\mathcal{H}_1(\alpha_1)$**. (b) The errors in the $\mathcal{H}_1$, $\mathcal{H}_2$, $\mathcal{H}_3$ and supplementary variable α_2 using **SVM-$\mathcal{H}_2(\alpha_2)$**. (c) The errors in the $\mathcal{H}_1$, $\mathcal{H}_2$, $\mathcal{H}_3$ and supplementary variable α_3 using **SVM-$\mathcal{H}_3(\alpha_3)$**. (d) The errors in the $\mathcal{H}_1$, $\mathcal{H}_2$, $\mathcal{H}_3$ and supplementary variables $(\alpha_1, \alpha_2, \alpha_3)$ using **SVM-all**(α_1, α_2, α_3).

consistent, solvable, and structural stable. Discretization from the extended system to a fully discretized system can be carried out without any additional constraints so long as it is consistent. The numerical results demonstrate that

the proposed numerical schemes based on the supplementary variable approach can predict coarsening dynamics of the Cahn–Hilliard model accurately and outperform **SAV-CN** in solution accuracy and **FICN** in efficiency. Here we simply present two convenient implementations of the SVM when developing energy-dissipation-rate preserving schemes for TCPDEs. There can be many other ways to implement the SVM guided by the paradigm to achieve energy stability, better computational efficiency and accuracy. In addition, it is very easy to develop the numerical schemes in preservation of single or multiple deduced equations or invariants. High-order structure-preserving algorithms using the SVM can be developed and will be reported in a sequel.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

‘

Acknowledgments

Research is partially supported by the Foundation of Jiangsu Key Laboratory for Numerical Simulation of Large Scale Complex Systems, China (202001, 202002), by the China Postdoctoral Science Foundation through Grant 2020M670116, the Natural Science Foundation of Jiangsu Province, China (award BK20180413) and the National Natural Science Foundation of China (award 11801269, 12071216, 11971051 and NSAF-U1930402). Qi Wang's research is partially supported by National Science Foundation of US (award DMS-1815921 and OIA-1655740), DOE DE-SC0020272 award and a GEAR award from SC EPSCoR/IDeA Program.

References

[1] L. Onsager, Reciprocal relations in irreversible processes. I, Phys. Rev. 37 (1931) 405–426.
[2] L. Onsager, Reciprocal relations in irreversible processes. II, Phys. Rev. 38 (1931) 2265–2279.
[3] X. Yang, J. Li, G. Forest, Q. Wang, Hydrodynamic theories for flows of active liquid crystals and the Generalized Onsager Principle, Entropy 18 (2016) 202.
[4] K. Feng, M. Qin, Symplectic Geometric Algorithms for Hamiltonian Systems, Springer Science & Business Media, 2010.
[5] E. Hairer, C. Lubich, G. Wanner, Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations, Vol. 31, Springer, 2006.
[6] O. Gonzalez, Time integration and discrete Hamiltonian systems, J. Nonlinear Sci. 6 (1996) 449–467.
[7] R. McLachlan, G. Quispel, N. Robidoux, Geometric integration using discrete gradients, Phil. Trans. R. Soc. A 357 (1999) 1021–1045.
[8] D. Furihata, T. Matsuo, in: C. Lai, F. Magoules (Eds.), Discrete Variational Derivative Method: A Structure-Preserving Numerical Method for Partial Differential Equations, CRC Press, 2010.
[9] M. Dahlby, B. Owren, A general framework for deriving integral preserving numerical methods for PDEs, SIAM J. Sci. Comput. 33 (5) (2011) 2318–2340.
[10] G.R. Quispel, D.I. McLaren, A new class of energy-preserving numerical integration methods, J. Phys. A 41 (2008) 045206.
[11] E. Celledoni, V. Grimm, R. McLachlan, D. McLaren, D. O'Neale, B. Owren, G. Quispel, Preserving energy resp. dissipation in numerical PDEs using the "average vector field" method, J. Comput. Phys. 231 (2012) 6770–6789.
[12] C. Wang, X. Wang, S. Wise, Unconditionally stable schemes for equations of thin film epitaxy, Discrete Contin. Dyn. Syst. 28 (1) (2010) 405–423.
[13] C. Wang, S. Wise, An energy stable and convergent finite-difference scheme for the modified phase field crystal equation, SIAM J. Numer. Anal. 49 (2011) 945–969.
[14] J. Shen, C. Wang, X. Wang, S. Wise, Second-order convex splitting schemes for gradient flows with Ehrlich-Schwoebel type energy: application to thin film epitaxy, SIAM J. Numer. Anal. 50 (2012) 105–125.
[15] J. Shin, H. Lee, J. Lee, Unconditionally stable methods for gradient flow using convex splitting Runge-Kutta scheme, J. Comput. Phys. 347 (2017) 367–381.
[16] D. Li, Z. Qiao, T. Tang, Characterizing the stabilization size for semi-implicit Fourier-spectral method to phase field equations, SIAM J. Numer. Anal. 54 (2016) 1653–1681.
[17] J. Shen, X. Yang, Numerical approximations of Allen-Cahn and Cahn-Hilliard equations, Discrete Contin. Dyn. Syst. A 28 (2010) 1669–1691.
[18] X. Feng, T. Tang, J. Yang, Stabilized Crank-Nicolson and Adams-Bashforth schemes for phase field models, East Asian J. Appl. Math. 3 (2013) 59–80.
[19] S. Badia, F. Guillén-González, J.V. Gutiérrez-Santacreu, Finite element approximation of nematic liquid crystal flows using a saddle-point structure, J. Comput. Phys. 230 (2011) 1686–1706.
[20] F. Guillén-González, G. Tierra, On linear schemes for a Cahn-Hilliard diffuse interface model, J. Comput. Phys. 234 (2013) 140–171.
[21] X. Yang, J. Zhao, Q. Wang, Numerical approximations for the molecular beam epitaxial growth model based on the invariant energy quadratization method, J. Comput. Phys. 333 (2017) 104–127.

[22] J. Zhao, Q. Wang, X. Yang, Numerical approximations for a phase field dendritic crystal growth model based on the invariant energy quadratization approach, Internat. J. Numer. Methods Engrg. 110 (2017) 279–300.
[23] J. Shen, J. Xu, J. Yang, The scalar auxiliary variable (SAV) approach for gradient flows, J. Comput. Phys. 353 (2018) 407–416.
[24] J. Zhao, X. Yang, Y. Gong, X. Zhao, X. Yang, J. Li, Q. Wang, A general strategy for numerical approximations of non-equilibrium models–part i thermodynamical systems, Int. J. Numer. Anal. Model. 15 (2018) 884–918.
[25] J. Shen, J. Xu, J. Yang, A new class of efficient and robust energy stable schemes for gradient flows, SIAM Rev. 61 (2019) 474–506.
[26] L. Chen, J. Zhao, X. Yang, Regularized linear schemes for the molecular beam epitaxy model with slope selection, Appl. Numer. Math. 128 (2018) 138–156.
[27] Z. Xu, X. Yang, H. Zhang, Z. Xie, Efficient and linear schemes for anisotropic Cahn-Hilliard model using the stabilized-invariant energy quadratization (s-IEQ) approach, Comput. Phys. Comm. 238 (2019) 36–49.
[28] D. Hou, M. Azaiez, C. Xu, A variant of scalar auxiliary variable approaches for gradient flows, J. Comput. Phys. 395 (2019) 307–332.
[29] Z. Yang, S. Dong, A roadmap for discretely energy-stable schemes for dissipative systems based on a generalized auxiliary variable with guaranteed positivity, J. Comput. Phys. 404 (2020) 109121.
[30] Y. Gong, J. Zhao, Q. Wang, Arbitrarily high-order unconditionally energy stable schemes for thermodynamically consistent gradient flow models, SIAM J. Sci. Comput. 42 (2020) B135–B156.
[31] Y. Gong, J. Zhao, Energy-stable Runge-Kutta schemes for gradient flow models using the energy quadratization approach, Appl. Math. Lett. 94 (2019) 224–231.
[32] Y. Gong, J. Zhao, Q. Wang, Arbitrarily high-order linear energy stable schemes for gradient flow models, J. Comput. Phys. 419 (2020) 109610.
[33] G. Akrivis, B. Li, D. Li, Energy-decaying extrapolated RK-SAV methods for the Allen-Cahn and Cahn-Hilliard equations, SIAM J. Sci. Comput. 41 (2019) A3703–A3727.
[34] Y. Gong, J. Zhao, Q. Wang, Arbitrarily high-order unconditionally energy stable SAV schemes for gradient flow models, Comput. Phys. Comm. 249 (2020) 107033.
[35] J. Xu, Y. Li, S. Wu, A. Bousquet, On the stability and accuracy of partially and fully implicit schemes for phase field modeling, Comput. Methods Appl. Mech. Engrg. 345 (2019) 826–853.
[36] Q. Du, R. Nicolaides, Numerical analysis of a continuum model of phase transition, SIAM J. Numer. Anal. 28 (1991) 1310–1322.
[37] Z. Zhang, Z. Qiao, An adaptive time-stepping strategy for the Cahn-Hilliard equation, Commun. Comput. Phys. 11 (2012) 1261–1278.
[38] Q. Cheng, C. Liu, J. Shen, A new Lagrange multiplier approach for gradient flows, Comput. Methods Appl. Mech. Engrg. 367 (2020) 113070.
[39] M. Calvo, M.P. Laburta, J.I. Montijano, L. Randez, Projection methods preserving Lyapunov functions, BIT Numer. Math. 50 (2010) 223–241.
[40] M. Calvo, D. Hernandez-Abreu, J.I. Montijano, L. Randez, On the preservation of invariants by explicit Runge-Kutta methods, SIAM J. Sci. Comput. 28 (2006) 868–885.
[41] M. Qin, Y. Wang, Structure-Preserving Algorithms for Partial Differential Equation (*Chinese*), Zhejiang Publishing United Group, Zhejiang Science and Technology Publishing House, 2011.
[42] J. Swift, P. Hohenberg, Hydrodynamic fluctuations at the convective instability, Phys. Rev. A 15 (1977) 319.
[43] P. Zhao, M. Qin, Multisymplectic geometry and multisymplectic preissmann scheme for the KdV equation, J. Phys. A 33 (2000) 3613–3626.