Stability Estimate for a Second Order Operator Splitting Time Stepping for Diffusive Systems

Chung-Lin Tseng [1]
Department of Mathematics
National Central University
Hsinchu, Hsinchu 300
TAIWAN

and

Wei-Cheng Wang[2]
Department of Mathematics
National Tsing Hua University
Hsinchu, Hsinchu 300
TAIWAN

Abstract

We provide a rigorous stability analysis for a splitting-based second order time stepping method for linear self-adjoint diffusive equations. The scheme is based on a suitably chosen stabilizing splitting, combined with an implicit second order discretization, also known as stabilized predictor-corrector method. We obtain sufficient conditions on the stabilizing term for the scheme to be unconditionally stable. The proof utilizes discrete energy estimate together with a key observation that the composite linear system corresponding to the time stepping is in fact symmetrizable.

[1]E-mail: tseng0920@gapp.nthu.edu.tw
[2]Corresponding author. E-mail: wangwc@math.nthu.edu.tw

1 Introduction

Many dissipative systems can be modeled by

$$u_t + \mathcal{L}(u) = f, \tag{1}$$

evo1

where $\mathcal{L}$ is a non-negative elliptic operator. For the purpose of stable and efficient time discretization of (1), a conventional wisdom is to perform a suitable operator splitting

$$\mathcal{L}(u) = \mathcal{L}_0(u) + \big(\mathcal{L}(u) - \mathcal{L}_0(u)\big), \tag{2}$$

evo2

and treat $\mathcal{L}_0(u)$ and $\mathcal{L}(u) - \mathcal{L}_0(u)$ separately. Typically, $\mathcal{L}_0$ is a dominant linear operator. To stabilize the time stepping, $\mathcal{L}_0(u)$ is discretized implicitly, while $\mathcal{L}(u) - \mathcal{L}_0(u)$ is discretized explicitly. so that the time marching is stable with fast direct solvers for the resulting linear system:

$$\frac{u^{n+1} - u^n}{\Delta t} + \mathcal{L}_0(u^{n+1}) = \mathcal{L}_0(u^n) - \mathcal{L}(u^n) + f^n, \tag{3}$$

be1

For example, if $\mathcal{L} = -\nabla \cdot (\kappa(\boldsymbol{x})\nabla u)$ is a variable coefficient elliptic operator, a well know splitting takes $\mathcal{L}_0 = -C\Delta$ where Δ is standard Laplacian and $C > 0$ is large enough to stabilize the time marching (see, for example, [6, 3]). Similar stabilization techniques have been adapted in various applications, such as degenerate diffusion system [35], phase field models [33, 25, 24, 26, 27] and nonlinear iteration of steady state computation [14].

There have been several generalization of the stabilization to higher order schemes. A well known 2nd order time discretization (see [29, 15], for example) combines Crank-Nicolson method for $\mathcal{L}_0(u)$ together with 2nd order Adam-Bashforth for $\mathcal{L}(u) - \mathcal{L}_0(u)$ to get (CN-AB2)

$$\frac{u^{n+1} - u^n}{\Delta t} + \mathcal{L}_0\Big(\frac{u^{n+1} + u^n}{2}\Big) = \frac{3}{2}\big(\mathcal{L}_0(u^n) - \mathcal{L}(u^n)\big) - \frac{1}{2}\big(\mathcal{L}_0(u^{n-1}) - \mathcal{L}(u^{n-1})\big) + f^{n+\frac{1}{2}}. \tag{4}$$

cnab2

This scheme was proposed, for example, in [35] to simulate surface diffusion along a moving interface. Up to 3rd order time discretizations based on (stabilized) Backward Differencing Formula and Extrapolation (BD/EP) were proposed in [33] for epitaxial growth models, and proved to be unconditionally stable for 1st and 2nd order BD/EP. A 3rd order BD/EP scheme with a new stabilizing term is proposed in [12] for the no-slope-selection epitaxial thin film growth model and proved to be energy stable and convergent.

In this paper, we consider an alternative stabilized semi-implicit second order scheme (also known as predictor-corrector scheme [21]):

$$\begin{cases} \dfrac{u^{n+\frac{1}{2}} - u^n}{\frac{\Delta t}{2}} + \mathcal{L}_0(u^{n+\frac{1}{2}}) = (\mathcal{L}_0 - \mathcal{L})(u^n) + f^n \\ \dfrac{u^{n+1} - u^n}{\Delta t} + \mathcal{L}_0\Big(\dfrac{u^{n+1} + u^n}{2}\Big) = (\mathcal{L}_0 - \mathcal{L})(u^{n+\frac{1}{2}}) + f^{n+\frac{1}{2}}. \end{cases} \tag{5}$$

rk2

Due to the presence of intermediate step $u^{n+\frac{1}{2}}$ in the scheme (5), sharp stability estimate is difficult to obtain in general. It is well believed that the system is stabilized as long as the stabilizing term $\mathcal{L}_0$ is large enough. On the other hand, we have observed that the overall absolute error also increases with $\mathcal{L}_0$. In this paper, we start with the the simplest linear nonnegative self-adjoint case. One of the byproduct of our analysis is a precise sufficient condition on the stabilizing term $\mathcal{L}_0$ for the scheme (5) to be unconditional stable. See Theorem 2.

For general (possibly nonlinear or nonlocal) elliptic operator $\mathcal{L}$, the second order scheme (5) is a natural generalization of (3) and retains all the numerical advantages. Both steps in (5) are uniquely solvable with fast solvers provided $\mathcal{L}_0$ is a suitable chosen constant coefficient elliptic operator. The stability of (5) is not clear in general. numerical evidence in [21] (and rigorous proof in some cases) suggests that (5) is robust and accurate for gradient flows with strong anisotropic free energy.

Our interest in (5) originated from previous study on efficient preconditioners for linear fractional diffusion equations [28]. One can take $\mathcal{L}_0$ to be a preconditioner of $\mathcal{L}$ whenever it is available. We have observed that although both (4) and (5) are unconditionally stable for constant coefficient fractional diffusion operators (corresponding to linear self-adjoint $\mathcal{L}$), (5) appears to be more robust and stable than (4) for variable coefficient fractional diffusion operators (corresponding to linear non-self-adjoint $\mathcal{L}$). Similar performances have been observed for the problems studied in [21]. We believe this is due to lack of extrapolation steps in (5).

We remark here that when both $\mathcal{L}$ and $\mathcal{L}_0$ are linear, self-adjoint and $\mathcal{L}_0$ is large enough so that

$$\mathcal{L}_0 \geq 0, \qquad \mathcal{L}_0 - \mathcal{L} \geq 0, \tag{6}$$

we can write

$$\mathcal{L}(u) = \frac{\partial}{\partial u}\mathcal{E}(u) = \frac{\partial}{\partial u}(\mathcal{E}_c(u) - \mathcal{E}_e(u)), \tag{7}$$

where both $\mathcal{E}_c(u) = \frac{1}{2}(u, \mathcal{L}_0 u)$ and $\mathcal{E}_e(u) = \frac{1}{2}(u, (\mathcal{L}_0 - \mathcal{L})u)$ are convex. In this case, (3) belongs to the class of convex splitting scheme for gradient flows ([8, 10, 30, 31, 32], assuming $f = 0$ for simplicity):

$$\frac{u^{n+1} - u^n}{\Delta t} = -\Big(\frac{\partial}{\partial u}\mathcal{E}_c(u^{n+1}) - \frac{\partial}{\partial u}\mathcal{E}_e(u^n)\Big). \tag{8}$$

The scheme (8) is equipped with discrete energy law $\mathcal{E}(u^{n+1}) \leq \mathcal{E}(u^n)$ (therefore unconditionally stable) provided $\mathcal{E} = \mathcal{E}_c - \mathcal{E}_e$ and both $\mathcal{E}_c$ and $\mathcal{E}_e$ are convex. The convex splitting scheme (8) in its current form is 1st order accurate in time, and can be improved to 2nd order in some cases [19, 2, 34].

In addition to the splitting (6), a new class of energy stable methods, the Single Auxiliary Variable (SAV) method [20, 22, 4, 5] inspired by the Lagrangian multiplier approach [1, 11, 36, 39], are based on the splitting

$$\mathcal{E}(u) = \frac{1}{2}(u, \mathcal{L}_0 u) + \mathcal{E}_1(u), \tag{9}$$

where $\mathcal{L}_0$ is symmetric nonnegative linear operator and $\mathcal{E}_1$ is bounded below. This splitting, when applied to the current case amounts to requiring

$$\mathcal{L}_0 \geq 0, \qquad \mathcal{L} - \mathcal{L}_0 \geq 0, \tag{10}$$ split2

in contrast to (6). See [23] for a more thorough introduction of SAV.

In addition to the operator splitting methods mentioned above, the classical Strang splitting is one of the most well known and popular operator splitting methods. See [16, 38, 17], on recent progress on stability and convergence analysis for the Strang splitting scheme applied to various nonlinear PDEs. See also [37] on an iterative high order time discretization schemes using lower order operator splitting methods as preconditioners.

The rest of the paper is organized as follows. In section 2, we give the fundamental stability analysis for the linear problem. The novelty of our approach is to recast the 2nd order Stabilized Predictor-Corrector scheme (26) in a way that is structurally identical to the stabilized first order scheme and obtain similar energy estimates that leads to stability and convergence analysis for linear problems. In section 3, we extend our analysis to a class of semilinear diffusion equations including the generalized Allen-Cahn equation. The proof is based on the energy estimate developed for linear problems, combined with discrete Gronwall's inequality and leads to. finite time stability and 2nd order in time convergence rate results. Finally, we conduct various numerical experiments in Section 4 to support the theoretical analysis. We

2 The Schemes and Main Results for Linear Problems

main

We first rewrite (1), after spatial discretization, in matrix and vector notations:

$$\boldsymbol{u}_t + s\mathbb{M}\boldsymbol{u} = (s\mathbb{M} - \mathbb{A})\boldsymbol{u} + \boldsymbol{f}. \tag{11}$$ evo3

Here $\boldsymbol{u}$ and $\boldsymbol{f}$ are vector valued grid functions in $\mathbf{R}^p$. Denote by $\mathcal{L}_h$ and $\mathcal{L}_{0,h}$ the spatial discretization of $\mathcal{L}$ and $\mathcal{L}_{0,h}$ and their matrix representation by

$$\mathbb{A}: \text{ matrix representation of } \mathcal{L}_h, \qquad s\mathbb{M}: \text{ matrix representation of } \mathcal{L}_{0,h}$$

$\mathbb{M}$ and $\mathbb{A}$ are real symmetric $p \times p$ matrices. By assumption $\mathbb{M} = \mathbb{M}^{\mathrm{T}} > 0, \mathbb{A} = \mathbb{A}^{\mathrm{T}} > 0$. In practice, we can take $\mathbb{M}$ to be a preconditioner of $\mathbb{A}$, if available and s is a constant to be determined.

For two real symmetric matrices $\mathbb{P}$ and $\mathbb{Q}$, we write $\mathbb{P} > \mathbb{Q}$ provided $\boldsymbol{u}^{\mathrm{T}}\mathbb{P}\boldsymbol{u} > \boldsymbol{u}^{\mathrm{T}}\mathbb{Q}\boldsymbol{u}$ for all $\boldsymbol{u} \neq \mathbf{0}$. Similarly for $\mathbb{P} \geq \mathbb{Q}$. We also denote the standard L^2 and weighted inner products and norms in $\mathbf{R}^p$ for any $\mathbb{W} = \mathbb{W}^{\mathrm{T}} \geq 0$ by

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{v}, \qquad \langle \boldsymbol{u}, \boldsymbol{v} \rangle_{\mathbb{W}} = \boldsymbol{u}^{\mathrm{T}}\mathbb{W}\boldsymbol{v}, \tag{12}$$

and

$$\|\boldsymbol{u}\|^2 = \langle \boldsymbol{u}, \boldsymbol{u} \rangle, \qquad \|\boldsymbol{u}\|^2_{\mathbb{W}} = \langle \boldsymbol{u}, \boldsymbol{u} \rangle_{\mathbb{W}}. \tag{13}$$

aos1

2.1 Stabilized Backward Euler Scheme

In the matrix vector notation, the stabilized Backward Euler method (3) reads

$$\frac{\boldsymbol{u}^{n+1}-\boldsymbol{u}^n}{\Delta t} + s\mathbb{M}\boldsymbol{u}^{n+1} = (s\mathbb{M}-\mathbb{A})\boldsymbol{u}^n + \boldsymbol{f}^n, \tag{14}$$ be2

We will show that (14) are stable both in energy norm and L^2 norm. The argument is straight forward, see also [33], for example. We include it here for readers convenience in reading section 2.2.

thmbe1 **Theorem 1.** *If $s\mathbb{M} \geq \frac{1}{2}\mathbb{A}$, then the stabilized backward Euler scheme (14) is unconditionally stable and satisfies*

$$\|\boldsymbol{u}^N\|_{\mathbb{A}}^2 \leq \|\boldsymbol{u}^0\|_{\mathbb{A}}^2 + \frac{\Delta t}{2}\sum_{n=1}^{N-1}\|\boldsymbol{f}^n\|^2, \tag{15}$$ aest1

and

$$\|\boldsymbol{u}^N\|^2 + \Delta t\|\boldsymbol{u}^N\|_{s\mathbb{M}-\frac{\mathbb{A}}{2}}^2 \leq \|\boldsymbol{u}^0\|^2 + \Delta t\|\boldsymbol{u}^0\|_{s\mathbb{M}-\frac{\mathbb{A}}{2}}^2 + \frac{\Delta t}{2}\sum_{n=0}^{N-1}\|\boldsymbol{f}^n\|_{\mathbb{A}^{-1}}^2. \tag{16}$$ l2est1

Proof. Denote by

$$\mathbb{S}_1 \equiv \Big(\frac{1}{\Delta t}\mathbb{I} + s\mathbb{M}\Big), \tag{17}$$ defQ

and rewrite (14)

$$\mathbb{S}_1(\boldsymbol{u}^{n+1}-\boldsymbol{u}^n) + \mathbb{A}\boldsymbol{u}^n = \boldsymbol{f}^n, \tag{18}$$

and therefore

$$\Big(\mathbb{S}_1 - \frac{\mathbb{A}}{2}\Big)(\boldsymbol{u}^{n+1}-\boldsymbol{u}^n) + \frac{\mathbb{A}}{2}(\boldsymbol{u}^{n+1}+\boldsymbol{u}^n) = \boldsymbol{f}^n. \tag{19}$$ BE3

Take $\langle \boldsymbol{u}^{n+1}-\boldsymbol{u}^n, \bullet\rangle$ on both sides of (19), we see that

$$\|\boldsymbol{u}^{n+1}-\boldsymbol{u}^n\|_{\mathbb{S}_1-\frac{\mathbb{A}}{2}}^2 + \frac{1}{2}\left(\|\boldsymbol{u}^{n+1}\|_{\mathbb{A}}^2 - \|\boldsymbol{u}^n\|_{\mathbb{A}}^2\right) = \langle \boldsymbol{u}^{n+1}-\boldsymbol{u}^n, \boldsymbol{f}^n\rangle. \tag{20}$$ qmha11

From (17), we have

$$\|\cdot\|_{\mathbb{S}_1-\frac{\mathbb{A}}{2}}^2 = \frac{1}{\Delta t}\|\cdot\|^2 + \|\cdot\|_{s\mathbb{M}-\frac{\mathbb{A}}{2}}^2, \tag{21}$$ qmha2

therefore

$$\frac{1}{\Delta t}\|\boldsymbol{u}^{n+1}-\boldsymbol{u}^n\|^2 + \|\boldsymbol{u}^{n+1}-\boldsymbol{u}^n\|_{s\mathbb{M}-\frac{\mathbb{A}}{2}}^2 + \frac{1}{2}\left(\|\boldsymbol{u}^{n+1}\|_{\mathbb{A}}^2 - \|\boldsymbol{u}^n\|_{\mathbb{A}}^2\right) \leq \frac{1}{\Delta t}\|\boldsymbol{u}^{n+1}-\boldsymbol{u}^n\|^2 + \frac{\Delta t}{4}\|\boldsymbol{f}^n\|^2. \tag{22}$$

By assumption, $s\mathbb{M} - \frac{\mathbb{A}}{2} \geq 0$, therefore

$$\frac{1}{2}\left(\|\boldsymbol{u}^{n+1}\|_{\mathbb{A}}^2 - \|\boldsymbol{u}^n\|_{\mathbb{A}}^2\right) \leq \frac{\Delta t}{4}\|\boldsymbol{f}^n\|^2, \tag{23}$$ be_a1

which leads to (15) after summing over n.

The energy norm estimate (15) along is sufficient to assert unconditional stability of (14). For sake of completeness, we proceed with the L^2 estimate (16). Take the standard inner product with $\boldsymbol{u}^{n+1}+\boldsymbol{u}^n$ on both sides of (19) to get

$$\|\boldsymbol{u}^{n+1}\|^2_{\mathbb{S}_1-\frac{\mathbb{A}}{2}} - \|\boldsymbol{u}^{n}\|^2_{\mathbb{S}_1-\frac{\mathbb{A}}{2}} + \frac{1}{2}\|\boldsymbol{u}^{n+1}+\boldsymbol{u}^n\|^2_{\mathbb{A}} = \langle \boldsymbol{u}^{n+1}+\boldsymbol{u}^n, \boldsymbol{f}^n\rangle \le \frac{1}{2}\|\boldsymbol{u}^{n+1}+\boldsymbol{u}^n\|^2_{\mathbb{A}} + \frac{1}{2}\|\boldsymbol{f}^n\|^2_{\mathbb{A}^{-1}}. \tag{24}$$

qmha1

From (24) and (21), we have

$$\Big(\|\boldsymbol{u}^{n+1}\|^2 + \Delta t\|\boldsymbol{u}^{n+1}\|^2_{s\mathbb{M}-\frac{\mathbb{A}}{2}}\Big) - \Big(\|\boldsymbol{u}^{n}\|^2 + \Delta t\|\boldsymbol{u}^{n}\|^2_{s\mathbb{M}-\frac{\mathbb{A}}{2}}\Big) \le \frac{\Delta t}{2}\|\boldsymbol{f}^n\|^2_{\mathbb{A}^{-1}}, \tag{25}$$

which leads directly to (16) upon summing over n. This completes the proof of (16)

□

2.2 Stabilized Predictor-Corrector Scheme

aos2

In this section, we will give a rigorous stability estimate of the stabilized predictor-corrector scheme:

$$\begin{cases} \dfrac{\boldsymbol{u}^{n+\frac{1}{2}}-\boldsymbol{u}^n}{\frac{\Delta t}{2}} + s\mathbb{M}\boldsymbol{u}^{n+\frac{1}{2}} = (s\mathbb{M}-\mathbb{A})\boldsymbol{u}^n + \boldsymbol{f}^n \\ \dfrac{\boldsymbol{u}^{n+1}-\boldsymbol{u}^n}{\Delta t} + s\mathbb{M}\dfrac{\boldsymbol{u}^{n+1}+\boldsymbol{u}^n}{2} = (s\mathbb{M}-\mathbb{A})\boldsymbol{u}^{n+\frac{1}{2}} + \boldsymbol{f}^{n+\frac{1}{2}}, \end{cases} \tag{26}$$

OSRK2

where $\mathbb{M}$ and $\mathbb{A}$ are as in section 2.1. Here $s\mathbb{M}$ is the matrix representation of the stabilizing term $\mathcal{L}_0$ and s is another undetermined constant.

The discrete energy estimate for the stabilized Backward Euler method (14) in section 2.1 can probably be extended to higher order schemes including the stabilized BD2/EP2 in [33]. and stabilized CN-AB2 in [19] and [9]. See also [18] for energy estimate of up to BDF5 scheme for Stokes equations. The authors in [21] took a different approach and obtained L^2 error estimate of (5) applied to the isotropic Cahn-Hilliard equation.

Our goal is to establish similar estimates as in Theorem 1 for the second order scheme (26). To this end, we first eliminate the intermediate variable $\boldsymbol{u}^{n+\frac{1}{2}}$ by rewriting the first equation of (26) as

$$\Big(\frac{2}{\Delta t}\mathbb{I} + s\mathbb{M}\Big)\boldsymbol{u}^{n+\frac{1}{2}} = \Big(\frac{2}{\Delta t}\mathbb{I} + s\mathbb{M} - \mathbb{A}\Big)\boldsymbol{u}^n + \boldsymbol{f}^n, \tag{27}$$

and substitute it back to get

$$\begin{aligned}\Big(\frac{1}{\Delta t}\mathbb{I} + \frac{s\mathbb{M}}{2}\Big)\boldsymbol{u}^{n+1} = \Big(\frac{1}{\Delta t}\mathbb{I} + \frac{s\mathbb{M}}{2} - \mathbb{A} - (s\mathbb{M}-\mathbb{A})\big(\frac{2}{\Delta t}\mathbb{I} + s\mathbb{M}\big)^{-1}\mathbb{A}\Big)\boldsymbol{u}^n \\ + (s\mathbb{M}-\mathbb{A})\big(\frac{2}{\Delta t}\mathbb{I} + s\mathbb{M}\big)^{-1}\boldsymbol{f}^n + \boldsymbol{f}^{n+\frac{1}{2}}.\end{aligned} \tag{28}$$

evol1

To simplify the expression, we denote by

$$\mathbb{Q} \equiv \frac{2}{\Delta t}\mathbb{I} + s\mathbb{M} = \mathbb{Q}^{\mathrm{T}} > 0, \tag{29}$$

deftQ

$$\mathbb{T} \equiv (s\mathbb{M} - \mathbb{A}) = \mathbb{T}^{\mathrm{T}}, \tag{30}$$

defT

and rewrite (28) as

$$\begin{aligned}
\frac{1}{2}\mathbb{Q}\boldsymbol{u}^{n+1} &= \Big(\frac{1}{2}\mathbb{Q} - \mathbb{A} - \mathbb{T}\mathbb{Q}^{-1}\mathbb{A}\Big)\boldsymbol{u}^n + \mathbb{T}\mathbb{Q}^{-1}\boldsymbol{f}^n + \boldsymbol{f}^{n+\frac{1}{2}} \\
&= \Big(\frac{1}{2}\mathbb{Q} - \big(\mathbb{I} + \mathbb{T}\mathbb{Q}^{-1}\big)\mathbb{A}\Big)\boldsymbol{u}^n + \mathbb{T}\mathbb{Q}^{-1}\boldsymbol{f}^n + \boldsymbol{f}^{n+\frac{1}{2}} \\
&= \Big(\frac{1}{2}\mathbb{Q} - \big(\mathbb{Q} + \mathbb{T}\big)\mathbb{Q}^{-1}\mathbb{A}\Big)\boldsymbol{u}^n + \mathbb{T}\mathbb{Q}^{-1}\boldsymbol{f}^n + \boldsymbol{f}^{n+\frac{1}{2}},
\end{aligned} \tag{31}$$

or

$$\begin{aligned}
\boldsymbol{u}^{n+1} &= (\mathbb{I} - 2\mathbb{Q}^{-1}\big(\mathbb{Q} + \mathbb{T}\big)\mathbb{Q}^{-1}\mathbb{A})\boldsymbol{u}^n + 2\mathbb{Q}^{-1}\mathbb{T}\mathbb{Q}^{-1}\boldsymbol{f}^n + 2\mathbb{Q}^{-1}\boldsymbol{f}^{n+\frac{1}{2}} \\
&= (\mathbb{I} - \mathbb{S}_2^{-1}\mathbb{A})\boldsymbol{u}^n + 2\mathbb{Q}^{-1}\mathbb{T}\mathbb{Q}^{-1}\boldsymbol{f}^n + 2\mathbb{Q}^{-1}\boldsymbol{f}^{n+\frac{1}{2}},
\end{aligned} \tag{32}$$

2ndeq1

where

$$\mathbb{S}_2 \equiv \frac{1}{2}\mathbb{Q}(\mathbb{Q} + \mathbb{T})^{-1}\mathbb{Q}. \tag{33}$$

defS

It is crucial to note that

$$\mathbb{S}_2 = \mathbb{S}_2^{\mathrm{T}} > 0, \tag{34}$$

sst

therefore (26) can be symmetrized and put in a form that completely resembles the 1st order scheme (19):

$$\mathbb{S}_2(\boldsymbol{u}^{n+1} - \boldsymbol{u}^n) + \mathbb{A}\boldsymbol{u}^n = \mathbb{G}\boldsymbol{f}^n + \mathbb{H}\boldsymbol{f}^{n+\frac{1}{2}}, \tag{35}$$

2ndeq2old

or

$$\big(\mathbb{S}_2 - \frac{\mathbb{A}}{2}\big)(\boldsymbol{u}^{n+1} - \boldsymbol{u}^n) + \frac{\mathbb{A}}{2}(\boldsymbol{u}^{n+1} + \boldsymbol{u}^n) = \mathbb{G}\boldsymbol{f}^n + \mathbb{H}\boldsymbol{f}^{n+\frac{1}{2}}, \tag{36}$$

2ndeq2

where

$$\mathbb{G} \equiv 2\mathbb{S}_2\mathbb{Q}^{-1}\mathbb{T}\mathbb{Q}^{-1}, \quad \mathbb{H} \equiv 2\mathbb{S}_2\mathbb{Q}^{-1}. \tag{37}$$

defg

Upon comparing (19) with (32), and inspecting the left hand side of (20) and (24), it is clear that the following estimates are crucial for the estimates of the 2nd order scheme (36):

lemma1 **Lemma 1.** *If* $\mathbb{A} = \mathbb{A}^{\mathrm{T}} > 0$, $\mathbb{M} = \mathbb{M}^{\mathrm{T}} > 0$ *and* $s\mathbb{M} \geq \mathbb{A}$, *then*

(i)

$$\frac{1}{\Delta t}\mathbb{I} + \frac{\mathbb{A}}{2} \leq \mathbb{S}_2, \tag{38}$$

SmhA

(ii)

$$\mathbb{S}_2 \leq \frac{\mathbb{Q}}{2} = \frac{1}{\Delta t}\mathbb{I} + \frac{s\mathbb{M}}{2}, \tag{39}$$

ShtQ

where $\mathbb{S}_2$, $\mathbb{Q}$ *are defined by* (33), (29) *and* (30).

Proof. **Part (i)**: Denote by $\mathbb{B} = \mathbb{A} + \dfrac{2}{\Delta t}\mathbb{I}$ and $\lambda_{\max}$, $\mu_{\max}$ the maximal eigenvalue of

$$\mathbb{B}\boldsymbol{v} = \lambda\mathbb{Q}\boldsymbol{v} \quad \text{and} \quad \mathbb{B}\boldsymbol{v} = \mu\mathbb{S}_2\boldsymbol{v}, \tag{40}$$

eig1

respectively. We will show that

$$\mathbb{A} \le s\mathbb{M} \iff \mathbb{B} \le \mathbb{Q} \implies \lambda_{\max} \le 1 \implies \mu_{\max} \le 2. \tag{41}$$

eig2

The statements in (41) are obvious except the last implication. To verify it, we first note that the pencils in (40) are isospectral to

$$\mathbb{B}^{\frac{1}{2}}\mathbb{Q}^{-1}\mathbb{B}^{\frac{1}{2}}\boldsymbol{w} = \lambda\boldsymbol{w} \quad \text{and} \quad \mathbb{B}^{\frac{1}{2}}\mathbb{S}_2^{-1}\mathbb{B}^{\frac{1}{2}}\boldsymbol{w} = \mu\boldsymbol{w}, \tag{42}$$

eig3

respectively, where $\boldsymbol{w} = \mathbb{B}^{\frac{1}{2}}\boldsymbol{v}$. Therefore

$$\begin{aligned}
\mu_{\max} &= \max_{\boldsymbol{x}\in\mathbf{R}^p\setminus\{\mathbf{0}\}} 2\,\frac{\boldsymbol{x}^{\mathrm{T}}\mathbb{B}^{\frac{1}{2}}\mathbb{S}_2^{-1}\mathbb{B}^{\frac{1}{2}}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}} \\
&= \max_{\boldsymbol{x}\in\mathbf{R}^p\setminus\{\mathbf{0}\}} 2\,\frac{\boldsymbol{x}^{\mathrm{T}}\mathbb{B}^{\frac{1}{2}}\mathbb{Q}^{-1}(2\mathbb{Q}-\mathbb{B})\mathbb{Q}^{-1}\mathbb{B}^{\frac{1}{2}}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}} \qquad (\text{since } \mathbb{Q}+\mathbb{T} = 2\mathbb{Q}-\mathbb{B}) \\
&= \max_{\boldsymbol{x}\in\mathbf{R}^p\setminus\{\mathbf{0}\}} 2\,\frac{\boldsymbol{x}^{\mathrm{T}}\left(\mathbb{B}^{\frac{1}{2}}\mathbb{Q}^{-1}\mathbb{B}^{\frac{1}{2}}\right)\left(\mathbb{B}^{-\frac{1}{2}}(2\mathbb{Q}-\mathbb{B})\mathbb{B}^{-\frac{1}{2}}\right)\left(\mathbb{B}^{\frac{1}{2}}\mathbb{Q}^{-1}\mathbb{B}^{\frac{1}{2}}\right)\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}} \\
&= \max_{\boldsymbol{x}\in\mathbf{R}^p\setminus\{\mathbf{0}\}} 2\,\frac{\boldsymbol{x}^{\mathrm{T}}\widehat{\mathbb{B}}(2\widehat{\mathbb{B}}^{-1}-1)\widehat{\mathbb{B}}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}, \qquad \text{where} \quad \widehat{\mathbb{B}} \equiv \mathbb{B}^{\frac{1}{2}}\mathbb{Q}^{-1}\mathbb{B}^{\frac{1}{2}}.
\end{aligned} \tag{43}$$

eig4

From (42) and the definition of $\widehat{\mathbb{B}}$, we see that

$$\mu_{\max} \le 2\max_j \lambda_j\Big(2\frac{1}{\lambda_j}-1\Big)\lambda_j = \max_j\left(4\lambda_j - 2\lambda_j^2\right) = \max_j 2\Big(1-\big(1-\lambda_j\big)^2\Big). \tag{44}$$

tmp7

Since $\lambda_j \le \lambda_{\max} \le 1$, it follows that $\mu_{\max} \le 2$. This completes the proof of (41), hence (38).
Part (ii): From (33),

$$\mathbb{S}_2 = \frac{1}{2}\mathbb{Q}(\mathbb{Q}+\mathbb{T})^{-1}\mathbb{Q} = \mathbb{Q}\Big(\frac{4}{\Delta t}\mathbb{I} + 4s\mathbb{M} - 2\mathbb{A}\Big)^{-1}\mathbb{Q}. \tag{45}$$

qs1

Since $s\mathbb{M} \ge \mathbb{A}$, we see that

$$\frac{4}{\Delta t}\mathbb{I} + 4s\mathbb{M} - 2\mathbb{A} \ge \frac{4}{\Delta t}\mathbb{I} + 2s\mathbb{M} = 2\mathbb{Q}. \tag{46}$$

qs2

Therefore (39) follows from (45) and (46). □

Theorem 2. *Under the same assumptions in Lemma 1, the stabilized predictor-corrector scheme (26) (or (36)) is unconditionally stable and admits the following estimates:*

$$\|\boldsymbol{u}^N\|_{\mathbb{A}}^2 \le \|\boldsymbol{u}^0\|_{\mathbb{A}}^2 + \Delta t \sum_{n=0}^{N-1} \big(\|\boldsymbol{f}^n\|^2 + \|\boldsymbol{f}^{n+\frac{1}{2}}\|^2\big), \tag{47}$$

and

$$\|\boldsymbol{u}^N\|^2 \le \|\boldsymbol{u}^0\|_{\widetilde{\mathbb{I}}}^2 + \ \Delta t \sum_{n=0}^{N-1} \big(\|\boldsymbol{f}^n\|_{\mathbb{A}^{-1}}^2 + \|\boldsymbol{f}^{n+\frac{1}{2}}\|_{\mathbb{A}^{-1}}^2\big), \tag{48}$$

where

$$\widetilde{\mathbb{I}} = \frac{\Delta t}{2}\mathbb{Q} = \mathbb{I} + \frac{\Delta t}{2} s\mathbb{M}. \tag{49}$$

Proof. **Proof of** (47): Take standard inner product with $\boldsymbol{u}^{n+1} - \boldsymbol{u}^n$ on both sides of (36), we see that

$$\|\boldsymbol{u}^{n+1} - \boldsymbol{u}^n\|_{\mathbb{S}_2 - \frac{\mathbb{A}}{2}}^2 + \frac{1}{2}(\|\boldsymbol{u}^{n+1}\|_{\mathbb{A}}^2 - \|\boldsymbol{u}^n\|_{\mathbb{A}}^2) = \langle \boldsymbol{u}^{n+1} - \boldsymbol{u}^n, \mathbb{G}\boldsymbol{f}^n + \mathbb{H}\boldsymbol{f}^{n+\frac{1}{2}}\rangle. \tag{50}$$

With (38) and the inequality

$$\langle \boldsymbol{u}^{n+1} - \boldsymbol{u}^n, \mathbb{G}\boldsymbol{f}^n + \mathbb{H}\boldsymbol{f}^{n+\frac{1}{2}}\rangle \le \frac{1}{\Delta t}\|\boldsymbol{u}^{n+1} - \boldsymbol{u}^n\|^2 + \frac{\Delta t}{4}\|\mathbb{G}\boldsymbol{f}^n + \mathbb{H}\boldsymbol{f}^{n+\frac{1}{2}}\|^2, \tag{51}$$

we see from (38) that

$$\frac{1}{2}(\|\boldsymbol{u}^{n+1}\|_{\mathbb{A}}^2 - \|\boldsymbol{u}^n\|_{\mathbb{A}}^2) \le \frac{\Delta t}{4}\|\mathbb{G}\boldsymbol{f}^n + \mathbb{H}\boldsymbol{f}^{n+\frac{1}{2}}\|^2 \le \frac{\Delta t}{2}\|\mathbb{G}\boldsymbol{f}^n\|^2 + \frac{\Delta t}{2}\|\mathbb{H}\boldsymbol{f}^{n+\frac{1}{2}}\|^2. \tag{52}$$

To estimate the right hand side of (52), we note from (37) that

$$\|\mathbb{G}\boldsymbol{f}^n\|^2 = \|2\mathbb{S}_2\mathbb{Q}^{-1}\mathbb{T}\mathbb{Q}^{-1}\boldsymbol{f}^n\|^2, \quad \|\mathbb{H}\boldsymbol{f}^{n+\frac{1}{2}}\|^2 = \|2\mathbb{S}_2\mathbb{Q}^{-1}\boldsymbol{f}^{n+\frac{1}{2}}\|^2. \tag{53}$$

From (39) and the following inequality

$$\mathbb{T} = s\mathbb{M} - \mathbb{A} \le s\mathbb{M} \le \mathbb{Q}, \tag{54}$$

we have

$$\|\mathbb{G}\boldsymbol{f}^n\|^2 = \|2\mathbb{S}_2\mathbb{Q}^{-1}\mathbb{T}\mathbb{Q}^{-1}\boldsymbol{f}^n\|^2 \le \|\mathbb{T}\mathbb{Q}^{-1}\boldsymbol{f}^n\|^2 \le \|\boldsymbol{f}^n\|^2. \tag{55}$$

Similarly,

$$\|\mathbb{H}\boldsymbol{f}^{n+\frac{1}{2}}\|^2 = \|2\mathbb{S}_2\mathbb{Q}^{-1}\boldsymbol{f}^{n+\frac{1}{2}}\|^2 \le \|\boldsymbol{f}^{n+\frac{1}{2}}\|^2. \tag{56}$$

From (52), (55), (56) we obtain

$$\|\boldsymbol{u}^{n+1}\|_{\mathbb{A}}^2 \le \|\boldsymbol{u}^n\|_{\mathbb{A}}^2 + \Delta t\big(\|\boldsymbol{f}^n\|^2 + \|\boldsymbol{f}^{n+\frac{1}{2}}\|^2\big), \tag{57}$$

which completes the proof of (47) after summing over n.

The energy norm estimate (47) alone is sufficient for unconditional stability of (26). We proceed with the L^2 estimate (48) for sake of completeness.

Proof of (48):

Recall the derivation that leads to (24) and apply it to (36), we get the analogue of (24) for (36):

$$\|\boldsymbol{u}^{n+1}\|^2_{\mathbb{S}_2-\frac{\mathbb{A}}{2}} - \|\boldsymbol{u}^{n}\|^2_{\mathbb{S}_2-\frac{\mathbb{A}}{2}} \le \frac{1}{2}\|\mathbb{G}\boldsymbol{f}^n + \mathbb{H}\boldsymbol{f}^{n+\frac{1}{2}}\|^2_{\mathbb{A}^{-1}} \le \|\mathbb{G}\boldsymbol{f}^n\|^2_{\mathbb{A}^{-1}} + \|\mathbb{H}\boldsymbol{f}^{n+\frac{1}{2}}\|^2_{\mathbb{A}^{-1}}. \tag{58}$$

From (39), we see that that $\mathbb{A}^{-\frac{1}{2}}(2\mathbb{S}_2)\mathbb{A}^{-\frac{1}{2}} \le \mathbb{A}^{-\frac{1}{2}}\mathbb{Q}\mathbb{A}^{-\frac{1}{2}}$. Therefore

$$\begin{aligned}\|\mathbb{H}\boldsymbol{f}^{n+\frac{1}{2}}\|^2_{\mathbb{A}^{-1}} =&\|\mathbb{A}^{-\frac{1}{2}}2\mathbb{S}_2\mathbb{Q}^{-1}\boldsymbol{f}^{n+\frac{1}{2}}\|^2 = \|\mathbb{A}^{-\frac{1}{2}}2\mathbb{S}_2\mathbb{A}^{-\frac{1}{2}}(\mathbb{A}^{-\frac{1}{2}}\mathbb{Q}\mathbb{A}^{-\frac{1}{2}})^{-1}\mathbb{A}^{-\frac{1}{2}}\boldsymbol{f}^{n+\frac{1}{2}}\|^2\\ \le&\|\mathbb{A}^{-\frac{1}{2}}\boldsymbol{f}^{n+\frac{1}{2}}\|^2 = \|\boldsymbol{f}^{n+\frac{1}{2}}\|^2_{\mathbb{A}^{-1}}.\end{aligned} \tag{59}$$

Similarly, from (39) and (54),

$$\|\mathbb{G}\boldsymbol{f}^n\|^2_{\mathbb{A}^{-1}} = \|\mathbb{A}^{-\frac{1}{2}}2\mathbb{S}_2\mathbb{Q}^{-1}\mathbb{T}\mathbb{Q}^{-1}\boldsymbol{f}^n\|^2 \le \|\mathbb{A}^{-\frac{1}{2}}\mathbb{T}\mathbb{Q}^{-1}\boldsymbol{f}^n\|^2 \le \|\mathbb{A}^{-\frac{1}{2}}\boldsymbol{f}^n\|^2 = \|\boldsymbol{f}^n\|^2_{\mathbb{A}^{-1}}. \tag{60}$$

After summing (58) over n and applying the estimate (60), (59), we get

$$\|\boldsymbol{u}^N\|^2_{\mathbb{S}_2-\frac{\mathbb{A}}{2}} \le \|\boldsymbol{u}^0\|^2_{\mathbb{S}_2-\frac{\mathbb{A}}{2}} + \sum_{n=0}^{N-1}\big(\|\boldsymbol{f}^n\|^2_{\mathbb{A}^{-1}} + \|\boldsymbol{f}^{n+\frac{1}{2}}\|^2_{\mathbb{A}^{-1}}\big). \tag{61}$$

From (38) and (39) and (61), we have

$$\frac{1}{\Delta t}\|\boldsymbol{u}^N\|^2 \le \frac{1}{2}\big(\frac{2}{\Delta t}\|\boldsymbol{u}^0\|^2 + s\|\boldsymbol{u}^0\|^2_{\mathbb{M}}\big) + \sum_{n=0}^{N-1}\big(\|\boldsymbol{f}^n\|^2_{\mathbb{A}^{-1}} + \|\boldsymbol{f}^{n+\frac{1}{2}}\|^2_{\mathbb{A}^{-1}}\big). \tag{62}$$

Thus (48) follows from multiplying (62) by Δt. □

Remark 1: (Alternative estimates for convergence proof)

An alternative estimate to (55) is given by

$$\|\mathbb{G}\boldsymbol{f}^n\|^2 = \|2\mathbb{S}_2\mathbb{Q}^{-1}\mathbb{T}\mathbb{Q}^{-1}\boldsymbol{f}^n\|^2 \le \|\mathbb{T}\mathbb{Q}^{-1}\boldsymbol{f}^n\|^2 \le \|s\mathbb{M}(\frac{2}{\Delta t}\widetilde{\mathbb{I}})^{-1}\boldsymbol{f}^n\|^2 \le \|\frac{s}{2}\mathbb{M}(\Delta t\boldsymbol{f}^n)\|^2. \tag{63}$$

where we have used the estimates (38), (54) and (49).

Similarly, in addition to (60), we also have

$$\begin{aligned}\|\mathbb{G}\boldsymbol{f}^n\|^2_{\mathbb{A}^{-1}} =&\|\mathbb{A}^{-\frac{1}{2}}2\mathbb{S}_2\mathbb{Q}^{-1}\mathbb{T}\mathbb{Q}^{-1}\boldsymbol{f}^n\|^2 \le \|\mathbb{A}^{-\frac{1}{2}}\mathbb{T}\mathbb{Q}^{-1}\boldsymbol{f}^n\|^2\\ \le&\|\mathbb{A}^{-\frac{1}{2}}s\mathbb{M}(\frac{2}{\Delta t}\widetilde{\mathbb{I}})^{-1}\boldsymbol{f}^n\|^2 \le \|\frac{s}{2}\mathbb{M}(\Delta t\boldsymbol{f}^n)\|^2_{\mathbb{A}^{-1}}.\end{aligned} \tag{64}$$

We will give a rigorous error estimate for (26) in the next subsection. There the extra Δt factor in (63) and (64) is essential to compensate for first order local truncation error resulting from the first step of (26). See section 2.3 for details.

Remark 2: (Relaxing the assumption $s\mathbb{M} \geq \mathbb{A}$)

The assumption $s\mathbb{M} \geq \mathbb{A}$ in Lemma 1 and Theorem 2 can be relaxed to $s\mathbb{M} \geq \dfrac{1}{2}\mathbb{A}$, resulting in estimates slightly different from (47) and (48). To see this, we denote by $\theta_{\max}$ the maximal eigenvalue of

$$\mathbb{A}\boldsymbol{v} = \theta s\mathbb{M}\boldsymbol{v}. \tag{65}$$

Our original assumption $s\mathbb{M} \geq \mathbb{A}$ corresponds to $\theta_{\max} \leq 1$, which eventually leads to (41) and hence (38). We now claim that

$$s\mathbb{M} \geq \frac{\mathbb{A}}{2} \iff \theta_{\max} \leq 2 \implies \lambda_{\max} < 2 \implies \mu_{\max} \leq 2. \tag{41'}$$

To see this, we note that the if and only if part in (41') is obvious, and the first implication follows from

$$\mathbb{A} \leq 2s\mathbb{M} \implies \mathbb{B} = \mathbb{A} + \frac{2}{\Delta t}\mathbb{I} < 2(s\mathbb{M} + \frac{2}{\Delta t}\mathbb{I}) = 2\mathbb{Q}, \tag{66}$$

As to the last implication, it is not difficult to see the estimates (42) and (43) remain valid. Therefore we conclude from (44) that

$$\mu_{\max} \leq \max_{0<\lambda<2} 2\Big(1 - \big(1-\lambda\big)^2\Big) = 2. \tag{44'}$$

It follows that (38) remains valid under the assumption $s\mathbb{M} \geq \dfrac{1}{2}\mathbb{A}$.

As to the counter part of (39), we note that $s\mathbb{M} \geq \dfrac{1}{2}\mathbb{A}$ leads to

$$\frac{4}{\Delta t}\mathbb{I} + 4s\mathbb{M} - 2\mathbb{A} \geq \frac{4}{\Delta t}\mathbb{I}. \tag{46'}$$

It follows from (38), (49), (45) and (46') that

$$\frac{1}{\Delta t}\mathbb{I} + \frac{\mathbb{A}}{2} \leq \mathbb{S}_2 \leq \frac{\Delta t}{4}\mathbb{Q}^2 = \frac{1}{2}\widetilde{\mathbb{I}}\mathbb{Q} = \frac{1}{\Delta t}\mathbb{I} + s\mathbb{M} + \frac{\Delta t}{4}(s\mathbb{M})^2. \tag{39'}$$

As a result, we have

$$\begin{aligned}
\|\mathbb{G}\boldsymbol{f}^n\|^2 =& \|2\mathbb{S}_2\mathbb{Q}^{-1}\mathbb{T}\mathbb{Q}^{-1}\boldsymbol{f}^n\|^2 = \|2\mathbb{S}_2\mathbb{Q}^{-1}\Big((s\mathbb{M} - \frac{\mathbb{A}}{2}) - \frac{\mathbb{A}}{2}\Big)\mathbb{Q}^{-1}\boldsymbol{f}^n\|^2 \\
\leq& 2\|2\mathbb{S}_2\mathbb{Q}^{-1}\big(s\mathbb{M} - \frac{\mathbb{A}}{2}\big)\mathbb{Q}^{-1}\boldsymbol{f}^n\|^2 + 2\|2\mathbb{S}_2\mathbb{Q}^{-1}\frac{\mathbb{A}}{2}\mathbb{Q}^{-1}\boldsymbol{f}^n\|^2 \\
\leq& 4\|2\mathbb{S}_2\mathbb{Q}^{-1}(s\mathbb{M})\mathbb{Q}^{-1}\boldsymbol{f}^n\|^2 \leq 4\|\widetilde{\mathbb{I}}(s\mathbb{M}\mathbb{Q}^{-1})\boldsymbol{f}^n\|^2 \leq 4\|\widetilde{\mathbb{I}}\boldsymbol{f}^n\|^2,
\end{aligned} \tag{55'}$$

and

$$\|\mathbb{H}\boldsymbol{f}^{n+\frac{1}{2}}\|^2 = \|2\mathbb{S}_2\mathbb{Q}^{-1}\boldsymbol{f}^{n+\frac{1}{2}}\|^2 \leq \|\widetilde{\mathbb{I}}\boldsymbol{f}^{n+\frac{1}{2}}\|^2. \tag{56'}$$

Accordingly, the estimates (47) now becomes

$$\|\boldsymbol{u}^N\|_{\mathbb{A}}^2 \le \|\boldsymbol{u}^0\|_{\mathbb{A}}^2 + \Delta t \sum_{n=0}^{N-1} \left(4\|\widetilde{\mathbb{I}}\boldsymbol{f}^n\|^2 + \|\widetilde{\mathbb{I}}\boldsymbol{f}^{n+\frac{1}{2}}\|^2\right). \tag{47'}$$

Anormbound'

Similarly, the revised L^2 estimate reads:

$$\|\boldsymbol{u}^N\|^2 \le \|\widetilde{\mathbb{I}}\boldsymbol{u}^0\|^2 + \ \Delta t \sum_{n=0}^{N-1} \left(4\|\widetilde{\mathbb{I}}\boldsymbol{f}^n\|_{\mathbb{A}^{-1}}^2 + \|\widetilde{\mathbb{I}}\boldsymbol{f}^{n+\frac{1}{2}}\|_{\mathbb{A}^{-1}}^2\right) \tag{48'}$$

l2bound'

under the relaxed assumption $s\mathbb{M} \ge \frac{1}{2}\mathbb{A}$.

We note that (47) and (48) are direct analogue of the energy estimates for the PDE

$$\int_\Omega u\mathcal{L}(u)\ dx\ \Big|_0^T \le \frac{1}{2}\int_0^T \int_\Omega f^2 dxdt, \tag{67}$$

evo31

and

$$\int_\Omega u^2\ dx\ \Big|_0^T \le \frac{1}{2}\int_0^T \int_\Omega f\mathcal{L}^{-1}(f) dxdt. \tag{68}$$

evo32

On the other hand, the right hand sides of (47') and (48') involve higher order spatial derivatives and make subsequent estimates more complicated. For simplicity of presentation, we shall only focus on the case $s\mathbb{M} \ge \mathbb{A}$ in the rest of this paper. Nevertheless, numerical evidence confirms that $s\mathbb{M} \ge \frac{1}{2}\mathbb{A}$ is sufficient (and necessary) for unconditional stability and 2nd order in time convergence for both linear and the class of semilinear problems considered in section 3.

err_lin

2.3 Error Estimate for Linear Problems

Let u_e be the exact solution of the linear parabolic equation

$$\begin{aligned} &\partial_t u_e + \mathcal{L}(u_e) = f \quad \text{in } \Omega, \quad 0 \le t \le T, \\ &u_e(x,0) = u^0(x), \quad u_e \equiv 0 \quad \text{on } \partial\Omega. \end{aligned} \tag{69}$$

evo4

We first derive the error equation satisfied by $\boldsymbol{e} = \boldsymbol{u} - \boldsymbol{u}_e$. Firstly, we have

$$\partial_t u_e + \mathcal{L}_h(u_e) = f + \tau_{(x)}, \tag{70}$$

evo5

where $\tau_{(x)} = \mathcal{L}_h(u_e) - \mathcal{L}(u_e)$ is the spatial local truncation error. Rewriting (70) in matrix vector notation, we get

$$\partial_t \boldsymbol{u}_e + \mathbb{A}\boldsymbol{u}_e = \boldsymbol{f} + \boldsymbol{\tau}_{(x)}, \tag{71}$$

evo6

and

$$\left\{\begin{aligned} &\frac{\boldsymbol{u}_e^{n+\frac{1}{2}} - \boldsymbol{u}_e^n}{\frac{\Delta t}{2}} + s\mathbb{M}\boldsymbol{u}_e^{n+\frac{1}{2}} = (s\mathbb{M} - \mathbb{A})\boldsymbol{u}_e^n + \boldsymbol{f}^n + \boldsymbol{\tau}_{(x)}^n + \boldsymbol{\tau}_{(t)}^n \\ &\frac{\boldsymbol{u}_e^{n+1} - \boldsymbol{u}_e^n}{\Delta t} + s\mathbb{M}\frac{\boldsymbol{u}_e^{n+1} + \boldsymbol{u}_e^n}{2} = (s\mathbb{M} - \mathbb{A})\boldsymbol{u}_e^{n+\frac{1}{2}} + \boldsymbol{f}^{n+\frac{1}{2}} + \boldsymbol{\tau}_{(x)}^{n+\frac{1}{2}} + \boldsymbol{\tau}_{(t)}^{n+\frac{1}{2}}, \end{aligned}\right. \tag{72}$$

eest1

where the temporal local truncation errors are given by

$$\boldsymbol{\tau}_{(t)}^{n} = \frac{\boldsymbol{u}_e^{n+\frac{1}{2}} - \boldsymbol{u}_e^{n}}{\frac{\Delta t}{2}} - \partial_t \boldsymbol{u}_e^{n} + s\mathbb{M}(\boldsymbol{u}_e^{n+\frac{1}{2}} - \boldsymbol{u}_e^{n}), \tag{73}$$

and

$$\boldsymbol{\tau}_{(t)}^{n+\frac{1}{2}} = \frac{\boldsymbol{u}_e^{n+1} - \boldsymbol{u}_e^{n}}{\Delta t} - \partial_t \boldsymbol{u}_e^{n+\frac{1}{2}} + \frac{1}{2}s\mathbb{M}(\boldsymbol{u}_e^{n+1} - 2\boldsymbol{u}_e^{n+\frac{1}{2}} + \boldsymbol{u}_e^{n}). \tag{74}$$

After applying the Taylor formula

$$\begin{aligned} f(b) =& f(a) + f'(a)(b-a) + \int_a^b f''(a)(b-t)\ dt \\ =& f(a) + f'(a)(b-a) + \frac{1}{2}f''(a)(b-a)^2 + \frac{1}{2}\int_a^b f'''(a)(b-t)^2\ dt, \end{aligned} \tag{75}$$

on the intervals $(a,b) = (t^{n+\frac{1}{2}}, t^n)$ and $t^{n+\frac{1}{2}}, t^{n+1})$, it is easy to see that

$$\boldsymbol{\tau}_{(t)}^{n} = \frac{2}{\Delta t}\int_{t^n}^{t^{n+\frac{1}{2}}} \partial_t^2 \boldsymbol{u}_e(t)(t^{n+\frac{1}{2}} - t)\ dt + s\mathbb{M}\int_{t^n}^{t^{n+\frac{1}{2}}} \partial_t \boldsymbol{u}_e(t)\ dt, \tag{76}$$

ttn

$$\boldsymbol{\tau}_{(t)}^{n+\frac{1}{2}} = \frac{1}{2\Delta t}\int_{t^n}^{t^{n+1}} \partial_t^3 \boldsymbol{u}_e(t)\Big(\frac{\Delta t}{2} - |t - t^{n+\frac{1}{2}}|\Big)^2\ dt + \frac{1}{2}s\mathbb{M}\int_{t^n}^{t^{n+1}} \partial_t^2 \boldsymbol{u}_e(t)\Big(\frac{\Delta t}{2} - |t - t^{n+\frac{1}{2}}|\Big)\ dt, \tag{77}$$

ttnh

and conclude that $\boldsymbol{\tau}_{(t)}^{n} = O(\Delta t)$, $\boldsymbol{\tau}_{(t)}^{n+\frac{1}{2}} = O(\Delta t^2)$.

From (26) and (72), we can derive the equation for the error $\boldsymbol{e} = \boldsymbol{u} - \boldsymbol{u}_e$:

$$\left\{ \begin{aligned} &\frac{\boldsymbol{e}^{n+\frac{1}{2}} - \boldsymbol{e}^{n}}{\frac{\Delta t}{2}} + s\mathbb{M}\boldsymbol{e}^{n+\frac{1}{2}} = (s\mathbb{M} - \mathbb{A})\boldsymbol{e}^{n} - \boldsymbol{\tau}_{(x)}^{n} - \boldsymbol{\tau}_{(t)}^{n} \\ &\frac{\boldsymbol{e}^{n+1} - \boldsymbol{e}^{n}}{\Delta t} + s\mathbb{M}\frac{\boldsymbol{e}^{n+1} + \boldsymbol{e}^{n}}{2} = (s\mathbb{M} - \mathbb{A})\boldsymbol{e}^{n+\frac{1}{2}} - \boldsymbol{\tau}_{(x)}^{n+\frac{1}{2}} - \boldsymbol{\tau}_{(t)}^{n+\frac{1}{2}}. \end{aligned} \right. \tag{78}$$

eest2

Since (78) is structurally identical to (26), we can directly apply the estimates (47) together with (37) to get

$$\begin{aligned} \|\boldsymbol{e}^N\|_{\mathbb{A}}^2 \leq& \|\boldsymbol{e}^0\|_{\mathbb{A}}^2 + \Delta t \sum_{n=0}^{N-1} \Big(\big\|\mathbb{G}(\boldsymbol{\tau}_{(x)}^{n} + \boldsymbol{\tau}_{(t)}^{n})\big\|^2 + \big\|\mathbb{H}(\boldsymbol{\tau}_{(x)}^{n+\frac{1}{2}} + \boldsymbol{\tau}_{(t)}^{n+\frac{1}{2}})\big\|^2 \Big) \\ \leq& \|\boldsymbol{e}^0\|_{\mathbb{A}}^2 + \Delta t \sum_{n=0}^{N-1} 2\Big(\big\|\frac{s}{2}\mathbb{M}(\Delta t \boldsymbol{\tau}_{(t)}^{n})\big\|^2 + \|\boldsymbol{\tau}_{(x)}^{n}\|^2 + \|\boldsymbol{\tau}_{(x)}^{n+\frac{1}{2}}\|^2 + \|\boldsymbol{\tau}_{(t)}^{n+\frac{1}{2}}\|^2 \Big). \end{aligned} \tag{79}$$

eA

Here we have used (63), (55) and (56) in the last inequality. Similarly, from (48), (64), (60) and (59), we also have

$$\|\boldsymbol{e}^N\|^2 \leq \|\boldsymbol{e}^0\|_{\widetilde{\mathbb{I}}}^2 + \Delta t \sum_{n=0}^{N-1} 2\Big(\big\|\frac{s}{2}\mathbb{M}(\Delta t \boldsymbol{\tau}_{(t)}^{n})\big\|_{\mathbb{A}^{-1}}^2 + \|\boldsymbol{\tau}_{(x)}^{n}\|_{\mathbb{A}^{-1}}^2 + \|\boldsymbol{\tau}_{(x)}^{n+\frac{1}{2}}\|_{\mathbb{A}^{-1}}^2 + \|\boldsymbol{\tau}_{(t)}^{n+\frac{1}{2}}\|_{\mathbb{A}^{-1}}^2 \Big). \tag{80}$$

el2

In (79) and (80), The first order local truncation error $\boldsymbol{\tau}^n_{(t)}$ is compensated by the extra Δt factor. More precisely, from (76), we have

$$
\begin{aligned}
\Big\|\frac{s}{2}\mathbb{M}(\Delta t\boldsymbol{\tau}^n_{(t)})\Big\|^2 =&\Big\|s\mathbb{M}\int_{t^n}^{t^{n+\frac12}}\partial_t^2\boldsymbol{u}_e(t)(t^{n+\frac12}-t)\ dt+\frac{\Delta t}{2}(s\mathbb{M})^2\int_{t^n}^{t^{n+\frac12}}\partial_t\boldsymbol{u}_e(t)\ dt\Big\|^2\\
\le&2\Big\|\int_{t^n}^{t^{n+\frac12}}s\mathbb{M}\partial_t^2\boldsymbol{u}_e(t)(t^{n+\frac12}-t)\ dt\Big\|^2+\frac{(\Delta t)^2}{2}\Big\|\int_{t^n}^{t^{n+\frac12}}(s\mathbb{M})^2\partial_t\boldsymbol{u}_e(t)\ dt\Big\|^2 \qquad (81)\\
\le&\frac{(\Delta t)^3}{12}\int_{t^n}^{t^{n+\frac12}}\|s\mathbb{M}\partial_t^2\boldsymbol{u}_e(t)\|^2dt+\frac{(\Delta t)^3}{4}\int_{t^n}^{t^{n+\frac12}}\|(s\mathbb{M})^2\partial_t\boldsymbol{u}_e(t)\|^2dt.
\end{aligned}
$$

Similarly,

$$
\|\boldsymbol{\tau}^{n+\frac12}_{(t)}\|^2\le\frac{(\Delta t)^3}{320}\int_{t^n}^{t^{n+1}}\|\partial_t^3\boldsymbol{u}_e(t)\|^2dt+\frac{(\Delta t)^3}{48}\int_{t^n}^{t^{n+1}}\|s\mathbb{M}\partial_t^2\boldsymbol{u}_e(t)\|^2dt. \qquad (82)
$$

The estimates for $\|\boldsymbol{\tau}^n_{(t)}\|^2_{\mathbb{A}^{-1}}$ and $\|\boldsymbol{\tau}^{n+\frac12}_{(t)}\|^2_{\mathbb{A}^{-1}}$ are similar. We can now summarize these estimates as the following

Theorem 3. *Let u_e be the exact solution of the PDE (69). Denote by $\boldsymbol{u}_e$ the grid values of u_e and $\boldsymbol{u}^n$ the numerical solution of (26) at time $t^n=n\Delta t$, $0\le n\le N$, $\Delta t=\frac{T}{N}$. If $\mathbb{A}=\mathbb{A}^{\mathrm{T}}>0$, $\mathbb{M}=\mathbb{M}^{\mathrm{T}}>0$ and $s\mathbb{M}\ge\mathbb{A}$, then*

$$
\begin{aligned}
\|\boldsymbol{u}^N-\boldsymbol{u}_e(T)\|^2_{\mathbb{A}}\le&\|\boldsymbol{u}^0-\boldsymbol{u}_e(0)\|^2_{\mathbb{A}}+\Delta t\sum_{n=0}^{N-1}2\big(\|\boldsymbol{\tau}^n_{(x)}\|^2+\|\boldsymbol{\tau}^{n+\frac12}_{(x)}\|^2\big)\\
&+C(\Delta t)^4\int_0^T\big(\|\partial_t^3\boldsymbol{u}_e(t)\|^2+\|\mathbb{M}\partial_t^2\boldsymbol{u}_e(t)\|^2+\|\mathbb{M}^2\partial_t\boldsymbol{u}_e(t)\|^2\big)dt
\end{aligned}
\qquad (83)
$$

and

$$
\begin{aligned}
\|\boldsymbol{u}^N-\boldsymbol{u}_e(T)\|^2\le&\|\boldsymbol{u}^0-\boldsymbol{u}_e(0)\|^2_{\tilde{\mathbb{I}}}+\Delta t\sum_{n=0}^{N-1}2\big(\|\boldsymbol{\tau}^n_{(x)}\|^2_{\mathbb{A}^{-1}}+\|\boldsymbol{\tau}^{n+\frac12}_{(x)}\|^2_{\mathbb{A}^{-1}}\big)\\
&+C(\Delta t)^4\int_0^T\big(\|\partial_t^3\boldsymbol{u}_e(t)\|^2_{\mathbb{A}^{-1}}+\|\mathbb{M}\partial_t^2\boldsymbol{u}_e(t)\|^2_{\mathbb{A}^{-1}}+\|\mathbb{M}^2\partial_t\boldsymbol{u}_e(t)\|^2_{\mathbb{A}^{-1}}\big)dt
\end{aligned}
\qquad (84)
$$

for some constant C.

The estimates in Theorem 3 are generic. They can be characterized more specifically in terms of u_e when $\mathbb{A}$ and $\mathbb{M}$ are explicitly given.
Example: Let $\mathbb{A}$ be the matrix representation of $\mathcal{L}_h=-\nabla_h\cdot(\kappa(\boldsymbol{x})\nabla_h)$, the second order central difference discretization of the elliptic operator $\mathcal{L}=-\nabla(\kappa(\boldsymbol{x})\nabla)$ with $\kappa(\boldsymbol{x})$ smooth and positive, and zero Dirichlet boundary condition. We take $\mathcal{L}_{0,h}=-\Delta_h$ to be the discrete Laplacian and denote by $\mathbb{M}$ its matrix representation. It follows that

$$
|\tau_{(x)}(\boldsymbol{x},t)|\le C_1(\Delta x)^2\|u_e\|_{C^4_{\boldsymbol{x}}(\Omega)}(t), \qquad (85)
$$

$$|\mathcal{L}_{0,h}\partial_t^j u_e(\boldsymbol{x},t)| \le C_1 \|\partial_t^j u_e\|_{C^2_{\boldsymbol{x}}(\Omega)}(t). \tag{86}$$

Denote by $\boldsymbol{e}^n = \boldsymbol{u}^n - \boldsymbol{u}_e(t^n)$ and rewrite (83) as

$$\begin{aligned}\|\boldsymbol{e}^N\|_{\mathbb{A}}^2 \le & \|\boldsymbol{e}^0\|_{\mathbb{A}}^2 + C_2|\Omega|(\Delta x)^4 \Delta t \sum_{n=0}^{N-1}\Big(\|u_e\|^2_{C^4_{\boldsymbol{x}}(\Omega)}(t^n) + \|u_e\|^2_{C^4_{\boldsymbol{x}}(\Omega)}(t^{n+\frac{1}{2}})\Big) \\ & + C_2|\Omega|(\Delta t)^4 \int_0^T \big(\|u_e\|^2_{C^3_t C^0_{\boldsymbol{x}}(\Omega)} + \|u_e\|^2_{C^2_t C^2_{\boldsymbol{x}}(\Omega)} + \|u_e\|^2_{C^1_t C^4_{\boldsymbol{x}}(\Omega)}\big)(t)\ dt.\end{aligned} \tag{87}$$

The estimate (87) together with the Poincaré inequality

$$\|\boldsymbol{e}\|^2 \le C_{\mathrm{P}} \|\boldsymbol{e}\|^2_{\mathbb{A}}, \quad \big(C_{\mathrm{P}} \equiv \big(\lambda_{\min}(\mathbb{A})\big)^{-1} = O(1)\big) \tag{88}$$

results in optimal second order convergence rate in discrete H^1 norm:

$$\|\boldsymbol{e}^N\|^2_{H^1_h} = \|\boldsymbol{e}^N\|^2 + \|\boldsymbol{e}^N\|^2_{\mathbb{M}} \le \|\boldsymbol{e}^0\|^2_{H^1_h} + C\big((\Delta x)^2 + (\Delta t)^2\big)^2,$$

where the constant C depends on $\kappa_{\min} > 0$ and the norms and constants in (87) and (88).

3 Application to Semilinear Problems

We now generalize the estimates in sections 2.2 and 2.3 to the semilinear equation

$$v_t + \mathcal{L}(v) = g(v, \ell(v), x, t), \tag{89}$$

with Dirichlet boundary condition, where $\mathcal{L}(\cdot)$ is linear self-adjoint and $\ell(\cdot)$ is a lower order differential operator satisfying

$$\langle\, \ell(v), \ell(v)\,\rangle \le C_\ell \langle\, v, \mathcal{L}(v)\,\rangle \tag{90}$$

together with the discrete analogue

$$\|\boldsymbol{\ell}_h(\boldsymbol{v})\|^2 \le C_\ell \|\boldsymbol{v}\|^2_{\mathbb{A}} \tag{91}$$

for some fixed constant C_ℓ. We further assume that g is Lifchitz continuous in the first two arguments:

$$|g(v_1, p_1, x, t) - g(v_2, p_2, x, t)|^2 \le L^2\big((v_1 - v_2)^2 + \|p_1 - p_2\|^2\big), \tag{92}$$

where L is a fixed constant.

A typical example satisfying these assumption is given by

$$\mathcal{L}(v) = -\nabla(\kappa(x)\nabla v) + c(x)v, \quad g(v, \ell(v), x, t) = -w(x,t)\cdot\nabla v - h(v) + g_0(x,t), \tag{93}$$

with $|h'| \le L$ and $\kappa > 0$, $c > 0$, w and g_0 sufficiently smooth.

Assuming $s\mathbb{M} \geq \mathbb{A}$, the 2nd order stabilized predictor-corrector scheme for semilinear equation (89) is similar to the linear one:

$$\begin{cases} \dfrac{\boldsymbol{v}^{n+\frac{1}{2}} - \boldsymbol{v}^n}{\frac{\Delta t}{2}} + s\mathbb{M}\boldsymbol{v}^{n+\frac{1}{2}} = (s\mathbb{M} - \mathbb{A})\boldsymbol{v}^n + \boldsymbol{g}^n \\ \dfrac{\boldsymbol{v}^{n+1} - \boldsymbol{v}^n}{\Delta t} + s\mathbb{M}\dfrac{\boldsymbol{v}^{n+1} + \boldsymbol{v}^n}{2} = (s\mathbb{M} - \mathbb{A})\boldsymbol{v}^{n+\frac{1}{2}} + \boldsymbol{g}^{n+\frac{1}{2}}, \end{cases} \tag{94}$$
Nl2

where

$$\boldsymbol{g}^n \equiv \boldsymbol{g}(\boldsymbol{v}^n, \boldsymbol{\ell}_h(\boldsymbol{v}^n), t^n), \quad \boldsymbol{g}^{n+\frac{1}{2}} \equiv \boldsymbol{g}(\boldsymbol{v}^{n+\frac{1}{2}}, \boldsymbol{\ell}_h(\boldsymbol{v}^{n+\frac{1}{2}}), t^{n+\frac{1}{2}}). \tag{95}$$
Nl3

Following the derivation in Theorem 2, we get the same estimate as (52):

$$\frac{1}{\Delta t}\big(\|\boldsymbol{v}^{n+1}\|_{\mathbb{A}}^2 - \|\boldsymbol{v}^n\|_{\mathbb{A}}^2\big) \leq \|\mathbb{G}\boldsymbol{g}^n\|^2 + \|\mathbb{H}\boldsymbol{g}^{n+\frac{1}{2}}\|^2, \tag{96}$$
Anorm4

where $\mathbb{G}$ and $\mathbb{H}$ are defined in (37). To estimate the right hand side of (96), we decompose $\boldsymbol{g}^n$ as

$$\boldsymbol{g}^n = \delta_{\boldsymbol{0}}^1\boldsymbol{g}^n + \delta_{\boldsymbol{0}}^2\boldsymbol{g}^n + \boldsymbol{g}_{\boldsymbol{0}}^n, \tag{97}$$
df1

where

$$\begin{aligned} \delta_{\boldsymbol{0}}^1\boldsymbol{g}^n &\equiv \boldsymbol{g}(\boldsymbol{v}^n, \boldsymbol{\ell}_h(\boldsymbol{v}^n), t^n) - \boldsymbol{g}(\boldsymbol{0}, \boldsymbol{\ell}_h(\boldsymbol{v}^n), t^n), \\ \delta_{\boldsymbol{0}}^2\boldsymbol{g}^n &\equiv \boldsymbol{g}(\boldsymbol{0}, \boldsymbol{\ell}_h(\boldsymbol{v}^n), t^n) - \boldsymbol{g}(\boldsymbol{0}, \boldsymbol{0}, t^n), \\ \boldsymbol{g}_{\boldsymbol{0}}^n &\equiv \boldsymbol{g}(\boldsymbol{0}, \boldsymbol{0}, t^n). \end{aligned} \tag{98}$$
df2

From (92) and (91), we can estimate the first term in right hand side of (96):

$$\begin{aligned} \|\mathbb{G}\boldsymbol{g}^n\|^2 \leq \|\boldsymbol{g}^n\|^2 &\leq 3\Big(\|\delta_{\boldsymbol{0}}^1\boldsymbol{g}^n\|^2 + \|\delta_{\boldsymbol{0}}^2\boldsymbol{g}^n\|^2 + \|\boldsymbol{g}_{\boldsymbol{0}}^n\|^2\Big) \\ &\leq 3\Big(L^2\|\boldsymbol{v}^n\|^2 + L^2\|\boldsymbol{\ell}_h(\boldsymbol{v}^n)\|^2 + \|\boldsymbol{g}_{\boldsymbol{0}}^n\|^2\Big) \\ &\leq 3\Big(L^2(C_{\mathrm{P}} + C_\ell)\|\boldsymbol{v}^n\|_{\mathbb{A}}^2 + \|\boldsymbol{g}_{\boldsymbol{0}}^n\|^2\Big). \end{aligned} \tag{99}$$
Lif2n

Here we have used the discrete Poincaré inequality (88) in the last inequality. Similarly,

$$\|\mathbb{H}\boldsymbol{g}^{n+\frac{1}{2}}\|^2 \leq \|\boldsymbol{g}^{n+\frac{1}{2}}\|^2 \leq 3\Big(L^2(C_{\mathrm{P}} + C_\ell)\|\boldsymbol{v}^{n+\frac{1}{2}}\|_{\mathbb{A}}^2 + \|\boldsymbol{g}_{\boldsymbol{0}}^{n+\frac{1}{2}}\|^2\Big). \tag{100}$$
Lif2nh

Upon comparing the first equation of (94) with the first order scheme (14), we immediately obtain the following estimate from (23):

$$\|\boldsymbol{v}^{n+\frac{1}{2}}\|_{\mathbb{A}}^2 \leq \|\boldsymbol{v}^n\|_{\mathbb{A}}^2 + \frac{\Delta t}{4}\|\boldsymbol{g}^n\|^2. \tag{101}$$
Anorm5

From (96), (99), (100) and (101) that

$$\frac{1}{\Delta t}\big(\|\boldsymbol{v}^{n+1}\|_{\mathbb{A}}^2 - \|\boldsymbol{v}^n\|_{\mathbb{A}}^2\big) \leq 6L^2(C_{\mathrm{P}} + C_\ell)\Big(\|\boldsymbol{v}^n\|_{\mathbb{A}}^2 + \|\boldsymbol{g}_{\boldsymbol{0}}^n\|^2 + \|\boldsymbol{g}_{\boldsymbol{0}}^{n+\frac{1}{2}}\|^2\Big). \tag{102}$$
Anorm6

After applying the discrete Gronwall's inequality to (102), we obtain the following thm4 **Theorem 4.** *Let $\boldsymbol{v}^n$ be the numerical solution of the stabilized predictor-corrector scheme (94) at time $t^n = n\Delta t$, $0 \le n \le N$, $\Delta t = \frac{T}{N}$. If $\mathbb{A} = \mathbb{A}^{\mathrm{T}} > 0$ satisfies the discrete Poincaré inequality (88), $\mathbb{M} = \mathbb{M}^{\mathrm{T}} > 0$ and $s\mathbb{M} \ge \mathbb{A}$, then under the conditions (92) and (91), the scheme (94) is unconditionally stable and satisfies*

$$\|\boldsymbol{v}^N\|_{\mathbb{A}}^2 \le e^{CT}\Big(\|\boldsymbol{v}^0\|_{\mathbb{A}}^2 + C\Delta t \sum_{n=0}^{N-1}\big(\|\boldsymbol{g}_{\boldsymbol{0}}^n\|^2 + \|\boldsymbol{g}_{\boldsymbol{0}}^{n+\frac{1}{2}}\|^2\big)\Big) \tag{103}$$ Anorm7

for some constant C.

An a priori bound for $\|\boldsymbol{v}^N\|^2$ can be obtained similarly from (58) following the same procedure as above, or by applying the discrete Poincaré inequality (88) to (103) directly.

We now proceed to give an error estimate for (94). Let v_e be the solution of the semilinear equation (89),

$$\partial_t v_e + \mathcal{L}(v_e) = g(v_e, \ell(v_e), x, t). \tag{104}$$ nevo2

After spatial discretization, we have

$$\partial_t v_e + \mathcal{L}_h(v_e) = g(v_e, \ell_h(v_e), x, t) + \tau_{(x)}, \tag{105}$$ nevo3

where

$$\tau_{(x)} = \mathcal{L}_h(v_e) - \mathcal{L}(v_e) - \big(g(v_e, \ell_h(v_e), x, t) - g(v_e, \ell(v_e), x, t)\big) \tag{106}$$ nevo31

is the spatial local truncation error. Next, we recast (105) in matrix vector notation:

$$\partial_t \boldsymbol{v}_e + \mathbb{A}\boldsymbol{v}_e = \boldsymbol{g}_e + \boldsymbol{\tau}_{(x)}, \tag{107}$$ nevo4

which leads to

$$\begin{cases} \dfrac{\boldsymbol{v}_e^{n+\frac{1}{2}} - \boldsymbol{v}_e^n}{\frac{\Delta t}{2}} + s\mathbb{M}\boldsymbol{v}_e^{n+\frac{1}{2}} = (s\mathbb{M} - \mathbb{A})\boldsymbol{v}_e^n + \boldsymbol{g}_e^n + \boldsymbol{\tau}_{(x)}^n + \boldsymbol{\tau}_{(t)}^n \\ \dfrac{\boldsymbol{v}_e^{n+1} - \boldsymbol{v}_e^n}{\Delta t} + s\mathbb{M}\dfrac{\boldsymbol{v}_e^{n+1} + \boldsymbol{v}_e^n}{2} = (s\mathbb{M} - \mathbb{A})\boldsymbol{v}_e^{n+\frac{1}{2}} + \boldsymbol{g}_e^{n+\frac{1}{2}} + \boldsymbol{\tau}_{(x)}^{n+\frac{1}{2}} + \boldsymbol{\tau}_{(t)}^{n+\frac{1}{2}}, \end{cases} \tag{108}$$ nleest1

where

$$\boldsymbol{g}_e^n \equiv \boldsymbol{g}(\boldsymbol{v}_e^n, \boldsymbol{\ell}_h(\boldsymbol{v}_e^n), t^n), \quad \boldsymbol{g}_e^{n+\frac{1}{2}} \equiv \boldsymbol{g}(\boldsymbol{v}_e^{n+\frac{1}{2}}, \boldsymbol{\ell}_h(\boldsymbol{v}_e^{n+\frac{1}{2}}), t^{n+\frac{1}{2}}), \tag{109}$$

and the temporal local truncation errors are given by

$$\boldsymbol{\tau}_{(t)}^n = \frac{\boldsymbol{v}_e^{n+\frac{1}{2}} - \boldsymbol{v}_e^n}{\frac{\Delta t}{2}} - \partial_t \boldsymbol{v}_e^n + s\mathbb{M}(\boldsymbol{v}_e^{n+\frac{1}{2}} - \boldsymbol{v}_e^n), \tag{110}$$

and

$$\boldsymbol{\tau}_{(t)}^{n+\frac{1}{2}} = \frac{\boldsymbol{v}_e^{n+1} - \boldsymbol{v}_e^n}{\Delta t} - \partial_t \boldsymbol{v}_e^{n+\frac{1}{2}} + \frac{1}{2}s\mathbb{M}(\boldsymbol{v}_e^{n+1} - 2\boldsymbol{v}_e^{n+\frac{1}{2}} + \boldsymbol{v}_e^n). \tag{111}$$

Subtract (94) from (108), we obtain the equation for the error $\boldsymbol{e}=\boldsymbol{v}-\boldsymbol{v}_e$:

$$\left\{\begin{aligned} &\frac{\boldsymbol{e}^{n+\frac{1}{2}}-\boldsymbol{e}^n}{\frac{\Delta t}{2}}+s\mathbb{M}\boldsymbol{e}^{n+\frac{1}{2}}=(s\mathbb{M}-\mathbb{A})\boldsymbol{e}^n+\delta\boldsymbol{g}^n\\ &\frac{\boldsymbol{e}^{n+1}-\boldsymbol{e}^n}{\Delta t}+s\mathbb{M}\frac{\boldsymbol{e}^{n+1}+\boldsymbol{e}^n}{2}=(s\mathbb{M}-\mathbb{A})\boldsymbol{e}^{n+\frac{1}{2}}+\delta\boldsymbol{g}^{n+\frac{1}{2}}, \end{aligned}\right. \tag{112}$$ `nleest2`

where

$$\begin{aligned} \delta\boldsymbol{g}^n=&\big(\boldsymbol{g}^n-\boldsymbol{g}_e^n\big)-\boldsymbol{\tau}_{(x)}^n-\boldsymbol{\tau}_{(t)}^n=\delta_e^1\boldsymbol{g}^n+\delta_e^2\boldsymbol{g}^n-\boldsymbol{\tau}_{(x)}^n-\boldsymbol{\tau}_{(t)}^n,\\ \delta\boldsymbol{g}^{n+\frac{1}{2}}=&\big(\boldsymbol{g}^{n+\frac{1}{2}}-\boldsymbol{g}_e^{n+\frac{1}{2}}\big)-\boldsymbol{\tau}_{(x)}^{n+\frac{1}{2}}-\boldsymbol{\tau}_{(t)}^{n+\frac{1}{2}}=\delta_e^1\boldsymbol{g}^{n+\frac{1}{2}}+\delta_e^2\boldsymbol{g}^{n+\frac{1}{2}}-\boldsymbol{\tau}_{(x)}^{n+\frac{1}{2}}-\boldsymbol{\tau}_{(t)}^{n+\frac{1}{2}}, \end{aligned} \tag{113}$$ `tmp0`

with

$$\begin{aligned} \delta_e^1\boldsymbol{g}^n\equiv&\boldsymbol{g}(\boldsymbol{v}^n,\boldsymbol{\ell}_h(\boldsymbol{v}^n),t^n)-\boldsymbol{g}(\boldsymbol{v}_e(t^n),\boldsymbol{\ell}_h(\boldsymbol{v}^n),t^n),\\ \delta_e^2\boldsymbol{g}^n\equiv&\boldsymbol{g}(\boldsymbol{v}_e(t^n),\boldsymbol{\ell}_h(\boldsymbol{v}^n),t^n)-\boldsymbol{g}(\boldsymbol{v}_e(t^n),\boldsymbol{\ell}_h(\boldsymbol{v}_e(t^n)),t^n), \end{aligned} \tag{114}$$

and similarly for $\delta_e^1\boldsymbol{g}^{n+\frac{1}{2}}$ and $\delta_e^2\boldsymbol{g}^{n+\frac{1}{2}}$.

`thm5` **Theorem 5.** *Let v_e be the exact solution of the semilinear equation (89), $\boldsymbol{v}_e$ the grid values of v_e and $\boldsymbol{v}^n$ the numerical solution of (94), (95) at time $t^n=n\Delta t$, $0\le n\le N$, $\Delta t=\frac{T}{N}$. Then under the same assumptions in Theorem 4, the error $\boldsymbol{v}-\boldsymbol{v}_e$ satisfies*

$$\begin{aligned} \|\boldsymbol{v}^N-\boldsymbol{v}_e(T)\|_{\mathbb{A}}^2\le& e^{CT}\Bigg(\|\boldsymbol{v}^0-\boldsymbol{v}_e(0)\|_{\mathbb{A}}^2+C\Big(\Delta t\sum_{n=0}^{N-1}\big(\|\boldsymbol{\tau}_{(x)}^n\|^2+\|\boldsymbol{\tau}_{(x)}^{n+\frac{1}{2}}\|^2\big)\\ &+(\Delta t)^4\int_0^T\big(\|\partial_t^3\boldsymbol{v}_e(t)\|^2+\|\mathbb{M}\partial_t^2\boldsymbol{v}_e(t)\|^2+\|\mathbb{M}^2\partial_t\boldsymbol{v}_e(t)\|^2\big)dt\Big)\Bigg) \end{aligned} \tag{115}$$ `nleest`

for some constant C.

Proof: We first derive the error equation. In view of the similarity between (94) and (112), we see from (96) that

$$\frac{1}{\Delta t}\big(\|\boldsymbol{e}^{n+1}\|_{\mathbb{A}}^2-\|\boldsymbol{e}^n\|_{\mathbb{A}}^2\big)\le\|\mathbb{G}\delta\boldsymbol{g}^n\|^2+\|\mathbb{H}\delta\boldsymbol{g}^{n+\frac{1}{2}}\|^2. \tag{116}$$ `nleest21`

We now proceed with the estimates on the right hand side of (116). From (99), (100), (92), (88) and (91), we have

$$\begin{aligned} \|\mathbb{G}\delta\boldsymbol{g}^n\|^2&\le4\Big(\|\mathbb{G}\delta_e^1\boldsymbol{g}^n\|^2+\|\mathbb{G}\delta_e^2\boldsymbol{g}^n\|^2+\|\mathbb{G}\boldsymbol{\tau}_{(x)}^n\|^2+\|\mathbb{G}\boldsymbol{\tau}_{(t)}^n\|^2\Big)\\ &\le4\Big(\|\delta_e^1\boldsymbol{g}^n\|^2+\|\delta_e^2\boldsymbol{g}^n\|^2+\|\boldsymbol{\tau}_{(x)}^n\|^2+\|\frac{s}{2}\mathbb{M}(\Delta t\boldsymbol{\tau}_{(t)}^n)\|^2\Big)\\ &\le4\Big(L^2(C_{\mathrm{P}}+C_\ell)\|\boldsymbol{e}^n\|_{\mathbb{A}}^2+\|\boldsymbol{\tau}_{(x)}^n\|^2+\|\frac{s}{2}\mathbb{M}(\Delta t\boldsymbol{\tau}_{(t)}^n)\|^2\Big), \end{aligned} \tag{117}$$ `nleest2n`

$$\|\mathbb{H}\delta \boldsymbol{g}^{n+\frac{1}{2}}\|^2 \leq \|\delta \boldsymbol{g}^{n+\frac{1}{2}}\|^2 \leq 4\Big(L^2(C_{\mathrm{P}}+C_\ell)\|\boldsymbol{e}^{n+\frac{1}{2}}\|_{\mathbb{A}}^2 + \|\boldsymbol{\tau}_{(x)}^{n+\frac{1}{2}}\|^2 + \|\boldsymbol{\tau}_{(t)}^{n+\frac{1}{2}}\|^2\Big). \tag{118}$$ nleest2nh

It remains to estimate $\|\boldsymbol{e}^{n+\frac{1}{2}}\|_{\mathbb{A}}^2$ in terms of $\|\boldsymbol{e}^n\|_{\mathbb{A}}^2$. To this end, we rewrite the first equation of (94) as

$$\big(\mathbb{I}+\frac{\Delta t}{2}s\mathbb{M}\big)\boldsymbol{e}^{n+\frac{1}{2}} = \big(\mathbb{I}+\frac{\Delta t}{2}(s\mathbb{M}-\mathbb{A})\big)\boldsymbol{e}^n + \frac{\Delta t}{2}\delta \boldsymbol{g}^n, \tag{119}$$ tmp1

or

$$\widetilde{\mathbb{I}}\,\boldsymbol{e}^{n+\frac{1}{2}} = \big(\widetilde{\mathbb{I}} - \frac{\Delta t}{2}\mathbb{A}\big)\boldsymbol{e}^n + \frac{\Delta t}{2}\delta \boldsymbol{g}^n, \tag{120}$$ tmp2

where $\widetilde{\mathbb{I}} = \mathbb{I} + \dfrac{\Delta t}{2}s\mathbb{M}$. It follows from (120) that

$$\|\boldsymbol{e}^{n+\frac{1}{2}}\|_{\mathbb{A}}^2 \leq 2\big\|\widetilde{\mathbb{I}}^{-1}\big(\widetilde{\mathbb{I}}-\frac{\Delta t}{2}\mathbb{A}\big)\boldsymbol{e}^n\big\|_{\mathbb{A}}^2 + 2\big\|\frac{\Delta t}{2}\widetilde{\mathbb{I}}^{-1}\delta \boldsymbol{g}^n\big\|_{\mathbb{A}}^2 \equiv (\mathrm{I}) + (\mathrm{II}). \tag{121}$$ nleest3

Since $s\mathbb{M} \geq \mathbb{A}$ by assumption, we have

$$\widetilde{\mathbb{I}}^{-1} \leq \big(\widetilde{\mathbb{I}} - \frac{\Delta t}{2}\mathbb{A}\big)^{-1} \leq \mathbb{I}, \qquad \widetilde{\mathbb{I}}^{-1} \leq \big(\frac{\Delta t}{2}\mathbb{A}\big)^{-1} \quad \Big(\implies \frac{\Delta t}{2}\mathbb{A}^{\frac{1}{2}}\widetilde{\mathbb{I}}^{-1} \leq \mathbb{A}^{-\frac{1}{2}}\Big). \tag{122}$$ tmp3

From (122), (113), (92) and (91),

$$(\mathrm{I}) = 2\big\|\mathbb{A}^{\frac{1}{2}}\widetilde{\mathbb{I}}^{-1}\big(\widetilde{\mathbb{I}}-\frac{\Delta t}{2}\mathbb{A}\big)\boldsymbol{e}^n\big\|^2 \leq 2\|\mathbb{A}^{\frac{1}{2}}\boldsymbol{e}^n\|^2 = 2\|\boldsymbol{e}^n\|_{\mathbb{A}}^2. \tag{123}$$ nleest4

Note that $\boldsymbol{\tau}_{(t)}^n$ is only $O(\Delta t)$ and needs to be treated separately:

$$(\mathrm{II}) \leq 4\big\|\frac{\Delta t}{2}\widetilde{\mathbb{I}}^{-1}\big(\delta_e^1\boldsymbol{g}^n + \delta_e^2\boldsymbol{g}^n - \boldsymbol{\tau}_{(x)}^n\big)\big\|_{\mathbb{A}}^2 + 4\big\|\frac{\Delta t}{2}\widetilde{\mathbb{I}}^{-1}\boldsymbol{\tau}_{(t)}^n\big\|_{\mathbb{A}}^2 \equiv (\mathrm{II}_1) + (\mathrm{II}_2), \tag{124}$$ nleest5

$$\begin{aligned}(\mathrm{II}_1) =& 4\big\|\frac{\Delta t}{2}\mathbb{A}^{\frac{1}{2}}\widetilde{\mathbb{I}}^{-1}\big(\delta_e^1\boldsymbol{g}^n + \delta_e^2\boldsymbol{g}^n - \boldsymbol{\tau}_{(x)}^n\big)\big\|_{\mathbb{A}}^2\\ \leq& 4\big\|\mathbb{A}^{-\frac{1}{2}}\big(\delta_e^1\boldsymbol{g}^n + \delta_e^2\boldsymbol{g}^n - \boldsymbol{\tau}_{(x)}^n\big)\big\|^2 \leq 4C_{\mathrm{P}}\big\|\delta_e^1\boldsymbol{g}^n + \delta_e^2\boldsymbol{g}^n - \boldsymbol{\tau}_{(x)}^n\big\|^2\\ \leq& 12C_{\mathrm{P}}\big(C_{\mathrm{P}}L^2\|\boldsymbol{e}^n\|_{\mathbb{A}}^2 + C_\ell L^2\|\boldsymbol{e}^n\|_{\mathbb{A}}^2 + \|\boldsymbol{\tau}_{(x)}^n\|^2\big),\end{aligned} \tag{125}$$ nleest6

$$(\mathrm{II}_2) = \|\mathbb{A}^{\frac{1}{2}}\widetilde{\mathbb{I}}^{-1}(\Delta t\boldsymbol{\tau}_{(t)}^n)\|^2 \leq \big\|\mathbb{A}^{\frac{1}{2}}(\Delta t\boldsymbol{\tau}_{(t)}^n)\big\|^2 \leq \|C_{\mathrm{P}}^{\frac{1}{2}}s\mathbb{M}(\Delta t\boldsymbol{\tau}_{(t)}^n)\|^2. \tag{126}$$ nleest7

We can further estimate the $\boldsymbol{\tau}_{(t)}$ terms as (see the derivation of (81) and (82))

$$\|\mathbb{M}(\Delta t\boldsymbol{\tau}_{(t)}^n)\|^2 \leq C(\Delta t)^3\int_{t^n}^{t^{n+\frac{1}{2}}}\Big(\|\mathbb{M}\partial_t^2\boldsymbol{v}_e(t)\|^2 + \|\mathbb{M}^2\partial_t\boldsymbol{v}_e(t)\|^2\Big)dt, \tag{127}$$ eA1n

$$\|\boldsymbol{\tau}_{(t)}^{n+\frac{1}{2}}\|^2 \leq C(\Delta t)^3\int_{t^n}^{t^{n+1}}\Big(\|\partial_t^3\boldsymbol{v}_e(t)\|^2 + \|\mathbb{M}\partial_t^2\boldsymbol{v}_e(t)\|^2\Big)dt. \tag{128}$$ eA2n

From (116), (117), (118), (121) and (123)- (128), we have

$$\frac{1}{\Delta t}\big(\|\boldsymbol{e}^{n+1}\|_{\mathbb{A}}^2 - \|\boldsymbol{e}^n\|_{\mathbb{A}}^2\big) \le C_{\mathrm{G}}\Big(\|\boldsymbol{e}^n\|_{\mathbb{A}}^2 + \|\boldsymbol{\tau}_{(x)}^n\|^2 + \|\boldsymbol{\tau}_{(x)}^{n+\frac{1}{2}}\|^2 + (\Delta t)^3\int_{t^n}^{t^{n+1}}\big(\|\partial_t^3\boldsymbol{v}_e(t)\|^2 + \|\mathbb{M}\partial_t^2\boldsymbol{v}_e(t)\|^2 + \|\mathbb{M}^2\partial_t\boldsymbol{v}_e(t)\|^2\big)dt\Big). \tag{129}$$

Therefore (115) follows from Gronwall's inequality.

4 Numerical Results

In this section, we conduct various numerical experiments to verify the stability and error estimates.

Example 1: In the first example, we consider the linear diffusion equation

$$u_t = (\kappa(x)u_x)_x + f, \qquad x \in (-1,1), \quad t \in [0,T], \tag{130}$$

with degenerate diffusion coefficient $\kappa(x) = (1-x^2)$. For this test problem, we take the exact solution to be

$$u_e(x,t) = e^{\sin(t)}\cos\big(\frac{\pi x}{2}\big) \tag{131}$$

and use u_e to generate the source term f, the initial value $u(x,0)$ and the boundary values $u(\pm 1, t)$. Here $\mathbb{A}$ is the matrix representation of $\mathcal{L}_h$:

$$\mathcal{L}_h(u)_i = -D_x\big(\kappa D_x u\big)_i = -\frac{1}{\Delta x}\left(\kappa_{i+\frac{1}{2}}\Big(\frac{u_{i+1}-u_i}{\Delta x}\Big) - \kappa_{i-\frac{1}{2}}\Big(\frac{u_i - u_{i-1}}{\Delta x}\Big)\right). \tag{132}$$

We can choose the stabilizing term to be $s\mathbb{M}$, where $s > 0$ is the stabilizing constant and $\mathbb{M}$ is the matrix representation of the standard discrete Laplician $\mathcal{L}_{h,0}$:

$$\mathcal{L}_{0,h}(u)_i = -D_x^2 u_i = -\frac{u_{i+1} - 2u_i + u_{i-1}}{(\Delta x)^2}. \tag{133}$$

From Remark 2 in Section 2.2, $s\mathbb{M} \ge \frac{1}{2}\mathbb{A}$ is sufficient for stability of (26) applied to linear problems. Since $\kappa(x) \le 1$, we have $\mathbb{M} \ge \mathbb{A}$. Therefore $s \ge 0.5$ will do. In Table 4, we present the supnorm errors and orders of convergence of the stabilized predictor-corrector scheme (SPC) for both $s = 1$ and $s = 0.5$ at time $T = 1$. In addition, we also include the result of Crank-Nicolson scheme (CN) for comparison.

As we can see from Table 4, the performances of CN and SPC are comparable, exhibiting near or full 2nd order accuracy. Also note that in this example, smaller (but large enough) stabilizing term ($s = 0.5$) results in better performance. This is consistent with similar observations reported in the literature.

Example 2:

Table 1: Errors and orders of convergence for Example 1 at $T = 1$ using Crank-Nicolson method (CN), Stabilized Predictor-Corrector method (26) (SPC) with $s = 1$ and $s = 0.5$.

	$\Delta t (= \Delta x)$	1/64	1/128	1/256	1/512
CN	L^∞ error	6.717e-4	1.659e-4	4.111e-5	1.020e-5
	order	–	2.02	2.01	2.01
SPC	L^∞ error	6.884e-4	2.038e-4	5.764e-5	1.580e-6
$s = 1$	order	–	1.76	1.82	1.87
SPC	L^∞ error	5.435e-4	1.276e-4	3.027e-5	7.249e-6
$s = 0.5$	order	–	2.09	2.08	2.06

In the second example, we consider the following 1D semilinear equation:

$$v_t + c_1 w(x) v_x = c_2(\kappa(x) v_x)_x - c_3 h(v) + g_0(x,t), \qquad x \in (-1,1), \quad t \in [0,T] \tag{134}$$

with various parameters c_1, c_2 and c_3. From the proof of Theorem 4, we know the constant C in (103) depends mostly on L and C_ℓ, while

$$\frac{L = O(\max c_3, c_1) \quad \text{and } C_\ell = O(c_1^2}{c_2|)} \tag{135}$$

from (91) and (92). In the following tests, we choose the time scale with $\max(\frac{c_1^2}{c_2}, c_1, c_3) = O(1)$ in order to resolve the time evolution properly.

Example 2a: $w(x) = \sin(x)$, $\kappa(x) = 1$, $h(v) = v^3 - v$.

Example 2b: $w(x) = \sin(\pi x)$, $\kappa(x) = 1 - x^2$, $h(v) = \sinh(v)$.

Although the nonlinear part $h(v) = (v^3 - v)$ does not satisfy the Lifchitz assumption (92), it is known that the generalized Allen-Cahn equation (without the source term g_0) satisfies the maximum principle:

$$|v(x,0)| \le 1 \implies |v(x,t)| \le 1 \text{ for any } t > 0. \tag{136}$$

Therefore the assumption (92) is practically satisfied with the effective Lifchitz constant

$$L = \max_{|v| \le 1} |h'(v)| = 2. \tag{137}$$

This is equivalent to replacing $h(v)$ by

$$\tilde{h}(v) = \begin{cases} 2(v-1), & v > 1 \\ (v^3 - v), & v \in [-1,1] \\ 2(v+1), & v < -1 \end{cases} \tag{138}$$

in the scheme without affecting the solution.

For this test problem, we take the exact solution to be

$$v_e(x,t) = \sin(t) \cdot (1 - x^2) \tag{139}$$

and use it to generate the initial value, boundary value and source term $g_0(x,t)$. Since the leading order linear part $\epsilon^2 \partial_x^2$ has constant coefficient, our analysis shows that we can take

$$\mathbb{A} = \text{ matrix representation of } - \epsilon^2 D_x^2. \tag{140}$$

That is, $\mathbb{A}$ need not include the transport term v_x. As a result, we can simply take $\mathbb{M} = \mathbb{A}$. In view of Remark 2, we test it with $s = 0.5001$ using the Stabilized Predictor-Corrector scheme (94), (95) for semilinear equations. The result of the convergence test with $\Delta x = \Delta t$, $T = 1$ and various ϵ^2 is listed in Table 2.

Table 2: Results of Example 2 at $T = 1$ using Stabilized Predictor-Corrector scheme (94), (95) with $\mathbb{M} = \mathbb{A}$, $s = 0.5001$ and various ϵ.

	$\Delta t (= \Delta x)$	1/32	1/64	1/128	1/256	1/512	1/1024
$\epsilon^2 = 1$	L^∞ error	3.622e-4	8.931e-5	2.216e-5	5.517e-6	1.376e-6	3.437e-7
	order	–	2.02	2.01	2.01	2.00	2.00
$\epsilon^2 = 0.1$	L^∞ error	3.432e-4	8.619e-5	2.157e-5	5.393e-6	1.349e-6	3.372e-7
	order	–	1.99	2.00	2.00	2.00	2.00
$\epsilon^2 = 0.01$	L^∞ error	8.642e-4	1.840e-4	3.779e-5	9.417e-6	2.356e-6	5.893e-7
	order	–	2.23	2.28	2.00	2.00	2.00

5 Conclusion and discussion

In this paper, we give a rigorous stability and convergence proof of the Stabilized Predictor-Corrector method applied to linear diffusion equations in Section 2 and a class of semilinear diffusion equations described in Section (3). For simplicity of representation, these results are proved under the assumption $s\mathbb{M} \geq \mathbb{A}$ where $s\mathbb{M}$ is the artificially splitted stabilizing term and $\mathbb{A}$ is the matrix representation of the leading order linear part of the diffusion operator. The main advantage of the stabilization is the freedom to choose the splitting operator $s\mathbb{M}$ so that the resulting linear system is easier to solve, for example, using an FFT package on simple domains. In addition to the main Theorems, our analysis the linear problems shows that $s\mathbb{M} \geq \frac{1}{2}\mathbb{A}$ is sufficient.

Limitations of our approach include:

1. The analysis relies on the assumption $\mathbb{A} = \mathbb{A}^{\mathrm{T}} > 0$, which is not satisfied when the underlying boundary condition is Neumann or periodic.

2. The class of semilinear diffusion equations considered in Section 3 does not include some well known equations such as the isotropic Cahn-Hillard equations or more complicated gradient flows.

3. The discrete energy estimate is based on the new symmetrized formulation (35). It is not clear whether similar formulation exists for other 2nd order or higher order stabilized schemes of predictor-corrector type.

These problems are currently under consideration and will be reported elsewhere, if successful.

References

BaGuGu11 [1] S. Badia, F. Guillén-González, and J. V. Gutiérrez-Santacreu, Finite element approximation of nematic liquid crystal flows using a saddle-point structure, J. Comput. Phys. 230.4 (2011): 1686-1706.

BaLoWaWi13 [2] A. Baskaran, J. S. Lowengrub, C. Wang, and S. M. Wise, Convergence analysis of a second order convex splitting scheme for the modified phase field crystal equation, SIAM J. Numer. Anal. 51 (2013): 2851-2873.

CeHo01 [3] H. D. Ceniceros, and T. Y. Hou, An efficient dynamically adaptive mesh for potentially singular solutions, J. Comput. Phys. 172.2 (2001): 609-639.

ChShYa19 [4] Q. Cheng, J. Shen, and X. Yang, Highly efficient and accurate numerical schemes for the epitaxial thin film growth models by using the SAV approach, J. Sci. Comput. 78.3 (2019): 1467-1487.

ChYa19 [5] C. Chen, and X. Yang, Fast, provably unconditionally energy stable, and second-order accurate algorithms for the anisotropic Cahn-Hilliard model, Comput. Methods Appl. Mech. Eng. 351 (2019): 35-59.

Du [6] T. Dupont, Private communication with authors of [3]

ElSt93 [7] C. M. Elliott and A. M. Stuart, The global dynamics of discrete semilinear parabolic equations, SIAM J. Numer. Anal. 30.6 (1993): 1622-1663.

Ey98 [8] D. J. Eyre, Unconditionally gradient stable time marching the Cahn-Hilliard equation, in: Computational and Mathematical Models of Microstructural Evolution, San Francisco, CA, 1998, in Mater. Res. Soc. Symp. Proc., vol.529, MRS, Warrendale, PA, 1998, pp. 39-46

FeTaYa13 [9] X. Feng, T. Tang, and J. Yang, Stabilized Crank-Nicolson-Adams-Bashforth schemes for phase field models East Asian J. Appl. Math. 3.1 (2013): 59-80.

GuLoWaWi14 [10] Z. Guan, J. S. Lowengrub, C. Wang, and M. S. Wise, Second order convex splitting schemes for periodic nonlocal Cahn-Hilliard and Allen-Cahn equations, J. Comput. Phys. 277 (2014): 48-71.

GuTi13 [11] F. Guillén-González, and G. Tierra, On linear schemes for a Cahn-Hilliard diffuse interface model, J. Comput. Phys. 234 (2013): 140-171.

HaHuWa21 [12] Y. Hao, Q. Huang, and C. Wang, A third order BDF energy stable linear scheme for the no-slope-selection thin film model, Commun. Comput. Phys. 29.3 (2021): 905-929.

HoAzXu19 [13] D. Hou, M. Azaiez, and C. Xu, A variant of scalar auxiliary variable approaches for gradient flows, J. Comput. Phys. 395 (2019): 307-332.

HoTaYa17 [14] T. Hou, T. Tang, and J. Yang, Numerical analysis of fully discretized Crank¡Nicolson scheme for fractional-in-space Allen-Cahn equations, J. Sci. Comput. 72.3 (2017): 1214-1231.

JoLi04 [15] H. Johnston and J.-G. Liu, Accurate, stable and efficient Navier¡Stokes solvers based on explicit treatment of the pressure term, J. Comput. Phys. 199.1 (2004): 221-259.

LiQiZh17 [16] X. Li, Z. Qiao, and H. Zhang, Convergence of a fast explicit operator splitting method for the epitaxial growth model with slope selection, SIAM J. Numer. Anal. 55 (2017): 265-285.

LiWaWa21 [17] C. Liu, C. Wang, and Y. Wang, A structure-preserving, operator splitting scheme for reaction-diffusion equations with detailed balance, J. Comput. Phys. 436 (2021): 110253.

Li13 [18] J. Liu, Simple and efficient ALE methods with provable temporal accuracy up to fifth order for the Stokes equations on time varying domains, SIAM J. Numer. Anal. 51.2 (2013): 743-772.

ShWaWaWi12 [19] J. Shen, C. Wang, X. Wang, and S. M. Wise, Second-order convex splitting schemes for gradient flows with Ehrlich-Schnabel type energy: application to thin film epitaxy, SIAM J. Numer. Anal. 50.1 (2012): 105-125.

ShXu18 [20] J. Shen, and J. Xu, Convergence and error analysis for the scalar auxiliary variable (SAV) schemes to gradient flows, SIAM J. Numer. Anal. 56.5 (2018): 2895-2912.

ShXu18b [21] J. Shen and J. Xu, Stabilized predictor-corrector schemes for gradient flows with strong anisotropic free energy, Commun. Comput. Phys., 24 (2018): 635-654.

ShXuYa18 [22] J. Shen, J. Xu, and J. Yang, The scalar auxiliary variable (SAV) approach for gradient flows, J. Comput. Phys. 353 (2018): 407-416.

ShXuYa19 [23] J. Shen, J. Xu, and J. Yang, A new class of efficient and robust energy stable schemes for gradient flows, SIAM Rev. 61.3 (2019): 474-506.

ShTaYa16 [24] J. Shen, T. Tang, and J. Yang, On the maximum principle preserving schemes for the generalized Allen-Cahn equation, Commun. Math. Sci. 14.6 (2016): 1517-1534.

ShYa10 [25] J. Shen, and X. Yang, Numerical approximations of Allen-Cahn and Cahn-Hilliard equations, Disc. Cont. Dyn. Syst. 28.4 (2010): 1669.

TaYa16 [26] T. Tang, and J. Yang, Implicit-explicit scheme for the Allen-Cahn equation preserves the maximum principle, J. Comput. Math. 34.5 (2016): 451.

TaYuZh19 [27] T. Tang, H. Yu, and T. Zhou, On energy dissipation theory and numerical stability for time-fractional phase-field equations, SIAM J. Sci. Comput. 41.6 (2019): A3757-A3778.

Ts19 [28] C. L. Tseng, An Efficient Solver for Fractional Diffusion Equations PhD thesis, National Tsing Hua University.

WaLi00 [29] C. Wang, J.-G. Liu, Convergence of gauge method for incompressible flow, Math. Comput. 69.232 (2000): 1385-1407.

WaWaWi10 [30] C. Wang, X. Wang, and S. M. Wise, Unconditionally stable schemes for equations of thin film epitaxy, Disc. Cont. Dyn. Syst. 28.1 (2010): 405-423.

WaWi11 [31] C. Wang, and S. M. Wise, An energy stable and convergent finite-difference scheme for the modified phase field crystal equation, SIAM J. Numer. Anal. 49.3 (2011): 945-969.

WiWaLo09 [32] S. M. Wise, C. Wang, and J. S. Lowengrub, An energy-stable and convergent finite-difference scheme for the phase field crystal equation, SIAM J. Numer. Anal. 47.3 (2009): 2269-2288.

XuTa06 [33] C. Xu, and T. Tang, Stability analysis of large time-stepping methods for epitaxial growth models, SIAM J. Numer. Anal. 44.4 (2006): 1759-1779.

WuZwZe14 [34] X. Wu, G. J. van Zwieten, and K. G. van der Zee, Stabilized second-order convex splitting schemes for Cahn-Hilliard models with application to diffuse-interface tumor-growth models, Int. J. Numer. Methods Biomed. Eng., 30 (2014): 180-203.

XuZh03 [35] J. J. Xu, and H. K. Zhao, An Eulerian formulation for solving partial differential equations along a moving interface, J. Sci. Comput. 19.1 (2003): 573-594.

Ya16 [36] X. Yang, Linear, first and second-order, unconditionally energy stable numerical schemes for the phase field model of homopolymer blends, J. Comput. Phys. 327 (2016): 294-316.

ZhHuWaYu18 [37] C. Zhang, J. Huang, C. Wang, and X. Yue, On the operator splitting and integral equation preconditioned deferred correction methods for the “Good” Boussinesq equation, J. Sci. Comput. 75 (2018): 687-712.

ZhWaHuWaYu17 [38] C. Zhang, H. Wang, J. Huang, C. Wang, and X. Yue, A second order operator splitting numerical scheme for the “Good” Boussinesq equation, Appl. Numer. Math. 119 (2017): 179-193.

ZhWaYa17 [39] J. Zhao, Q. Wang, and X. Yang, Numerical approximations for a phase field dendritic crystal growth model based on the invariant energy quadratization approach, Int. J. Numer. Methods Eng. 110.3 (2017): 279-300.