MATHEMATICS OF COMPUTATION
Volume 91, Number 338, November 2022, Pages 2501–2515
https://doi.org/10.1090/mcom/3768
Article electronically published on July 28, 2022

WHY LARGE TIME-STEPPING METHODS FOR THE CAHN-HILLIARD EQUATION IS STABLE

DONG LI

Abstract. We consider the Cahn-Hilliard equation with standard double-well potential. We employ a prototypical class of first order in time semi-implicit methods with implicit treatment of the linear dissipation term and explicit extrapolation of the nonlinear term. When the dissipation coefficient is held small, a conventional wisdom is to add a judiciously chosen stabilization term in order to afford relatively large time stepping and speed up the simulation. In practical numerical implementations it has been long observed that the resulting system exhibits remarkable stability properties in the regime where the stabilization parameter is $\mathcal{O}(1)$, the dissipation coefficient is vanishingly small and the size of the time step is moderately large. In this work we develop a new stability theory to address this perplexing phenomenon.

1. Introduction

The Cahn-Hilliard equation was introduced in [1] to describe the phase separation and coarsening phenomena (i.e. formation of domains) in binary systems. If c denotes the concentration difference of the two components, then the Cahn-Hilliard equation can be written as

$$\partial_t c = D\Delta\mu = D\Delta(c^3 - c - \nu\Delta c), \tag{1.1}$$

where D is the diffusion coefficient, μ denotes the chemical potential and $\sqrt{\nu}$ characterizes the length scale of the transition regions between the domains. In a typical non-dimensionalized form, we take $D = 1$ and rewrite c as u. Then

$$\begin{cases} \partial_t u = \Delta(f(u)) - \nu\Delta^2 u, \quad \boxed{f(u) = u^3 - u}, \quad (t,x) \in (0,\infty)\times\Omega; \\ u\Big|_{t=0} = u_0. \end{cases} \tag{1.2}$$

For convenience we take the spatial domain Ω to be the 2π-periodic torus $\Omega = [-\pi,\pi]^d$ in physical dimensions $d = 1,2,3$. With some adjustments our analysis can be generalized to other boundary conditions. The system (1.2) admits a free energy given by

$$\mathcal{E}(u) = \int_\Omega \Big(\frac{1}{2}\nu|\nabla u|^2 + \frac{1}{4}(u^2-1)^2\Big)dx. \tag{1.3}$$

For smooth solutions the energy dissipation law takes the form

$$\frac{d}{dt}\Big(\mathcal{E}(u)\Big) = -\int_\Omega |\partial_t\mu|^2 dx. \quad \mu = u^3 - u - \nu\Delta u. \tag{1.4}$$

Received by the editor March 3, 2022, and, in revised form, March 20, 2022, April 2, 2022, and May 22, 2022.

2020 *Mathematics Subject Classification.* Primary 35Q35.

Key words and phrases. Cahn-Hilliard, maximum principle, stabilization.

This simple balance relation is quite natural since the system (1.2) corresponds to the gradient flow of $\mathcal{E}(u)$ in H^{-1}. It is not difficult to check that the average of u is preserved in time. For convenience we shall tacitly assume u has mean zero in our analysis. Whilst the a priori control (1.4) yields strong H^1 bounds on the solution, the lack of maximum principle renders it a nontrivial task to obtain $\mathcal{O}(1)$ bounds on the maximum norm of the solution. In the numerical context this issue turns out to be nontrivial even in the parabolic setting (cf. [2,4]).

In the past decades, there has been a lot of progress on designing efficient, accurate and stable numerical schemes to resolve the plethora of vastly different temporal and spatial scales in phase field models such as Cahn-Hilliard and Allen-Cahn. Many powerful numerical methods such as the convex-splitting scheme [10–12], the stabilization scheme [13,14], the scalar auxiliary variable (SAV) methods [15], semi-implicit/implicit-explicit (IMEX) schemes [4–7] are introduced in order to track accurately the dynamical evolution of the phase field variable. However many fundamental questions still remain unsolved concerning the analysis of these schemes. In this work we consider a class of semi-implicit schemes which were considered by He, Liu and Tang in [3]. In a semi-discrete formulation, it reads

$$\frac{u^{n+1}-u^n}{\tau} = -\nu\Delta^2 u^{n+1} + A\Delta(u^{n+1}-u^n) + \Delta(f(u^n)), \quad n \geq 0, \tag{1.5}$$

where $\tau > 0$ is the time step, and $A > 0$ is the coefficient for the $\mathcal{O}(\tau)$ regularization term. In [3], He, Liu and Tang showed that (see Theorem 1 therein) if

$$A \geq \max_{x\in\Omega}\{\frac{1}{2}|u^n(x)|^2 + \frac{1}{4}|u^{n+1}(x)+u^n(x)|^2\} - \frac{1}{2}, \quad \forall\, n \geq 0, \tag{1.6}$$

then $\mathcal{E}(u^n) \leq \mathcal{E}(u^0)$ for all $n \geq 0$. Note that the condition (1.6) is not satisfactory since the right hand side (RHS) depends also on A. An even more startling observation is that one can even take relatively large time stepping for moderately large A and miniscule dissipation coefficient ν. For example (see Table 1 in [3]), numerically one has the following list of admissible tuple of (ν, A, τ_c) where τ_c is the maximal time step for which energy decay holds monotonically in time:

ν	A	τ_c
$\nu = 0.01$	$A = 0$	$\tau_c \approx 0.02$
	$A = 0.5$	$\tau_c \approx 0.2$
	$A = 1$	$\tau_c \approx 0.2$
$\nu = 0.001$	$A = 0$	$\tau_c \approx 0.003$
	$A = 0.5$	$\tau_c \approx 0.013$
	$A = 1$	$\tau_c \approx 0.03$

In particular, for $\nu = 0.001$, $A = 1$, one can take large time step $\tau \approx 0.03$ whilst not losing energy dissipation! As far as we know, no existing theory can address this rather perplexing phenomenon. The purpose of this work is to develop a new stability theory to clarify this issue. Our first result reveals a deep connection between the stabilization parameter A and the maximum norm of the numerical solution.

Theorem 1.1 (Uniform in time L^∞ bound for (1.5))**.** *Let the spatial domain Ω be the 2π-periodic torus in physical dimensions $d = 1, 2, 3$, i.e. $\Omega = [-\pi, \pi]^d$. Let $\nu > 0$. Consider* (1.5) *with $u^0 \in L^\infty(\Omega)$ having zero mean. Assume $A \geq A_{\rm cr} := 1 + 2\sqrt{1 + \frac{4}{3}\cdot\frac{\nu}{\tau}}$. If the initial data u^0 satisfies (here $\|u^0\|_\infty := \|u^0\|_{L^\infty(\Omega)}$)*

$\|u^0\|_\infty \le M = \sqrt{\frac{A+1}{3}}$, *then in* (1.5),

$$\|u^n\|_\infty \le M, \qquad \forall\, n \ge 1.$$

Remark 1.1. Note that the threshold value $A_{\rm cr}$ is *not* inversely proportional to the diffusion coefficient ν. Our L^∞ bound here explains why (1.5) is stable when the diffusion coefficient ν is small and large time step τ is taken. For example, if $\nu = 0.001$ and $\tau = 0.03$, then $A_{\rm cr} \approx 3.04$ which is $\mathcal{O}(1)$! Of course some further nontrivial work is needed to achieve the optimal stabilization parameter $A \approx 1$.

Remark 1.2. We should point out that the maximal time step restriction in the aforementioned table is due to the fact that the choice of A does not satisfy the condition $A \ge A_{cr}$ in Theorem 1.1. Under the condition $A \ge A_{\rm cr} > 1$, there is no maximal time step restriction, whereas for $0 \le A \le 1$ there exists some threshold τ_c. In this connection further research is needed to calibrate the precise relationship between the maximal time step $\tau_{\rm c}$ and stabilization parameter A_{cr} for moderately small values of A_{cr}.

Remark 1.3. There is some flexibility in choosing the upper bound M. See Lemma 2.1 where one can choose any number $M \in [M_0, M_1]$ with $M_1 = \sqrt{\frac{A+1}{3}}$ and $M_0 = 2\sqrt{\frac{1+(\frac{1}{2}-\sqrt{\frac{1}{4}-\frac{1}{A^2}\cdot\frac{\nu}{\tau}})A}{3}}$.

Remark 1.4. For general initial data with nonzero mean, it is possible that one can develop a corresponding version of the maximum principle. However we shall not dwell on this rather technical point here in this work.

Remark 1.5. We should point out that the solvability of the iterative scheme (1.5) is not an issue even under the rough data assumption that $u^0 \in L^\infty(\Omega)$ with zero mean. This is due to the fact that one can recast (1.5) into the form:

$$(1 + \tau\nu\Delta^2 - A\tau\Delta)u^{n+1} = u^n - \tau A\Delta u^n + \tau\Delta(f(u^n)), \tag{1.7}$$

or equivalently

$$u^{n+1} = (1 + \tau\nu\Delta^2 - A\tau\Delta)^{-1}(1 - A\tau\Delta)u^n + (1 + \tau\nu\Delta^2 - A\tau\Delta)^{-1}\tau\Delta(f(u^n)). \tag{1.8}$$

For $\tau > 0$, the operators $(1+\tau\nu\Delta^2-A\tau\Delta)^{-1}(1-A\tau\Delta)$ and $(1+\tau\nu\Delta^2-A\tau\Delta)^{-1}\tau\Delta$ roughly correspond to the Fourier multiplier $|k|^{-2}$ for wave number $|k| \gg 1$. As such it leads to regularity upgrade in the same spirit of the usual elliptic theory.

Theorem 1.1 elucidates the appearance of L^∞ bound due to time discretization. On the other hand, in practical numerical computations, the bi-harmonic and Laplacian operators on the RHS of (1.5) would also have to be computed numerically. In this situation the L^∞ bound on the numerical solution certainly needs to be proved as well. To keep some generality, we denote the numerical approximation of Δ by $\Delta^{\rm NUM}$. For example, on a 1D uniform mesh with mesh size Δx, a function f is represented by numerical sequence f_i, and a typical central difference scheme on mesh vertex i (away from the boundary) takes the form

$$(\Delta^{\rm NUM} f)_i = \frac{f_{i+1} + f_{i-1} - 2f_i}{(\Delta x)^2}.$$

In the literature, Δ^{NUM} is sometimes called the graph Laplacian as it acts on functions defined a discrete graph with suitable weights on the edges. We need some "stability" property of the graph Laplacian Δ^{NUM}. This is illustrated by Definition 1.1.

Definition 1.1. We say a graph Laplacian Δ^{NUM} on a graph X obeys a sharp L^∞ estimate if the following holds for any constant $k > 0$: for any bounded $f : X \to \mathbb{R}$, there exists a unique function $u : X \to \mathbb{R}$ solving the equation

$$u - k\Delta^{\mathrm{NUM}} u = f; \tag{1.9}$$

moreover

$$\|u\|_\infty \le \|f\|_\infty.$$

In yet other words, for all $k > 0$, we have

$$\|(I - k\Delta^{\mathrm{NUM}})^{-1} f\|_\infty \le \|f\|_\infty.$$

Remark. One can certainly consider a more general operator (not necessarily the graph Laplacian) and introduce the notion of sharp L^∞ estimates in more abstract settings. However we do not pursue this generality here.

Remark. For Δ^{NUM} introduced via typically finite difference schemes, one can easily verify the the solvability of (1.9) and the sharp L^∞ estimate. See Section 2 for some examples.

We now consider the following fully discretized (in both space and time) scheme:

$$\frac{u^{n+1} - u^n}{\tau} = -\nu(\Delta^{\mathrm{NUM}})^2 u^{n+1} + A\Delta^{\mathrm{NUM}}(u^{n+1} - u^n) + \Delta^{\mathrm{NUM}}(f(u^n)), \quad n \ge 0. \tag{1.10}$$

Corollary 1.1. *Assume Δ^{NUM} satisfies the sharp L^∞ estimate in the sense of Definition* 1.1. *Let $A \ge A_{\mathrm{cr}} := 1 + 2\sqrt{1 + \frac{4}{3} \cdot \frac{\nu}{\tau}}$. If the initial data u^0 satisfies $\|u^0\|_\infty \le M = \sqrt{\frac{A+1}{3}}$, then in* (1.10),

$$\|u^n\|_\infty \le M, \qquad \forall\, n \ge 1.$$

We state Corollary 1.1 as a conditional result just to keep some generality. On the other hand, as was already mentioned earlier, the condition on Δ^{NUM} can be easily checked for typical finite difference schemes (see Section 2). Corollary 1.2 records this fact.

Corollary 1.2. *The graph Laplacian Δ^{NUM} introduced by typical finite difference schemes satisfies the sharp L^∞ estimate in the sense of Definition* 1.1. *Therefore Corollary* 1.1 *holds for* (1.10) *with corresponding Δ^{NUM}.*

Theorem 1.2. *Consider* (1.5) *in physical dimensions $d \le 3$ and assume $u^0 \in L^\infty(\Omega) \cap H^1(\Omega)$ with zero mean. Recall*

$$\mathcal{E}(u) = \int_\Omega \Big(\frac{1}{2}\nu|\nabla u|^2 + F(u)\Big)dx,$$

where $F(u) = \frac{1}{4}(u^2-1)^2$. *Assume (as in Theorem* 1.1*)* $A \geq A_{\rm cr} = 1 + 2\sqrt{1 + \frac{4}{3}\cdot\frac{\nu}{\tau}}$ *and the initial data* u^0 *satisfies* $\|u^0\|_\infty \leq \sqrt{\frac{A+1}{3}}$. *Then*

$$\begin{aligned}&\mathcal{E}(u^{n+1}) + \frac{A}{2}\|u^{n+1}-u^n\|_2^2\\ &\quad + \tau\|\nabla\big(-\nu\Delta u^{n+1} + A(u^{n+1}-u^n) + f(u^n)\big)\|_2^2\\ &\quad \leq \mathcal{E}(u^n), \qquad \forall\, n \geq 0.\end{aligned}$$

In particular

$$\mathcal{E}(u^{n+1}) \leq \mathcal{E}(u^n), \qquad \forall\, n \geq 0.$$

Remark. To understand the role of the stabilization term $A\Delta(u^{n+1}-u^n)$, it is useful to consider the general case

$$\frac{u^{n+1}-u^n}{\tau} = -\nu\Delta^2 u^{n+1} + B(u^{n+1}-u^n) + \Delta f(u^n),$$

where B is an operator to be determined. Taking the L^2 inner product with $(-\Delta)^{-1}(u^{n+1}-u^n)$ on both sides, one arrives at (see (1.12) for the definition of $|\nabla|^s = (-\Delta)^{s/2}$)

$$\begin{aligned}&\frac{1}{\tau}\||\nabla|^{-1}(u^{n+1}-u^n)\|_2^2 + E_{n+1} - E_n + \frac{\nu}{2}\|\nabla(u^{n+1}-u^n)\|_2^2\\ &\qquad + (B(u^{n+1}-u^n), (-\Delta)^{-1}(u^{n+1}-u^n))\\ &\leq \frac{L}{2}\|u^{n+1}-u^n\|_2^2,\end{aligned}$$

where $L = \sup_{0\leq s\leq 1}\|f'(u^n + s(u^{n+1}-u^n))\|_\infty$ and we have denoted $E_n = \mathcal{E}(u^n)$. It should be noted here the rough estimate of f' makes no use of the spectral information around linearization of the continuous PDE solution. Clearly if $B \equiv 0$, then to ensure $E_{n+1} \leq E_n$, one must enforce

$$\frac{1}{\tau}\||\nabla|^{-1}(u^{n+1}-u^n)\|_2^2 + \frac{\nu}{2}\|\nabla(u^{n+1}-u^n)\|_2^2 \geq \frac{L}{2}\|u^{n+1}-u^n\|_2^2.$$

In view of the interpolation inequality (for mean-zero functions)

$$\|g\|_2 \leq \||\nabla|^{-1}g\|_2^{\frac{1}{2}}\|\nabla g\|_2^{\frac{1}{2}}$$

and Cauchy-Schwartz, we deduce the constraint

$$2\sqrt{\frac{\nu}{2\tau}} \geq \frac{L}{2} \Rightarrow \tau \leq \frac{8\nu}{L^2}.$$

This is the main reason why small time step τ is needed when ν is small and no stabilization term is present. On the other hand, from the above computation, one can also see the necessity of having the operator $B = \text{const}\cdot\Delta$: it is precisely used to balance out the term $\frac{L}{2}\|u^{n+1}-u^n\|_2$ on the RHS.

We gather below some notation used in this work.

Notation. Throughout this work we denote by $\Omega = [-\pi, \pi]^d$ the usual 2π-periodic torus in physical dimensions $d \le 3$. For a real-valued Borel measurable function $u : \Omega \to \mathbb{R}$, we denote by

$$\|u\|_p := \|u\|_{L^p(\Omega)} = \begin{cases} \Big(\int_\Omega |u(x)|dx\Big)^{\frac{1}{p}}, & 1 \le p < \infty; \\ \operatorname{esssup}_{x\in\Omega}|u(x)|, & p = \infty, \end{cases} \tag{1.11}$$

the usual Lebesgue L^p-norm of u. For smooth periodic $u : \Omega \to \mathbb{R}$ with zero mean (i.e. $\hat{u}(0) = 0$) and $s \in \mathbb{R}$, we denote $|\nabla|^s = (-\Delta)^{s/2}$ as the Fourier multiplier $|k|^s$, i.e.

$$\widehat{|\nabla|^s u}(k) = |k|^s \widehat{u}(k), \qquad 0 \ne k \in \mathbb{Z}^d, \tag{1.12}$$

where $\widehat{u}(k) = \int_\Omega u(x)e^{-ik\cdot x}dx$. For smooth periodic $u : \Omega \to \mathbb{R}$ with zero mean, the usual $\dot{H}^s$-norm is defined as

$$\|u\|_{\dot{H}^s(\Omega)} := \||\nabla|^s u\|_2. \tag{1.13}$$

The rest of this paper is organized as follows. In Section 2 we give the proof of the main result Theorem 1.1. In Section 3 we give a resolvent bound. In the last section we prove Theorem 1.2.

2. Proof of Theorem 1.1, Corollary 1.1 and 1.2

Proof of Theorem 1.1. Write

$$u^{n+1} - u^n = -\nu\tau\Delta^2 u^{n+1} + A\tau\Delta(u^{n+1} - u^n) + \tau\Delta(f(u^n)).$$

Let $\beta > 0$ be a parameter whose value will be chosen later. Then

$$\begin{aligned}(1 - \beta A\tau\Delta)(u^{n+1} - u^n) &= -\nu\tau\Delta^2 u^{n+1} + (1-\beta)A\tau\Delta(u^{n+1} - u^n) + \tau\Delta(f(u^n)) \\ &= \tau\Delta\big((1-\beta)A - \nu\Delta\big)u^{n+1} + \tau\Delta\big(f(u^n) - (1-\beta)Au^n\big).\end{aligned}$$

Now choose β such that

$$\frac{1}{\beta A\tau} = \frac{(1-\beta)A}{\nu}$$

or simply

$$\beta(1-\beta) = \frac{\nu}{A^2\tau}.$$

The existence of β is out of question since $\nu/(A^2\tau) \le 1/4$ by assumption (see below).

Clearly by the definition of β (using $\nu = (1-\beta)A \cdot \beta A\tau$), we have

$$\begin{aligned}\frac{\tau\Delta\Big((1-\beta)A - \nu\Delta) - \nu\Delta\Big)}{1 - \beta A\tau\Delta} =& \frac{\tau\Delta\Big((1-\beta)A - (1-\beta)A\cdot\beta A\tau\Delta\Big)}{1 - \beta A\tau\Delta} \\ =&(1-\beta)A\tau\Delta.\end{aligned}$$

It follows that

$$\begin{aligned}&u^{n+1} - u^n \\ =&(1-\beta)A\tau\Delta u^{n+1} + (1 - \beta A\tau\Delta)^{-1}\tau\Delta\big(f(u^n) - (1-\beta)Au^n\big).\end{aligned}$$

Rearranging the terms, we get

$$\begin{aligned}&\big(1-(1-\beta)A\tau\Delta\big)u^{n+1}\\&=u^n+(1-\beta A\tau\Delta)^{-1}\tau\Delta\big(f(u^n)-(1-\beta)Au^n\big).\end{aligned}$$

The proof of Theorem 1.1 then follows from Lemma 2.1.

Lemma 2.1. *Let $k=\nu/\tau$. Define $A_{\rm cr}=1+2\sqrt{1+\frac{4}{3}k}$. If $A\geq A_{\rm cr}$, then the following hold:*

- *$\nu/A^2\tau\leq\frac{1}{4}$ and $\beta=\frac{1}{2}+\sqrt{\frac{1}{4}-\frac{1}{A^2}k}\in[\frac{1}{2},1)$.*
- *Define $M_0=2\sqrt{\frac{1+(1-\beta)A}{3}}$, $M_1=\sqrt{\frac{A+1}{3}}$. Then $M_0\leq M_1$.*
- *For any M with $M_0\leq M\leq M_1$, if $\|u^n\|_\infty\leq M$, then*

$$\|u^n+\tau\Delta(1-\beta A\tau\Delta)^{-1}\big(f(u^n)-(1-\beta)Au^n\big)\|_\infty\leq M;$$

and consequently $\|u^{n+1}\|_\infty\leq M$.

Remark 2.1. Lemma 2.1 shows that for $k>0$, the nonlocal operator $(1-k\Delta)^{-1}\Delta$ exhibits some form of maximum principle. Interestingly there exist also some "inverse Sobolev" type equalities for this operator, see [8] for more details.

To complete the proof of Lemma 2.1, we need the following simple lemma which in a sense identifies the "invariant region" of certain auxiliary cubic polynomials.

Lemma 2.2. *Let $\alpha_1>0$ and $f_1(x)=x^3-\alpha_1 x$. If $L\geq 2\sqrt{\frac{\alpha_1}{3}}$, then*

$$\max_{|x|\leq L}|f_1(x)|\leq f_1(L). \tag{2.1}$$

Similarly let $\alpha_2>0$ and $f_2(x)=-x^3+\alpha_2 x$. If $0<L\leq\sqrt{\frac{\alpha_2}{3}}$, then

$$\max_{|x|\leq L}|f_2(x)|\leq f_2(L). \tag{2.2}$$

Proof of Lemma 2.2. For $f_1(x)$, calculating $f_1'(x)=0$ yields that $x_1=\pm\sqrt{\alpha_1/3}$. It is then easy to check that at $L=2x_1$, $f_1(L)\geq|f_1(x_1)|$. An inspection of the graph of f_1 easily gives (2.1). For (2.2), one just need to notice that $f_2'\geq 0$ for $|x|\leq\sqrt{\frac{\alpha_2}{3}}$. □

Proof of Lemma 2.1. First note that

$$\begin{aligned}\tau\Delta(1-\beta A\tau\Delta)^{-1}&=(\tau\Delta-\frac{1}{\beta A}+\frac{1}{\beta A})(1-\beta A\tau\Delta)^{-1}\\&=-\frac{1}{\beta A}+\frac{1}{\beta A}(1-\beta A\tau\Delta)^{-1}.\end{aligned}$$

Thus

$$\begin{aligned}&u^n+\tau\Delta(1-\beta A\tau\Delta)^{-1}\big(f(u^n)-(1-\beta)Au^n\big)\\=&u^n-\frac{1}{\beta A}(f(u^n)-(1-\beta)Au^n)+\frac{1}{\beta A}(1-\beta A\tau\Delta)^{-1}\big(f(u^n)-(1-\beta)Au^n\big)\\=&\frac{1}{\beta A}\Big(\underbrace{-(u^n)^3+(A+1)u^n}_{:=f_2(u^n)}+(1-\beta A\tau\Delta)^{-1}\big(\underbrace{(u^n)^3-((1-\beta)A+1)u^n}_{:=f_1(u^n)}\big)\Big),\end{aligned} \tag{2.3}$$

where in the last equality above, we plugged in $f(u)=u^3-u$.

By Lemma 2.2, we have if $\|u^n\|_\infty \le M$, then

$$\begin{aligned}&\|(1-\beta A\tau\Delta)^{-1}\big(f_1(u^n)\big)\|_\infty\\ \le &\max_{|z|\le M}|f_1(z)| \le f_1(M),\end{aligned}$$

provided $M \ge 2\sqrt{\frac{1+(1-\beta)A}{3}}$.

Then under the condition $\|u^n\|_\infty \le M$ and for $M \le \sqrt{\frac{A+1}{3}}$ (by using Lemma 2.2),

$$\begin{aligned}\|\text{RHS of (2.3)}\|_\infty &\le \frac{1}{\beta A}\Big(\max_{|z|\le M}|f_2(z)| + f_1(M)\Big)\\ &\le \frac{1}{\beta A}\Big(f_2(M)+f_1(M)\Big) = M.\end{aligned}$$

Collecting all the inequalities, we get the following:

Under the conditions

- $\beta(1-\beta) = \frac{\nu}{A^2\tau} \le \frac{1}{4}$, $0<\beta<1$;
- $2\sqrt{\frac{1+(1-\beta)A}{3}} \le \sqrt{\frac{A+1}{3}}$,

if $\|u^n\|_\infty \le M$ ($M \le \sqrt{\frac{A+1}{3}}$), then $\|u^{n+1}\|_\infty \le M$.

It is then easy to deduce the condition $A \ge A_{\rm cr}$. □

2.1. **Proof of Corollary 1.1 and 1.2.** We first note that in view of (1.9), the proof of Corollary 1.1 is a repetition of that of Theorem 1.1 (with Δ simply replaced by $\Delta^{\rm NUM}$). Therefore we only focus on Corollary 1.2. This amounts to checking Definition 1.1 for typical finite difference schemes. We present several illustrative examples.

Example 2.1. 1D central difference with periodic boundary condition. Let $N \ge 2$ be an integer and $\Delta x > 0$. Let $u = (u_0, \cdots, u_{N-1})$ and define

$$(\Delta^{\rm NUM}u)_i = \frac{u_{i+1}+u_{i-1}-2u_i}{(\Delta x)^2}.$$

Here $u_{i+N} = u_i$. With data $f = (f_i)$, we need to examine solvability to the equation

$$u_i - k(\Delta^{\rm NUM}u)_i = f_i \tag{2.4}$$

and prove the estimate

$$\|u\|_\infty \le \|f\|_\infty. \tag{2.5}$$

First we note that (2.5) follows from a simple maximum principle argument: if $i_1 = \text{argmax}(u_i)$, then obviously $(\Delta^{\rm NUM}u)_{i_1} \le 0$, and $u_{i_1} \le f_{i_1}$. To show existence, we can rewrite (2.4) as

$$u_i = (Tu)_i := \frac{\theta}{2}(u_{i-1}+u_{i+1}) + (1-\theta)f_i, \tag{2.6}$$

where $\theta = \frac{2k}{2k+(\Delta x)^2}$. Since $0<\theta<1$, easy to check that T is a contraction operator (in l^∞-norm) and the existence follows from the standard fixed point theorem.[1]

[1] Actually from (2.6) one can also directly deduce the estimate $\|u\|_\infty \le \|f\|_\infty$ without appealing to the maximum principle.

Example 2.2. 1D central difference with Dirichlet boundary condition. This is similar to Example 2.1 except that the boundary condition is modified to $u_{-1} = u_N = 0$. Easy to check that in this case Δ^{NUM} still satisfies Definition 1.1.

Example 2.3. Graph Laplacian with special weights. Let X be a finite set with cardinality $|X| = N$. Without loss of generality we identify X as $\{0, \cdots, N-1\}$. Let w_{ij}, $0 \le i, j \le N-1$ be nonnegative numbers such that $w_{ii} = \sum_{j \ne i} w_{ij}$, for all i. For any $u : X \to \mathbb{R}$, define

$$(\Delta^{\mathrm{NUM}} u)_j = -w_{ii} u_i + \sum_{j \ne i} w_{ij} u_j. \tag{2.7}$$

Then Δ^{NUM} satisfies Definition 1.1. Indeed the equation $u - k\Delta^{\mathrm{NUM}} u = f$ can be rewritten as

$$u_i = (Tu)_i := \sum_{j \ne i} \frac{k w_{ij}}{1 + k w_{ii}} u_j + \frac{1}{1 + k w_{ii}} f_i. \tag{2.8}$$

Easy to check that $\|T(u-v)\|_\infty \le \theta \|u - v\|_\infty$ with

$$\theta = \max_{1 \le i \le N} \frac{k w_{ii}}{1 + k w_{ii}} < 1.$$

The estimate $\|u\|_\infty \le \|f\|_\infty$ is also obvious.

Remark. Example 2.3 includes many finite difference schemes as special cases. For example, on a 2D mesh with mesh size h, the usual five-point stencil discretized Laplacian has the form

$$\begin{aligned} &(\Delta^{\mathrm{NUM}} u)(x_1, x_2) \\ &= \frac{u(x_1 - h, x_2) + u(x_1 + h, x_2) + u(x_1, x_2 - h) + u(x_1, x_2 + h) - 4u(x_1, x_2)}{h^2}. \end{aligned}$$

This certainly can be rewritten in the style of (2.7).

3. Improved resolvent bounds

The resolvent bound $\|(I - k\Delta^{\mathrm{NUM}})^{-1} f\|_\infty \le \|f\|_\infty$ discussed in the previous section is generally optimal, as can been seen by taking f to be a constant function. On the other hand, for Cahn-Hilliard type equations, we usually work with functions with mean zero. As it turns out, for discretized Laplacians, one can refine the resolvent bound slightly if we restrict to the class of mean-zero functions.

Proposition 3.1. *Consider* (2.6). *There exists a constant* $0 < \epsilon < 1$ *(possibly depending on* θ *and* N*) such that*

$$\|u\|_\infty \le \epsilon \|f\|_\infty,$$

for any f *with mean zero, i.e.* $\sum_{i=0}^{N-1} f_i = 0$.

Remark 3.1. To see why Proposition 3.1 should hold, one can consider the special case $N = 3$. In this case by using $u_0 + u_1 + u_2 = 0$, one can explicitly solve u_i in terms of f_i as

$$u_i = \frac{1 - \theta}{1 + \frac{\theta}{2}} f_i.$$

Obviously $\|u\|_\infty \le \frac{1-\theta}{1+\frac{\theta}{2}} \|f\|_\infty$.

To prove Proposition 3.1, we need a simple lemma. The subtlety lies in the incorporation of the mean-zero constraint.

Lemma 3.1. *Let $N \geq 2$ be an integer. Suppose $0 \leq c_0 \leq c_1 \cdots \leq c_{N-1}$. Let*

$$X = \big\{\sigma = (\sigma_0, \cdots, \sigma_{N-1}) : \max_j |\sigma_j| \leq 1, \sum_{j=0}^{N-1} \sigma_j = 0.\big\}.$$

Then

$$\max_{\sigma \in X}(c \cdot \sigma) = \sum_{j=N-[\frac{N}{2}]}^{N-1} c_j - \sum_{j=0}^{[\frac{N}{2}]-1} c_j.$$

Here $[x]$ denotes the integer part of any real number x, for example $[3/2] = 1$.

Remark. If N is even, then the maximum of $c \cdot \sigma$ is achieved by

$$\sigma = (-1, -1, \cdots, -1, 1, \cdots, 1)$$

with equal number of 1s and -1s. If N is odd, then this is achieved by

$$\sigma = (-1, -1, \cdots, -1, 0, 1, \cdots, 1)$$

with $(N-1)/2$ ones and minus ones.

Proof of Lemma 3.1. Consider the function $f(\sigma) = c \cdot \sigma$. Since X is a compact set, the maximum of f must be attained at some point $\tilde{\sigma} = (\tilde{\sigma}_0, \cdots, \tilde{\sigma}_{N-1})$. Since $0 \leq c_0 \leq \cdots c_{N-1}$ and $\sum_j \tilde{\sigma}_j = 0$, we can assume $\tilde{\sigma}_0 \leq \cdots \tilde{\sigma}_{j_1} \leq 0 \leq \tilde{\sigma}_{j_1+1} \leq \cdots \leq \tilde{\sigma}_{N-1}$. By a simple optimization argument,[2] one can further assume that $\tilde{\sigma}$ has three possible forms:

- $\tilde{\sigma} = (-1, \cdots, -1, \sigma_{j_1}, \sigma_{j_1+1}, 1, \cdots, 1)$, where $-1 < \sigma_{j_1} \leq 0$ and $0 \leq \sigma_{j_1+1} < 1$. Now since $c_{j_1} \leq c_{j_1+1}$, for $\epsilon > 0$, we have

  $$c_{j_1}\sigma_{j_1} + c_{j_1+1}\sigma_{j_1+1} \leq c_{j_1}(\sigma_{j_1} - \epsilon) + c_{j_1+1}(\sigma_{j_1+1} + \epsilon).$$

  By using this argument together with the fact $\sum_j \tilde{\sigma}_j = 0$, it is easy to see that we can change $\tilde{\sigma}$ to $\tilde{\sigma} = (-1, \cdots, -1, 1, \cdots, 1)$ and the value of $c \cdot \tilde{\sigma}$ does not decrease.
- $\tilde{\sigma} = (-1, \cdots, -1, \tilde{\sigma}_{j_1}, 1, \cdots, 1)$ where $-1 < \tilde{\sigma}_{j_1} \leq 0$. Since $\sum_j \tilde{\sigma}_j = 0$, easy to see that in this case we must have $\tilde{\sigma}_{j_1} = 0$.
- $\tilde{\sigma} = (-1, \cdots, -1, \tilde{\sigma}_{j_1}, 1, \cdots, 1$ where $0 \leq \sigma_{j_1} < 1$. Easy to see that $\tilde{\sigma}_1 = 0$ again due to $\sum_j \tilde{\sigma}_j = 0$.

The rest of the argument is now obvious. One just need to discuss separately the case N is even and the case N is odd. □

Proof of Proposition 3.1.

Step 1. We first show that there exists $c = (c_0, \cdots, c_{N-1})$, such that

$$u_k = (c * f)_k = \sum_j c_{k-j} f_j,$$

[2]One can fix the sum $\sum_{l=0}^{j_1} \tilde{\sigma}_l$ and maximize $\sum_{l=0}^{j_1} \tilde{\sigma}_l \cdot c_l$. Similarly fix $\sum_{l=j_1+1}^{N-1} \tilde{\sigma}_l$ and maximize $\sum_{l=j_1+1}^{N-1} \sigma_l \cdot c_l$. Also observe that one can assume without loss of generality that there is at most one zero in $\tilde{\sigma}$.

with the identification that $c_{k\pm N} = c_k$. This follows easily from the discrete Fourier transform, which we briefly recall here. For a sequence of numbers $a_0, \ldots, a_{N-1}$, define

$$\hat{a}_j = \sum_{k=0}^{N-1} a_k e^{-\frac{2\pi ijk}{N}}.$$

Then a_k can be reproduced from $\hat{a}_j$ by

$$a_k = \frac{1}{N} \sum_{j=0}^{N-1} \hat{a}_j e^{\frac{2\pi ijk}{N}}.$$

For any two sequences $a = (a_0, \cdots, a_{N-1})$ and $b = (b_0, \cdots, b_{N-1})$, easy to check that

$$\widehat{(a * b)}_k = \hat{a}_k \hat{b}_k.$$

Now return to (2.6). Clearly

$$(1 - \theta \cos(\frac{2\pi k}{N}))\hat{u}_k = (1 - \theta)\hat{f}_k.$$

Thus

$$u_j = (c * f)_j,$$

where

$$c_j = \frac{1}{N} \sum_{k=0}^{N-1} \frac{1 - \theta}{1 - \theta \cos(\frac{2\pi k}{N})} e^{\frac{2\pi ijk}{N}}.$$

Step 2. We show that $\sum_{j=0}^{N-1} c_j = 1$ and

$$\min_{0 \le j \le N-1} c_j > 0. \tag{3.1}$$

By Step 1, if we solve

$$u_j = \frac{\theta}{2}(u_{j-1} + u_{j+1}) + (1 - \theta) f_j, \tag{3.2}$$

with $f = (1, 0, \cdots, 0)$. Then $u_j = c_{j-1}$. By a simple maximum principle argument we have $u_j \ge 0$ for all j. Now assume $u_{j_*} = 0$ for some j_*. Then from (3.2) evaluated at $j = j_*$, we get $u_{j_*-1} = u_{j_*+1} = 0$. Iterating this argument a couple of times, we get $u_j = 0$ for all j which is obviously impossible. Thus $\min u_j > 0$ and (3.1) holds. The fact $\sum_j u_j = 1$ is obvious from summing j on both sides of (3.2).

Step 3. Define

$$X = \big\{\tilde{f} = (\tilde{f}_0, \cdots, \tilde{f}_{N-1}) : \max_j |\tilde{f}_j| \le 1, \sum_j \tilde{f}_j = 0\big\}.$$

By Lemma 3.1 and Step 2, we have

$$\max_{\tilde{f} \in X} |c \cdot \tilde{f}| \le \begin{cases} 1 - 2\sum_{j=0}^{\frac{N}{2}-1} c_j, & \text{if } N \text{ is even,} \\ 1 - c_{\frac{N-1}{2}} - 2\sum_{j=0}^{\frac{N-1}{2}-1} c_j, & \text{if } N \text{ is odd.} \end{cases}$$

Thus

$$\max_{\tilde{f} \in X} |c \cdot \tilde{f}| \le 1 - N \min_j c_j.$$

Therefore

$$\|c * f\|_\infty \le \epsilon \|f\|_\infty,$$

where

$$\epsilon \le 1 - N \min_j c_j < 1. \qquad \square$$

Remark. By Lemma 3.1, one can get the sharp constant

$$\epsilon = \sum_{j=N-[\frac{N}{2}]}^{N-1} c_j - \sum_{j=0}^{[\frac{N}{2}]-1} c_j.$$

On the other hand, to get the bound $\|c * f\|_\infty \le (1 - N \min_j c_j)\|f\|_\infty$, one could just argue directly without using Lemma 3.1. Let $\epsilon_0 = \min_j c_j$ and define $\tilde{c}_j = c_j - \epsilon_0 \ge 0$. Then since f has mean zero, we have $c * f = \tilde{c} * f$. Thus

$$\begin{aligned} \|c * f\|_\infty &\le \|\tilde{c}\|_1 \|f\|_\infty \\ &= (1 - N\epsilon_0)\|f\|_\infty. \end{aligned}$$

A similar perturbation idea is exploited in recent [9] to show some generalized Poincaré inequalities.

We record below the generalization of Proposition 3.1.

Proposition 3.2. *Consider* (2.8). *There exists a constant* $0 < \epsilon < 1$ *such that*

$$\|u\|_\infty \le \epsilon \|f\|_\infty,$$

for any f *with mean zero.*

Proof of Proposition 3.2. This is similar to the proof of Proposition 3.1 and we only point out the needed modifications. First let δ_{li} be the usual Kronecker delta function and let $c_i^{(l)}$ solves (see (2.8))

$$c_i^{(l)} = \sum_{j \ne i} \frac{k w_{ij}}{1 + k w_{ii}} c_j^{(l)} + \frac{1}{1 + k w_{ii}} \delta_{li}.$$

Then clearly the solution to (2.8) can be represented by

$$u_i = \sum_l c_i^{(l)} f_l.$$

Easy to check that $\epsilon_0 = \min_{j,l} c_i^{(l)} > 0$. Furthermore (by taking f to be a constant function) easy to check that $\sum_{l=0}^{N-1} c_i^{(l)} = 1$ for any i. Using the fact that f has mean zero, clearly we have

$$|u_i| = \Big| \sum_{l=0}^{N-1} (c_i^{(l)} - \epsilon_0) f_l \Big| \le (1 - N\epsilon_0)\|f\|_\infty,$$

i.e. $\|u\|_\infty \le \epsilon \|f\|_\infty$ for $\epsilon = 1 - N\epsilon_0 < 1$. $\square$

4. Proof of Theorem 1.2

In this proof we denote by $(\cdot,\cdot)$ the usual L^2 inner product for real-valued functions. Denote

$$H = -\nu\Delta u^{n+1} + A(u^{n+1} - u^n) + f(u^n).$$

Here we suppress the notational dependence of H on n for simplicity. The scheme (1.5) simply reads as

$$\frac{u^{n+1} - u^n}{\tau} = \Delta H.$$

Clearly then

$$(u^{n+1} - u^n, H) = \tau(\Delta H, H) = -\tau\|\nabla H\|_2^2.$$

We now evaluate $(u^{n+1} - u^n, H)$ by examining the contribution of each term in H. First

$$\begin{aligned}
&(u^{n+1} - u^n, -\nu\Delta u^{n+1}) \\
&= -\nu\big((u^{n+1}, \Delta u^{n+1}) - (u^n, \Delta u^{n+1})\big) \\
&= \nu\big(\|\nabla u^{n+1}\|_2^2 - (\nabla u^n, \nabla u^{n+1})\big) \\
&\geq \nu\big(\frac{1}{2}\|\nabla u^{n+1}\|_2^2 - \frac{1}{2}\|\nabla u^n\|_2^2\big).
\end{aligned}$$

Here we used the simple inequality $a^2 + ab \geq \frac{1}{2}a^2 - \frac{1}{2}b^2$ for any $a, b \in \mathbb{R}$.

Next observe

$$(u^{n+1} - u^n, A(u^{n+1} - u^n)) = A\|u^{n+1} - u^n\|_2^2.$$

Finally

$$(u^{n+1} - u^n, f(u^n)) = (f(u^n)(u^{n+1} - u^n), 1),$$

where 1 denotes the constant function with value 1 on Ω. By the Fundamental Theorem of Calculus, we have

$$\begin{aligned}
F(u^{n+1}) - F(u^n) &= f(u^n)(u^{n+1} - u^n) + \int_{u^n}^{u^{n+1}} (u^{n+1} - s)f'(s)ds \\
&= f(u^n)(u^{n+1} - u^n) + \int_{u^n}^{u^{n+1}} (u^{n+1} - s)(3s^2 - 1)ds \\
&= f(u^n)(u^{n+1} - u^n) + 3\int_{u^n}^{u^{n+1}} (u^{n+1} - s)s^2 ds \\
&\quad - \frac{1}{2}(u^{n+1} - u^n)^2.
\end{aligned}$$

By using Theorem 1.1, we have $\|u^n\|_\infty \leq M = \sqrt{\frac{A+1}{3}}$, $\forall\, n \geq 0$. This gives

$$\left|3\int_{u^n}^{u^{n+1}} (u^{n+1} - s)s^2 ds\right| \leq \frac{3}{2}|u^{n+1} - u^n|^2 \cdot M^2.$$

Thus

$$\begin{aligned}&(u^{n+1}-u^n, f(u^n))\\ \geq &\int_\Omega F(u^{n+1})dx - \int_\Omega F(u^n)dx + \frac{1}{2}\|u^{n+1}-u^n\|_2^2\\ &-\frac{3}{2}\|u^{n+1}-u^n\|_2^2\cdot M^2.\end{aligned}$$

Collecting all the estimates, we get

$$\begin{aligned}&(u^{n+1}-u^n, H)\\ \geq &\mathcal{E}(u^{n+1})-\mathcal{E}(u^n)+(A+\frac{1}{2}-\frac{3}{2}M^2)\|u^{n+1}-u^n\|_2^2\\ = &\mathcal{E}(u^{n+1})-\mathcal{E}(u^n)+\frac{A}{2}\|u^{n+1}-u^n\|_2^2.\end{aligned}$$

The desired inequality follows easily.

References

[1] J. W. Cahn and J. E. Hilliard, *Free energy of a nonuniform system. I*, Interfacial energy free energy, J. Chem. Phys. **28** (1958) 258–267.

[2] B. Li, J. Yang, and Z. Zhou, *Arbitrarily high-order exponential cut-off methods for preserving maximum principle of parabolic equations*, SIAM J. Sci. Comput. **42** (2020), no. 6, A3957–A3978, DOI 10.1137/20M1333456. MR4186541

[3] Y. He, Y. Liu, and T. Tang, *On large time-stepping methods for the Cahn-Hilliard equation*, Appl. Numer. Math. **57** (2007), no. 5-7, 616–628, DOI 10.1016/j.apnum.2006.07.026. MR2322435

[4] D. Li, *Effective maximum principles for spectral methods*, Ann. Appl. Math. **37** (2021), no. 2, 131–290, DOI 10.4208/aam.oa-2021-0003. MR4294331

[5] D. Li, C. Quan, and T. Tang, *Stability and convergence analysis for the implicit-explicit method to the Cahn-Hilliard equation*, Math. Comp. **91** (2022), no. 334, 785–809, DOI 10.1090/mcom/3704. MR4379976

[6] D. Li and T. Tang, *Stability of the semi-implicit method for the Cahn-Hilliard equation with logarithmic potentials*, Ann. Appl. Math. **37** (2021), no. 1, 31–60, DOI 10.4208/aam.OA-2020-0003. MR4284064

[7] H. Song and C.-W. Shu, *Unconditional energy stability analysis of a second order implicit-explicit local discontinuous Galerkin method for the Cahn-Hilliard equation*, J. Sci. Comput. **73** (2017), no. 2-3, 1178–1203, DOI 10.1007/s10915-017-0497-5. MR3719623

[8] D. Li, X. Yu, and Z. Zhai, *On the Euler-Poincaré equation with non-zero dispersion*, Arch. Ration. Mech. Anal. **210** (2013), no. 3, 955–974, DOI 10.1007/s00205-013-0662-4. MR3116009

[9] D. Li, *On a frequency localized Bernstein inequality and some generalized Poincaré-type inequalities*, Math. Res. Lett. **20** (2013), no. 5, 933–945, DOI 10.4310/MRL.2013.v20.n5.a9. MR3207362

[10] W. Chen, S. Conde, C. Wang, X. Wang, and S. M. Wise, *A linear energy stable scheme for a thin film model without slope selection*, J. Sci. Comput. **52** (2012), no. 3, 546–562, DOI 10.1007/s10915-011-9559-2. MR2948706

[11] D. J. Eyre, *Unconditionally gradient stable time marching the Cahn-Hilliard equation*, Computational and mathematical models of microstructural evolution (San Francisco, CA, 1998), Mater. Res. Soc. Sympos. Proc., vol. 529, MRS, Warrendale, PA, 1998, pp. 39–46, DOI 10.1557/PROC-529-39. MR1676409

[12] C. Wang, X. Wang, and S. M. Wise, *Unconditionally stable schemes for equations of thin film epitaxy*, Discrete Contin. Dyn. Syst. **28** (2010), no. 1, 405–423, DOI 10.3934/dcds.2010.28.405. MR2629487

[13] J. Shen and X. Yang, *Numerical approximations of Allen-Cahn and Cahn-Hilliard equations*, Discrete Contin. Dyn. Syst. **28** (2010), no. 4, 1669–1691, DOI 10.3934/dcds.2010.28.1669. MR2679727

[14] C. Xu and T. Tang, *Stability analysis of large time-stepping methods for epitaxial growth models*, SIAM J. Numer. Anal. **44** (2006), no. 4, 1759–1779, DOI 10.1137/050628143. MR2257126
[15] J. Shen, J. Xu, and J. Yang, *The scalar auxiliary variable (SAV) approach for gradient flows*, J. Comput. Phys. **353** (2018), 407–416, DOI 10.1016/j.jcp.2017.10.021. MR3723659

SUSTech International Center for Mathematics, and Department of Mathematics, Southern University of Science and Technology, Shenzhen, People's Republic of China
Email address: lid@sustech.edu.cn