January 2015

ON SEVERAL EFFICIENT ALGORITHMS FOR SOME PARTIAL DIFFERENTIAL EQUATIONS

Hee Jun Choi
Purdue University

Graduate School Form 30
Updated 1/15/2015

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Hee Jun Choi

Entitled
ON SEVERAL EFFICIENT ALGORITHMS FOR SOME PARTIAL DIFFERENTIAL EQUATIONS

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

Jie Shen
Chair

Min Chen

Zhiqiang Cai

Suchuan Dong

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy of Integrity in Research" and the use of copyright material.

Approved by Major Professor(s): Jie Shen

Approved by: David Goldberg 06/19/2015

Head of the Departmental Graduate Program Date

ON SEVERAL EFFICIENT ALGORITHMS FOR SOME PARTIAL DIFFERENTIAL EQUATIONS

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Heejun Choi

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

August 2015

Purdue University

West Lafayette, Indiana

I dedicate this dissertation,

to my beloved family for their unconditional support.

ACKNOWLEDGMENTS

I would like to express the deepest gratitude to my advisor professor Jie Shen, without whom this thesis would be impossible to accomplish. I am extremely grateful for his patience and encouragement, helping me overcome the challenges in my research topics. His dedication towards research and his profound knowledge has always been great inspiration and guidance to me.

I also want to thank my committee members, professor Min Chen, Suchuan (Steven) Dong and Zhiqiang Cai for taking their time and energy to go through my thesis work and coordinate with me.

I would like to thank my friends who shared exciting mathematical experience on every other Saturday when I was in Korea. I won't be here without them. I would like to thank sincerely to my friends in my lab, Yongyong Cai, Lina Ma, Yingwei Wang and many more. They provided me a lot of help in both my research and life. Also I would like to thank my friends, Sungwon Ahn, Minji Ha, Joon Ho and many more for their warm encouragement.

Last but not least, I would like to thank sincerely my parents and sister for their unconditional support.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Choi, Heejun Ph.D., Purdue University, August 2015. On several efficient algorithms for some partial differential equations. Major Professor: Jie Shen.

This thesis focuses on the development and the analysis of high-order method for Partial Differential Equations (PDEs), the Magneto-HydroDynamics (MHD) equation, the Cahn-Hilliard phase-field equation and the Allen-Cahn phase-field equation and Ordinary Differential Equations (ODEs).

For the fluid related equations, we focus on the stability and the error estimates. We suggest four unconditionally stable discretizations of the MHD equation and perform the error analysis. As an application, we develop an adaptive scheme and carry out numerical experiments to see the effectiveness. We carry out the error analysis of the convex-splitting scheme and the stabilized scheme of the Cahn-Hilliard equation.

We develop the spectral method for complex geometries which is based on the fictitious domain method. For the ODEs, we develop a second-order defect correction method. The main tool for the defect correction is the Schur decomposition and the scheme is A-stable.

1. INTRODUCTION

1.1 Overview of thesis

A Partial Differential Equation (PDE) is an equation which consists of partial derivatives of an unknown function. It has been used widely to describe various phenomenon. The behavior of a heated metal is governed by the heat equation. The motion of viscous fluid substance is described by the Navier-Stokes Equation (NSE) . Also there are many important PDEs, the wave equation, the Korteweg-de Vries equation, the Allen-Cahn equation, describing other physical systems. The solution of a PDE helps us to have better understanding of the nature. Hence finding the solution of a PDE is very important. The analytic solution of a PDE gives us the precise information of the nature. However, obtaining the analytic solution of a PDE is not possible most of the time. Hence we use numerical methods to obtain an approximate solution of a PDE. In this thesis, we develop and analyze numerical methods which is stable and highly accurate.

The theory of a stable discretization of a PDE began in a paper of Courant, Friedrichs, and Lewy [1]. It was pointed out that a consistent discretization of a PDE is not enough for the convergence. In [2], the authors showed that for a consistent finite difference discretization the stability implies the convergence. Hence developing a stable discretization is essential to obtain an accurate numerical solution. In Chapter 2 and Chapter 3, we focus on developing and analyzing stable discretizations of the fluid related PDEs which are based on the projection methods.

In Chapter 2, we introduce four unconditionally stable numerical discretizations of the Magneto-HydroDynamics (MHD) equation. We also provide an adaptive time-stepping strategy so that small time steps are needed only for the accuracy. The

MHD equation describes the motion of conducting fluid under the magnetic field. The example includes Plasmas, liquid metals and salt water.

The NSE is a part of the MHD and there has been much research to apply the projection method which is an efficient algorithm to solve NSE [3–5]. We improve existing numerical discretizations of the MHD equation in two ways. One is a rotational pressure-correction form and the other is an adaptive time-stepping. It is known that the projection method suffers from the numerical layer, decreasing the accuracy for the pressure [6]. Hence we apply the rotational correction form to avoid the numerical boundary layer. We also suggest the an adaptive time-stepping strategy. By the strategy, we can choose the size of time steps small when the solution changes quickly and small when solution changes slowly. The numerical experiment shows this strategy is efficient in terms of number of time steps.

In Chapter 3, we perform the error analysis of the stabilized scheme and the convex-splitting scheme of the Cahn-Hilliard phase-field model and the Allen-Cahn phase-field model. The Cahn-Hilliard phase-field model and the Allen-Cahn phase-field mode describe the flow with two constitutive components. The stabilized scheme and the convex-splitting for the Cahn-Hilliard phase-field equation were developed in [7] and [8]. Two schemes mimic the energy dissipation law of the PDE, hence they are unconditionally energy stable. In Chapter 3, we presents detailed error analysis of two schemes for the Cahn-Hilliard phase-field equation. The error analysis of the Allen-Cahn phase-field equation can be carried out by using essentially the same arguments.

In Chapter 4 and Chapter 5, we focus on developing high-order schemes for PDEs and Ordinary Differential Equations (ODEs). In Chapter 4 we develop spectral methods for complex geometries. There are many numerical methods to solve the PDEs, finite element, finite difference, spectral element, spectral method. The finite element method subdivide the domain and approximate the solution using locally supported functions. Hence finite element method would be a good choice when the domain is complex. But the matrix system due to the finite element is huge and the convergence rate is an algebraic function of the number of unknowns. The spectral method approximate the solution using global function and accuracy shows exponential decay. Hence spectral method is a good choice when the solution is smooth. However, the disadvantage of the spectral method is that its domain needs to be simple (rectangle, disk. ...).

In Chapter 4, we introduce two algorithms for the Helmholtz equation which show spectral accuracy which are based on spectral method and fictitious domain method. One method uses the weak formulation on the extended domain and imposes boundary condition as a constraint. We can avoid the derivative discontinuity by imposing no boundary condition on the extended domain. The other method is based on the splitting of original equation into two equations which was introduced in [9]. In [9], the author uses Fourier spectral method to solve the extended problem which requires periodic extension of a function. The algorithm that we suggest is based on non-periodic and smooth extension. Hence the Gibbs phenomenon can be avoided. For both algorithms we assumed that smooth extension of a function is available and obtaining a smooth extension would be a further research topic.

In Chapter 5 we presents a numerical scheme to solve an ODE to obtain a high order of accuracy. The theories to obtain highly accurate numerical solution of the ODEs have been researched extensively [10–13]. There are mainly two strategies to obtain a high order of accuracy. One strategy is to construct a numerical scheme with a high order of consistency error. The examples of these are the Runge-Kutta type or the multi-step methods. The other direction to obtain a high order of accuracy is to solve a low order scheme repeatedly. It is called a defect correction type scheme.

The defect correction type method was extensively studied in seventies [13]. In [12], the authors improved the defect correction method and the new method is called a Spectral Deffered Correction (SDC) method. SDC improved the classical defect correction method in two ways. The authors in [12] discretized the interval using Gauss points, rather than equidistance points. Hence the Runge's phenomenon could be avoided. The authors also changed the differential equation into a Picard form so that the numerical integration could be applied rather than differentiation. As a result, SDC obtains larger stability region than the classical defect correction.

In Chapter 5, we suggest a different type of corrector. Our new scheme improves the SDC in two ways. One is the order of accuracy and the other is the A-stability. The order of accuracy of the SDC improves by one at each correction. However, we obtain two orders at each correction. The stability region of the SDC does not cover left-half plane. But our scheme is the same as the collocation method if the underlying problem is a linear, constant-coefficient problem. Hence the scheme is A-stable which enables us to take larger time steps. However the scheme is complicated than the SDC and it is not easy to program.

1.2 Notations

We consider a finite time interval $[0, T]$ and a domain $\Omega \subset \mathbb{R}^d$ $(d = 2, 3)$ which is connected, bounded and open with a sufficiently smooth boundary $\partial\Omega$ such that the following Sobolev inequalities hold:

$$\begin{aligned} \|g\|^2_{L^4(\Omega)} &\leq C(2,\Omega)\|g\|_{L^2(\Omega)}\|g\|_{H^1(\Omega)}, \qquad d = 2, \\ \|g\|^2_{L^3(\Omega)} &\leq C(3,\Omega)\|g\|_{L^2(\Omega)}\|g\|_{H^1(\Omega)}, \qquad d = 3. \end{aligned} \tag{1.1}$$

Let δt be the time-step size. For a sequence of functions $\varphi^0, \varphi^1, \ldots, \varphi^N$ in some Hilbert space E, we denote the sequence by $\varphi_{\delta t}$ and define the following discrete norms for $\varphi_{\delta t}$:

$$\|\varphi_{\delta t}\|_{l^2(E)} = \left(\delta t \sum_{n=0}^{N} \|\varphi^n\|^2_E\right), \qquad \|\varphi_{\delta t}\|_{l^\infty(E)} = \max_{0\leq n\leq N} \left(\|\varphi\|_E\right). \tag{1.2}$$

Let $\|\cdot\|_k$ denote the usual norm in $H^k(\Omega)$. In particular, $\|\cdot\|$ and $(,)$ denote $L^2(\Omega)$ norm and the associated inner product, respectively.

Denote

$$X = H^1_0(\Omega)^d, \quad M = L^2_0(\Omega) = \left\{q \in L^2(\Omega) : \int_\Omega q = 0\right\}, \tag{1.3}$$

and introduce the following spaces of incompressible vector fields:

$$\begin{aligned} H &= \left\{ v \in L^2(\Omega)^d : \nabla \cdot v = 0;\ v \cdot \mathbf{n}|_{\partial\Omega} = 0 \right\}, \\ V &= \left\{ v \in H^1(\Omega)^d : \nabla \cdot v = 0;\ v|_{\partial\Omega} = 0 \right\}, \end{aligned} \tag{1.4}$$

where $\mathbf{n}$ is the outward normal vector of $\partial\Omega$ and the following holds (cf. for instance [14]):

$$L^2(\Omega)^d = H \oplus \nabla(H^1(\Omega)). \tag{1.5}$$

Define P_H as the L^2-orthogonal projector in H, i.e.,

$$(u - P_H u, v) = 0, \qquad \forall u \in L^2(\Omega)^d, \quad v \in H, \tag{1.6}$$

and P_H is stable in H^1, i.e., $\|P_H(u)\|_{H^1} \leq c\|u\|_{H^1}$ [14].

In the following, we define the inverse Stokes operator $S : H^{-1}(\Omega)^d \to V$. For all $v \in H^{-1}(\Omega)^d$, $(S(v), r) \in V \times L_0^2(\Omega)$ is the solution to the following problem:

$$\begin{cases} (\nabla S(v), \nabla w) - (r, \nabla \cdot w) = \langle v, w \rangle, \qquad \forall w \in H_0^1(\Omega)^d, & (1.7) \\ (q, \nabla \cdot S(v)) = 0, \qquad \forall q \in L_0^2(\Omega), & (1.8) \end{cases}$$

where $\langle \cdot, \cdot \rangle$ denote the paring between $H^{-1}(\Omega)^d$ and $H_0^1(\Omega)^d$. It is well-known that the following H^2 regularity results hold [14]:

$$\|S(v)\|_2 + \|\nabla r\| \leq c\|v\|, \quad \forall v \in L^2(\Omega)^d. \tag{1.9}$$

2. AN EFFICIENT NUMERICAL SCHEME FOR MAGNETO-HYDRODYNAMIC FLOWS

We study in this chapter a numerical approximation of the magneto-hydrodynamics of a viscous incompressible fluid. We construct an unconditionally stable, semi-discretized scheme which requires solving, at each time step, a coupled, linear, positive-definite system for the velocity and magnetic fields, and a Poisson equation for the pressure.

2.1 Introduction

The Magneto-HydroDynamics (MHD) of incompressible, viscous, resistive, conducting fluid describes the motion of fluid under the magnetic field. Examples are plasmas, liquid metal or salt water. MHD is governed by following MHD equation:

$$u_t + (u \cdot \nabla)u - \nu\Delta u + \nabla p - \alpha(\nabla \times b) \times b = f, \quad \text{in} \quad \Omega, \tag{2.1a}$$

$$b_t - \eta\Delta b + \nabla \times (b \times u) = 0, \quad \text{in} \quad \Omega, \tag{2.1b}$$

$$\text{div } u = 0, \quad \text{in} \quad \Omega, \tag{2.1c}$$

$$\text{div } b = 0, \quad \text{in} \quad \Omega, \tag{2.1d}$$

$$u|_{\partial\Omega} = 0, \tag{2.1e}$$

$$n \cdot b|_{\partial\Omega} = 0, n \times (\nabla \times b)|_{\partial\Omega} = 0, \tag{2.1f}$$

with the given initial data

$$u(\cdot, 0) = u^0(\cdot),\ b(\cdot, 0) = b^0,\ p(\cdot, 0) = p^0. \tag{2.2}$$

Ω is an open and bounded domain in $\mathbb{R}^d$, $(d = 2, 3)$, n is unit outward normal of $\partial\Omega$, the unknowns are the fluid velocity field u, magnetic field b and pressure p, ν, α

and η are respectively; the viscosity, magnetic diffusivity and $\alpha = 1/(4\pi\mu\rho)$, where μ is magnetic permeability and ρ is fluid density. The fluid is governed by the Navier Stokes equation (NSE) and the magnetic field is governed by the Maxwell equation. Interaction between the fluid and magnetic field is expressed by the Lorentz force. Since NSE is a part of the MHD equation, the currently existing efficient numerical scheme for NSE would be a very good building block for developing a numerical scheme for the MHD equation.

The projection method is an efficient numerical scheme to solve a time-dependent NSE. After the development of the projection method [15,16], there has been extensive research to improve the projection method. [17] contains an inclusive review of the projection method. One of the difficulties in developing a numerical scheme for NSE is caused by the coupling of fluid velocity and pressure. Because of this coupling, spatial discretization for fluid velocity and pressure need to be carefully chosen so that they satisfy the inf-sup condition. The projection method decouples fluid velocity and pressure using Helmholtz decomposition and it makes the projection method very attractive for the discretization of NSE. Hence there has been much research studies applying the projection method for the discretization of the MHD equation.

A time semi-discrete scheme based on a consistent splitting scheme [18] was developed in [3]. In [3], the authors proved local-time well posedness. In [4], the author developed three fully discrete finite element discretizations. One scheme features fluid velocity, magnetic field and pressure all coupled and unconditionally stable. The other two schemes are fluid velocity and magnetic field decoupled and conditionally stable. In [5], the authors developed three unconditionally stable algorithms. One is a second-order method based on pressure-correction and Crank-Nicholson. The other two schemes are fully decoupled, first-order schemes based on perturbation of the fluid velocity or magnetic field.

In this chapter, we introduce four unconditionally time semi-discrete stable schemes based on the projection method. Two schemes are based on the pressure-correction scheme and two schemes are based on the rotational pressure-correction scheme. For the first-order pressure-correction scheme, we carry out rigorous error analysis. We improve previously existing methods in two ways: rotational pressure-correction form and the adaptive time-stepping. It is well known that the pressure-correction projection method suffers from a numerical boundary layer, decreasing accuracy for pressure [6]. The rotational pressure-correction for NSE improves pressure approximation, and the rotational pressure-correction form for the MHD equation, also it improves pressure estimation. The main advantage of an unconditionally stable scheme is that the size of time step does not need to be maintained small when solution changes slowly. Based on the unconditionally stable scheme, we develop an adaptive time-stepping scheme.

This chapter is organized as follows: we introduce a few mathematical preliminaries in Section 2.2. In Section 2.3, we introduce an first-order scheme and provide stability and error analysis. In Section 2.4, we introduce high-order numerical schemes and rotational form. Since we have developed a time semi-discrete scheme, we develop spatial discretization in Section 2.5 based on Legendre-Galerkin method. The advantage of the unconditional scheme is that the time step needs to be maintained small only for accuracy. Therefore larger time-stepping can be chosen to reduce the cost if the solution does not change fast. Hence, we develop an adaptive time-stepping strategy in Section 2.6 based on the numerical scheme introduced in Section 2.3. In Section 2.7, we perform numerical tests to verify the order of accuracy and performance of the adaptive scheme. Conclusion are drawn in 2.8.

2.2 Preliminary

In this section, we summarize the basic mathematical properties of the MHD equation described in [19] for completeness.

2.2.1 Function spaces and weak formulation

To describe properties of the MHD equation, we first define a few functional spaces. We denote $\mathbb{L}(\Omega) = (L(\Omega))^d$, $\mathbb{H}^s(\Omega) = (H^s(\Omega))^d$ for $s \in \mathbb{N}$ and $\mathbb{H}_0^1(\Omega) = (H_0^1(\Omega))^d$. $d = 2, 3$ cases will be considered in this chapter. $\|\cdot\|_k$ denotes the usual $H^k(\Omega)$ norm and we define $\|\cdot\| = \|\cdot\|_0$. Let $T > 0$ and X be a Banach space with the norm $\|\cdot\|_X$. Then $L^p(0, T; X)$ is all measurable functions $u : [0, T] \to X$ with

$$\|u\|_{L^p(0,T;X)} = (\int_0^T \|u\|_X^p dt)^{1/p} < \infty, \tag{2.3}$$

for $1 \leq p < \infty$, and

$$\|u\|_{L^\infty(0,T;X)} = esssup_{0\leq t\leq T}\|u\|_X < \infty. \tag{2.4}$$

The following spaces are needed to define the weak formulation of MHD equation:

$$\begin{aligned}
\mathcal{V}_1 &= \{v \in (C_c^\infty(\Omega))^d, \nabla \cdot v = 0\}, \\
V_1 &= \{v \in \mathbb{H}_0^1(\Omega), \nabla \cdot v = 0\}, \\
D(\mathcal{A}) &= V_1 \cap \mathbb{H}^2, \\
H_1 &= \{v \in \mathbb{L}^2(\Omega), \nabla \cdot v = 0 \text{ and } v \cdot n|_{\partial\Omega} = 0\}, \\
\mathcal{V}_2 &= \{C \in (C^\infty(\bar{\Omega}))^d, \nabla \cdot C = 0 \text{ and } C \cdot n|_{\partial\Omega} = 0\}, \\
V_2 &= \{C \in \mathbb{H}^1(\Omega), \nabla \cdot C = 0 \text{ and } C \cdot n|_{\partial\Omega} = 0\}, \\
H_2 &= H_1, \\
V &= V_1 \times V_2, \\
H &= H_1 \times H_2, \\
&V_1', \text{ the dual space of } V_1.
\end{aligned} \tag{2.5}$$

The following trilinear form on $L^1(\Omega) \times W^{1,1}(\Omega) \times L^1(\Omega)$ is useful to define the weak formulation of the MHD equation:

$$t(u, v, w) = \sum_{i,j=1}^d \int_\Omega u_i D_i v_j w_j dx. \tag{2.6}$$

Then it is known in [20] that

$$t(u,v,w) \le \begin{cases} c\|u\|_1\|v\|_1\|w\|_1 \\ c\|u\|_1\|v\|_2\|w\| \\ c\|u\|\|v\|_2\|w\|_1 \end{cases} \tag{2.7}$$

for $d = 2$ or 3. And we have similar inequalities for curl:

$$((\nabla \times u) \times v, w) \le \begin{cases} c\|u\|_1\|v\|_1\|w\|_1 \\ c\|u\|_1\|v\|_2\|w\| \\ c\|u\|_2\|v\|\|w\|_1 \end{cases} \tag{2.8}$$

Let u and b be smooth solution of (2.1a) - (2.1f). Using the identity,

$$\begin{aligned} (\nabla \times b) \times b &= (b \cdot \nabla)b - \frac{1}{2}\nabla(b \cdot b), \\ \nabla \times (b \times u) &= b(\nabla \cdot u) - u(\nabla \cdot b) + (u \cdot \nabla)b - (b \cdot \nabla)u \\ &= (u \cdot \nabla)b - (b \cdot \nabla)u, \\ \Delta b &= \nabla(\nabla \cdot b) - \nabla \times \nabla \times b, \end{aligned} \tag{2.9}$$

(2.1a) and (2.1b) can be written as follows:

$$u_t + (u \cdot \nabla)u - \nu\Delta u + \nabla p + \alpha(\tfrac{1}{2}\nabla(b \cdot b) - (b \cdot \nabla)b) = f, \tag{2.10}$$

$$b_t + \eta\nabla \times \nabla \times b + (u \cdot \nabla)b - (b \cdot \nabla)u = 0. \tag{2.11}$$

If we take the inner product (2.10) with $v \in \mathcal{V}_1$, we have

$$\frac{d}{dt}(u,v) + \nu(\nabla u, \nabla v) + t(u,u,v) - \alpha t(b,b,v) = (f,v). \tag{2.12}$$

Accordingly, if we take the inner product (2.11) with $c \in \mathcal{V}_2$, we have

$$\frac{d}{dt}(b,c) + \eta(\nabla \times b, \nabla \times c) + t(u,b,c) - t(b,u,c) = 0. \tag{2.13}$$

From (2.12) and (2.13), we can define the following weak formulation of the MHD equation:

Problem (Weak)

For $\Omega \subset \mathbb{R}^d$, for $d = 2$ or 3, for f in $L^2(0,T;V_1')$ and $(u^0, b^0) \in H$, to find $u \in L^2(0,T;\mathbb{V}_1)$ and $b \in L^2(0,T;\mathbb{V}_2)$ satisfying

$$\frac{d}{dt}(u,v) + \frac{d}{dt}(b,c) + \nu(\nabla u, \nabla v) + \eta(\nabla \times b, \nabla \times c) + t(u,u,v) - \alpha t(b,b,v)$$

$$+t(u,b,c) - t(b,u,c) = (f,v), \qquad \text{for all } (v,c) \in V, \tag{2.14}$$

$$u(0,\cdot) = u^0(\cdot) \text{ and } b(0,\cdot) = b^0(\cdot). \tag{2.15}$$

If $f \in L^2(0,T;H), (u^0, b^0) \in V$, we call the solution of Problem (Weak) a strong solution, if it satisfies $(u,b) \in L^2(0,T;D(\mathcal{A})) \cap L^\infty(0,T;V))$.

2.2.2 Existence, uniqueness and regularity result

We have the following existence and uniqueness result:

Theorem 2.2.1 *For $f \in L^2(0,T;V_1'), (u^0, b^0) \in H$,*

Problem (Weak) has a solution $\Phi = (u,b) \in L^2(0,T;V) \cap L^\infty(0,T;H)$.

Furthermore, if $d = 2$, Φ is unique and

$$\Phi' \in L^2(0,T;V') \text{ and } \Phi \in C([0,T];H). \tag{2.16}$$

If $d = 3$, there is, at most, one solution of Problem (Weak) satisfying

$$\Phi \in L^4(0,T;V). \tag{2.17}$$

Theorem 2.2.2 *For $f \in L^\infty(0,T;H), (u^0, b^0) \in V$,*

if $d = 2$, the solution ϕ satisfies

$$\Phi \in L^2(0,T;D(\mathcal{A})) \cap L^\infty(0,T;V). \tag{2.18}$$

If $d = 3$, there exists $T_* > 0$,(depending on $\Omega, f, (u^0, b^0)$) and, on $[0,T_*]$, there exists a unique solution Φ of Problem (Weak) with

$$\Phi \in L^2(0,T_*;D(\mathcal{A})) \cap L^\infty(0,T_*;V). \tag{2.19}$$

We have the following regularity for the time derivative:

Theorem 2.2.3 *Let $\Phi_0 \in H$ and f be functions satisfying*

$$f^{(j)} \in L^\infty(0,T;H_1) \text{ for } j = 0,\ldots,j_0, \tag{2.20}$$

where $0 < T \leq \infty$ and $j_0 \in \mathbb{N}$. If $d = 2$, the solution Φ of Problem (Weak) satisfies, for every $\alpha_0 > 0$,

$$\begin{aligned} &\Phi^{(j)} \in L^\infty(\alpha_0,T;D(\mathcal{A})) \quad \text{for } j = 0,\ldots,j_0-1, \\ &\Phi^{(j_0)} \in L^\infty(\alpha_0,T;V). \end{aligned} \tag{2.21}$$

If $d = 3$, if the strong solution Φ of Problem (Weak) is such that

$$\Phi \in L^\infty(0,T;V), \tag{2.22}$$

then Φ satisfies (2.21).

Now we are about to describe the regularity of the spatial variable for a special case. For the rest of this subsection, we assume f is independent of t and belongs to H^1. Let $S(t)$ be the semi-group associated with the strong solution of Problem (Weak), i.e. for $\Phi_0 \in V$ and $t > 0$, $S(t)\Phi_0 = \Phi(t) \in V$ where $\Phi \in C([0,t];V)$ is the solution of Problem (Weak).

Definition 2.2.1 *A funcional invariant set for Problem (Weak) is a set $X \subset V$ which satisfies the following properties:*

1. *For every $\Phi_0 \in X$, Problem (Weak) as a strong solution in $[0,\infty)$.*

2. *S(t)X = X, for all $t > 0$.*

3. *X possesses an open neighborhood $\omega($ in V or $H)$, and for every $u_0 \in \omega, S(t)u_0$ tends to X, in V or H, as $t \to \infty$.*

Note that if Φ is a stationary solution, $\{\Phi\}$ is a functional invariant set. Then we have following regularity result.

Theorem 2.2.4 *For $d = 2$ or 3, if $f \in (C^\infty(\bar{\Omega}))^d \cap H^1$. Then any functional invariant set for Problem (Weak) is contained in $(C^\infty(\bar{\Omega}))^{2d}$.*

2.2.3 Energy dissipation

The MHD equation has many conservations laws (i.e. mass, momentum, angular momentum, energy). In this subsection, we describe the conservation law of energy. This is important, especially when we develop the semi-discrete scheme of the MHD equation. When there is no volume force f in (2.1a), we can derive the energy law as follows. If we take the inner product u with (2.1a) and αb with (2.1b), we obtain

$$\frac{1}{2}\frac{d}{dt}\|u\|^2 + \nu(\nabla u, \nabla u) - \alpha((\nabla \times b) \times b, u) = 0, \tag{2.23}$$

and

$$\frac{\alpha}{2}\frac{d}{dt}\|b\|^2 + \alpha\eta(\nabla b, \nabla b) + \alpha(\nabla \times (b \times u), b) = 0. \tag{2.24}$$

Taking the sum of (2.23) and (2.24), we have the following energy dissipation:

$$\frac{1}{2}\frac{d}{dt}\|u\|^2 + \frac{\alpha}{2}\frac{d}{dt}\|b\|^2 \leq -\nu\|\nabla u\|^2 - \alpha\eta\|\nabla b\|^2. \tag{2.25}$$

We will develop first- and second-order time-stepping that mimics the property (2.25) in subsequent sections.

2.3 Stability and error estimates for a first-order scheme

In this section, we introduce an unconditionally stable first-order scheme for (2.1a) - (2.1f) based on the pressure-correction method and perform stability and error analysis.

The scheme consists of two steps. Given $(u^n, b^n, \nabla p^n)$, we find $(\tilde{u}^{n+1}, b^{n+1})$ solving (2.26) and (2.27) together. And we obtain $(u^{n+1}, \nabla p^{n+1})$ solving (2.28) where

$$\begin{cases} \dfrac{\tilde{u}^{n+1} - u^n}{\delta t} + (u^n \cdot \nabla)\tilde{u}^{n+1} - \nu\Delta\tilde{u}^{n+1} + \nabla p^n - \alpha(\nabla \times b^{n+1}) \times b^n = 0, & (2.26) \\ \dfrac{b^{n+1} - b^n}{\delta t} - \eta\Delta b^{n+1} + \nabla \times (b^n \times \tilde{u}^{n+1}) = 0, & (2.27) \\ \tilde{u}^{n+1}|_{\partial\Omega} = 0, & \\ b^{n+1} \cdot n|_{\partial\Omega} = 0, n \times (\nabla \times b^{n+1})|_{\partial\Omega} = 0; & \end{cases}$$

and

$$\begin{cases} \dfrac{u^{n+1} - \tilde{u}^{n+1}}{\delta t} + \nabla(p^{n+1} - p^n) = 0, \\ u^{n+1} \cdot n|_{\partial\Omega} = 0, \mathrm{div} u^{n+1} = 0. \end{cases} \tag{2.28}$$

In Section 2.5, we will discuss how to obtain a solution of (2.26)-(2.28) with Legendre-Galerkin spatial discretization.

2.3.1 Stability

For the stability analysis, we can define the following discrete energy:

$$E(u, b, p, \delta t) = \|u\|^2 + \alpha\|b\|^2 + \delta t^2\|\nabla p\|^2. \tag{2.29}$$

We can obtain the following discrete energy law using E.

Theorem 2.3.1 *The scheme* (2.26)-(2.28) *is unconditionally energy stable in the sense that:*

$$E(u^{n+1}, b^{n+1}, p^{n+1}, \delta t) \leq E(u^n, b^n, p^n, \delta t) \qquad n \geq 0. \tag{2.30}$$

Proof Taking the inner product (2.26) with $2\delta t\tilde{u}^{n+1}$ implies

$$\begin{aligned} &\|\tilde{u}^{n+1}\|^2 - \|u^n\|^2 + \|\tilde{u}^{n+1} - u^n\|^2 + 2\nu\delta t\|\nabla\tilde{u}^{n+1}\|^2 \\ &\qquad + 2\delta t(\nabla p^n, \tilde{u}^{n+1}) - 2\delta t\alpha((\nabla \times b^{n+1}) \times b^n, \tilde{u}^{n+1}) = 0, \end{aligned} \tag{2.31}$$

where we used the fact $((u \cdot \nabla)v, v) = 0$ if $\nabla \cdot u = u \cdot n|_{\partial\Omega} = 0$ and $v \in H_0^1(\Omega)^d$. Taking the inner product (2.27) with $2\delta t\alpha b^{n+1}$ implies

$$\alpha\|b^{n+1}\|^2 - \alpha\|b^n\|^2 + \alpha\|b^{n+1} - b^n\|^2 + 2\delta t\alpha\eta\|\nabla b^{n+1}\|^2 + 2\delta t\alpha(\nabla \times (b^n \times \tilde{u}^{n+1}), b^{n+1}) = 0. \tag{2.32}$$

Equation (2.28) can be written as follows:

$$\frac{1}{\delta t}u^{n+1} + \nabla p^{n+1} = \frac{1}{\delta t}\tilde{u}^{n+1} + \nabla p^n.$$

Taking the inner product to itself we have

$$(\frac{1}{\delta t}u^{n+1} + \nabla p^{n+1}, \frac{1}{\delta t}u^{n+1} + \nabla p^{n+1}) = (\frac{1}{\delta t}\tilde{u}^{n+1} + \nabla p^{n}, \frac{1}{\delta t}\tilde{u}^{n+1} + \nabla p^{n}),$$
$$\frac{1}{\delta t^2}\|u^{n+1}\|^2 + \|\nabla p^{n+1}\|^2 = \frac{1}{\delta t^2}\|\tilde{u}^{n+1}\|^2 + \|\nabla p^{n}\|^2 + 2\frac{1}{\delta t}(\nabla p^n, \tilde{u}^{n+1}).$$

Hence

$$2\delta t(\nabla p^n, \tilde{u}^{n+1}) = \|u^{n+1}\|^2 - \|\tilde{u}^{n+1}\|^2 + \delta t^2(\|\nabla p^{n+1}\|^2 - \|\nabla p^n\|^2). \tag{2.33}$$

Using integration by parts on the curl, we have the following:

$$(\nabla \times (b^n \times \tilde{u}^{n+1}), b^{n+1}) = ((b^n \times \tilde{u}^{n+1}), \nabla \times b^{n+1}) = ((\nabla \times b^{n+1}) \times b^n, \tilde{u}^{n+1}). \tag{2.34}$$

Taking the sum of (2.31) and (2.32) and using (2.33), (2.34), we have,

$$\begin{aligned}&\|u^{n+1}\|^2 - \|u^n\|^2 + \|\tilde{u}^{n+1} - u^n\|^2 + 2\delta t\nu\|\nabla\tilde{u}^{n+1}\|^2 + \delta t^2(\|\nabla p^{n+1}\|^2 - \|\nabla p^n\|^2)\\ &+ \alpha\|b^{n+1}\|^2 - \alpha\|b^n\|^2 + \alpha\|b^{n+1} - b^n\|^2 + 2\delta t\alpha\eta\|\nabla b^{n+1}\|^2 = 0,\end{aligned} \tag{2.35}$$

which implies the desired result. ■

2.3.2 Error estimates

With the assumption of smoothness of the exact solution, we can prove following error estimates:

Theorem 2.3.2 *If the exact solutions* (u, b, p) *are smooth, when* $\delta t \leq \tau_0$ *for some* $\tau_0 > 0$*, solution of the scheme* (2.26)-(2.28) (u^n, b^n, p^n) $(0 \leq n \leq \frac{T}{\delta t})$ *satisfies the following error estimates:*

$$\begin{aligned}&\|e_{u,\delta t}\|_{l^\infty(L^2(\Omega)^d)} + \|e_{b,\delta t}\|_{l^\infty(L^2(\Omega)^d)}\\ &\qquad + \|\tilde{e}_{u,\delta t}\|_{l^2(H^1(\Omega)^d)} + \|e_{b,\delta t}\|_{l^2(H^1(\Omega)^d)} + \|q_{\delta t}\|_{l^2(L^2(\Omega)^d)} \lesssim \delta t, \qquad (2.36)\\ &\|\tilde{e}_{u,\delta t}\|_{l^\infty(H^1(\Omega)^d)} + \|e_{b,\delta t}\|_{l^\infty(H^1(\Omega)^d)} + \|q_{\delta t}\|_{l^\infty(L^2(\Omega)^d)} \lesssim \delta t^{1/2}.\end{aligned}$$

The proof consists of three steps. In Lemma 1, we prove error estimates for fluid velocity and magnetic field. In Lemma 2, we prove error estimates for the time derivative of fluid velocity. And we prove the error estimate for pressure which completes the proof.

We first define the truncation error R_u^n ($n = 0, 1, \ldots, N-1$) for the velocity field (2.1a) and R_b^n for the magnetic field (2.1b) as follows:

$$R_u^{n+1} = \frac{u(t^{n+1}) - u(t^n)}{\delta t} + (u(t^n)\cdot\nabla)u(t^{n+1}) - \nu\Delta u(t^{n+1}) + \nabla p(t^n) - \alpha(\nabla\times b(t^{n+1}))\times b(t^n), \tag{2.37}$$

$$R_b^{n+1} = \frac{b(t^{n+1}) - b(t^n)}{\delta t} - \eta\Delta b(t^{n+1}) + \nabla\times(b(t^n)\times u(t^{n+1})). \tag{2.38}$$

We also define

$$R_p^{n+1} = \frac{u(t^{n+1}) - u(t^n)}{\delta t} + \nabla(p(t^{n+1}) - p(t^n)). \tag{2.39}$$

It is clear that we have

$$\|R_{u,\delta t}\|_{l^\infty(H^1(\Omega))} + \|R_{b,\delta t}\|_{l^\infty(H^1(\Omega))} + \|R_{p,\delta t}\|_{l^\infty(L^2(\Omega))} \le c_R\delta t, \tag{2.40}$$

where $c_R > 0$ is independent of δt.

Subtracting (2.26), (2.27), (2.28) from (2.37), (2.38), (2.39), respectively, we get the following error equations for $n \ge 0$:

$$\begin{aligned}
&\frac{\tilde{e}_u^{n+1} - e_u^n}{\delta t} + ((u(t^n)\cdot\nabla)u(t^{n+1}) - (u^n\cdot\nabla)\tilde{u}^{n+1}) - \nu\Delta\tilde{e}_u^{n+1} \\
&\qquad + \nabla q^n - \alpha(\nabla\times b(t^{n+1}))\times e_b^n - \alpha(\nabla\times e_b^{n+1})\times b^n = R_u^{n+1}, && (2.41)\\
&\frac{e_b^{n+1} - e_b^n}{\delta t} - \eta\Delta e_b^{n+1} + \nabla\times(e_b^n\times u(t^{n+1})) + \nabla\times(b^n\times\tilde{e}_u^{n+1}) = R_b^{n+1}, && (2.42)\\
&\frac{e_u^{n+1} - \tilde{e}_u^{n+1}}{\delta t} + \nabla q^{n+1} - \nabla q^n = R_p^{n+1}, && (2.43)
\end{aligned}$$

where

$$\begin{aligned}
\tilde{e}_u^n &= u(t^n) - \tilde{u}^n,\\
e_u^n &= u(t^n) - u^n,\\
e_b^n &= b(t^n) - b^n,\\
q^n &= p(t^n) - p^n.
\end{aligned} \tag{2.44}$$

Lemma 1 *Under the assumptions of Theorem 2.3.2, when* $\delta t \le \tau_0$ *for some* $\tau_0 > 0$, *the solution of the scheme* (2.26)-(2.28) (u^n, b^n, p^n) $(0 \le n \le \frac{T}{\delta t})$ *satisfies the following error estimate:*

$$\begin{aligned} &\|e_{u,\delta t}\|_{l^\infty(L^2(\Omega)^d)} + \|e_{b,\delta t}\|_{l^\infty(L^2(\Omega)^d)} + \|\tilde{e}_{u,\delta t}\|_{l^2(H^1(\Omega)^d)} + \|e_{b,\delta t}\|_{l^2(H^1(\Omega)^d)} \lesssim \delta t, \\ &\|\tilde{e}_{u,\delta t}\|_{l^\infty(H^1(\Omega)^d)} + \|e_{b,\delta t}\|_{l^\infty(H^1(\Omega)^d)} \lesssim \delta t^{1/2}. \end{aligned} \tag{2.45}$$

Proof Taking an inner product (2.41) with $2\delta t \tilde{e}_u^{n+1}$, we obtain

$$\begin{aligned} &\|\tilde{e}_u^{n+1}\|^2 - \|e_u^n\|^2 + \|\tilde{e}_u^{n+1} - e_u^n\|^2 + 2\delta t((u(t^n)\cdot\nabla)u(t^{n+1}) - (u^n\cdot\nabla)\tilde{u}^{n+1}, \tilde{e}_u^{n+1}) \\ &\quad + 2\delta t\nu\|\nabla\tilde{e}_u^{n+1}\|^2 + 2\delta t(\nabla q^n, \tilde{e}_u^{n+1}) - 2\delta t\alpha((\nabla\times b(t^{n+1}))\times e_b^n, \tilde{e}_u^{n+1}) \\ &\quad - 2\delta t\alpha((\nabla\times e_b^{n+1})\times b^n, \tilde{e}_u^{n+1}) = 2\delta t(R_u^{n+1}, \tilde{e}_u^{n+1}). \end{aligned} \tag{2.46}$$

The convection term can be rewritten as follows:

$$\begin{aligned} &((u(t^n)\cdot\nabla)u(t^{n+1}) - (u^n\cdot\nabla)\tilde{u}^{n+1}, \tilde{e}_u^{n+1}) \\ &= ((u(t^n)\cdot\nabla)\tilde{e}_u^{n+1} + (e_u^n\cdot\nabla)\tilde{u}^{n+1}, \tilde{e}_u^{n+1}) \\ &= ((u(t^n)\cdot\nabla)\tilde{e}_u^{n+1} - (e_u^n\cdot\nabla)\tilde{e}_u^{n+1} \\ &+ (e_u^n\cdot\nabla)u(t^{n+1}), \tilde{e}_u^{n+1}) = ((e_u^n\cdot\nabla)u(t^{n+1}), \tilde{e}_u^{n+1}), \end{aligned} \tag{2.47}$$

where we used the fact that $((u\cdot\nabla)v, v) = 0$ for $u \in V$ and $v \in H_0^1(\Omega)^d$. It follows that the term containing pressure has the following expression:

$$\begin{aligned} 2\delta t(\nabla q^n, \tilde{e}_u^{n+1}) &= 2\delta t(\nabla q^n, e_u^{n+1} + \delta t(\nabla q^{n+1} - \nabla q^n) - \delta t R_p^{n+1}) \\ &= \delta t^2(\|\nabla q^{n+1}\|^2 - \|\nabla q^n\|^2 - \|\nabla q^{n+1} - \nabla q^n\|^2) - 2\delta t^2(\nabla q^n, R_p^{n+1}), \end{aligned} \tag{2.48}$$

and

$$\begin{aligned} &\|\frac{1}{\delta t}e_u^{n+1} + \nabla q^{n+1} - \nabla q^n\|^2 = \|R_p^{n+1} + \frac{1}{\delta t}\tilde{e}_u^{n+1}\|^2. \\ &\qquad \|\nabla q^{n+1} - \nabla q^n\|^2 = \frac{1}{\delta t^2}(\|\tilde{e}_u^{n+1}\|^2 - \|e_u^{n+1}\|^2) + \|R_p^{n+1}\|^2 + \frac{2}{\delta t}(R_p^{n+1}, \tilde{e}_u^{n+1}), \end{aligned}$$

where we used the fact that $e_u^{n+1} \in H$. Hence we have

$$2\delta t(\nabla q^n, \tilde{e}_u^{n+1}) = \|e_u^{n+1}\|^2 - \|\tilde{e}_u^{n+1}\|^2 + \delta t^2(\|\nabla q^{n+1}\|^2 - \|\nabla q^n\|^2) - \delta t^2\|R_p^{n+1}\|^2 - 2\delta t(R_p^{n+1}, \tilde{e}_u^{n+1}) - 2\delta t^2(\nabla q^n, R_p^{n+1}). \tag{2.49}$$

Taking an inner product (2.42) with $2\alpha\delta t e_b^{n+1}$ we obtain

$$\alpha(\|e_b^{n+1}\|^2 - \|e_b^n\|^2 + \|e_b^{n+1} - e_b^n\|^2) + 2\alpha\delta t\eta\|\nabla e_b^{n+1}\|^2 + 2\alpha\delta t(\nabla \times (e_b^n \times u(t^{n+1})), e_b^{n+1}) + 2\alpha\delta t(\nabla \times (b^n \times \tilde{e}_u^{n+1}), e_b^{n+1}) = 2\alpha\delta t(R_b^{n+1}, e_b^{n+1}) \tag{2.50}$$

Using the integration by parts, we obtain

$$2\alpha\delta t(\nabla \times (b^n \times \tilde{e}_u^{n+1}), e_b^{n+1}) = 2\alpha\delta t((b^n \times \tilde{e}_u^{n+1}, \nabla \times e_b^{n+1}) = 2\delta t\alpha((\nabla \times e_b^{n+1}) \times b^n, \tilde{e}_u^{n+1}). \tag{2.51}$$

Taking the sum of (2.46) and (2.50) and using (2.51), (2.49) and (2.47), we obtain

$$\begin{aligned}
&\|e_u^{n+1}\|^2 - \|e_u^n\|^2 + \|\tilde{e}_u^{n+1} - e_u^n\|^2 + 2\delta t\nu\|\nabla \tilde{e}_u^{n+1}\|^2 \\
&+ \alpha(\|e_b^{n+1}\|^2 - \|e_b^n\|^2 + \|e_b^{n+1} - e_b^n\|^2) + 2\alpha\delta t\eta\|\nabla e_b^{n+1}\|^2 \\
&+ \delta t^2(\|\nabla q^{n+1}\|^2 - \|\nabla q^n\|^2) - \delta t^2\|R_p^{n+1}\|^2 - 2\delta t(R_p^{n+1}, \tilde{e}_u^{n+1}) - 2\delta t^2(\nabla q^n, R_p^{n+1}) \\
&+ 2\delta t((e_u^n \cdot \nabla)u(t^{n+1}), \tilde{e}_u^{n+1}) - 2\delta t\alpha((\nabla \times b(t^{n+1})) \times e_b^n, \tilde{e}_u^{n+1}) \\
&+ 2\alpha\delta t(\nabla \times (e_b^n \times u(t^{n+1})), e_b^{n+1}) = 2\delta t(R_u^{n+1}, \tilde{e}_u^{n+1}) + 2\alpha\delta t(R_b^{n+1}, e_b^{n+1}).
\end{aligned} \tag{2.52}$$

Each inner product term can be bounded as follows:

$$\begin{aligned}
|2\delta t\alpha((\nabla \times b(t^{n+1})) \times e_b^n, \tilde{e}_u^{n+1})| &\le c\delta t\|(\nabla \times b(t^{n+1})) \times e_b^n\|\|\tilde{e}_u^{n+1}\| \\
&\le c\delta t\|e_b^n\|\|\tilde{e}_u^{n+1}\| \\
&\le c\delta t\|e_b^n\|\|\nabla \tilde{e}_u^{n+1}\| \\
&\le c\delta t\|e_b^n\|^2 + \frac{\nu}{8}\delta t\|\nabla \tilde{e}_u^{n+1}\|^2,
\end{aligned} \tag{2.53}$$

$$\begin{aligned}
|2\alpha\delta t(\nabla \times (e_b^n \times u(t^{n+1})), e_b^{n+1})| &= |2\alpha\delta t(e_b^n \times u(t^{n+1}), \nabla \times e_b^{n+1})| \\
&\le c\delta t\|e_b^n \times u(t^{n+1})\|\|\nabla \times e_b^{n+1}\| \\
&\le c\delta t\|e_b^n\|\|\nabla e_b^{n+1}\| \\
&\le c\delta t\|e_b^n\|^2 + \delta t\frac{\alpha\eta}{8}\|\nabla e_b^{n+1}\|^2,
\end{aligned} \tag{2.54}$$

$$\delta t^2 \|R_p^{n+1}\|^2 \le \delta t^4, \tag{2.55}$$

$$\begin{aligned} |2\delta t(R_p^{n+1}, \tilde{e}_u^{n+1})| &\le 2\delta t\|R_p^{n+1}\|\|\tilde{e}_u^{n+1}\| \\ &\le c\delta t^3 + \delta t\frac{\nu}{8}\|\nabla \tilde{e}_u^{n+1}\|^2, \end{aligned} \tag{2.56}$$

$$\begin{aligned} |2k((e_u^n \cdot \nabla)u(t^{n+1}), \tilde{e}_u^{n+1})| &\le 2\delta t\|(e_u^n \cdot \nabla)u(t^{n+1})\|\|\tilde{e}_u^{n+1}\| \\ &\le c\delta t\|e_u^n\|\|\tilde{e}_u^{n+1}\| \\ &\le c\delta t\|e_u^n\|^2 + \delta t\frac{\nu}{8}\|\nabla \tilde{e}_u^{n+1}\|^2, \end{aligned} \tag{2.57}$$

$$\begin{aligned} 2\delta t(R_b^{n+1}, e_b^{n+1}) &\le 2\delta t\|R_b^{n+1}\|\|e_b^{n+1}\| \\ &\le c\delta t^3 + \delta t\eta\frac{\alpha}{8}\|\nabla e_b^{n+1}\|^2, \end{aligned} \tag{2.58}$$

$$\begin{aligned} |2\delta t^2(\nabla q^n, R_p^{n+1})| &= 2\delta t^2\|\nabla q^n\|\|R_p^{n+1}\| \\ &\le c\delta t^3\|\nabla q^n\|^2 + c\delta t^3. \end{aligned} \tag{2.59}$$

Using (2.53), (2.54), (2.56), (2.57), (2.58) and (2.59), (2.60) becomes

$$\begin{aligned} &\|e_u^{n+1}\|^2 - \|e_u^n\|^2 + \|\tilde{e}_u^{n+1} - e_u^n\|^2 + \delta t\nu\|\nabla \tilde{e}_u^{n+1}\|^2 \\ &+ \alpha(\|e_b^{n+1}\|^2 - \|e_b^n\|^2 + \|e_b^{n+1} - e_b^n\|^2) + \alpha\delta t\eta\|\nabla e_b^{n+1}\|^2 + \delta t^2(\|\nabla q^{n+1}\|^2 - \|\nabla q^n\|^2) \\ &\le c(\delta t^4 + \delta t^3) + c\delta t(\|e_u^n\|^2 + \|e_b^n\|^2 + \delta t^2\|\nabla q^n\|^2). \end{aligned} \tag{2.60}$$

Taking the sum of (2.60) for $n = 0, \cdots, N$, we obtain

$$\begin{aligned} &\|e_u^{N+1}\|^2 + \alpha\|e_b^{N+1}\|^2 + \delta t^2\|\nabla q^{N+1}\|^2 + \\ &\sum_{n=0}^{N}(\|\tilde{e}_u^{n+1} - e_u^n\|^2 + \delta t\nu\|\nabla \tilde{e}_u^{n+1}\|^2 + \alpha\delta t\eta\|\nabla e_b^{n+1}\|^2 + \alpha\|e_b^{n+1} - e_b^n\|^2) \\ &\le c(\delta t^3 + \delta t^2) + c\delta t\sum_{n=0}^{N}(\|e_u^n\|^2 + \|e_b^n\|^2 + \delta t^2\|\nabla q^n\|^2). \end{aligned} \tag{2.61}$$

Using discrete Gronwall's inequality, we obtain the desired result. ■

Lemma 2 *Under the assumptions of Theorem 2.3.2, when $\delta t \le \tau_0$ for some $\tau_0 > 0$, solution of the scheme* (2.26)-(2.28) (u^n, b^n, p^n) *$(0 \le n \le \frac{T}{\delta t})$ satisfies the following error estimate,*

$$\|\delta_t e_u^n\| \lesssim \delta t^2. \tag{2.62}$$

for all $n = 1, \cdots, [T/\delta t]$.

Proof We work on the equations for the time increment $\delta_t e_u^n$, $\delta_t e_b^n$. Denote

$$\epsilon_u^n = \delta_t e_u^n, \quad \tilde{\epsilon}_u^n = \delta_t \tilde{e}_u^n, \quad \epsilon_b^n = \delta_t e_b^n, \quad \psi^n = \delta_t q^n, \quad n \geq 1. \tag{2.63}$$

Applying δ_t operator to the equation (2.41), (2.42), (2.43) for $n \geq 1$ and using the following identity

$$\begin{aligned} \delta_t(a^n b^n) &= (\delta_t a^n) b^n + a^{n-1} \delta_t b^n && (2.64) \\ &= a^n (\delta_t b^n) + (\delta_t a^n) b^{n-1}, && (2.65) \end{aligned}$$

we have

$$\begin{cases} \dfrac{\tilde{\epsilon}_u^{n+1} - \epsilon_u^n}{\delta t} - \nu \Delta \tilde{\epsilon}_u^{n+1} + \nabla \psi^n = \delta_t R_u^{n+1} - R_{u,u}^{n+1} + \alpha R_{b,b}^{n+1}, & (2.66) \\ \dfrac{\epsilon_b^{n+1} - \epsilon_b^n}{\delta t} - \eta \Delta \epsilon_b^{n+1} = \delta_t R_b^{n+1} - R_{u,b}^{n+1}, & (2.67) \\ \dfrac{\epsilon_u^{n+1} - \tilde{\epsilon}_u^{n+1}}{\delta t} + \nabla(\psi^{n+1} - \psi^n) = \delta_t R_p^{n+1}, & (2.68) \end{cases}$$

where

$$\begin{aligned} R_{u,u}^{n+1} =& \delta_t((u(t^n) \cdot \nabla) u(t^{n+1}) - (u^n \cdot \nabla) \tilde{u}^{n+1}) \\ =& \delta_t((e_u^n \cdot \nabla) u(t^{n+1}) + (u^n \cdot \nabla) \tilde{e}_u^{n+1}) \\ =& e_u^n \cdot \nabla \delta_t u(t^{n+1}) + \epsilon_u^n \cdot \nabla u(t^n) + u^n \cdot \nabla \tilde{\epsilon}_u^{n+1} + \delta_t u^n \cdot \nabla \tilde{e}_u^n, \\ R_{b,b}^{n+1} =& \delta_t((\nabla \times b(t^{n+1})) \times b(t^n) - (\nabla \times b^{n+1}) \times b^n) \\ =& \delta_t((\nabla \times b(t^{n+1})) \times e_b^n + (\nabla \times e_b^{n+1}) \times b^n) \\ =& (\nabla \times \delta b(t^{n+1})) \times e_b^n + (\nabla \times b(t^n)) \times \epsilon_b^n + (\nabla \times \epsilon_b^{n+1}) \times b^n + (\nabla \times e_b^n) \times \delta b^n, \\ R_{u,b}^{n+1} =& \delta_t(\nabla \times (b(t^n) \times u(t^{n+1})) - \nabla \times (b^n \times \tilde{u}^{n+1})) \\ =& \delta_t(\nabla \times (e_b^n \times u(t^{n+1})) + \nabla \times (b^n \times \tilde{e}_u^{n+1})) \\ =& \nabla \times (\epsilon_b^n \times u(t^{n+1})) + \nabla \times (e_b^{n-1} \times \delta u(t^{n+1})) \\ & + \nabla \times (b^n \times \tilde{\epsilon}_u^{n+1}) + \nabla \times (\delta b^n \times \tilde{e}_u^n). \end{aligned}$$

To find out the estimation of ϵ_u^1 and ϵ_b^1, we let $n = 0$ in (2.41) and (2.42) and we obtain

$$\begin{cases} \dfrac{\tilde{e}_u^1}{\delta t} + (u(t^0)\cdot\nabla)\tilde{e}_u^1 - \nu\Delta\tilde{e}_u^1 - \alpha(\nabla\times e_b^1)\times b(t^0) = R_u^1, & (2.69)\\ \dfrac{e_b^1}{\delta t} - \eta\Delta e_b^1 + \nabla\times(b(t^0)\times\tilde{e}_u^1) = R_b^1, & (2.70)\\ \dfrac{1}{\delta t}(e_u^1 - \tilde{e}_u^1) + \nabla q^1 = R_p^1. & (2.71) \end{cases}$$

Taking the inner product (2.69) with $\delta t\tilde{e}_u^1$ and (2.70) with $\alpha\delta t e_b^1$ and adding them together, we have

$$\begin{aligned} &\|\tilde{e}_u^1\|^2 + \nu\delta t\|\nabla\tilde{e}_u^1\|^2 + \alpha\|e_b^1\|^2 + \alpha\eta\delta t\|\nabla e_b^1\|^2 - \alpha\delta t((\nabla\times e_b^1)\times b(t^0), \tilde{e}_u^1)\\ &\quad + \alpha\delta t((\nabla\times(b(t^0)\times\tilde{e}_u^1), e_b^1) = \delta t(R_u^1, \tilde{e}_u^1) + \alpha\delta t(R_b^1, e_b^1). \end{aligned} \tag{2.72}$$

Using integration by parts on curl, we have

$$(\nabla\times(b(t^0)\times\tilde{e}_u^1), e_b^1) = (b(t^0)\times\tilde{e}_u^1, \nabla\times e_b^1) = ((\nabla\times e_b^1)\times b(t^0), \tilde{e}_u^1). \tag{2.73}$$

Using (2.73), (2.72) becomes

$$\|\tilde{e}_u^1\|^2 + \nu\delta t\|\nabla\tilde{e}_u^1\|^2 + \alpha\|e_b^1\|^2 + \alpha\eta\delta t\|\nabla e_b^1\|^2 = \delta t(R_u^1, \tilde{e}_u^1) + \alpha\delta t(R_b^1, e_b^1). \tag{2.74}$$

RHS of (2.74) can be bounded as follows:

$$\begin{aligned} \delta t(R_u^1, \tilde{e}_u^1) + \alpha\delta t(R_b^1, e_b^1) &\le \delta t\|R_u^1\|\|\tilde{e}_u^1\| + \alpha\delta t\|R_b^1\|\|e_b^1\|\\ &\le c\delta t^2\|R_u^1\|^2 + \frac{1}{2}\|\tilde{e}_u^1\|^2 + c\delta t^2\|R_b^1\|^2 + \frac{\alpha}{2}\|e_b^1\|^2. \end{aligned} \tag{2.75}$$

From (2.74) and (2.75), we obtain

$$\|\tilde{e}_u^1\|^2 + \|e_b^1\|^2 + \delta t\|\nabla\tilde{e}_u^1\|^2 \le c\delta t^4. \tag{2.76}$$

Using $P_H(\tilde{e}_u^1) = e_u^1$ to (2.76), we have

$$\|e_u^1\|^2 + \|\tilde{e}_u^1\|^2 + \|e_b^1\|^2 + \delta t\|\nabla\tilde{e}_u^1\|^2 \le c\delta t^4. \tag{2.77}$$

From (2.71) we have

$$\begin{aligned}\|\nabla q^1\| &= \|R_p^1 + \frac{e_u^1 - \tilde{e}_u^1}{\delta t}\| \\ &\le \|R_p^1\| + \frac{1}{\delta t}(\|e_u^1\| + \|\tilde{e}_u^1\|) \\ &\le c\delta t.\end{aligned} \tag{2.78}$$

Using (2.78) to (2.77), we obtain

$$\|\tilde{e}_u^1\|^2 + \|e_b^1\|^2 + \delta t^2\|\nabla q^1\|^2 \le c\delta t^4. \tag{2.79}$$

This proves the lemma for $n = 1$ case. We can prove the lemma for $n \ge 2$ by taking the inner product of (2.66) with $2\delta t\tilde{\epsilon}_u^{n+1}$ and (2.67) with $2\alpha\delta t\tilde{\epsilon}_b^{n+1}$, the sum of which is

$$\begin{aligned}&\|\tilde{\epsilon}_u^{n+1}\|^2 - \|\epsilon_u^n\|^2 + \|\tilde{\epsilon}_u^{n+1} - \epsilon_u^n\|^2 + 2\delta t\|\nabla\tilde{\epsilon}^{n+1}\|^2 + 2\delta t(\nabla\psi^n, \tilde{\epsilon}_u^{n+1}) \\ &\quad + \alpha\|\epsilon_b^{n+1}\|^2 - \alpha\|\epsilon_b^n\|^2 + \alpha\|\epsilon_b^{n+1} - \epsilon_b^n\|^2 + 2\alpha\eta\delta t\|\nabla\epsilon_b^{n+1}\|^2 \\ &\quad = 2\delta t(\delta_t R_u^{n+1}, \tilde{\epsilon}_u^{n+1}) - 2\delta t(R_{u,u}^{n+1}, \tilde{\epsilon}_u^{n+1}) + 2\alpha\delta t(R_{b,b}^{n+1}, \tilde{\epsilon}_u^{n+1}) - 2\alpha\delta t(R_{u,b}^{n+1}, \epsilon_b^{n+1}).\end{aligned} \tag{2.80}$$

Each term of RHS can be bounded as follows. Using the inequality (2.7) and Poincare inequality and Lemma 1, we obtain

$$\begin{aligned}&|2\delta t(R_{u,u}^{n+1}, \tilde{\epsilon}_u^{n+1})| \\ &= \|2\delta t(\delta_t u^n \cdot \nabla\tilde{e}_u^n + e_u^n \cdot \nabla\delta_t u(t^{n+1}) + \epsilon_u^n \cdot \nabla u(t^n), \tilde{\epsilon}_u^{n+1})\| \\ &\le c\delta t(\|\delta_t u^n\|_1\|\nabla\tilde{e}_u^n\|\|\tilde{\epsilon}_u^{n+1}\|_1 + \|e_u^n\|\|\nabla\delta_t u(t^{n+1})\|_{L^\infty}\|\tilde{\epsilon}_u^{n+1}\| + \|\epsilon_u^n\|\|\nabla u(t^n)\|_2\|\tilde{\epsilon}_u^{n+1}\|_1) \\ &\le c\delta t((\delta t + \|\tilde{\epsilon}_u^n\|_1)\|\nabla\tilde{e}_u^n\|\|\tilde{\epsilon}_u^{n+1}\|_1 + \delta t^2\|\nabla\tilde{\epsilon}_u^{n+1}\| + \|\epsilon_u^n\|\|\tilde{\epsilon}_u^{n+1}\|_1) \\ &\le c\delta t(\delta t\|\nabla\tilde{e}_u^n\|\|\nabla\tilde{\epsilon}_u^{n+1}\| + \delta t^{1/2}\|\nabla\tilde{\epsilon}_u^n\|\|\nabla\tilde{\epsilon}_u^{n+1}\| + \delta t^2\|\nabla\tilde{\epsilon}_u^{n+1}\| + \|\epsilon_u^n\|\|\tilde{\epsilon}_u^{n+2}\|_1) \\ &\le c\delta t^{3/2}\|\nabla\tilde{e}_u^n\|\delta t^{1/2}\|\nabla\tilde{\epsilon}_u^{n+1}\| + \delta t^{3/2}\|\nabla\tilde{\epsilon}_u^n\|\|\nabla\tilde{\epsilon}_u^{n+1}\| + c\delta t^{5/2}\delta t^{1/2}\|\nabla\tilde{\epsilon}_u^{n+1}\| \\ &\quad + c\delta t\|\epsilon_u^n\|\|\nabla\tilde{\epsilon}_u^{n+1}\| \\ &\le c\delta t^5 + c\delta t^3\|\nabla\tilde{e}_u^n\|^2 + \frac{\nu}{16}\delta t\|\nabla\tilde{\epsilon}_u^{n+1}\|^2 + \frac{\nu}{16}\delta t\|\nabla\tilde{\epsilon}_u^n\|^2 + c\delta t\|\epsilon_u^n\|^2,\end{aligned} \tag{2.81}$$

where we also used the fact that $e_u^n = P_H(\tilde{e}_u^n)$ and the projection P_H is stable in H^1. Using the similar argument as (2.73), we have

$$\begin{aligned}(R_{u,b}^{n+1}, \epsilon_b^{n+1}) =&(\epsilon_b^n \times u(t^{n+1}), \nabla \times \epsilon_b^{n+1}) + ((\nabla \times \epsilon_b^{n+1}) \times e_b^{n-1}, \delta u(t^{n+1}))\\ &+ ((\nabla \times \epsilon_b^{n+1}) \times b^n, \tilde{\epsilon}_u^{n+1}) + ((\nabla \times \epsilon_b^{n+1}) \times \delta b^n, \tilde{e}_u^n).\end{aligned} \tag{2.82}$$

Using (2.82), the sum of the third and fourth terms of RHS of (2.80) is as follows:

$$\begin{aligned}&2\alpha\delta t((R_{b,b}^{n+1}, \tilde{\epsilon}_u^{n+1}) - 2\alpha\delta t(R_{u,b}^{n+1}, \epsilon_b^{n+1})\\ &\quad=2\alpha\delta t(((\nabla \times \delta b(t^{n+1})) \times e_b^n, \tilde{\epsilon}_u^{n+1}) + ((\nabla \times b(t^n)) \times \epsilon_b^n, \tilde{\epsilon}_u^{n+1})\\ &\quad+ (\nabla \times e_b^n) \times \delta b^n, \tilde{\epsilon}_u^{n+1}) - (\epsilon_b^n \times u(t^{n+1}), \nabla \times \epsilon_b^{n+1})\\ &\quad- ((\nabla \times \epsilon_b^{n+1}) \times e_b^{n-1}, \delta u(t^{n+1}))\\ &\quad- ((\nabla \times \epsilon_b^{n+1}) \times \delta b^n, \tilde{e}_u^n)).\end{aligned} \tag{2.83}$$

Each term of (2.83) can be bounded as follows:

$$\begin{aligned}|\delta t((\nabla \times \delta b(t^{n+1})) \times e_b^n, \tilde{\epsilon}_u^{n+1})| &\le c\delta t|((\nabla \times \delta b(t^{n+1})) \times e_b^n, \tilde{\epsilon}_u^{n+1})|\\ &\le c\delta t\|\nabla \times \delta b(t^{n+1})\|_{L^\infty}\|e_b^n\|\|\tilde{\epsilon}_u^{n+1}\|\\ &\le c\delta t^3\|\nabla\tilde{\epsilon}_u^{n+1}\|\\ &\le \frac{\eta}{16}\delta t\|\nabla\tilde{\epsilon}_u^{n+1}\|^2 + c\delta t^5,\end{aligned} \tag{2.84}$$

$$\begin{aligned}|\delta t((\nabla \times b(t^n)) \times \epsilon_b^n, \tilde{\epsilon}_u^{n+1})| &\le c\delta t\|\nabla \times b(t^n)\|_{L^\infty}\|\epsilon_b^n\|\|\tilde{\epsilon}_u^{n+1}\|\\ &\le c\delta t\|\epsilon_b^n\|^2 + \frac{\nu}{16}\delta t\|\nabla\tilde{\epsilon}_u^{n+1}\|^2,\end{aligned} \tag{2.85}$$

$$\begin{aligned}&|\delta t((\nabla \times e_b^n) \times \delta b^n, \tilde{\epsilon}_u^{n+1})|\\ &\quad\le c\delta t|((\nabla \times e_b^n) \times \epsilon_b^n, \tilde{\epsilon}_u^{n+1})| + c\delta t|((\nabla \times e_b^n) \times \delta b(t^n), \tilde{\epsilon}_u^{n+1})|\\ &\quad\le c\delta t\|\nabla e_b^n\|\|\nabla\epsilon_b^n\|\|\nabla\tilde{\epsilon}_u^{n+1}\| + c\delta t\|\nabla e_b^n\|\|\delta b(t^n)\|_{L^\infty}\|\tilde{\epsilon}_u^{n+1}\|\\ &\quad\le c\delta t\|\nabla e_b^n\|\|\nabla\epsilon_b^n\|\|\nabla\tilde{\epsilon}_u^{n+1}\| + c\delta t^2\|\nabla e_b^n\|\|\nabla\tilde{\epsilon}_u^{n+1}\|\\ &\quad\le c\delta t^{3/2}\|\nabla\epsilon_b^n\|\|\nabla\tilde{\epsilon}_u^{n+1}\| + c\delta t^3\|\nabla e_b^n\|^2 + c\frac{\nu}{16}\delta t\|\nabla\tilde{\epsilon}_u^{n+1}\|^2\\ &\quad\le \frac{\eta\alpha}{16}\delta t\|\nabla\epsilon_b^n\|^2 + \frac{\nu}{16}\delta t\|\nabla\tilde{\epsilon}_u^{n+1}\|^2 + c\delta t^3\|\nabla e_b^n\|^2 + \frac{\nu}{16}\delta t\|\nabla\tilde{\epsilon}_u^{n+1}\|^2,\end{aligned} \tag{2.86}$$

$$|\delta t(\epsilon_b^n \times u(t^{n+1}), \nabla \times \epsilon_b^{n+1})| \leq c\|\epsilon_b^n\|\|u(t^{n+1})\|_{L^\infty}\|\nabla \times \epsilon_b^{n+1}\|$$
$$\leq c\delta t\|\epsilon_b^n\|\|\nabla\epsilon_b^{n+1}\| \tag{2.87}$$
$$\leq c\delta t\|\epsilon_b^n\|^2 + \frac{\eta\alpha}{16}\delta t\|\nabla\epsilon_b^{n+1}\|^2,$$

$$|\delta t((\nabla \times \epsilon_b^{n+1}) \times e_b^{n-1}, \delta u(t^{n+1}))| \leq \delta t\|\nabla\epsilon_b^{n+1}\|\|e_b^{n-1}\|\|\delta u(t^{n+1})\|_{L^\infty}$$
$$\leq \delta t^3\|\nabla\epsilon_b^{n+1}\| \tag{2.88}$$
$$\leq \frac{\eta\alpha}{16}\delta t\|\nabla\epsilon_b^{n+1}\|^2 + c\delta t^5,$$

$$|\delta t((\nabla \times \epsilon_b^{n+1}) \times \delta b^n, \tilde{e}_u^n)|$$
$$\leq c\delta t|((\nabla \times \epsilon_b^{n+1}) \times \epsilon_b^n, \tilde{e}_u^n| + c\delta t|((\nabla \times \epsilon_b^{n+1}) \times \delta b(t^n), \tilde{e}_u^n)|$$
$$\leq c\delta t\|\nabla\epsilon_b^{n+1}\|\|\nabla\epsilon_b^n\|\|\nabla\tilde{e}_u^n\| + c\delta t\|\nabla\epsilon_b^{n+1}\|\|\nabla\delta b(t^n)\|\|\nabla\tilde{e}_u^n\|$$
$$\leq c\delta t^{3/2}\|\nabla\epsilon_b^{n+1}\|\|\nabla\epsilon_b^n\| + c\delta t^2\|\nabla\epsilon_b^{n+1}\|\|\nabla\tilde{e}_u^n\|$$
$$\leq \frac{\eta\alpha}{16}\delta t\|\nabla\epsilon_b^{n+1}\|^2 + \frac{\eta\alpha}{16}\delta t\|\nabla\epsilon_b^n\|^2 + \frac{\eta\alpha}{16}\delta t\|\nabla\epsilon_b^{n+1}\|^2 + c\delta t^3\|\nabla\tilde{e}_u^n\|^2.$$

It remains to control $2\delta t(\nabla\psi^n, \tilde{\epsilon}_u^{n+1})$ on LHS of (2.80), and we have the following result analogous to Lemma 1:

$$\begin{aligned} 2\delta t(\nabla\psi^n, \tilde{\epsilon}_u^{n+1}) =& \delta t^2(\|\nabla\psi^{n+1}\|^2 - \|\nabla\psi^n\|^2) + (\|\epsilon_u^{n+1}\|^2 - \|\tilde{\epsilon}_u^{n+1}\|^2) \\ &+ 2\delta t^2(\delta_t R_p^{n+1}, \nabla\psi^n) - \delta t(\delta_t R_p^{n+1}, \tilde{\epsilon}_u^{n+1}), \end{aligned} \tag{2.89}$$

where the last two terms can be easily bounded by the Cauchy inequality.

Using the above inequalities and (2.89), (2.80) becomes

$$\begin{aligned} &\|\epsilon_u^{n+1}\|^2 - \|\epsilon_u^n\|^2 + \|\tilde{\epsilon}_u^{n+1} - \epsilon_u^n\|^2 + \nu\delta t\|\nabla\tilde{\epsilon}_u^{n+1}\| + \delta t^2(\|\nabla\psi^{n+1}\|^2 - \|\nabla\psi^n\|^2) \\ &+ \alpha(\|\epsilon_b^{n+1}\|^2 - \|\epsilon_b^n\|^2 + \|\epsilon_b^{n+1} - \epsilon_b^n\|^2 + \eta\delta t\|\nabla\epsilon_b^{n+1}\|^2) \\ &\leq c\delta t^5 + c\delta t^3\|\nabla\tilde{e}_u^n\|^2 + c\delta t^3\|\nabla e_b^n\|^2 + c\delta t\|\epsilon_u^n\|^2 + c\delta t\|\epsilon_b^n\|^2 + c\delta t^3\|\nabla\psi^n\|^2. \end{aligned} \tag{2.90}$$

Taking a sum of (2.90) for $n = 1, \ldots, m$, we have the following inequality:

$$\begin{aligned}
&\|\epsilon_u^{m+1}\|^2 + \alpha\|\epsilon_b^{m+1}\|^2 + \delta t^2\|\nabla\psi^{m+1}\|^2 + \sum_{k=1}^{m}(\|\tilde{\epsilon}_u^{k+1} - \epsilon_u^k\|^2 + \alpha\|\epsilon_b^{k+1} - \epsilon_b^k\|^2) \\
&+ \delta t \sum_{k=1}^{m}(\|\nabla\tilde{\epsilon}_u^{k+1}\|^2 + \alpha\eta\|\epsilon_b^{k+1}\|^2) \\
&\le c\delta t^4 + c\delta t \sum_{k=1}^{m}(\|\epsilon_u^k\|^2 + \|\epsilon_b^k\|^2 + \delta t^2\|\nabla\psi^k\|^2) + \|\epsilon_u^1\|^2 + \alpha\|\epsilon_b^1\|^2 + \delta t^2\|\nabla\psi^1\|^2 \\
&\le c\delta t^4 + c\delta t \sum_{k=1}^{m}(\|\epsilon_u^k\|^2 + \|\epsilon_b^k\|^2 + \delta t^2\|\nabla\psi^k\|^2).
\end{aligned} \tag{2.91}$$

Applying Grownwall's inequality to (2.91) , we obtain the desired result. ■

Proof [**Proof of Theorem 2.3.2**] Adding (2.41) and (2.43), we get

$$\begin{aligned}
-\nu\Delta\tilde{e}_u^{n+1} + \nabla q^{n+1} &= h^{n+1}, \\
\nabla \cdot \tilde{e}^{n+1} = g^{n+1}, &\quad \tilde{e}_u^{n+1}|_{\partial\Omega} = 0,
\end{aligned} \tag{2.92}$$

where

$$\begin{aligned}
h^{n+1} =& \tilde{h}^{n+1} - \frac{e_u^{n+1} - e_u^n}{\delta t}, \\
\tilde{h}^{n+1} =& R_u^n + R_p^n - ((u(t^n)\cdot\nabla)u(t^{n+1}) - (u^n\cdot\nabla)\tilde{u}^{n+1}) + \alpha(\nabla\times b(t^{n+1}))\times e_b^n \\
&+ \alpha(\nabla\times e_b^{n+1})\times b^n, \\
g^{n+1} =& \delta t\Delta(p^{n+1} - p^n).
\end{aligned} \tag{2.93}$$

Using the similar arguments in Lemma 1, we find

$$\|g^{n+1}\| = \|\nabla\cdot\tilde{e}_u^{n+1}\| \le \|\nabla\tilde{e}_u^{n+1}\| \lesssim \delta t^{1/2}, \quad \|\tilde{h}^{n+1}\|_{-1} \lesssim \delta t^{1/2}. \tag{2.94}$$

Therefore, we have

$$\|h^{n+1}\|_{-1} \le \|\tilde{h}^{n+1}\|_{-1} + \|\frac{e_u^{n+1} - e_u^n}{\delta t}\|_{-1} \tag{2.95}$$

and it follows that

$$\|h_{\delta t}\|_{l^2(H^{-1}(\Omega))^d} \lesssim \|\tilde{h}_{\delta t}\|_{l^2(H^{-1}(\Omega)^d)} + \frac{1}{\delta t}\|(\delta_t e_u)_{\delta t}\|_{l^2(L^2(\Omega)^d)} \lesssim \delta t. \tag{2.96}$$

Applying the standard stability results for inhomogeneous Stokes system [20] to (2.92), it turns out

$$\|\tilde{e}_u^{n+1}\|_1 + \|q^{n+1}\| \lesssim \|h^{n+1}\|_{-1} + \|g^{n+1}\|, \tag{2.97}$$

and we obtain

$$\|q_{\delta t}\|_{l^2(L^2(\Omega))} \lesssim \delta t. \tag{2.98}$$

The proof is complete. ■

2.4 Other schemes with improved accuracy

It is known that the pressure-correction scheme suffers from a numerical boundary layer and that rotational form could reduce the effect of the boundary layer for NSE. In this section we develop a rotational pressure-correction scheme for the MHD equation and perform a stability analysis. We also introduce a second-order numerical scheme which is based on pressure-correction and rotational pressure-correction scheme.

2.4.1 First-order rotational scheme and stability

The following is the rotational pressure-correction discretization of the MHD equation. Given $(u^n, b^n, \nabla p^n)$, we find $(\tilde{u}^{n+1}, b^{n+1})$ by solving (2.99) and (2.100) together. And we obtain $(u^{n+1}, \nabla p^{n+1})$ solving (2.101) where

$$\begin{cases} \dfrac{\tilde{u}^{n+1} - u^n}{\delta t} + (u^n \cdot \nabla)\tilde{u}^{n+1} - \nu\Delta\tilde{u}^{n+1} + \nabla p^n - \alpha(\nabla \times b^{n+1}) \times b^n = 0, & (2.99) \\ \dfrac{b^{n+1} - b^n}{\delta t} - \eta\Delta b^{n+1} + \nabla \times (b^n \times \tilde{u}^{n+1}) = 0, & (2.100) \\ \tilde{u}^{n+1}|_{\partial\Omega} = 0, \\ b^{n+1} \cdot n|_{\partial\Omega} = 0, n \times (\nabla \times b^{n+1})|_{\partial\Omega} = 0, \end{cases}$$

and

$$\begin{cases} \dfrac{u^{n+1} - \tilde{u}^{n+1}}{\delta t} + \nabla(p^{n+1} - p^n + \nu\nabla \cdot \tilde{u}^{n+1}) = 0, \\ u^{n+1} \cdot n|_{\partial\Omega} = 0, \mathrm{div} u^{n+1} = 0. \end{cases} \tag{2.101}$$

The scheme (2.99) - (2.101) is stable in the following sense.

Theorem 2.4.1 *The scheme* (2.99)-(2.101) *is unconditionally energy stable in the following sense:*

$$\begin{aligned}&\|u^{n+1}\|^2+\alpha\|b^{n+1}\|^2+\delta t^2\|\nabla\psi^{n+1}\|^2+\frac{\delta t}{\nu}\|q^{n+1}\|^2\\&\le\|u^n\|^2+\alpha\|b^n\|^2+\delta t^2\|\nabla\psi^n\|^2+\frac{\delta t}{\nu}\|q^n\|^2,\quad n\ge 0,\end{aligned}\tag{2.102}$$

where

$$\begin{cases}q^{n+1}=q^n-\nu\nabla\cdot\tilde{u}^{n+1},\\ p^{n+1}=\psi^{n+1}+q^{n+1},\end{cases}\tag{2.103}$$

for $n\ge 0$ *and* $q^0=0$ *and* $\psi^0=p^0$.

Proof Using $\{\psi^n\}$ and $\{q^n\}$, we can rewrite (2.101) as the following two steps:

$$\begin{cases}\dfrac{u^{n+1}-\tilde{u}^{n+1}}{\delta t}+\nabla(\psi^{n+1}-\psi^n)=0,\\ u^{n+1}\cdot n|_{\partial\Omega}=0,\mathrm{div}u^{n+1}=0,\end{cases}\tag{2.104}$$

and

$$\begin{cases}q^{n+1}=q^n-\nu\nabla\cdot\tilde{u}^{n+1},\\ p^{n+1}=\psi^{n+1}+q^{n+1}.\end{cases}\tag{2.105}$$

We can check that (2.104) and (2.105) are equivalent to (2.101) because

$$p^{n+1}-p^n+\nu\nabla\cdot\tilde{u}^{n+1}=p^{n+1}-p^n-(q^{n+1}-q^n)=\psi^{n+1}-\psi^n.\tag{2.106}$$

Hence we are going to prove the stability of the scheme (2.99), (2.100), (2.104) and (2.105) with the initial condition u^0,b^0,p^0 and $q^0=0,\psi^0=p^0$. Taking the inner product of (2.99) with $2\delta t\tilde{u}^{n+1}$ and (2.100) with $2\alpha\delta t b^{n+1}$ and adding them together, we obtain

$$\begin{aligned}&\|\tilde{u}^{n+1}\|^2-\|u^n\|^2+\|\tilde{u}^{n+1}-u^n\|^2+2\delta t\nu\|\nabla\tilde{u}^{n+1}\|^2+2\delta t(\nabla(\psi^n+q^n),\tilde{u}^{n+1})\\&+\alpha\|b^{n+1}\|^2-\alpha\|b^n\|^2+\alpha\|b^{n+1}-b^n\|^2+2\delta t\alpha\eta\|\nabla b^{n+1}\|^2=0,\end{aligned}\tag{2.107}$$

where curl terms are canceled as in Theorem 2.3.1. Taking the inner product (2.101) with $2\delta t u^{n+1}$, we obtain

$$\|u^{n+1}\|^2-\|\tilde{u}^{n+1}\|^2+\|u^{n+1}-\tilde{u}^{n+1}\|^2=0.\tag{2.108}$$

We can do the estimation of pressure as follows:

$$
\begin{aligned}
&2\delta t(\nabla\psi^n + \nabla q^n, \tilde{u}^{n+1}) \\
&=2\delta t(\nabla\psi^n, \tilde{u}^{n+1}) + 2\delta t(\nabla q^n, \tilde{u}^{n+1}) \\
&=2\delta t(\nabla\psi^n, u^{n+1} + \delta t(\nabla\psi^{n+1} - \nabla\psi^n)) - 2\delta t(q^n, \nabla\cdot\tilde{u}^{n+1}) \\
&=\delta t^2(\|\nabla\psi^{n+1}\|^2 - \|\nabla\psi^n\|^2 - \|\nabla\psi^{n+1} - \nabla\psi^n\|^2) + \frac{2\delta t}{\nu}(q^n, q^{n+1} - q^n) \\
&=\delta t^2(\|\nabla\psi^{n+1}\|^2 - \|\nabla\psi^n\|^2 - \frac{1}{\delta t^2}\|u^{n+1} - \tilde{u}^{n+1}\|^2) \\
&\quad + \frac{\delta t}{\nu}(\|q^{n+1}\|^2 - \|q^n\|^2 - \|q^{n+1} - q^n\|^2) \\
&=\delta t^2(\|\nabla\psi^{n+1}\|^2 - \|\nabla\psi^n\|^2) - \|u^{n+1} - \tilde{u}^{n+1}\|^2 \\
&\quad + \frac{\delta t}{\nu}(\|q^{n+1}\|^2 - \|q^n\|^2 - \nu^2\|\nabla\cdot\tilde{u}^{n+1}\|^2).
\end{aligned} \tag{2.109}
$$

We can add (2.107) and (2.108) and substitute (2.109). Then, using the identity $\|\nabla u\|^2 = \|\nabla\times u\|^2 + \|\nabla\cdot u\|^2$, we obtain

$$
\begin{aligned}
&\|u^{n+1}\|^2 - \|u^n\|^2 + \|\tilde{u}^{n+1} - u^n\|^2 + 2\delta t\nu\|\nabla\times\tilde{u}^{n+1}\|^2 + \nu\delta t\|\nabla\cdot\tilde{u}^{n+1}\|^2 \\
&+ \delta t^2(\|\nabla\psi^{n+1}\|^2 - \|\nabla\psi^n\|^2) + \frac{\delta t}{\nu}(\|q^{n+1}\|^2 - \|q^n\|^2) \\
&+ \alpha\|b^{n+1}\|^2 - \alpha\|b^n\|^2 + \alpha\|b^{n+1} - b^n\|^2 + 2\delta t\alpha\eta\|\nabla b^{n+1}\|^2 = 0,
\end{aligned} \tag{2.110}
$$

which implies the desired result. ■

2.4.2 Second-order standard scheme

We consider the following second-order scheme for solving the system (2.1a)-(2.1f). First we find $(u^1, b^1, \nabla p^1)$ using (2.26)-(2.28). For $n \geq 1$ we do the following.: given $(u^n, u^{n-1}, b^n, b^{n-1}, \nabla p^n)$, find $(\tilde{u}^{n+1}, b^{n+1})$.

$$
\begin{cases}
\dfrac{3\tilde{u}^{n+1} - 4u^n + u^{n-1}}{2\delta t} + ((2u^n - u^{n-1})\cdot\nabla)\tilde{u}^{n+1} - \nu\Delta\tilde{u}^{n+1} \\
\quad +\nabla p^n - \alpha(\nabla\times b^{n+1})\times(2b^n - b^{n-1}) = 0, & (2.111) \\
\dfrac{3b^{n+1} - 4b^n + b^{n-1}}{2\delta t} - \eta\Delta b^{n+1} + \nabla\times((2b^n - b^{n-1})\times\tilde{u}^{n+1}) = 0, & (2.112) \\
\tilde{u}^{n+1}|_{\partial\Omega} = 0, & (2.113) \\
b^{n+1}\cdot n|_{\partial\Omega} = 0, n\times(\nabla\times b^{n+1})|_{\partial\Omega} = 0. & (2.114)
\end{cases}
$$

After $(\tilde{u}^{n+1}, b^{n+1})$ is obtained, we update the pressure and velocity field via Helmholtz decomposition.

Given $(\tilde{u}^{n+1}, \nabla p^n)$, find $(u^{n+1}, \nabla p^{n+1})$ by solving

$$\begin{cases} \dfrac{3u^{n+1} - 3\tilde{u}^{n+1}}{2\delta t} + \nabla(p^{n+1} - p^n) = 0, \\ u^{n+1} \cdot n|_{\partial\Omega} = 0, \nabla \cdot u^{n+1} = 0. \end{cases} \tag{2.115}$$

Then we can prove the following theorem.

Theorem 2.4.2 *The scheme* (2.111)-(2.115) *is unconditionally energy stable in the following sense: for any* δt *there is* $c > 0$ *such that*

$$\|u^n\|^2 + \alpha\|b^n\|^2 \le c(\|u^0\|^2 + \|b^0\|^2 + \|\nabla p^0\|^2), \quad \textit{for } n \ge 1. \tag{2.116}$$

Proof Taking the inner product of (2.111) with $4\delta t\tilde{u}^{n+1}$, we derive

$$\begin{aligned} (3\tilde{u}^{n+1} - 4u^n + u^{n-1}, 2\tilde{u}^{n+1}) + 4\delta t\nu\|\nabla\tilde{u}^{n+1}\|^2 + 4\delta t(\nabla p^n, \tilde{u}^{n+1}) \\ - 4\alpha\delta t((\nabla \times b^{n+1}) \times (2b^n - b^{n-1}), \tilde{u}^{n+1}) = 0, \end{aligned} \tag{2.117}$$

where we used the fact $((u \cdot \nabla)v, v) = 0$ for $u \in H$ and $v \in H^1(\Omega)^d$.

Taking the inner product of (2.112) with $4\delta t\alpha b^{n+1}$, we derive

$$\alpha(3b^{n+1} - 4b^n + b^{n-1}, 2b^{n+1}) + 4\eta\alpha\delta t\|\nabla b^{n+1}\|^2 + 4\alpha\delta t(\nabla \times ((2b^n - b^{n-1}) \times \tilde{u}^{n+1}), b^{n+1}) = 0. \tag{2.118}$$

Using integration by parts, we obtain

$$\begin{aligned} (\nabla \times ((2b^n - b^{n-1}) \times \tilde{u}^{n+1}), b^{n+1}) &= ((2b^n - b^{n-1}) \times \tilde{u}^{n+1}, \nabla \times b^{n+1}) \\ &= ((\nabla \times b^{n+1}) \times (2b^n - b^{n-1}), \tilde{u}^{n+1}). \end{aligned} \tag{2.119}$$

Taking the sum of (2.117) and (2.118) and using (2.119), we obtain

$$\begin{aligned} &(3\tilde{u}^{n+1} - 4u^n + u^{n-1}, 2\tilde{u}^{n+1}) + \alpha(3b^{n+1} - 4b^n + b^{n-1}, 2b^{n+1}) \\ &\qquad + 4\delta t\nu\|\nabla\tilde{u}^{n+1}\|^2 + 4\eta\alpha\delta t\|\nabla b^{n+1}\|^2 + 4\delta t(\nabla p^n, \tilde{u}^{n+1}) = 0. \end{aligned} \tag{2.120}$$

Let

$$\begin{aligned} I_1 &= (3\tilde{u}^{n+1} - 4u^n + u^{n-1}, 2\tilde{u}^{n+1}), \\ I_2 &= \alpha(3b^{n+1} - 4b^n + b^{n-1}, 2b^{n+1}), \\ I_3 &= 4\delta t(\nabla p^n, \tilde{u}^{n+1}). \end{aligned} \tag{2.121}$$

$$\begin{aligned} I_1 =&(3\tilde{u}^{n+1} - 4u^n + u^{n-1}, 2\tilde{u}^{n+1}) \\ =&(3u^{n+1} - 4u^n + u^{n-1}, 2\tilde{u}^{n+1}) - (3u^{n+1} - 3\tilde{u}^{n+1}, 2\tilde{u}^{n+1}) \\ =&(3u^{n+1} - 4u^n + u^{n-1}, 2u^{n+1}) - (3u^{n+1} - 4u^n + u^{n-1}, 2(u^{n+1} - \tilde{u}^{n+1})) \\ &- (3u^{n+1} - 3\tilde{u}^{n+1}, 2\tilde{u}^{n+1}) \\ =&I_{11} - I_{12} - I_{13}. \end{aligned} \tag{2.122}$$

We can simplify $I_{11}, I_{12}, I_{13}, I_2$ as follows:

$$\begin{aligned} I_{11} &= (3u^{n+1} - 4u^n + u^{n-1}, 2u^{n+1}) \\ &= (u^{n+1} - 2u^n + u^{n-1}, 2u^{n+1}) + (2u^{n+1} - 2u^n, 2u^{n+1}) \\ &= \|u^{n+1}\|^2 - \|2u^n - u^{n-1}\|^2 + \|\delta_t^2 u^{n+1}\|^2 + \|2u^{n+1} - u^n\|^2 - \|u^n\|^2, \end{aligned} \tag{2.123}$$

where we used two identities

$$(a - b, 2a) = |a|^2 - |b|^2 + |a - b|^2,$$

$$(2a - 2b, 2a) = |2a - b|^2 - |b|^2.$$

$$\begin{aligned} I_{12} &= (3u^{n+1} - 4u^n + u^{n-1}, 2(u^{n+1} - \tilde{u}^{n+1})) \\ &= -(3u^{n+1} - 4u^n + u^{n-1}, \frac{4\delta t}{3}\nabla(p^{n+1} - p^n) = 0. \end{aligned} \tag{2.124}$$

$$\begin{aligned} I_{13} &= (3u^{n+1} - 3\tilde{u}^{n+1}, 2\tilde{u}^{n+1}) \\ &= 3(\|u^{n+1}\|^2 - \|\tilde{u}^{n+1}\|^2 - \|u^{n+1} - \tilde{u}^{n+1}\|^2). \end{aligned} \tag{2.125}$$

$$\begin{aligned} I_2 &= \alpha(3b^{n+1} - 4b^n + b^{n-1}, 2b^{n+1}) \\ &= \alpha((b^{n+1} - 2b^n + b^{n-1}, 2b^{n+1}) + (2b^{n+1} - 2b^n, 2b^{n+1})) \\ &= \alpha(\|b^{n+1}\|^2 - \|2b^n - b^{n-1}\|^2 + \|\delta_t^2 b^{n+1}\|^2 + \|2b^{n+1} - b^n\|^2 - \|b^n\|^2). \end{aligned} \tag{2.126}$$

From the equation (2.115), we have

$$3u^{n+1} + 2\delta t \nabla p^{n+1} = 3\tilde{u}^{n+1} + 2\delta t \nabla p^n. \tag{2.127}$$

By taking the inner product to itself, we derive

$$\begin{aligned} 9\|u^{n+1}\|^2 + 4\delta t^2 \|\nabla p^{n+1}\|^2 &= 9\|\tilde{u}^{n+1}\|^2 + 12\delta t(\tilde{u}^{n+1}, \nabla p^n) + 4\delta t^2 \|\nabla p^n\|^2 \\ 3\|u^{n+1}\|^2 + \frac{4}{3}\delta t^2 \|\nabla p^{n+1}\|^2 &= 3\|\tilde{u}^{n+1}\|^2 + 4\delta t(\tilde{u}^{n+1}, \nabla p^n) + \frac{4\delta t^2}{3}\|\nabla p^n\|^2. \end{aligned} \tag{2.128}$$

Hence

$$\begin{aligned} I_3 &= 4\delta t(\tilde{u}^{n+1}, \nabla p^n) \\ &= 3\|u^{n+1}\| - 3\|\tilde{u}^{n+1}\| + \frac{4}{3}\delta t^2 (\|\nabla p^{n+1}\|^2 - \|\nabla p^n\|^2). \end{aligned} \tag{2.129}$$

By combining (2.122), (2.123), (2.124), (2.125), (2.126) and (2.129), we can rewrite (2.120) as follows:

$$\begin{aligned} &\|u^{n+1}\|^2 - \|u^n\|^2 + \|2u^{n+1} - u^n\|^2 - \|2u^n - u^{n-1}\|^2 + 4\delta t\nu \|\nabla \tilde{u}^{n+1}\|^2 \\ &\quad + 4\alpha\eta\delta t \|\nabla b^{n+1}\|^2 + \|\delta_t^2 u^{n+1}\|^2 + 3\|u^{n+1} - \tilde{u}^{n+1}\|^2 \\ &\quad + \alpha(\|b^{n+1}\|^2 - \|b^n\|^2 + \|2b^{n+1} - b^n\|^2 - \|2b^n - b^{n-1}\|^2 + \|\delta_t^2 b^n\|^2) \\ &\quad + \frac{4}{3}\delta t^2 (\|\nabla p^{n+1}\|^2 - \|\nabla p^n\|^2) = 0, \quad \text{for } n \geq 1. \end{aligned} \tag{2.130}$$

Taking the sum of (2.130) from $n = 1, \ldots, N$ for $N \geq 1$, we obtain

$$\begin{aligned} &\|u^{N+1}\|^2 - \|u^1\|^2 + \|2u^{N+1} - u^N\|^2 - \|2u^1 - u^0\|^2 \\ &+ \sum_{n=1}^{N} (4\delta t\nu \|\nabla \tilde{u}^{n+1}\|^2 + 4\alpha\eta\delta t \|b^{n+1}\|^2 + \|\delta_t^2 u^{n+1}\|^2 + 3\|u^{n+1} - \tilde{u}^{n+1}\|^2 + \|\delta_t^2 b^n\|^2) \\ &+ \alpha(\|b^{N+1}\|^2 - \|b^1\|^2 + \|2b^{N+1} - b^N\|^2 - \|2b^1 - b^0\|^2) \\ &+ \frac{4}{3}\delta t^2 (\|\nabla p^{N+1}\|^2 - \|\nabla p^1\|^2) = 0. \end{aligned} \tag{2.131}$$

(2.131) can be bounded as follows:

$$\begin{aligned} &\|u^{N+1}\|^2 + \alpha \|b^{N+1}\|^2 \\ &\leq \|u^1\|^2 + \|2u^1 - u^0\|^2 + \alpha \|b^1\|^2 + \alpha \|2b^1 - b^0\|^2 + \frac{4}{3}\delta t^2 \|\nabla p^1\|^2. \end{aligned} \tag{2.132}$$

$n = 1$ case of the conclusion can be drawn from $n = 1$ case of (2.35). By applying (2.35) for $n = 0$ to (2.132), we can derive the rest of the conclusion. ■

2.4.3 Second-order rotational scheme

We consider the following second-order rotational form for solving the system (2.1a)-(2.1f). First we find $(u^1, b^1, \nabla p^1)$ using (2.99)-(2.101). For $n \geq 1$, we do the following. Given $(u^n, u^{n-1}, b^n, b^{n-1}, \nabla p^n)$, find $(\tilde{u}^{n+1}, b^{n+1})$ by solving

$$\begin{cases} \dfrac{3\tilde{u}^{n+1} - 4u^n + u^{n-1}}{2\delta t} + ((2u^n - u^{n-1}) \cdot \nabla)\tilde{u}^{n+1} - \nu\Delta\tilde{u}^{n+1} \\ \quad +\nabla p^n - \alpha(\nabla \times b^{n+1}) \times (2b^n - b^{n-1}) = 0, & (2.133) \\ \dfrac{3b^{n+1} - 4b^n + b^{n-1}}{2\delta t} - \eta\Delta b^{n+1} + \nabla \times ((2b^n - b^{n-1}) \times \tilde{u}^{n+1}) = 0, & (2.134) \\ \tilde{u}^{n+1}|_{\partial\Omega} = 0, & (2.135) \\ b^{n+1} \cdot n|_{\partial\Omega} = 0, n \times (\nabla \times b^{n+1})|_{\partial\Omega} = 0. & (2.136) \end{cases}$$

After $(\tilde{u}^{n+1}, b^{n+1})$ is obtained, we update the pressure and velocity field via the Helmholtz decomposition.

Given $(\tilde{u}^{n+1}, b^{n+1})$, find $(u^{n+1}, \nabla p^{n+1})$ by solving

$$\begin{cases} \dfrac{3u^{n+1} - 3\tilde{u}^{n+1}}{2\delta t} + \nabla(p^{n+1} - p^n + \nu\nabla \cdot \tilde{u}^{n+1}) = 0, \\ u^{n+1} \cdot n|_{\partial\Omega} = 0, \nabla \cdot u^{n+1} = 0. \end{cases} \tag{2.137}$$

We can prove the similar stability result.

Theorem 2.4.3 *The scheme* (2.133)-(2.137) *is unconditionally energy stable in the following sense: for any δt there is $c > 0$ such that*

$$\|u^n\|^2 + \alpha\|b^n\|^2 \leq c(\|u^0\|^2 + \|b^0\|^2 + \|\nabla p^0\|^2), \quad \textit{for } n \geq 1. \tag{2.138}$$

2.5 Legendre-Galerkin method for the MHD equation

In this section we develop the Legendre-Galerkin method for $\Omega = (-1, 1)^2$ to solve (2.26) - (2.28). The second-order scheme and rotational pressure-correction scheme can be solved similarly. To describe the Galerkin method, we need to define a few spaces.

$$\mathbf{P_N} = P_N \otimes P_N,$$

where P_N is the polynomial of the degree equal or less than N and

$$\mathbf{X_N} = \{u \in (\mathbf{P_N})^2 : u|_{\partial\Omega} = 0\},$$
$$\mathbf{Y_N} = \{b \in (\mathbf{P_N})^2 : n \cdot b|_{\partial\Omega} = 0 \text{ and } n \times (\nabla \times b)|_{\partial\Omega} = 0\},$$
$$\mathbf{Z_N} = \{p \in (\mathbf{P_{N-2}}) : \int_\Omega p dx = 0\}.$$

Solving the scheme (2.26) - (2.27) consists of two steps.

- First step

The Galerkin formulation for solving (2.26) and (2.27) consists of finding $\tilde{u}^{n+1} \in \mathbf{X_N}$ and $b^{n+1} \in \mathbf{Y_N}$ which statisfy

$$(\frac{1}{\delta t}\tilde{u}^{n+1}, v) + ((u^n \cdot \nabla)\tilde{u}^{n+1}, v) - \nu(\Delta\tilde{u}^{n+1}, v) - \alpha((\nabla \times b^{n+1}) \times b^n, v) = (I_N f, v), \tag{2.139}$$

$$(\frac{\alpha}{\delta t}b^{n+1}, w) - \alpha\eta(\Delta b^{n+1}, w) + \alpha(\nabla \times (b^n \times \tilde{u}^{n+1}), w) = 0, \tag{2.140}$$

for all $v \in \mathbf{X}_N$ and $w \in \mathbf{Y_N}$ where $(\cdot, \cdot)$ denotes the L^2 inner product on Ω.

- Second step

Given ∇p^n and $\tilde{u}^{n+1}$, find ∇p^{n+1} which satisfies

$$(\nabla p^{n+1}, \nabla q) = (\nabla p^n + \frac{1}{\delta t}\tilde{u}^{n+1}, \nabla q) \tag{2.141}$$

for all $q \in \mathbf{Z_N}$.

We can consider the following operator:

$$\mathcal{A} = \begin{pmatrix} \frac{1}{\delta t}(\cdot) + (u^n \cdot \nabla)(\cdot) - \nu\Delta(\cdot) & -\alpha((\nabla \times \cdot) \times b^n \\ \alpha(\nabla \times (b^n \times \cdot) & \frac{\alpha}{\delta t}(\cdot) - \alpha\eta\Delta(\cdot) \end{pmatrix}. \tag{2.142}$$

Then

$$((\begin{matrix} u \\ b \end{matrix}), \mathcal{A}(\begin{matrix} u \\ b \end{matrix})) = \frac{1}{\delta t}\|u\|^2 + \nu\|\nabla u\|^2 + \frac{\alpha}{\delta t}\|b\|^2 + \alpha\eta\|\nabla b\|^2 \geq 0, \tag{2.143}$$

and (2.143) is zero if and only if both u and v are equal to zero. Hence the system is positive-definite, and it can be solved efficiently using the iterative solver BiCGSTAB. A proper preconditioner is essential for the implementation of an iterative method. We use the following decoupled equation as a preconditioner:

$$\mathcal{P} = \begin{pmatrix} \frac{1}{\delta t}(\cdot) - \nu\Delta(\cdot) & 0 \\ 0 & \frac{\alpha}{\delta t}(\cdot) - \alpha\eta\Delta(\cdot). \end{pmatrix}. \tag{2.144}$$

In the following subsection, we describe how to solve the preconditioner (2.159) efficiently.

2.5.1 Discretization of preconditioner

To solve the preconditioner $\mathcal{P}$, we need to solve the following two equations:

$$\begin{cases} \dfrac{1}{\delta t}u - \nu\Delta u = f, \\ u|_{\partial\Omega} = 0, \end{cases} \tag{2.145}$$

and

$$\begin{cases} \dfrac{\alpha}{\delta t}b - \alpha\eta\Delta b = f, \\ n \cdot b|_{\partial\Omega} = 0 \text{ and } n \times (\nabla \times b)|_{\partial\Omega} = 0. \end{cases} \tag{2.146}$$

Equation (2.145) can be solved efficiently by the Legendre-Galerkin method introduced in [21]. In this section we discuss how to solve (2.146).

The first boundary condition $n \cdot b|_{\partial\Omega} = 0$ can be written as follows:

$$b_1(\pm 1, y) = 0 \text{ and } b_2(x, \pm 1) = 0, \tag{2.147}$$

for $x, y \in [-1, 1]$. Note that

$$n \times (\nabla \times b) = n \times (0, 0, \partial_x b_2 - \partial_y b_1). \tag{2.148}$$

If $x = \pm 1$ (i.e $n = (\pm 1, 0)$), we have

$$\partial_x b_2(\pm 1, y) - \partial_y b_1(\pm 1, y) = 0.$$

$b_1(\pm 1, y) = 0$ implies $\partial_y b_1(\pm 1, y) = 0$. Hence we have $\partial_x b_2(\pm 1, y) = 0$. Similarly we have $\partial_y b_1(x, \pm 1) = 0$. Hence the boundary condition for b_1 and b_2 is completely decoupled. For b_1 we have

$$b_1(\pm 1, y) = 0, \partial_y b_1(x, \pm 1) = 0, \tag{2.149}$$

and for b_2 we have

$$b_2(x, \pm 1) = 0, \partial_x b_1(\pm 1, y) = 0. \tag{2.150}$$

We can now solve for b_1 and b_2 separately. Since the boundary conditions for b_1 and b_2 are essentially the same, we only discuss how to solve for b_1.

We define the proper test and trial space for the Galerkin method.

$$X_N = \{u \in \mathbf{P_N} | u \text{ satisfies the condition (2.149)}\}, \tag{2.151}$$

and its corresponding Galerkin formulation is:
find $b \in X_N$ such that

$$\frac{\alpha}{\delta t}(b, v) + \eta\alpha(\nabla b, \nabla v) = (I_N f, v) \tag{2.152}$$

for all $v \in X_N$.

Constructing a proper basis is essential for the efficient implementation of the Galerkin method. In this case we can develop an efficient basis.

Define $\tilde{\phi}_i(x) = L_i(x) - L_{i+2}(x)$ where $L_i(x)$'s the ith Legendre polynomials and let

$$\phi_i(x) = \frac{1}{\sqrt{(\tilde{\phi}_i', \tilde{\phi}_i')}} \tilde{\phi}_i(x).$$

Then $\phi_i(\pm 1) = 0$ for $i \geq 0$ and $(\phi_i', \phi_j') = \delta_{ij}$. Similarly we can define a set of basis which satisfies the homogeneous Neumann boundary condition as follows. Define $\tilde{\psi}_i(x) = L_i(x) - \frac{i(i+1)}{(i+2)(i+3)} L_{i+2}(x)$.

$$\psi_i(x) = \frac{1}{\sqrt{(\tilde{\psi}', \tilde{\psi}')}} \tilde{\psi}(x).$$

Then $\psi_i'(\pm 1) = 0$ for all $i \geq 0$ and $(\psi_i', \psi_j') = \delta_{ij}$. Then

$$X_N = \text{span}\{\phi_i(x)\psi_j(y) | 0 \leq i, j \leq N-2\}.$$

We can write the solution of (2.152) as

$$b_N = \sum_{i=0}^{N-2} \sum_{j=0}^{N-2} b_{ij}\phi_i(x)\psi_j(y), \tag{2.153}$$

Then (2.152) has the following matrix representation:

$$\frac{\alpha}{\delta t} M_x B M_y^T + \eta\alpha(BM_y^T + M_x B) = F, \tag{2.154}$$

where

$$\begin{aligned} (M_x)_{ij} &= (\phi_j(x), \phi_i(x)), \\ (M_y)_{ij} &= (\psi_j(y), \psi_i(y)), \\ B_{ij} &= b_{ij}, \\ F_{ij} &= (I_N f, \phi_i(x)\psi_j(y)), \end{aligned} \tag{2.155}$$

for all $0 \leq i, j \leq N-2$. Since the equation (2.152) is separable, the discretized equation (2.154) can be solved efficiently using the diagonalization.

Since M_x and M_y are symmetric, they are diagonalized by orthonormal matrices. Say

$$\begin{aligned} M_x E_x &= E_x \Lambda_x, \\ M_y E_y &= E_y \Lambda_y. \end{aligned}$$

Let $\mathbf{B} = (E_x)^{-1} B {E_y}^{-T}$. Then (2.154) can be simplified as follows:

$$\begin{aligned} \frac{\alpha}{\delta t} M_x E_x \mathbf{B} E_y^T M_y^T + \eta\alpha(E_x \mathbf{B} E_y^T M_y^T + M_x E_x \mathbf{B} E_y^T) &= F, \\ \frac{\alpha}{\delta t} E_x \Lambda_x \mathbf{B} E_y^T M_y^T + \eta\alpha(E_x \mathbf{B} \Lambda_y^T E_y^T + E_x \Lambda_x \mathbf{B} E_y^T) &= F. \end{aligned}$$

We can multiply ${E_x}^{-1}$ from left and E_y^{-T} from the right. Then we have

$$\frac{\alpha}{\delta t} \Lambda_x \mathbf{B} \Lambda_y^T + \eta\alpha(\mathbf{B}\Lambda_y^T + \Lambda_x \mathbf{B}) = G \tag{2.156}$$

where $G = E_x^{-1} F E_y^{-T}$. Then

$$\mathbf{B}_{ij} = \frac{1}{\eta \Lambda_{xii} \Lambda_{yjj} + \eta\alpha(\Lambda_{xii} + \Lambda_{yjj})} G_{ij},$$

and we can find the coefficient matrix B as

$$B = E_x \mathbf{B} E_y^T.$$

2.6 Adaptive Implementation

In this section, we develop an adaptive time-stepping strategy for the MHD equation based on the first-order pressure-correction scheme. In [22], the authors developed an adaptive time-stepping of the phase field crystal model, and we apply it to the MHD equation. In previous sections, we developed unconditionally stable schemes. The main advantage of an unconditionally stable scheme is that step size need not remain small. Hence we can take the time step as large as is necessary. It is obvious that too large time step would degrade accuracy of the computed solution if the solution changes rapidly. Hence it would be a natural strategy that we take a small time step if motion changes fast and take a large time step if the solutions changes slowly. Based on the above discussion, we can take the following strategy:

- predetermine : $\delta t_{min}, \delta t_{max}, \gamma$.

The nth step size δt_n can be determined as follows:

$$\delta t_n = \begin{cases} \delta t_{min}, & \text{if } n = 1, \\ \max\{\delta t_{min}, \delta t_{max} / \sqrt{1 + \gamma \frac{E(u^n, b^n) - E(u^{n-1}, b^{n-1})}{\delta t_{n-1}}}\}, & \text{otherwise}, \end{cases} \tag{2.157}$$

where $\gamma > 0$ is a parameter chosen by experience and $E(u, b) = \|u\|^2 + \alpha\|b\|^2$.

The indicator of whether or not the solution changes rapidly or slowly is the discrete derivative of discrete energy. If the solution does not change(i.e.$u^n = u^{n-1}$ and $b^n = b^{n-1}$), δt_{max} will be chosen. If difference of energy is large, δt_{min} will be chosen as a time step. In the following section, we verify the performance of the adaptive time-stepping strategy.

2.7 Numerical Results

In this section, we perform two types of numerical experiments. In the first subsection, we test two second-order schemes with known exact solution. In the following subsection, we test efficiency of adaptive time-stepping with an arbitrary initial condition.

2.7.1 Analytic test solution

In this experiment, we check the order of accuracy of the solution of the scheme (2.111) - (2.115) and (2.133)-(2.137) with the exact solution,

$$u = (\sin(t)\sin(2\pi y)\sin(\pi x)^2, -\sin(t)\sin(2\pi x)\sin(\pi y)^2),$$

$$b = (\sin(t)\sin(\pi x)\cos(\pi y), -\sin(t)\sin(\pi y)\cos(\pi x)),$$

$$p = \sin(t)exp(x+y),$$

with coefficients $\nu = \alpha = \eta = 1$. The Legendre-Galerkin method developed in the previous section was applied for the spatial discretization. Forty-one points in each direction were used so that spatial error is negligible compared to the time discretization error. The tolerance for BiCGSTAB is 10^{-10}.

In Fig. 2.1, we plot error as a function of δt for second-order standard form. In Fig. 2.2, we plot error as a function of δt for second-order rotational form. Because the iterative solver BiCGSTAB was used, the number of iterations is very important. Hence we observe the iteration numbers of BiCGSTAB as a function of time with fixed δt for the case of standard form in Fig. 2.3.

For the NSE case, rotational form outperforms standard form in that it gives a better approximation of pressure. In [6], it is reported that standard and rotational form are second-second order in the L^2-norm. The improved pressure for rotational form accelerates the convergence of fluid velocity in H^1-norm. We observe the same

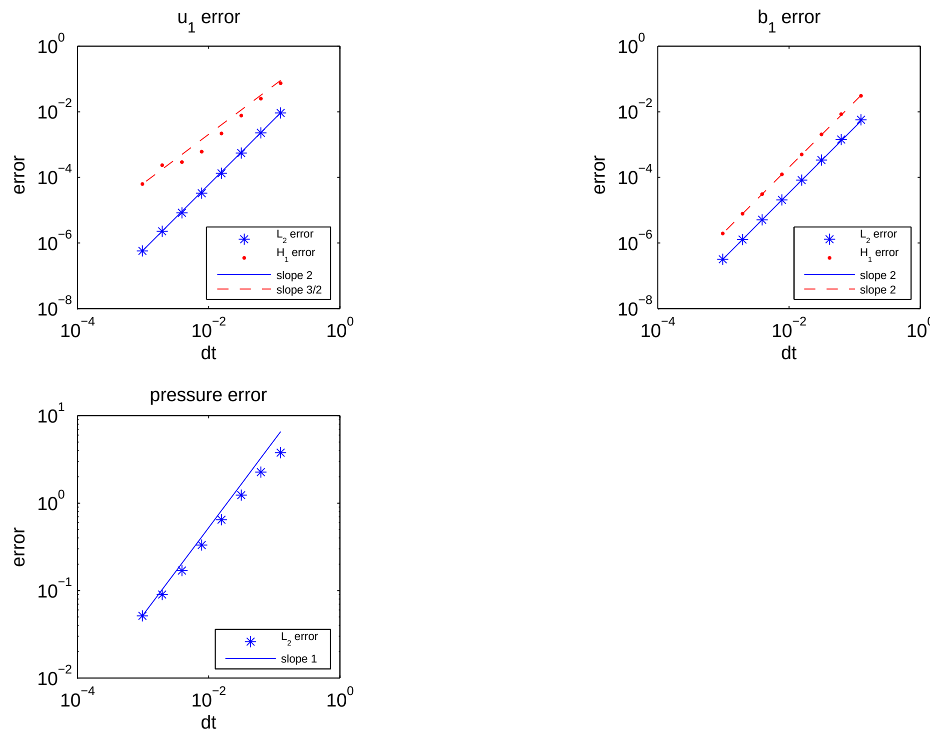

Figure 2.1. Errors of second-order standard form at $T = 1$

effect for the MHD equation in Fig. 2.1 and Fig. 2.2. The order for the pressure is first in standard form and 3/2 in the case of rotational form. As a result, we obtain higher order for H^1-norm of fluid velocity. And fluid velocity for L^2-norm is the same for both schemes. Both schemes show second order for the magnetic field. One reason for this is that the equation for the magnetic field does not contain pressure term and we do not observe an advantage for rotational form over standard form.

In Fig. 2.3, the iteration numbers for various δt are plotted. It is observed that the number of iterations has a period π. Let

$$\mathcal{A} = \begin{pmatrix} \frac{1}{\delta t}(\cdot) + ((2u^n - u^{n-1}) \cdot \nabla)(\cdot) - \nu\Delta(\cdot) & -\alpha((\nabla \times \cdot) \times (2b^n - b^{n-1}) \\ \alpha(\nabla \times ((2b^n - b^{n-1}) \times \cdot) & \frac{\alpha}{\delta t}(\cdot) - \alpha\eta\Delta(\cdot) \end{pmatrix} \tag{2.158}$$

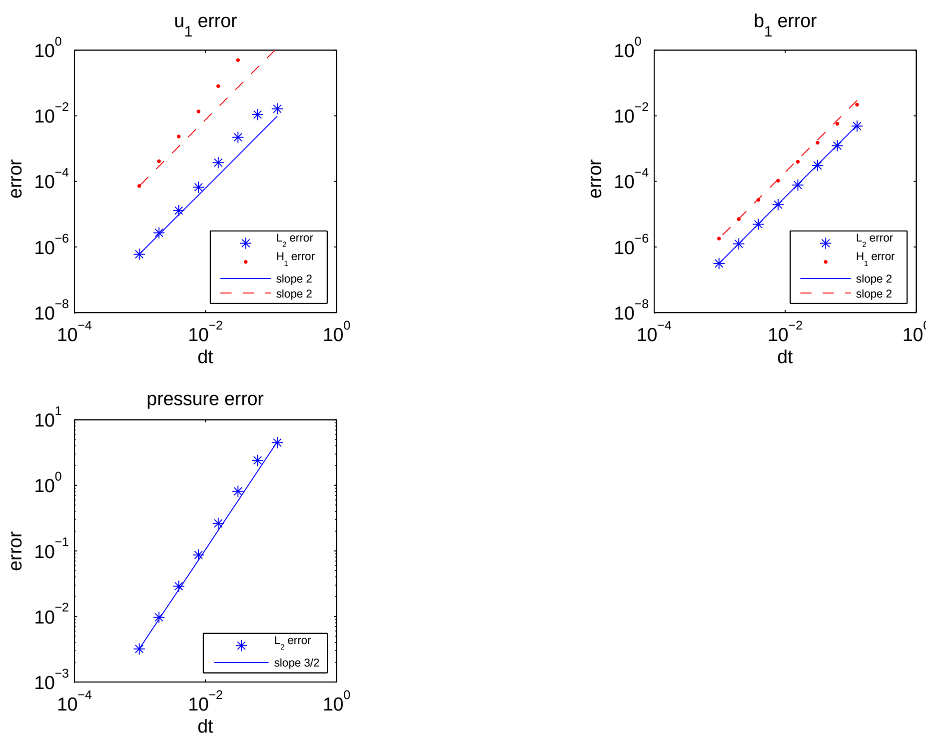

Figure 2.2. Errors of second-order rotational form at $T = 1$

and

$$\mathcal{P} = \begin{pmatrix} \frac{1}{\delta t}(\cdot) - \nu\Delta(\cdot) & 0 \\ 0 & \frac{\alpha}{\delta t}(\cdot) - \alpha\eta\Delta(\cdot) \end{pmatrix}. \tag{2.159}$$

Then the system we need to solve is

$$\mathcal{P}^{-1}\mathcal{A}\begin{pmatrix} \tilde{u}^{n+1} \\ b^{n+1} \end{pmatrix} = \mathcal{P}^{-1}\begin{pmatrix} f \\ g \end{pmatrix} \tag{2.160}$$

for some f and g. We can decompose $\mathcal{P}^{-1}\mathcal{A}$ into two parts:

$$\mathcal{P}^{-1}\mathcal{A} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} + \mathcal{P}^{-1}\begin{pmatrix} ((2u^n - u^{n-1}) \cdot \nabla)(\cdot) & -\alpha((\nabla \times \cdot) \times (2b^n - b^{n-1}) \\ \alpha(\nabla \times ((2b^n - b^{n-1}) \times \cdot) & 0 \end{pmatrix}. \tag{2.161}$$

By the choice of our exact solution, $\|2u^n - u^{n-1}\|$ and $\|2b^n - b^{n-1}\|$ are maximum around $t = k + \frac{\pi}{2}$ and minimum around $t = k\pi$ for $k = 0, 1, \cdots$. Hence we can expect that the second term of (2.161) becomes large for $t = k + \frac{\pi}{2}$ and small $t = k\pi$ for $k = 0, 1, \cdots$ and $\mathcal{P}^{-1}$ are independent of u^n and b^n. Since it is known that when a preconditioned system is close to identity, iterative method converges fast; therefore behavior like Fig. 2.3 can be explained.

2.7.2 Adaptive implementation

In this section, we implement the adaptive time-stepping scheme introduced in Section 2.6 with an arbitrary initial condition. Parameters are set up as follows: $\nu = .01$, $\eta = 1$, $\alpha = 10$ with initial condition $u_1 = u_2 = b_1 = b_2 = 1$ and $p_x = p_y = 0$. We use sixty-one points in each direction using the Legendre-Galerkin method so that spatial error is negligible compared to the time discretization error.. Here ν is selected as a small value because the solution converges to a stationary condition too quickly for a larger ν.

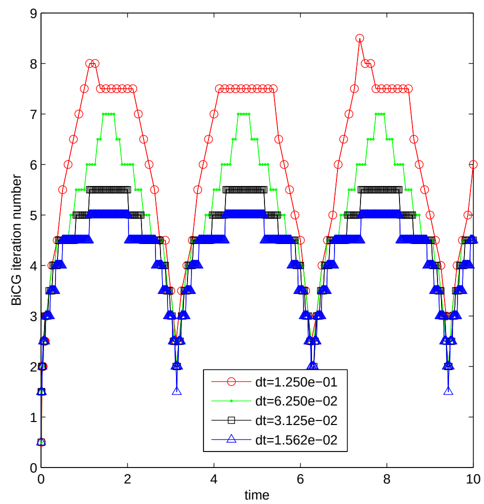

Figure 2.3. Number of iterations of BiCGSTAB with various δt

	t=0.1	t=1	t=2	t=10
dt=0.1	$0.002*10^4$	$0.011*10^4$	$0.015*10^4$	$0.035*10^4$
Adaptive	$0.02*10^4$	$0.14*10^4$	$0.15*10^4$	$0.16*10^4$
dt=0.001	$0.02*10^4$	$0.15*10^4$	$0.25*10^4$	$1.05*10^4$

Table 2.1.
Approximate numbers of BiCGSTAB iteration

For the adaptive time-stepping, we choose $\gamma = 100$, $\delta t_{min} = .001$, $\delta t_{max} = 0.1$ as parameters. We compare this result to the constant time step cases $\delta t = 0.1$ and $\delta t = 0.001$.

First we check the accuracy of the adaptive time-stepping. We can plot u_1 along $x = y$. In the first column of Fig. 2.4, the adaptive time-stepping solution and the $\delta t = 0.001$ solution are plotted together. They look almost identical. In the second column of Fig. 2.4, the $\delta t = 0.001$ solution , the adaptive time-stepping solution and the $\delta t = 0.1$ solution are plotted together. We can observe $\delta t = 0.01$ is too large to capture the dynamics. Hence we can conclude that the adaptive time-stepping solution is accurate and $\delta t = .01$ is too large to capture the dynamics. This also can be explained by the decay of energy. Fig. 2.6 shows that $\delta t = 0.1$ is too large compare to $\delta t = 0.001$. The next part we need to consider is whether the adaptive time-stepping strategy is efficient or not.

The efficiency of adaptive time-stepping scheme can be measured in terms of the number of time steps. It can be observed in Table 2.1. The size of the time step as time passes can be found in Fig. 2.7. Until $t = 1$, small time steps are maintained, but after $t = 2$, large time steps are maintained. The first graph of Fig.2.6 explains this phenomenon. Until $t = 1$, a rapid decay of energy is observed; after $t = 1$, no change in energy is observed.

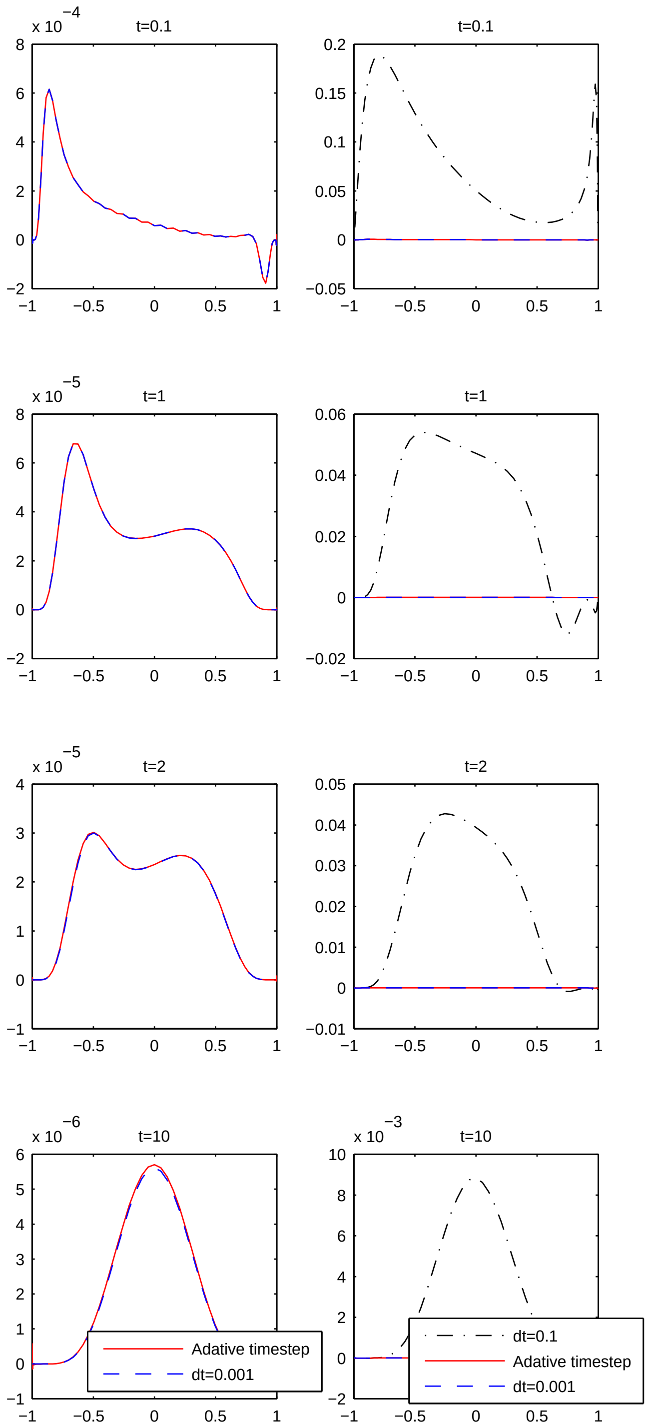

Figure 2.4. The graph of u_1 along the $x = y$

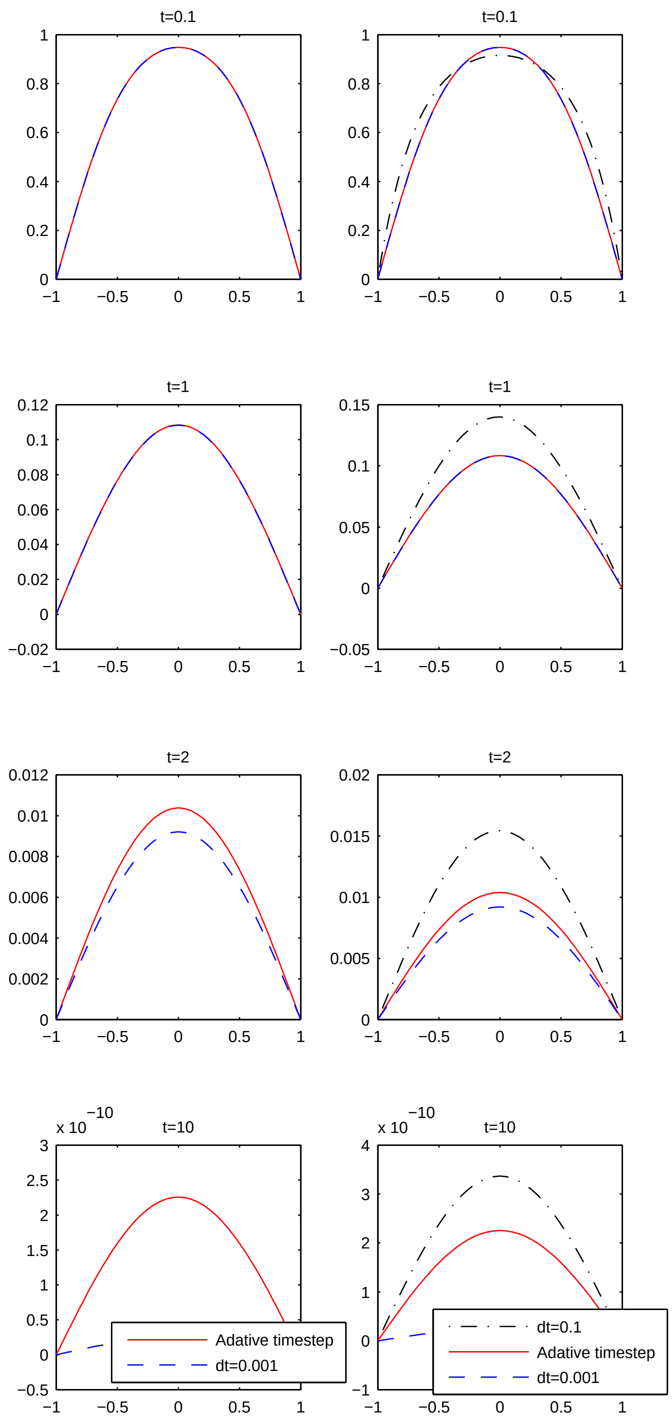

Figure 2.5. The graph of b_1 along the $x = y$

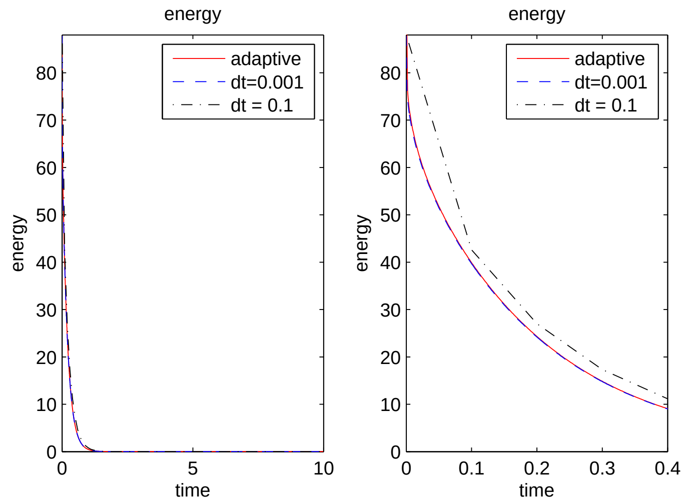

Figure 2.6. Energy decay

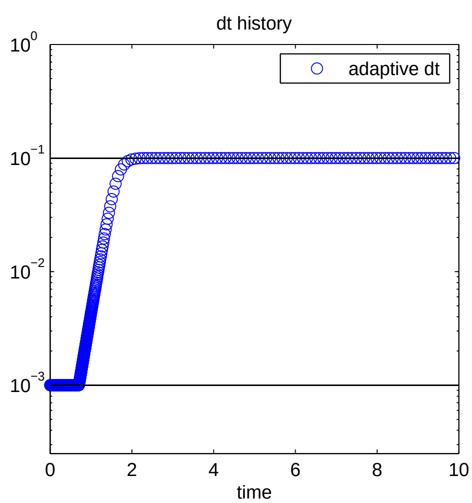

Figure 2.7. Size of time steps

2.8 Concluding remarks

In this chapter, we developed four unconditionally stable projection type discretizations for the MHD equation. We proved the unconditional stability of first- and second-order incremental projection method, first- and second-order rotational projection method. And in Section 2.3.2, we performed a rigorous error analysis for the first-order projection method. At each time step, we need to solve a linear, positive-definite system for the velocity and magnetic field, and a Poisson equation for the pressure. Hence the iterative solver like BICGSTAB would improve the performance of numerical simulation. The Legendre-Galerkin method was analyzed for the 2d rectangular domain case.

3. ERROR ESTIMATES FOR TIME DISCRETIZATIONS OF CAHN-HILLIARD AND ALLEN-CAHN PHASE-FIELD MODELS FOR TWO-PHASE INCOMPRESSIBLE FLOWS

In this chapter we carry out rigorous error analysis for some energy stable time discretization schemes developed in [23] for a Cahn-Hilliard phase-field model and in [24] for an Allen-Cahn phase-field model.

3.1 Introduction

Phase-field approach for multi-phase incompressible flows has attracted much attention in recent years (cf. [25–31] and the references therein). For two-phase incompressible flows, the phase-field models consist of either a Navier-Stokes-Cahn-Hilliard (NSCH) system or a Navier-Stokes-Allen-Cahn (NSAC) system. These are coupled nonlinear systems which are difficult to deal with numerically. Thus, designing efficient and accurate numerical methods for solving these coupled equations has been a great challenge to the scientific computing community.

While various convergence results and error estimates are available for the Navier-Stokes equations [14, 32, 33], there are only a few convergence results available for phase-field models of multi-phase flows. In [34], Feng proved the convergence of discrete finite element solutions for a Cahn-Hilliard phase-field model with matching density, and in [35] the authors established similar convergence results for an Allen-Cahn phase-field model with matching density. Most recently, Grün [36] proved convergent results for a scheme for the Cahn-Hilliard phase-field model with variable densities. However, the schemes considered in these papers are fully coupled with the pressure. From a computational point of view, it is more efficient to use a projection type method [33] to decouple the pressure from the velocity and phase function. In [23,24], some simple energy stable schemes are constructed, where the phase equations are discretized by the stabilized scheme [37, 38] or the convex splitting scheme [8, 39] and the Navier-Stokes (NS) equations are discretized by a projection scheme [33]. These schemes lead to, at each time step, a weakly coupled elliptic equations for the phase function and velocity, and a pressure Poisson equation for the pressure. Hence, they are very efficient and easy to implement.

Though various error estimates are available for projection type methods for Navier-Stokes equations [33] and for Cahn-Hilliard/Allen-Cahn equations [37, 40–42], it is highly non trivial to deal with systems which couple Navier-Stokes and Cahn-Hilliard/Allen-Cahn, since the splitting error in the projection step affects the whole system. The major difficulties arise from the velocity splitting step to deal with the incompressible constraint and from the coupling between the phase function and the velocity. To the best of our knowledge, error estimates for such schemes in semi-discrete-in-time or fully-discrete form is not yet available. Thus, the main purpose of this chapter is to provide a rigorous error analysis for these schemes in semi-discrete (in time) form. Note error analyses for semi-discrete (in time) schemes present special challenges since the often useful device of using discrete inverse inequalities with CFL-like conditions to control nonlinear terms is not available. On the other hand, the error analysis presented in this paper will provide some essential tools and procedures that can be used for error analysis of full discretized schemes for phase-field models.

This chapter is organized as follows. In Section 3.2, we describe the stabilized numerical scheme and the convex splitting scheme for the Cahn-Hilliard phase-field model. Section 3.3 is devoted to the error analysis, where we prove the error estimates for phase functions, velocity field and pressure. In Section 3.4, we extend the results to the Allen-Cahn phase-field model.

3.2 Cahn-Hilliard Phase-field model and its time discretization schemes

For the sake of simplicity, we shall only consider the two-phase flows with matching density, since one can expect that similar results hold for the case of non-matching densities but the actual proof will be much more tedious. The Cahn-Hilliard phase-field model for a two-phase incompressible flow with matching density reads (cf. for instance [24, 29]),

$$\begin{cases} \phi_t + u\cdot\nabla\phi - \gamma\Delta w = 0, \quad \text{in} \quad \Omega\subset\mathbb{R}^d, & (3.1)\\ w = -\Delta\phi + f(\phi), \quad \text{in} \quad \Omega\subset\mathbb{R}^d, & (3.2)\\ \rho_0(u_t + (u\cdot\nabla)u) - \mu_0\Delta u + \nabla p - \lambda w\nabla\phi = 0, \quad \text{in} \quad \Omega\subset\mathbb{R}^d, & (3.3)\\ \nabla\cdot u = 0, \quad \text{in} \quad \Omega\subset\mathbb{R}^d, & (3.4)\\ u|_{\partial\Omega} = 0, \dfrac{\partial\phi}{\partial\mathbf{n}}|_{\partial\Omega} = 0, \dfrac{\partial w}{\partial\mathbf{n}}|_{\partial\Omega} = 0, & (3.5) \end{cases}$$

with given initial data $u(0) = u_0$, $\phi(0) = \phi_0$. Here, ϕ is the phase function where $\phi \approx \pm 1$ corresponds to two different fluids, w is the chemical potential, u is the velocity field and p is the pressure. ρ_0 is the density of the fluids and we assume that the two fluids have the same density here; γ is the relaxation constant; λ is the mixing energy density, $f(\phi) = F'(\phi)$ where $F(\phi) = \frac{(1-\phi^2)^2}{4\varepsilon^2}$, and the parameter $\varepsilon > 0$ represents the interfacial thickness.

It is easy to show that the above system satisfies the following energy law (cf. for instance [24, 29]):

$$\frac{d}{dt}\int_\Omega (\frac{\rho_0}{2}|u|^2 + \frac{\lambda}{2}|\nabla\phi|^2 + \lambda F(\phi))dx = -\int_\Omega (\mu_0|\nabla u|^2 + \lambda\gamma|\nabla w|^2)dx. \tag{3.6}$$

We shall consider two time discretization schemes for (3.1)-(3.5), which were shown to be energy stable. The first one is the stabilized scheme introduced in [24]:

Given (u^n, ϕ^n, w^n, p^n), find $(\tilde{u}^{n+1}, \phi^{n+1}, w^{n+1})$ such that

$$\begin{cases} \dfrac{\phi^{n+1}-\phi^n}{\delta t} + (\tilde{u}^{n+1}\cdot\nabla)\phi^n - \gamma\Delta w^{n+1} = 0, & (3.7)\\ w^{n+1} - \dfrac{S}{\varepsilon^2}(\phi^{n+1}-\phi^n) = -\Delta\phi^{n+1} + f(\phi^n), & (3.8)\\ \rho_0(\dfrac{\tilde{u}^{n+1}-u^n}{\delta t} + (u^n\cdot\nabla)\tilde{u}^{n+1}) - \mu_0\Delta\tilde{u}^{n+1} + \nabla p^n - \lambda w^{n+1}\nabla\phi^n = 0, & (3.9)\\ \dfrac{\partial\phi^{n+1}}{\partial\mathbf{n}}|_{\partial\Omega} = 0, \quad \dfrac{\partial w^{n+1}}{\partial\mathbf{n}}|_{\partial\Omega} = 0, \quad \tilde{u}^{n+1}|_{\partial\Omega} = 0. & (3.10) \end{cases}$$

Given $(\tilde{u}^{n+1}, p^n)$, find (u^{n+1}, p^{n+1}) such that

$$\begin{cases} \rho_0\dfrac{u^{n+1}-\tilde{u}^{n+1}}{\delta t} + \nabla(p^{n+1}-p^n) = 0,\\ \nabla\cdot u^{n+1} = 0, \quad \mathbf{n}\cdot u^{n+1}|_{\partial\Omega} = 0. \end{cases} \tag{3.11}$$

In the above, S is a stabilizing constant to be specified. Since we are interested in the values of phase variable ϕ in the range of $[-1,1]$, it is a common practice to replace $F(\phi)$ by

$$F(\phi) = \begin{cases} \frac{1}{\varepsilon^2}(\phi-1)^2, & \phi > 1,\\ \frac{1}{4\varepsilon^2}(\phi^2-1)^2, & \phi\in[-1,1],\\ \frac{1}{\varepsilon^2}(\phi+1)^2, & \phi < -1, \end{cases} \tag{3.12}$$

so that we have

$$\max_{\phi\in\mathbb{R}} |f'(\phi)| \le \frac{2}{\varepsilon^2}. \tag{3.13}$$

The following stability result is proved in [24] (Thm. 3.1):

Theorem 1 *Let $S \ge 1/2$. Then, the scheme* (3.7)-(3.11) *with $F(\phi)$ given by* (3.12) *is unconditionally stable, and satisfies the following energy law:*

$$\begin{aligned} &\left[\frac{\rho_0}{2}\|u^{n+1}\|^2 + \frac{\lambda}{2}\|\nabla\phi^{n+1}\|^2 + \lambda(F(\phi^{n+1}),1)\right] + \frac{\delta t^2}{2\rho_0}\|\nabla p^{n+1}\|^2\\ &\quad + \frac{\rho_0}{2}\|\tilde{u}^{n+1}-u^n\|^2 + \mu_0\delta t\|\nabla\tilde{u}^{n+1}\|^2 + \lambda\gamma\delta t\|\nabla w^{n+1}\|^2\\ &\le \left[\frac{\rho_0}{2}\|u^n\|^2 + \frac{\lambda}{2}\|\nabla\phi^n\|^2 + \lambda(F(\phi^n),1)\right] + \frac{\delta t^2}{2\rho_0}\|\nabla p^n\|^2, \quad n\ge 0. \end{aligned} \tag{3.14}$$

The second-scheme we consider is based on convex splitting [8, 39]. We assume that the nonlinear potential in the phase equation can be split-up as follows:

$$F(\cdot) = F_c(\cdot) - F_e(\cdot), \quad \text{with } F_c \text{ and } F_e \quad \text{being convex functions.} \tag{3.15}$$

We set $f(\cdot) = F'(\cdot)$, $f_c(\cdot) = F_c'(\cdot)$, $f_e(\cdot) = F_e'(\cdot)$. In the case of a double well potential $F(\phi) = \frac{1}{4\varepsilon^2}(\phi^2 - 1)^2$, one can choose, for instance,

$$F_c(\phi) = \frac{1}{4\varepsilon^2}(\phi^4 + 1), \quad F_e(\phi) = \frac{1}{2\varepsilon^2}\phi^2,$$

so that we have

$$f_c(\phi) = \frac{1}{\varepsilon^2}\phi^3, \quad f_e(\phi) = \frac{1}{\varepsilon^2}\phi. \tag{3.16}$$

Then, we consider the following convex splitting scheme:

Given (u^n, ϕ^n, w^n, p^n), find $(\tilde{u}^{n+1}, \phi^{n+1}, w^{n+1})$ from

$$\begin{cases} \dfrac{\phi^{n+1} - \phi^n}{\delta t} + (\tilde{u}^{n+1} \cdot \nabla)\phi^n - \gamma\Delta w^{n+1} = 0, & (3.17) \\ w^{n+1} = -\Delta\phi^{n+1} + f_c(\phi^{n+1}) - f_e(\phi^n), & (3.18) \\ \rho_0(\dfrac{\tilde{u}^{n+1} - u^n}{\delta t} + (u^n \cdot \nabla)\tilde{u}^{n+1}) - \mu_0\Delta\tilde{u}^{n+1} + \nabla p^n - \lambda w^{n+1}\nabla\phi^n = 0, & (3.19) \\ \dfrac{\partial\phi^{n+1}}{\partial\mathbf{n}}|_{\partial\Omega} = 0, \quad \dfrac{\partial w^{n+1}}{\partial\mathbf{n}}|_{\partial\Omega} = 0, \quad \tilde{u}^{n+1}|_{\partial\Omega} = 0. & (3.20) \end{cases}$$

Given $(\tilde{u}^{n+1}, p^n)$, find (u^{n+1}, p^{n+1}) from

$$\begin{cases} \rho_0(\dfrac{u^{n+1} - \tilde{u}^{n+1}}{\delta t}) + \nabla(p^{n+1} - p^n) = 0, \\ \nabla \cdot u^{n+1} = 0, \quad \mathbf{n} \cdot u^{n+1}|_{\partial\Omega} = 0. \end{cases} \tag{3.21}$$

By using essentially the same procedure as in the proof of Theorem 3.1 in [24], we can prove the following:

Theorem 2 *The scheme* (3.17)-(3.21) *is unconditionally stable, and satisfies the following energy law:*

$$\begin{aligned} &\left[\frac{\rho_0}{2}\|u^{n+1}\|^2 + \frac{\lambda}{2}\|\nabla\phi^{n+1}\|^2 + \lambda(F(\phi^{n+1}), 1)\right] + \frac{\delta t^2}{2\rho_0}\|\nabla p^{n+1}\|^2 \\ &\qquad + \mu_0\delta t\|\nabla\tilde{u}^{n+1}\|^2 + \lambda\gamma\delta t\|\nabla w^{n+1}\|^2 + \frac{\rho_0}{2}\|\tilde{u}^{n+1} - u^n\|^2 \\ &\le \left[\frac{\rho_0}{2}\|u^n\|^2 + \frac{\lambda}{2}\|\nabla\phi^n\|^2 + \lambda(F(\phi^n), 1)\right] + \frac{\delta t^2}{2\rho_0}\|\nabla p^n\|^2, \quad n \ge 0. \end{aligned} \tag{3.22}$$

3.3 Error estimates

Let $(u^n, p^n, \phi^n, w^n, \tilde{u}^n)$ be the numerical solution obtained from the scheme (3.7)-(3.11) or (3.17)-(3.21), and $(u(t^n), p(t^n), w(t^n), \phi(t^n))$ be the exact solution, we define the error functions for $n = 0, 1, 2, \ldots, N$ as

$$\begin{aligned} \tilde{e}_u^n = u(t^n) - \tilde{u}^n, \qquad e_u^n = u(t^n) - u^n, \qquad e_w^n = w(t^n) - w^n, \\ e_\phi^n = \phi(t^n) - \phi^n, \qquad q^n = p(t^n) - p^n, \end{aligned} \tag{3.23}$$

and denote by $\tilde{e}_{u,\delta t}$, $e_{w,\delta t}$, $e_{u,\delta t}$, $e_{\phi,\delta t}$, $q_{\delta t}$ the corresponding sequence of error functions.

Assumption A: We assume that the exact solutions (u, ϕ, w, p) are sufficiently smooth. More precisely

$$\begin{aligned} &\phi \in L^\infty(0,T;H^3(\Omega)) \cap W^{1,\infty}(0,T;H^2(\Omega)) \cap W^{2,\infty}(0,T;H^1(\Omega)) \cap W^{3,\infty}(0,T;L^2(\Omega)), \\ &u \in W^{1,\infty}(0,T;H^2(\Omega)^d) \cap W^{2,\infty}(0,T;H^1(\Omega)^d) \cap W^{3,\infty}(0,T;L^2(\Omega)^d), \\ &w \in L^\infty(0,T;H^2(\Omega)) \cap W^{1,\infty}(0,T;L^2(\Omega)), \quad p \in W^{2,\infty}(0,T;H^1(\Omega)). \end{aligned}$$

Our main results are stated in the following theorem.

Theorem 3 *Under Assumption A, we have the following error estimates for the schemes* (3.7)-(3.11) *and* (3.17)-(3.21)*:*

$$\begin{aligned} \|e_{\phi,\delta t}\|_{l^\infty(H^1(\Omega))} + \|e_{u,\delta t}\|_{l^\infty(L^2(\Omega)^d)} + \|\tilde{e}_{u,\delta t}\|_{l^\infty(L^2(\Omega)^d)} &\lesssim \delta t, \\ \|e_{w,\delta t}\|_{l^2(H^1(\Omega))} + \|e_{u,\delta t}\|_{l^2(H^1(\Omega)^d)} + \|\tilde{e}_{u,\delta t}\|_{l^2(H^1(\Omega)^d)}\| + \|q_{\delta t}\|_{l^2(L^2(\Omega))} &\lesssim \delta t, \\ \|e_{w,\delta t}\|_{l^\infty(H^1(\Omega))} + \|e_{u,\delta t}\|_{l^\infty(H^1(\Omega)^d)} + \|\tilde{e}_{u,\delta t}\|_{l^\infty(H^1(\Omega)^d)} + \|q_{\delta t}\|_{l^\infty(L^2(\Omega))} &\lesssim \delta t^{1/2}. \end{aligned}$$

The proof of the above theorem will be carried out in several steps. We first prove some first-order estimates for the phase function, chemical potential and velocity which will imply a sub-optimal half-order estimate for the pressure. Then, we shall prove some estimates for their time derivatives. Finally, we shall derive a first-order error estimate for the pressure.

3.3.1 First-order error estimates for (ϕ, w, u)

As the first step in the proof of Theorem 3, we shall establish the following:

Lemma 1 *Under the assumption A, there exists some* $\tau_0 > 0$ *such that when* $\delta t \le \tau_0$ *the solution* (u^n, p^n, ϕ^n, w^n) $(0 \le n \le \frac{T}{\delta t})$ *of scheme* (3.7)-(3.11) *or* (3.17)-(3.21) *satisfies the following error estimates*

$$\begin{aligned}
&\|e_{\phi,\delta t}\|_{l^\infty(H^1(\Omega))} + \|e_{u,\delta t}\|_{l^2(H^1(\Omega)^d)} + \|\tilde{e}_{u,\delta t}\|_{l^2(H^1(\Omega)^d)} + \|e_{w,\delta t}\|_{l^2(H^1(\Omega))} \lesssim \delta t, \\
&\|e_{w,\delta t}\|_{l^\infty(H^1(\Omega))} + \|e_{u,\delta t}\|_{l^\infty(H^1(\Omega)^d)} + \|e_{w,\delta t}\|_{l^\infty(H^1(\Omega))} + \|\tilde{e}_{u,\delta t}\|_{l^\infty(H^1(\Omega)^d)} \lesssim \delta t^{1/2}, \\
&\|e_{u,\delta t}\|_{l^\infty(L^2(\Omega)^d)} + \|\tilde{e}_{u,\delta t}\|_{l^\infty(L^2(\Omega)^d)} \lesssim \delta t, \quad \|\nabla q^n\|_{L^2(\Omega)^d} \lesssim 1.
\end{aligned} \tag{3.24}$$

Proof We notice that in the stabilized scheme (3.7)-(3.11), the nonlinear function f is assumed to be Lipschitz (cf. (3.13)), while f does not have to be Lipschitz (3.16) in the convex splitting scheme (3.17)-(3.21). Thus, the proof of Lemma 1 for the two schemes will be different. In particular, the results for convex splitting scheme (3.17)-(3.21) rely heavily on the structure of the nonlinear potential (3.16). We will prove Lemma 1 for these two schemes separately.

Proof of Lemma 1 for the stabilized scheme (3.7)-(3.11) To analyze the error for stabilized scheme (3.7)-(3.11), we define the local truncation error R_ϕ^n $(n = 0, 1, \ldots, N-1)$ for the phase equation (3.1):

$$R_\phi^{n+1} = \frac{1}{\delta t}(\phi(t^{n+1}) - \phi(t^n)) + u(t^{n+1}) \cdot \nabla\phi(t^n) - \gamma\Delta w(t^{n+1}), \tag{3.25}$$

the truncation error R_w^{n+1} $(n = 0, 1, \ldots, N-1)$ for the chemical potential equation (3.2):

$$R_w^{n+1} = w(t^{n+1}) - \frac{S}{\varepsilon^2}(\phi(t^{n+1}) - \phi(t^n)) + \Delta\phi(t^{n+1}) - f(\phi(t^n)), \tag{3.26}$$

the truncation error R_u^{n+1} $(n = 0, 1, \ldots, N-1)$ for the momentum equation (3.3):

$$\begin{aligned}
R_u^{n+1} =& \rho_0\left(\frac{u(t^{n+1}) - u(t^n)}{\delta t} + u(t^n) \cdot \nabla u(t^{n+1})\right) \\
&- \mu_0\Delta u(t^{n+1}) + \nabla p(t^n) - \lambda w(t^{n+1})\nabla\phi(t^n),
\end{aligned} \tag{3.27}$$

and the local truncation error R_p^{n+1} $(n = 0, 1, \ldots, N-1)$ for the pressure correction (3.11):

$$R_p^{n+1} = \rho_0\frac{u(t^{n+1}) - u(t^{n+1})}{\delta t} + \nabla(p(t^{n+1}) - p(t^n)) = \nabla(p(t^{n+1}) - p(t^n)). \tag{3.28}$$

With a standard procedure, it is easy to establish the following estimates for the truncation errors, provided that the exact solutions are sufficiently smooth.

Lemma 2 *Under Assumption A, the truncation errors satisfy*

$$\|R_{u,\delta t}\|_{l^\infty(L^2(\Omega)^d)} + \|R_{\phi,\delta t}\|_{l^\infty(H^1(\Omega))} + \|R_{w,\delta t}\|_{l^\infty(H^1(\Omega))} + \|R_{p,\delta t}\|_{l^\infty(L^2(\Omega))} \leq c\delta t, \tag{3.29}$$

where c is independent of δt.

Next, we derive the equations governing the error growth. Define

$$\begin{aligned}
\dot{\tilde{e}}_u^{n+1} =& \left(\frac{u(t^{n+1}) - u(t^n)}{\delta t} + u(t^n)\cdot\nabla u(t^{n+1})\right) - \left(\frac{\tilde{u}^{n+1} - u^n}{\delta t} + u^n \cdot \nabla \tilde{u}^{n+1}\right) \\
=& \frac{\tilde{e}_u^{n+1} - e_u^n}{\delta t} + u^n \cdot \nabla \tilde{e}_u^{n+1} + e_u^n \cdot \nabla u(t^{n+1}), \qquad (3.30) \\
G^n =& f(\phi(t^n)) - f(\phi^n), \quad n \geq 0. \qquad (3.31)
\end{aligned}$$

Subtracting (3.25), (3.26), (3.27) and (3.28) from (3.7), (3.8), (3.9) and (3.11), respectively, we get the following error equations for $n \geq 0$:

$$\begin{cases}
\dfrac{e_\phi^{n+1} - e_\phi^n}{\delta t} + (u(t^{n+1})\cdot\nabla\phi(t^n) - \tilde{u}^{n+1}\cdot\nabla\phi^n) - \gamma\Delta e_w^{n+1} = R_\phi^{n+1}, & (3.32) \\
e_w^{n+1} - \dfrac{S}{\varepsilon^2}(e_\phi^{n+1} - e_\phi^n) + \Delta e_\phi^{n+1} - G^n = R_w^{n+1}, & (3.33) \\
\rho_0 \dot{\tilde{e}}_u^{n+1} - \mu_0 \Delta \tilde{e}_u^{n+1} + \nabla q^n - \lambda(w(t^{n+1})\nabla\phi(t^n) - w^{n+1}\nabla\phi^n) = R_u^{n+1}, & (3.34) \\
\rho_0 \dfrac{e_u^{n+1} - \tilde{e}_u^{n+1}}{\delta t} + \nabla(q^{n+1} - q^n) = R_p^{n+1}, & (3.35)
\end{cases}$$

with the boundary conditions

$$\tilde{e}_u^{n+1}|_{\partial\Omega} = 0, \quad \partial_{\mathbf{n}} e_\phi^{n+1}|_{\partial\Omega} = 0, \quad \partial_{\mathbf{n}} e_w^{n+1}|_{\partial\Omega} = 0. \tag{3.36}$$

We first derive the following properties for the nonlinear term G^n:

Lemma 3 *Under Assumption A, for $0 \leq n \leq [T/\delta t] - 1$, we have*

$$\|G^n\| \lesssim \|e_\phi^n\|, \tag{3.37}$$

$$\|\nabla G^n\| \lesssim \|e_\phi^n\| + \|\nabla e_\phi^n\|. \tag{3.38}$$

Proof We rewrite G^n as

$$G^n = e_\phi^n \int_0^1 f'(s\phi(t^n) + (1-s)\phi^n)\, ds, \tag{3.39}$$

which implies (3.37). Taking gradient of (3.39), we arrive at

$$\begin{aligned}\nabla G^n =& f'(\phi(t^n))\nabla\phi(t^n) - f'(\phi^n)\nabla\phi^n \\ =& (f'(\phi(t^n)) - f'(\phi^n))\nabla\phi(t^n) + f'(\phi^n)\nabla e_\phi^n.\end{aligned}$$

Since f' is bounded and Lipschitz, under the assumption on the exact solution we have

$$\|\nabla G^n\| \lesssim \|e_\phi^n\|\|\phi(t^n)\|_{H^3} + \|\nabla e_\phi^n\|.$$

Combining all the results above, we obtain (3.38). ■

Taking inner product of (3.32) with $\lambda\delta t e_\phi^{n+1}$ and $\lambda\delta t e_w^{n+1}$, we obtain

$$\begin{aligned}&\frac{\lambda}{2}(\|e_\phi^{n+1}\|^2 - \|e_\phi^n\|^2 + \|e_\phi^{n+1} - e_\phi^n\|^2) + \lambda\delta t(u(t^{n+1})\cdot\nabla\phi(t^n) - \tilde{u}^{n+1}\cdot\nabla\phi^n, e_\phi^{n+1}) \\ &\qquad + \lambda\gamma\delta t(\nabla e_\phi^{n+1}, \nabla e_w^{n+1}) = \lambda\delta t(R_\phi^{n+1}, e_\phi^{n+1}),\end{aligned} \tag{3.40}$$

and

$$\begin{aligned}&\lambda(e_\phi^{n+1} - e_\phi^n, e_w^{n+1}) + \lambda\gamma\delta t\|\nabla e_w^{n+1}\|^2 \\ &+ \lambda\delta t(u(t^{n+1})\cdot\nabla\phi(t^n) - \tilde{u}^{n+1}\cdot\nabla\phi^n, e_w^{n+1}) = \lambda\delta t(R_\phi^{n+1}, e_w^{n+1}).\end{aligned} \tag{3.41}$$

Taking inner product of (3.33) with $\lambda\gamma\delta t e_w^{n+1}$ and $-\lambda(e_\phi^{n+1} - e_\phi^n)$, respectively, we have

$$\begin{aligned}\lambda\gamma\delta t\|e_w^{n+1}\|^2 - \frac{S\lambda\gamma}{\varepsilon^2}\delta t(e_\phi^{n+1} - e_\phi^n, e_w^{n+1}) - \lambda\gamma\delta t(\nabla e_\phi^{n+1}, \nabla e_w^{n+1}) \\ = \lambda\gamma\delta t(G^n, e_w^{n+1}) + \lambda\gamma\delta t(R_w^{n+1}, e_w^{n+1}),\end{aligned} \tag{3.42}$$

and

$$\begin{aligned}&-\lambda(e_\phi^{n+1} - e_\phi^n, e_w^{n+1}) + \frac{S\lambda}{\varepsilon^2}\|e_\phi^{n+1} - e_\phi^n\|^2 + \frac{\lambda}{2}(\|\nabla e_\phi^{n+1}\|^2 - \|\nabla e_\phi^n\|^2 + \|\nabla e_\phi^{n+1} - \nabla e_\phi^n\|^2) \\ &\qquad = -\lambda(G^n, e_\phi^{n+1} - e_\phi^n) - \lambda(R_w^{n+1}, e_\phi^{n+1} - e_\phi^n).\end{aligned} \tag{3.43}$$

Taking inner product of (3.34) with $\delta t \tilde{e}_u^{n+1}$, we get

$$\begin{aligned}&\frac{\rho_0}{2}(\|\tilde{e}_u^{n+1}\|^2 - \|e_u^n\|^2 + \|\tilde{e}_u^{n+1} - e_u^n\|^2) + \delta t(e_u^n \cdot \nabla u(t^{n+1}), \tilde{e}_u^{n+1}) + \mu_0 \delta t \|\nabla \tilde{e}_u^{n+1}\|^2 \\ &\quad + \delta t(\nabla q^n, \tilde{e}_u^{n+1}) - \lambda \delta t(w(t^{n+1})\nabla\phi(t^n) - w^{n+1}\nabla\phi^n, \tilde{e}_u^{n+1}) = \delta t(R_u^{n+1}, \tilde{e}_u^{n+1}).\end{aligned} \tag{3.44}$$

In addition, we know

$$\begin{aligned}\lambda\delta t(u(t^{n+1})\cdot\nabla\phi(t^n) - \tilde{u}^{n+1}\cdot\nabla\phi^n, e_\phi^{n+1}) &= \lambda\delta t(u(t^{n+1})\cdot\nabla e_\phi^n + \tilde{e}_u^{n+1}\cdot\nabla\phi^n, e_\phi^{n+1}) \\ &= \lambda\delta t(u(t^{n+1})\cdot\nabla e_\phi^n, e_\phi^{n+1}) + \lambda\delta t(\tilde{e}_u^{n+1}\cdot\nabla\phi^n, e_\phi^{n+1}),\end{aligned} \tag{3.45}$$

and

$$\begin{aligned}&\lambda\delta t(u(t^{n+1})\cdot\nabla\phi(t^n) - \tilde{u}^{n+1}\cdot\nabla\phi^n, e_w^{n+1}) - \lambda\delta t(w(t^{n+1})\nabla\phi(t^n) - w^{n+1}\nabla\phi^n, \tilde{e}_u^{n+1}) \\ =&\lambda\delta t(u(t^{n+1})\cdot\nabla e_\phi^n + \tilde{e}_u^{n+1}\cdot\nabla\phi^n, e_w^{n+1}) - \lambda\delta t(w(t^{n+1})\nabla e_\phi^n + e_w^{n+1}\nabla\phi^n, \tilde{e}_u^{n+1}) \\ =&\lambda\delta t(u(t^{n+1})\cdot\nabla e_\phi^n - \tilde{e}_u^{n+1}\cdot\nabla e_\phi^n, e_w^{n+1}) + \lambda\delta t(\tilde{e}_u^{n+1}\cdot\nabla\phi(t^n), e_w^{n+1}) \\ &- \lambda\delta t(w(t^{n+1})\cdot\nabla e_\phi^n - e_w^{n+1}\nabla e_\phi^n, \tilde{e}_u^{n+1}) - \lambda\delta t(e_w^{n+1}\nabla\phi(t^n), \tilde{e}_u^{n+1}) \\ =&\lambda\delta t(u(t^{n+1})\cdot\nabla e_\phi^n, e_w^{n+1}) - \lambda\delta t(w(t^{n+1})\cdot\nabla e_\phi^n, \tilde{e}_u^{n+1}).\end{aligned} \tag{3.46}$$

Combining (3.40)–(3.44) and using (3.45) and (3.46), we obtain

$$\begin{aligned}&\frac{\lambda}{2}(\|e_\phi^{n+1}\|^2 - \|e_\phi^n\|^2 + \|e_\phi^{n+1} - e_\phi^n\|^2) + \frac{\rho_0}{2}(\|\tilde{e}_u^{n+1}\|^2 - \|e_u^n\|^2 + \|\tilde{e}_u^{n+1} - e_u^n\|^2) \\ &+ \frac{\lambda}{2}(\|\nabla e_\phi^{n+1}\|^2 - \|\nabla e_\phi^n\|^2 + \|\nabla e_\phi^{n+1} - \nabla e_\phi^n\|^2) + \frac{S\lambda}{\varepsilon^2}\|e_\phi^{n+1} - e_\phi^n\|^2 \\ &+ \mu_0\delta t\|\nabla\tilde{e}_u^{n+1}\|^2 + \lambda\gamma\delta t\|e_w^{n+1}\|^2 + \lambda\gamma\delta t\|\nabla e_w^{n+1}\|^2 + \delta t(\nabla q^n, \tilde{e}_u^{n+1}) \\ &+ \delta t(e_u^n\cdot\nabla u(t^{n+1}), \tilde{e}_u^{n+1}) + \lambda\delta t(u(t^{n+1})\cdot\nabla e_\phi^n, e_\phi^{n+1}) + \lambda\delta t(\tilde{e}_u^{n+1}\cdot\nabla\phi^n, e_\phi^{n+1}) \\ &+ \lambda\delta t(u(t^{n+1})\cdot\nabla e_\phi^n, e_w^{n+1}) - \lambda\delta t(w(t^{n+1})\cdot\nabla e_\phi^n, \tilde{e}_u^{n+1}) \\ =&\frac{S\lambda\gamma}{\varepsilon^2}\delta t(e_\phi^{n+1} - e_\phi^n, e_w^{n+1}) + \lambda\delta t(R_\phi^{n+1}, e_\phi^{n+1}) + \lambda\delta t(R_\phi^{n+1}, e_w^{n+1}) + \lambda\gamma\delta t(G^n, e_w^{n+1}) \\ &+ \lambda\gamma\delta t(R_w^{n+1}, e_w^{n+1}) - \lambda(G^n, e_\phi^{n+1} - e_\phi^n) - \lambda(R_w^{n+1}, e_\phi^{n+1} - e_\phi^n) + \delta t(R_u^{n+1}, \tilde{e}_u^{n+1}).\end{aligned} \tag{3.47}$$

We now control each term in (3.47) as follows.

(i) Firstly, it is easy to check that

$$
\begin{aligned}
&\left|\lambda\delta t(R_\phi^{n+1}, e_\phi^{n+1})\right| \le c(\delta t^3 + \delta t\|e_\phi^{n+1}\|^2),\\
&\left|\lambda\delta t(R_\phi^{n+1}, e_w^{n+1})\right| \le \lambda\delta t\|R_\phi^{n+1}\|\|e_w^{n+1}\| \le c\delta t^3 + \frac{\lambda\gamma}{8}\delta t\|e_w^{n+1}\|^2,\\
&\left|\delta t(R_u^{n+1}, \tilde{e}_u^{n+1})\right| \le \delta t\|R_u^{n+1}\|\|\tilde{e}_u^{n+1}\| \le c\delta t^3 + \frac{\mu_0}{8}\delta t\|\nabla\tilde{e}_u^{n+1}\|^2,\\
&\left|\lambda\gamma\delta t(R_w^{n+1}, e_w^{n+1})\right| \le \lambda\gamma\delta t\|R_w^{n+1}\|\|e_w^{n+1}\| \le c\delta t^3 + \frac{\lambda\gamma}{8}\delta t\|e_w^{n+1}\|^2,
\end{aligned}
\tag{3.48}
$$

and

$$
\begin{aligned}
&\left|\lambda\delta t(u(t^{n+1})\cdot\nabla e_\phi^n, e_w^{n+1})\right| \le c\delta t\|\nabla e_\phi^n\|\|e_w^{n+1}\| \le c\delta t\|\nabla e_\phi^n\|^2 + \frac{\lambda\gamma}{8}\delta t\|e_w^{n+1}\|^2,\\
&\left|\lambda\delta t(w(t^{n+1})\cdot\nabla e_\phi^n, \tilde{e}_u^{n+1})\right| \le c\delta t\|\nabla e_\phi^n\|\|\tilde{e}_u^{n+1}\| \le c\delta t\|\nabla e_\phi^n\|^2 + \frac{\mu_0}{8}\delta t\|\nabla\tilde{e}_u^{n+1}\|^2,\\
&\left|\lambda\delta t(u(t^{n+1})\cdot\nabla e_\phi^n, e_\phi^{n+1})\right| \le c\delta t\|\nabla e_\phi^n\|\|e_\phi^{n+1}\| \le c\delta t\|\nabla e_\phi^n\|^2 + c\delta t\|e_\phi^{n+1}\|^2,\\
&\left|\frac{S\lambda\gamma}{\varepsilon^2}\delta t(e_\phi^{n+1} - e_\phi^n, e_w^{n+1})\right| \le \frac{S\lambda}{\varepsilon^2}\delta t\|e_\phi^{n+1} - e_\phi^n\|\|e_w^{n+1}\|\\
&\qquad\qquad \le c\delta t\|e_\phi^{n+1} - e_\phi^n\|^2 + \frac{\lambda\gamma}{8}\delta t\|e_w^{n+1}\|^2.
\end{aligned}
\tag{3.49}
$$

(ii) Next , we deal with terms involving G^n. In view of (3.37)-(3.38), we obtain

$$
\left|\lambda\delta t(G^n, e_w^{n+1})\right| \le \lambda\delta t\|G^n\|\|e_w^{n+1}\| \le c\delta t\|e_\phi^n\|^2 + \frac{\lambda\gamma}{8}\delta t\|e_w^{n+1}\|^2. \tag{3.50}
$$

Using (3.32) , (3.37)-(3.38) and Theorem 1, we get

$$
\begin{aligned}
\left|\lambda\gamma(G^n, e_\phi^{n+1} - e_\phi^n)\right| = &\left|\lambda\gamma\delta t(G^n, \frac{e_\phi^{n+1} - e_\phi^n}{\delta t})\right|\\
=&\lambda\gamma\delta t\left|(G^n, -u(t^{n+1})\cdot\nabla e_\phi^n - \tilde{e}_u^{n+1}\cdot\nabla\phi^n + \gamma\Delta e_w^{n+1} + R_\phi^{n+1})\right|\\
\le&\lambda\gamma\delta t\big(c(\|G^n\|^2 + \|\nabla e_\phi^n\|^2) + \left|(G^n, \tilde{e}_u^{n+1}\cdot\nabla\phi^n)\right|\\
&\quad + c\|\nabla G^n\|^2 + \frac{1}{8}\|\nabla e_w^{n+1}\|^2 + \|G^n\|^2 + \frac{1}{4}\|R_\phi^{n+1}\|^2\big)\\
\le&\lambda\gamma\delta t\big(c\|G^n\|^2 + c\|\nabla G^n\|^2 + c\|\nabla e_\phi^n\|^2 + \left|(G^n, \tilde{e}_u^{n+1}\cdot\nabla\phi^n)\right|\\
&\quad + \frac{1}{8}\|\nabla e_w^{n+1}\|^2 + c\delta t^2\big).
\end{aligned}
\tag{3.51}
$$

Noticing that $\tilde{e}_u^{n+1}|_{\partial\Omega} = 0$, so H^1 norm of $\tilde{e}_u^{n+1}$ is equivalent to $\|\nabla \tilde{e}_u^{n+1}\|$, there holds

$$
\begin{aligned}
\left|(G^n, \tilde{e}_u^{n+1} \cdot \nabla\phi^n)\right| \leq & \begin{cases} \|G^n\|_{L^4(\Omega)} \|\nabla\phi^n\|_{L^2(\Omega)^d} \|\tilde{e}_u^{n+1}\|_{L^4(\Omega)^d}, & d=2, \\ \|G^n\|_{L^3(\Omega)} \|\nabla\phi^n\|_{L^2(\Omega)^d} \|\tilde{e}_u^{n+1}\|_{L^6(\Omega)^d}, & d=3, \end{cases} \\
\leq & \begin{cases} c_2 c(2,\Omega) \|G^n\|^{1/2} \|G^n\|_{H^1(\Omega)}^{1/2} \|\tilde{e}_u^{n+1}\|^{1/2} \|\nabla \tilde{e}_u^{n+1}\|^{1/2}, & d=2, \\ c_2 c(3,\Omega) \|G^n\|^{1/2} \|G^n\|_{H^1(\Omega)}^{1/2} \|\nabla \tilde{e}_u^{n+1}\|, & d=3, \end{cases} \\
\leq & c_2 c_\Omega \left(\|G^n\|^2 + \|\nabla G^n\|^2 \right) + \frac{\mu_0}{8\lambda\gamma} \|\nabla \tilde{e}_u^{n+1}\|_{L^2(\Omega)^d}^2 \\
\leq & c \left(\|\nabla e_\phi^n\|^2 + \|e_\phi^n\|^2 \right) + \frac{\mu_0}{8\lambda\gamma} \|\nabla \tilde{e}_u^{n+1}\|^2,
\end{aligned} \tag{3.52}
$$

which implies

$$
\left|\lambda\gamma(G^n, e_\phi^{n+1} - e_\phi^n)\right| \leq c\delta t^3 + c\delta t \left(\|e_\phi^n\|^2 + \|\nabla e_\phi^n\|^2 \right) + \frac{\lambda\gamma}{8} \delta t \|\nabla e_w^{n+1}\|^2 + \frac{\mu_0}{8} \delta t \|\nabla \tilde{e}_u^{n+1}\|^2. \tag{3.53}
$$

Following (3.51) and (3.52) and using Lemma 2, we estimate

$$
\begin{aligned}
& \left|\lambda\gamma(R_w^{n+1}, e_\phi^{n+1} - e_\phi^n)\right| \\
\leq & \lambda\gamma\delta t \left(c\|R_w^{n+1}\|_{H^1(\Omega)}^2 + c\|\nabla e_\phi^n\|^2 + |(R_w^{n+1}, \tilde{e}_u^{n+1} \cdot \nabla\phi^n)| + \frac{1}{8} \|\nabla e_w^{n+1}\|^2 + \|R_\phi^{n+1}\|^2 \right) \\
\leq & \delta t \left(c\|e_\phi^n\|^2 + c\|\nabla e_\phi^n\|^2 + c\delta t^2 \right) + \frac{\lambda\gamma}{8} \delta t \|\nabla e_w^{n+1}\|^2 + \frac{\mu_0}{8} \delta t \|\nabla \tilde{e}_u^{n+1}\|^2.
\end{aligned} \tag{3.54}
$$

Again, similar to (3.52), we have

$$
\left|\lambda\delta t(\tilde{e}_u^{n+1} \cdot \nabla\phi^n, e_\phi^{n+1})\right| \leq c\delta t(\|e_\phi^{n+1}\|^2 + \|\nabla e_\phi^{n+1}\|^2) + \frac{\mu_0}{8} \delta t \|\nabla \tilde{e}_u^{n+1}\|^2. \tag{3.55}
$$

(iii) It remains to estimate the term involving pressure. Using (3.35), $e_u^{n+1} \cdot \mathbf{n}|_{\partial\Omega} = 0$ and $\operatorname{div} e_u^{n+1} = 0$, we get

$$
\begin{aligned}
\left(\nabla q^n, \tilde{e}_u^{n+1}\right) = & \left(\nabla q^n, \frac{\delta t}{\rho_0} (\nabla q^{n+1} - \nabla q^n) - \frac{\delta t}{\rho_0} R_p^{n+1} + e_u^{n+1} \right) \\
= & \frac{\delta t}{2\rho_0} \left(\|\nabla q^{n+1}\|^2 - \|\nabla q^n\|^2 - \|\nabla q^{n+1} - \nabla q^n\|^2 \right) - \frac{\delta t}{\rho_0} (\nabla q^n, R_p^{n+1}),
\end{aligned} \tag{3.56}
$$

and

$$\left\|\rho_0\frac{e_u^{n+1}}{\delta t}+\big(\nabla q^{n+1}-\nabla q^n\big)\right\|^2=\left\|R_p^{n+1}+\rho_0\frac{\tilde{e}_u^{n+1}}{\delta t}\right\|^2, \tag{3.57}$$

which implies

$$\|\nabla q^{n+1}-\nabla q^n\|^2=\frac{\rho_0^2}{\delta t^2}(\|\tilde{e}_u^{n+1}\|^2-\|e_u^{n+1}\|^2)+\|R_p^{n+1}\|^2+\frac{2\rho_0}{\delta t}(R_p^{n+1},\tilde{e}_u^{n+1}). \tag{3.58}$$

Hence, from (3.56) and (3.58), we find

$$\begin{aligned}\big(\nabla q^n,\tilde{e}_u^{n+1}\big)=&\frac{\delta t}{2\rho_0}\left(\|\nabla q^{n+1}\|^2-\|\nabla q^n\|^2\right)+\frac{\rho_0}{2\delta t}\left(\|e_u^{n+1}\|^2-\|\tilde{e}_u^{n+1}\|^2\right)\\&-\frac{\delta t}{2\rho_0}\|R_p^{n+1}\|^2-(R_p^{n+1},\tilde{e}_u^{n+1})-\frac{\delta t}{\rho_0}(\nabla q^n,R_p^{n+1}),\end{aligned} \tag{3.59}$$

where

$$\begin{aligned}&\frac{\delta t}{\rho_0}\left|(\nabla q^n,R_p^{n+1})\right|\le c(\delta t^2\|\nabla q^n\|^2+\delta t^2),\\&\left|(R_p^{n+1},\tilde{e}_u^{n+1})\right|\le c\delta t\|\tilde{e}_u^{n+1}\|\le c\delta t^2+\frac{\mu_0}{8}\|\nabla\tilde{e}_u^{n+1}\|^2.\end{aligned} \tag{3.60}$$

Lastly, denote

$$\begin{aligned}I_n=&\frac{\lambda}{2}\|e_\phi^n\|^2+\frac{\lambda}{2}\|\nabla e_\phi^n\|^2+\frac{\rho_0}{2}\|e_u^n\|^2+\frac{\delta t^2}{2\rho_0}\|\nabla q^n\|^2+(\frac{\lambda}{2}+\frac{S\lambda}{\varepsilon^2})\|e_\phi^n-e_\phi^{n-1}\|^2,\\S_n=&\sum_{k=1}^{n+1}\left[\frac{\lambda}{2}\|\nabla e_\phi^{k+1}-\nabla e_\phi^k\|^2+\frac{\rho_0}{2}\|\tilde{e}_u^{k+1}-e_u^k\|^2+\frac{\mu_0}{4}\delta t\|\nabla\tilde{e}_u^k\|^2\right.\\&\left.+\frac{3\lambda\gamma}{4}\delta t\|\nabla e_w^{k+1}\|^2+\frac{3\lambda\gamma}{8}\delta t\|e_w^{k+1}\|^2\right],\quad n\ge1,\end{aligned} \tag{3.61}$$

with $S_0=0$ and $e_\phi^{-1}=0$. Summing (3.47) together for time steps 1 to n and combining the results above, i.e.(3.48),(3.49),(3.50),(3.53),(3.54) and (3.60), we derive that for $1\le n\le[T/\delta t]-1$,

$$I_{n+1}+S_{n+1}\le c\delta t^3(n+1)+c\delta t\sum_{k=0}^{n+1}I_k\le cT\delta t^2+c\delta t\sum_{k=0}^{n+1}I_k. \tag{3.62}$$

Since the constants appearing in the above inequalities are independent of δt, we derive from Gronwall's inequality that there exists some constant c such that

$$I_{n+1}+S_{n+1}\le c\delta t^2,\quad 1\le n\le[T/\delta t]-1. \tag{3.63}$$

Thus (3.24) holds.

Finally, the estimates for $\|e_{u,\delta t}\|_{H^1}$ can be derived from the inequality $\|P_H v\|_1 \lesssim \|v\|_1$ (cf. [14]) and the fact that $e_u^{n+1} = P_H \tilde{e}_u^{n+1}$.

Next, we consider the convex splitting scheme.

Proof of Lemma 1 for the convex splitting scheme (3.17)-(3.21)**.**

We define the truncation error R_ϕ^n $(n = 0, 1, \dots, N-1)$ for the phase equation (3.1):

$$R_\phi^{n+1} = \frac{1}{\delta t}(\phi(t^{n+1}) - \phi(t^n)) + u(t^{n+1}) \cdot \nabla\phi(t^n) - \gamma\Delta w(t^{n+1}), \tag{3.64}$$

the truncation error R_w^{n+1} $(n = 0, 1, \dots, N-1)$ for the chemical potential equation (3.2):

$$R_w^{n+1} = w(t^{n+1}) + \Delta\phi(t^{n+1}) - f_c(\phi(t^{n+1})) + f_e(\phi(t^n)), \tag{3.65}$$

the truncation error R_u^{n+1} $(n = 0, 1, \dots, N-1)$ for the momentum equation (3.3):

$$\begin{aligned} R_u^{n+1} =& \rho_0 \left(\frac{u(t^{n+1}) - u(t^n)}{\delta t} + u(t^n) \cdot \nabla u(t^{n+1}) \right) \\ & - \mu_0 \Delta u(t^{n+1}) + \nabla p(t^n) - \lambda w(t^{n+1}) \nabla\phi(t^n), \end{aligned} \tag{3.66}$$

and the truncation error R_p^{n+1} $(n = 0, 1, \dots, N-1)$ for the pressure correction (3.21):

$$R_p^{n+1} = \rho_0 \frac{u(t^{n+1}) - u(t^{n+1})}{\delta t} + \nabla(p(t^{n+1}) - p(t^n)) = \nabla(p(t^{n+1}) - p(t^n)). \tag{3.67}$$

Under Assumption A, we have

$$\|R_{u,\delta t}\|_{l^\infty(H^1(\Omega)^d)} + \|R_{\phi,\delta t}\|_{l^\infty(H^1(\Omega))} + \|R_{w,\delta t}\|_{l^\infty(H^1(\Omega))} + \|R_{p,\delta t}\|_{l^\infty(L^2(\Omega))} \lesssim \delta t. \tag{3.68}$$

Denote

$$\dot{\tilde{e}}_u^{n+1} = \frac{\tilde{e}_u^{n+1} - e_u^n}{\delta t} + u^n \cdot \nabla \tilde{e}_u^{n+1} + e_u^n \cdot \nabla u(t^{n+1}), \tag{3.69}$$

$$G_c^{n+1} = f_c(\phi(t^{n+1})) - f_c(\phi^{n+1}), \quad G_e^n = f_e(\phi(t^n)) - f_e(\phi^n) \quad n \geq 0. \tag{3.70}$$

Given the special form of f_c and f_e, we have

$$G_e^n = \frac{1}{\varepsilon^2} e_\phi^n, \qquad G_c^{n+1} = \frac{1}{\varepsilon^2}\left[3(\phi(t^{n+1}))^2 e_\phi^{n+1} - 3\phi(t^{n+1})(e_\phi^{n+1})^2 + (e_\phi^{n+1})^3\right]. \tag{3.71}$$

Subtracting (3.64), (3.65), (3.66) and (3.67) from (3.17), (3.18), (3.19) and (3.21), respectively, we get the following error equations for $n \geq 0$,

$$
\begin{cases}
\dfrac{e_\phi^{n+1} - e_\phi^n}{\delta t} + (u(t^{n+1}) \cdot \nabla\phi(t^n) - \tilde{u}^{n+1} \cdot \nabla\phi^n) - \gamma\Delta e_w^{n+1} = R_\phi^{n+1}, & (3.72)\\
e_w^{n+1} + \Delta e_\phi^{n+1} - G_c^{n+1} + G_e^n = R_w^{n+1}, & (3.73)\\
\rho_0 \dot{\tilde{e}}_u^{n+1} - \mu_0 \Delta\tilde{e}_u^{n+1} + \nabla q^n - \lambda(w(t^{n+1})\nabla\phi(t^n) - w^{n+1}\nabla\phi^n) = R_u^{n+1}, & (3.74)\\
\rho_0 \dfrac{e_u^{n+1} - \tilde{e}_u^{n+1}}{\delta t} + \nabla(q^{n+1} - q^n) = R_p^{n+1}, & (3.75)
\end{cases}
$$

with the boundary conditions

$$\tilde{e}_u^{n+1}|_{\partial\Omega} = 0, \quad \partial_{\mathbf{n}} e_\phi^{n+1}|_{\partial\Omega} = 0, \quad \partial_{\mathbf{n}} e_w^{n+1}|_{\partial\Omega} = 0. \tag{3.76}$$

We first establish the following results:

Lemma 4 *For $n \leq \frac{T}{\delta t} - 1$, we have*

$$\|G_c^{n+1}\| \leq c\|e_\phi^{n+1}\|_{H^1}, \qquad \|G_e^n\| \leq \frac{1}{\varepsilon^2}\|e_\phi^n\|, \qquad \|\nabla G_e^n\| \leq \frac{1}{\varepsilon^2}\|\nabla e_\phi^n\|; \tag{3.77}$$

and

$$\|e_\phi^{n+1}\|_{H^2} \leq c(\|e_w^{n+1}\|_{L^2} + \|e_\phi^{n+1}\|_{H^1} + \|e_\phi^n\|_{L^2} + \delta t), \tag{3.78}$$

where c depends on the initial data, T and Ω.

Proof The part for G_e^n is trivial since it is linear in e_ϕ^n. For G_c^{n+1}, we use (3.71) and Theorem 2 to deduce that

$$
\begin{aligned}
\|G_c^{n+1}\| \leq& c(\|e_\phi^{n+1}\|\|\phi(t^{n+1})\|_{L^\infty} + \|e_\phi^{n+1}\|_{L^4}^2\|\phi(t^{n+1})\|_{L^\infty} + \|e_\phi^{n+1}\|_{L^6}^3)\\
\leq& c(\|e_\phi^{n+1}\|\|\phi(t^{n+1})\|_{L^\infty} + \|e_\phi^{n+1}\|_{H^1}^2\|\phi(t^{n+1})\|_{L^\infty} + \|e_\phi^{n+1}\|_{H^1}^3)\\
\leq& c\|e_\phi^{n+1}\|_{H^1},
\end{aligned}
$$

where we have used the *a priori* bound $\|\phi^n\|_{H^1} \lesssim 1$ $(n \leq \frac{T}{\delta t} - 1)$ implied by the stability result Theorem 2.

Using the H^2 regularity results for elliptic equations, we conclude from (3.73) that

$$\|e_\phi^{n+1}\|_{H^2} \leq c(\|e_\phi^{n+1}\|_{L^2} + \|\Delta e_\phi^{n+1}\|_{L^2}),$$

and the claim follows immediately once we notice that

$$\begin{aligned}\|e_w^{n+1} - G_c^{n+1} + G_e^n - R_w^{n+1}\|_{L^2} \leq & \|e_w^{n+1}\|_{L^2} + \|G_c^{n+1}\|_{L^2} + \|G_e^n\|_{L^2} + \|R_w^{n+1}\|_{L^2} \\ \leq & c(\delta t + \|e_w^{n+1}\|_{L^2} + \|e_\phi^{n+1}\|_{H^1} + \|e_\phi^n\|_{L^2}).\end{aligned}$$

■

As in the proof of the stabilized scheme (3.7)-(3.11), it is not difficult to obtain

$$\begin{aligned}&\frac{\lambda}{2}(\|e_\phi^{n+1}\|^2 - \|e_\phi^n\|^2 + \|e_\phi^{n+1} - e_\phi^n\|^2) + \frac{\rho_0}{2}(\|\tilde{e}_u^{n+1}\|^2 - \|e_u^n\|^2 + \|\tilde{e}_u^{n+1} - e_u^n\|^2) \\ &+ \frac{\lambda}{2}(\|\nabla e_\phi^{n+1}\|^2 - \|\nabla e_\phi^n\|^2 + \|\nabla e_\phi^{n+1} - \nabla e_\phi^n\|^2) \\ &+ \mu_0 \delta t \|\nabla \tilde{e}_u^{n+1}\|^2 + \lambda\gamma\delta t\|e_w^{n+1}\|^2 + \lambda\gamma\delta t\|\nabla e_w^{n+1}\|^2 + \delta t(\nabla q^n, \tilde{e}_u^{n+1}) \\ &+ \delta t(e_u^n \cdot \nabla u(t^{n+1}), \tilde{e}_u^{n+1}) + \lambda\delta t(u(t^{n+1}) \cdot \nabla e_\phi^n, e_\phi^{n+1}) + \lambda\delta t(\tilde{e}_u^{n+1} \cdot \nabla\phi^n, e_\phi^{n+1}) \\ &+ \lambda\delta t(u(t^{n+1}) \cdot \nabla e_\phi^n, e_w^{n+1}) - \lambda\delta t(w(t^{n+1}) \cdot \nabla e_\phi^n, \tilde{e}_u^{n+1}) \\ =&\lambda\delta t(R_\phi^{n+1}, e_\phi^{n+1}) + \lambda\delta t(R_\phi^{n+1}, e_w^{n+1}) + \lambda\gamma\delta t(G_c^{n+1} - G_e^n, e_w^{n+1}) + \lambda\gamma\delta t(R_w^{n+1}, e_w^{n+1}) \\ &- \lambda(G_c^n - G_e^n, e_\phi^{n+1} - e_\phi^n) - \lambda(R_w^{n+1}, e_\phi^{n+1} - e_\phi^n) + \delta t(R_u^{n+1}, \tilde{e}_u^{n+1}).\end{aligned} \tag{3.79}$$

All the terms in the above can be controlled easily by following the arguments for stabilized scheme (3.7)-(3.11) except the two terms involving $G_c^{n+1} - G_e^n$.

(i) Using Lemma 4, we have for any $\alpha > 0$,

$$\begin{aligned}\lambda\gamma\delta t(G_c^{n+1} - G_e^n, e_w^{n+1}) \leq & c\delta t(\|G_e^n\| + \|G_c^{n+1}\|)\|e_w^{n+1}\| \\ \leq & c\delta t(\|e_\phi^n\| + \|e_\phi^{n+1}\|_{H^1})\|e_w^{n+1}\| \\ \leq & c_\alpha \delta t(\|e_\phi^n\|^2 + \|e_\phi^{n+1}\|^2 + \|\nabla e_\phi^{n+1}\|^2) + \alpha\lambda\gamma\delta t\|e_w^{n+1}\|^2.\end{aligned} \tag{3.80}$$

(ii) Next,

$$\begin{aligned}&- \lambda(G_c^{n+1} - G_e^n, e_\phi^{n+1} - e_\phi^n) \\ =&- \frac{\lambda}{4\varepsilon^2}(\|e_\phi^{n+1}\|_{L^4}^4 - \|e_\phi^n\|_{L^4}^4 + \|(e_\phi^{n+1})^2 - (e_\phi^n)^2\|^2 + 2\|e_\phi^{n+1}(e_\phi^{n+1} - e_\phi^n)\|^2) \\ &- \frac{\lambda}{\varepsilon^2}\|e_\phi^{n+1} - e_\phi^n\|^2 - \frac{\lambda}{\varepsilon^2}(3(\phi(t^{n+1}))^2 e_\phi^{n+1} - 3\phi(t^{n+1})(e_\phi^{n+1})^2 - e_\phi^{n+1}, e_\phi^{n+1} - e_\phi^n),\end{aligned} \tag{3.81}$$

where we have made use of the identity

$$a^3(a-b) = \frac{1}{2}a^2(a^2 - b^2 + |a-b|^2) = \frac{1}{4}(a^4 - b^4 + |a^2 - b^2|^2 + 2a^2|a-b|^2).$$

To bound the last term of the RHS in (3.81), we denote

$$\tilde{G}^{n+1} = -\frac{\lambda}{\varepsilon^2}(3(\phi(t^{n+1}))^2 e_\phi^{n+1} - 3\phi(t^{n+1})(e_\phi^{n+1})^2 - e_\phi^{n+1}). \tag{3.82}$$

Similar to Lemma 4, we can get

$$\|\tilde{G}^{n+1}\| \leq c\|e_\phi^{n+1}\|_{H^1}. \tag{3.83}$$

Taking gradient of $\tilde{G}^{n+1}$, we obtain

$$\begin{aligned}\nabla\tilde{G}^{n+1} =& -\frac{\lambda}{\varepsilon^2}\Big[(3(\phi(t^{n+1}))^2 - 1)\nabla e_\phi^{n+1} + 6\phi(t^{n+1})e_\phi^{n+1}\nabla\phi(t^{n+1}) \\ &\qquad - 3(e_\phi^{n+1})^2\nabla\phi(t^{n+1}) - 6\phi(t^{n+1})e_\phi^{n+1}\nabla e_\phi^{n+1}\Big].\end{aligned}$$

Noticing $H^2(\Omega) \subset L^\infty(\Omega)$ in $d = 2, 3$ dimensions, and recalling the *a-priori* bound of $\|\nabla e_\phi^{n+1}\|_{L^2}$ implied by Theorem 2, we deduce that

$$\|e_\phi^{n+1}\nabla e_\phi^{n+1}\|_{L^2} \leq \|e_\phi^{n+1}\|_{L^\infty}\|\nabla e_\phi^{n+1}\|_{L^2} \leq c\|e_\phi^{n+1}\|_{H^2}. \tag{3.84}$$

In view of Lemma 4, we have

$$\begin{aligned}\|\nabla\tilde{G}^{n+1}\| \leq& c\Big[(\|\phi(t^{n+1})\|_{L^\infty}^2 + 1)\|\nabla e_\phi^{n+1}\|_{L^2} + \|\phi(t^{n+1})\|_{L^\infty}\|\nabla\phi(t^{n+1})\|_{L^3}\|e_\phi^{n+1}\|_{L^6} \\ &\qquad + \|\nabla\phi(t^{n+1})\|_{L^6}\|e_\phi^{n+1}\|_{L^6}^2 + \|\phi(t^{n+1})\|_{L^\infty}\|e_\phi^{n+1}\nabla e_\phi^{n+1}\|_{L^2}\Big] \\ \leq& c\Big[(\|\phi(t^{n+1})\|_{H^2}^2 + 1)\|\nabla e_\phi^{n+1}\|_{L^2} + \|\phi(t^{n+1})\|_{H^2}^2\|e_\phi^{n+1}\|_{H^1} \\ &\qquad + \|\phi(t^{n+1})\|_{H^2}\|e_\phi^{n+1}\|_{H^1}^2 + \|\phi(t^{n+1})\|_{H^2}\|e_\phi^{n+1}\nabla e_\phi^{n+1}\|_{L^2}\Big] \\ \leq& c(\|e_w^{n+1}\|_{L^2} + \|e_\phi^{n+1}\|_{H^1} + \|e_\phi^n\|_{L^2} + \delta t).\end{aligned}$$

Taking inner product of (3.17) with $\tilde{G}^n$, in view of (3.83) and (3.84) and the fact that $\tilde{e}_u^{n+1} \in H_0^1(\Omega)^d$, we get

$$
\begin{aligned}
&\left(\tilde{G}^{n+1}, e_\phi^{n+1} - e_\phi^n\right) \\
=&\delta t\left(\tilde{G}^{n+1}, -u(t^{n+1})\cdot\nabla e_\phi^n - \tilde{e}_u^{n+1}\cdot\nabla\phi^n + \gamma\Delta e_w^{n+1} + R_\phi^{n+1}\right) \\
\le&\delta t\|\tilde{G}^{n+1}\|(\|R_\phi^{n+1}\| + \|u(t^{n+1})\|_{L^\infty}\|\nabla e_\phi^n\|) + \gamma\|\nabla e_w^{n+1}\|\|\nabla\tilde{G}^{n+1}\| \\
&+ \delta t\|\nabla\phi^n\|\|\tilde{e}_u^{n+1}\|_{H^1}\|\tilde{G}^{n+1}\|_{H^1} \qquad (3.85) \\
\le&c\delta t(\delta t^2 + \|e_\phi^n\|^2 + \|e_\phi^{n+1}\|_{H^1}^2) + \varepsilon_1\mu_0\|\nabla\tilde{e}_u^{n+1}\|^2 + \varepsilon_2\lambda\gamma\|e_w^{n+1}\|^2 + \varepsilon_3\lambda\gamma\|\nabla e_w^{n+1}\|^2,
\end{aligned}
$$

for arbitrary $\varepsilon_1, \varepsilon_2, \varepsilon_3 > 0$ and c depends on the initial data, domain Ω, time T and $\varepsilon_1, \varepsilon_2, \varepsilon_3$.

(iii) Following the proof of Lemma 1 for the stabilized scheme (3.7)-(3.10), we can bound the other terms as

$$
\begin{aligned}
&\lambda\delta t(R_\phi^{n+1}, e_\phi^{n+1}) + \lambda\delta t(R_\phi^{n+1}, e_w^{n+1}) + \lambda\gamma\delta t(R_w^{n+1}, e_w^{n+1}) - \lambda(R_w^{n+1}, e_\phi^{n+1} - e_\phi^n) \\
&+ \delta t(R_u^{n+1}, \tilde{e}_u^{n+1}) - \delta t(e_u^n\cdot\nabla u(t^{n+1}), \tilde{e}_u^{n+1}) - \lambda\delta t(u(t^{n+1})\cdot\nabla e_\phi^n, e_\phi^{n+1}) \\
&- \lambda\delta t(\tilde{e}_u^{n+1}\cdot\nabla\phi^n, e_\phi^{n+1}) - \lambda\delta t(u(t^{n+1})\cdot\nabla e_\phi^n, e_w^{n+1}) + \lambda\delta t(w(t^{n+1})\cdot\nabla e_\phi^n, \tilde{e}_u^{n+1}) \qquad (3.86) \\
&\le c\delta t\left(\delta t^2 + \|e_u^n\|^2 + \sum_{m=n,n+1}(\|e_\phi^m\|^2 + \|\nabla e_\phi^m\|^2)\right) + \frac{\mu_0\delta t}{4}\|\nabla\tilde{e}_u^{n+1}\|^2 + \frac{\lambda\gamma\delta t}{4}\|e_w^{n+1}\|^2,
\end{aligned}
$$

and

$$
\begin{aligned}
-\delta t\left(\nabla q^n, \tilde{e}_u^{n+1}\right) \le& -\frac{\delta t^2}{2\rho_0}\left(\|\nabla q^{n+1}\|^2 - \|\nabla q^n\|^2\right) - \frac{\rho_0}{2}\left(\|e_u^{n+1}\|^2 - \|\tilde{e}_u^{n+1}\|^2\right) \\
&+ c\delta t(\delta t^2 + \delta t^2\|\nabla q^n\|^2) + \frac{\mu_0\delta t}{4}\|\nabla\tilde{e}_u^{n+1}\|^2. \qquad (3.87)
\end{aligned}
$$

Choosing $\alpha = 1/8$ in (3.80) and $\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = 1/8$ in (3.85), combining (3.86) and (3.87), we have

$$\begin{aligned}
&\frac{\lambda}{2}(\|e_\phi^{n+1}\|^2 - \|e_\phi^n\|^2 + \|e_\phi^{n+1} - e_\phi^n\|^2) + \frac{\rho_0}{2}(\|e_u^{n+1}\|^2 - \|e_u^n\|^2 + \|\tilde{e}_u^{n+1} - e_u^n\|^2) \\
&+ \frac{\lambda}{2}(\|\nabla e_\phi^{n+1}\|^2 - \|\nabla e_\phi^n\|^2 + \|\nabla e_\phi^{n+1} - \nabla e_\phi^n\|^2) + \frac{5\mu_0\delta t}{8}\|\nabla \tilde{e}_u^{n+1}\|^2 \\
&+ \frac{\lambda\gamma\delta t}{2}\|e_w^{n+1}\|^2 + \frac{7\lambda\gamma\delta t}{8}\|\nabla e_w^{n+1}\|^2 + \frac{\lambda}{4\varepsilon^2}(\|e_\phi^{n+1}\|_{L^4}^4 - \|e_\phi^n\|_{L^4}^4 + \|(e_\phi^{n+1})^2 - (e_\phi^n)^2\|^2) \\
&+ \frac{\lambda}{2\varepsilon^2}\|e_\phi^{n+1}(e_\phi^{n+1} - e_\phi^n)\|^2 + \frac{\lambda}{\varepsilon^2}\|e_\phi^{n+1} - e_\phi^n\|^2 \\
&\le c\delta t\left(\delta t^2 + \delta t^2\|\nabla q^n\|^2 + \|e_u^n\|^2 + \sum_{m=n,n+1}(\|e_\phi^m\|^2 + \|\nabla e_\phi^m\|^2)\right). \qquad (3.88)
\end{aligned}$$

Then the desired results can be derived from the Gronwall inequality and the initial error $\|e_\phi^0\|_1^2 + \|e_u^0\|^2 + \delta t^2\|\nabla p^0\|^2 \lesssim \delta t^2$. The proof of Lemma 1 is now complete. ■

Remark 3.3.1 *Under the same assumption in Theorem 3,* (3.63) *and* (3.88) *imply the following estimates for both the stabilized scheme* (3.7)-(3.11) *and the convex splitting scheme* (3.17)-(3.21)*: for* $0 \le k \le N = [T/\delta t] - 1$*, we have*

$$\|\tilde{e}_{u,\delta t}^{k+1} - e_{u,\delta t}^k\|_{L^2(\Omega)^d} \lesssim \delta t, \qquad \delta t\sum_{k=0}^{N}\|\tilde{e}_{u,}^{k+1} - e_u^k\|_{L^2(\Omega)^d}^2 \lesssim \delta t^3. \qquad (3.89)$$

In addition, we can estimate the H^2 norm of the phase function errors.

Lemma 5 *Under the assumption of Theorem 3, for both stabilized scheme* (3.7)-(3.11) *and convex splitting scheme* (3.17)-(3.21)*, we have*

$$\|e_\phi^{n+1}\|_2 \lesssim \delta t^{1/2}, \quad n \le [T/\delta t] - 1. \qquad (3.90)$$

Proof We shall only prove the case for stabilized scheme (3.7)-(3.11), the proof for convex splitting scheme (3.17)-(3.21) is similar.

Applying H^2 theory to equation (3.33), we have

$$\|e_\phi^{n+1}\|_2 \le c(\|e_\phi^{n+1}\| + \|e_w^{n+1}\| + \|e_\phi^{n+1} - e_\phi^n\| + \|R_w^{n+1}\| + \|G^n\|).$$

Since $\|G^n\| \lesssim \|e_\phi^n\|$, by the results obtained in Lemma 1, we conclude that

$$\|e_\phi^{n+1}\|_2 \le c(\|e_\phi^{n+1}\| + \|e_w^{n+1}\| + \|e_\phi^{n+1} - e_\phi^n\| + \|R_w^{n+1}\| + \|e_\phi^n\|) \lesssim \delta t^{1/2}.$$

■

3.3.2 Improved pressure estimates

With Lemma 1, we can establish the following sub-optimal error bound for the pressure:

Lemma 6 *Under the assumption of Theorem 3, for both stabilized scheme* (3.7)-(3.11) *and convex splitting scheme* (3.17)-(3.21)*, we have the error for pressure as*

$$\|q_{\delta t}\|_{l^2(L^2(\Omega))} \lesssim \delta t^{1/2}. \tag{3.91}$$

Proof As before, we only show the case for the stabilized scheme (3.7)-(3.11), as the convex splitting scheme (3.17)-(3.21) can be analyzed similarly. Remark 3.3.1 ensures that

$$\|(\delta_t e_u)_{\delta t}\|_{l^\infty(L^2(\Omega)^d)} \lesssim \delta t, \qquad \|(\delta_t e_u)_{\delta t}\|_{l^2(L^2(\Omega)^d)} \lesssim \delta t^{3/2}. \tag{3.92}$$

Adding (3.35) to (3.34), we get

$$\begin{aligned} -\mu_0 \nabla^2 \tilde{e}_u^{n+1} + \nabla q^{n+1} &= h^{n+1}, \\ \nabla \cdot \tilde{e}^{n+1} = g^{n+1}, \quad \tilde{e}^{n+1}|_{\partial\Omega} &= 0, \end{aligned} \tag{3.93}$$

where

$$\begin{aligned} h^{n+1} &= \tilde{h}^{n+1} - \rho_0 \frac{e_u^{n+1} - e_u^n}{\delta t}, \\ \tilde{h}^{n+1} &= R_u^n + R_p^n - e_u^n \cdot \nabla u(t^{n+1}) - u^n \cdot \nabla \tilde{e}_u^{n+1} + \lambda(w(t^{n+1}) \nabla e_\phi^n + e_w^{n+1} \nabla \phi^n), \\ g^{n+1} &= \frac{\delta t}{\rho_0} \nabla^2 (p^{n+1} - p^n). \end{aligned} \tag{3.94}$$

Using the similar arguments in Lemma 1, we find

$$\|g^{n+1}\| = \|\nabla \cdot \tilde{e}_u^{n+1}\| \le \|\nabla \tilde{e}_u^{n+1}\| \lesssim \delta t^{1/2}, \quad \|\tilde{h}^{n+1}\|_{-1} \lesssim \delta t^{1/2}. \tag{3.95}$$

Then, we have

$$\|h^{n+1}\|_{-1} \le \|\tilde{h}^{n+1}\|_{-1} + \rho_0 \|\frac{e_u^{n+1} - e_u^n}{\delta t}\|_{-1}, \tag{3.96}$$

and it is not difficult to see

$$\|h_{\delta t}\|_{l^2(H^{-1}(\Omega)^d)} \lesssim \|\tilde{h}_{\delta t}\|_{l^2(H^{-1}(\Omega)^d)} + \frac{1}{\delta t} \|(\delta_t e_u)_{\delta t}\|_{l^2(L^2(\Omega)^d)} \lesssim \delta t^{1/2}. \tag{3.97}$$

Applying stand stability results for inhomogeneous Stokes system [14] to (3.93), it turns out

$$\|\tilde{e}_u^{n+1}\|_1 + \|q^{n+1}\| \lesssim \|h^{n+1}\|_{-1} + \|g^{n+1}\|, \tag{3.98}$$

and we have

$$\|q_{\delta t}\|_{l^2(L^2(\Omega))} \lesssim \delta t^{1/2}. \tag{3.99}$$

The sub-optimal error estimate for pressure is derived. ■

In order to derive improved estimates for pressure, we shall estimate errors for the time increment, and use a similar procedure as in [43]. The idea is to use the stability results for Stokes system, and improved L^2 estimate for $\delta_t e_u^n$. For a sequence of functions $\varphi^0, \varphi^1, \dots, \varphi^k, \dots$, we set

$$\delta_t \varphi^k = \varphi^k - \varphi^{k-1}. \tag{3.100}$$

Lemma 7 *Under the assumptions of Theorem 3, for both the stabilized scheme* (3.7)-(3.11) *and the convex splitting scheme* (3.17)-(3.21), *we have*

$$\|\delta_t e_u^n\| \lesssim \delta t^{3/2}, \quad \|\delta_t e_{u,\delta t}\|_{l^2(L^2(\Omega)^d)} \lesssim \delta t^2, \quad 1 \le n \le [T/\delta t]. \tag{3.101}$$

Proof We only present the proof for stabilized scheme (3.7)-(3.11), as the proof for convex splitting scheme (3.17)-(3.21) is similar to the proof shown below by combining the previous arguments in proving Lemma 1 together.

We work on equations for the time increment $\delta_t e_u^n$, $\delta_t e_\phi^n$ and $\delta_t e_w^n$. Denote

$$\epsilon_u^n = \delta_t e_u^n, \quad \tilde{\epsilon}_u^n = \delta_t \tilde{e}_u^n, \quad \epsilon_\phi^n = \delta_t e_\phi^n, \quad \epsilon_w^n = \delta_t e_w^n, \quad \psi^n = \delta_t q^n. \tag{3.102}$$

Applying time increment operator δ_t to (3.32)-(3.35), we have for $n \ge 1$

$$\begin{cases} \dfrac{\epsilon_\phi^{n+1} - \epsilon_\phi^n}{\delta t} - \gamma \Delta \epsilon_w^{n+1} = \delta_t R_\phi^{n+1} - \tilde{R}_\phi^{n+1}, & (3.103) \\ \epsilon_w^{n+1} - \dfrac{S}{\varepsilon^2}(\epsilon_\phi^{n+1} - \epsilon_\phi^n) + \Delta \epsilon_\phi^{n+1} = \delta_t R_w^{n+1} + \delta_t G^n, & (3.104) \\ \rho_0 \dfrac{\tilde{\epsilon}_u^{n+1} - \epsilon_u^n}{\delta t} - \mu_0 \Delta \tilde{\epsilon}_u^{n+1} + \nabla \psi^n = \delta_t R_u^{n+1} - \tilde{R}_{u,u}^{n+1} - \tilde{R}_{u,\phi}^{n+1}, & (3.105) \\ \rho_0 \dfrac{\epsilon_u^{n+1} - \tilde{\epsilon}_u^{n+1}}{\delta t} + \nabla(\psi^{n+1} - \psi^n) = \delta_t R_p^{n+1}, & (3.106) \end{cases}$$

where

$$\tilde{R}_{\phi}^{n+1} = \tilde{e}_u^n \cdot \nabla \delta_t \phi^n + \delta_t u(t^{n+1}) \cdot \nabla e_{\phi}^n + \tilde{\epsilon}_u^{n+1} \cdot \nabla \phi^n + u(t^n) \cdot \nabla \epsilon_{\phi}^n, \tag{3.107}$$

$$\delta_t G^n = f(\phi(t^n)) - f(\phi^n) - f(\phi(t^{n-1})) + f(\phi^{n-1}), \tag{3.108}$$

$$\tilde{R}_{u,u}^{n+1} = \rho_0 \delta_t u^n \cdot \nabla \tilde{e}_u^n + \rho_0 e_u^n \cdot \nabla \delta_t u(t^{n+1}) + \rho_0 u^n \cdot \nabla \tilde{\epsilon}_u^{n+1} + \rho_0 \epsilon_u^n \cdot \nabla u(t^n), \tag{3.109}$$

$$\tilde{R}_{u,\phi}^{n+1} = -\lambda e_w^n \nabla \delta_t \phi^n - \lambda \delta_t w(t^{n+1}) \nabla e_{\phi}^n - \lambda w(t^n) \nabla \epsilon_{\phi}^n - \lambda \epsilon_w^{n+1} \nabla \phi^n. \tag{3.110}$$

Next, we proceed as in the proof of Lemma 1. Thanks to Lemma 5, we can avoid the estimates for $\|\epsilon_w^n\|_1$, which leads to a simplified proof.

(i) For $n = 1$, from (3.62), there holds

$$\|e_w^1\|_1 \lesssim \delta t, \quad \|e_{\phi}^1\|_1 \lesssim \delta t^{3/2}, \quad \|e_u^1\|_1 \lesssim \delta t. \tag{3.111}$$

Letting $n = 0$ in (3.32)-(3.35), we have

$$\begin{cases} \dfrac{e_{\phi}^1}{\delta t} + \tilde{e}_u^1 \cdot \nabla \phi^0 - \gamma \Delta e_w^1 = R_{\phi}^1, & (3.112) \\ e_w^1 - \dfrac{S}{\varepsilon^2} e_{\phi}^1 + \Delta e_{\phi}^1 = R_w^1, & (3.113) \\ \rho_0 \dfrac{\tilde{e}_u^1}{\delta t} + \rho_0 (u^0 \cdot \nabla) \tilde{e}_u^1 - \mu_0 \Delta \tilde{e}_u^1 - \lambda e_w^1 \nabla \phi^0 = R_u^1, & (3.114) \\ \dfrac{\rho_0}{\delta t}(e_u^1 - \tilde{e}_u^1) + \nabla q^1 = R_p^1. & (3.115) \end{cases}$$

Taking inner product of (3.112) with $\delta t e_{\phi}^1$, (3.113) with $\gamma \delta t \Delta e_w^1$, (3.114) with $\delta t \tilde{e}_u^1$, respectively, we have

$$\|e_{\phi}^1\|^2 + \gamma \delta t (\nabla e_w^1, \nabla e_{\phi}^1) = -\delta t (\tilde{e}_u^1 \cdot \nabla \phi^0, e_{\phi}^1) + \delta t (R_{\phi}^1, e_{\phi}^1), \tag{3.116}$$

$$-\gamma \delta t (\nabla e_w^1, \nabla e_{\phi}^1) + \frac{S \gamma \delta t}{\varepsilon^2} \|\nabla e_{\phi}^1\|^2 + \gamma \delta t \|\Delta e_{\phi}^1\| = \gamma \delta t (\Delta R_w^1, e_{\phi}^1), \tag{3.117}$$

$$\rho_0 \|\tilde{e}_u^1\|^2 + \mu_0 \delta t \|\nabla \tilde{e}_u^1\|^2 = \delta t (R_u^1, \tilde{e}_u^1) + \lambda \delta t (e_w^1 \nabla \phi^0, \tilde{e}_u^0). \tag{3.118}$$

Here we have used the fact that $\partial_{\mathbf{n}} e_w^1|_{\partial\Omega} = 0$ and $e_w^1 \in H^2(\Omega)$, since $e_w^1 = f(\phi(t^1)) - f(\phi^0) - \frac{S}{\varepsilon^2}(\phi(t^1) - \phi^0)$ and f' is Lipschitz. It is also clear that $\|e_w^1\|_2 \lesssim \delta t$ if $\phi_t(t) \in H^2(\Omega)$.

Combining (3.116)−(3.117)−(3.118), and applying Cauchy inequality and Sobolev inequality, we get

$$\begin{aligned}
&\|e_\phi^1\|^2 + \gamma\delta t\|\Delta e_\phi^1\|^2 + \frac{S\gamma\delta t}{\varepsilon^2}\|\nabla e_\phi^1\|^2 + \rho_0\|\tilde{e}_u^1\|^2 + \mu_0\delta t\|\nabla\tilde{e}_u^1\|^2\\
=& -\delta t(\tilde{e}_u^1\cdot\nabla\phi^0, e_\phi^1) + \delta t(R_\phi^1, e_\phi^1) + \gamma\delta t(\Delta R_w^1, e_\phi^1) + \delta t(R_u^1, \tilde{e}_u^1) + \lambda\delta t(e_w^1\nabla\phi^0, \tilde{e}_u^1)\\
\leq& c\delta t\|\nabla\phi^0\|_{L^3}\|e_\phi^1\|_1\|\tilde{e}_u^1\| + c(\delta t^2\|R_\phi^1\|^2 + \delta t^2\|R_w^1\|_2^2) + \frac{1}{2}\|e_\phi^1\|^2 + c\delta t^2\|R_u^1\|^2\\
&+\frac{\rho_0}{4}\|\tilde{e}_u^1\|^2 + c\delta t\|e_w^1\|_1\|\nabla\phi^0\|_{L^3}\|\tilde{e}_u^1\|\\
\leq& c\delta t^2(\|e_\phi^1\|_1^2 + \|e_w^1\|_1^2 + \|R_\phi^1\|^2 + \|R_w^1\|_2^2 + \|R_u^1\|^2) + \frac{1}{2}\|e_\phi^1\|^2 + \frac{\rho_0}{2}\|\tilde{e}_u^1\|^2.
\end{aligned}$$

In view of (3.111), we have $\|e_\phi^1\|^2 + \|\tilde{e}_u^1\|^2 \lesssim \delta t^4$. By using the property of projection P_H, we have $\|\nabla q^1\| \lesssim \|R_p^1\| + \frac{\|\tilde{e}_u^1\|}{\delta t} \leq \delta t$. Hence, we have

$$\|\epsilon_\phi^1\|^2 + \|\epsilon_u^1\|^2 + \|\tilde{\epsilon}_u^1\|^2 + \delta t^2\|\nabla\psi^1\|^2 + \delta t\|\tilde{\epsilon}_u^1\|_1^2 + \delta t\|\Delta\epsilon_\phi^1\|^2 \lesssim \delta t^4. \tag{3.119}$$

(ii) Taking inner product of (3.103) with $\lambda\delta t\epsilon_\phi^{n+1}$, we obtain

$$\begin{aligned}
&\frac{\lambda}{2}(\|\epsilon_\phi^{n+1}\|^2 - \|\epsilon_\phi^n\|^2 + \|\epsilon_\phi^{n+1} - \epsilon_\phi^n\|^2) + \lambda\gamma\delta t(\nabla\epsilon_\phi^{n+1}, \nabla\epsilon_w^{n+1})\\
&\quad=\lambda\delta t(\delta_t R_\phi^{n+1} + \tilde{R}_\phi^{n+1}, \epsilon_\phi^{n+1}).
\end{aligned} \tag{3.120}$$

Taking inner product of (3.104) with $\lambda\gamma\delta t\epsilon_w^{n+1}$ and $\lambda\gamma\delta t\Delta\epsilon_\phi^{n+1}$, we get

$$\begin{aligned}
&\lambda\gamma\delta t\|\epsilon_w^{n+1}\|^2 - \frac{S\lambda\gamma}{\varepsilon^2}\delta t(\epsilon_\phi^{n+1} - \epsilon_\phi^n, \epsilon_w^{n+1}) - \lambda\gamma\delta t(\nabla\epsilon_\phi^{n+1}, \nabla\epsilon_w^{n+1})\\
&\quad=\lambda\gamma\delta t(\delta_t G^n, \epsilon_w^{n+1}) + \lambda\gamma\delta t(\delta_t R_w^{n+1}, \epsilon_w^{n+1}),
\end{aligned} \tag{3.121}$$

and

$$\begin{aligned}
&-\lambda\gamma\delta t(\nabla\epsilon_w^{n+1}, \nabla\epsilon_\phi^{n+1}) - \frac{S\lambda\gamma}{\varepsilon^2}\delta t(\epsilon_\phi^{n+1} - \epsilon_\phi^n, \Delta\epsilon_\phi^{n+1}) + \lambda\gamma\delta t\|\Delta\epsilon_\phi^{n+1}\|^2\\
&\quad= \lambda\gamma\delta t(\delta_t G^n, \Delta\epsilon_\phi^{n+1}) + \lambda\gamma\delta t(\delta_t R_w^{n+1}, \Delta\epsilon_\phi^{n+1}).
\end{aligned} \tag{3.122}$$

Taking inner product of (3.105) with $\delta t\tilde{\epsilon}_u^{n+1}$, we have

$$\begin{aligned}
&\frac{\rho_0}{2}(\|\tilde{\epsilon}_u^{n+1}\|^2 - \|\epsilon_u^n\|^2 + \|\tilde{\epsilon}_u^{n+1} - \epsilon_u^n\|^2) + \mu_0\delta t\|\nabla\tilde{\epsilon}_u^{n+1}\|^2 + \delta t(\psi^n, \tilde{\epsilon}_u^{n+1})\\
&\quad= \delta t(\delta_t R_u^n + \tilde{R}_{u,u}^n + \tilde{R}_{u,\phi}^n, \tilde{\epsilon}_u^{n+1}).
\end{aligned} \tag{3.123}$$

Then, summing up (3.120)+$\frac{1}{2}$(3.121)+$\frac{1}{2}$(3.122)+(3.123), we derive

$$\begin{aligned}
&\frac{\lambda}{2}(\|\epsilon_\phi^{n+1}\|^2 - \|\epsilon_\phi^n\|^2 + \|\epsilon_\phi^{n+1} - \epsilon_\phi^n\|^2) + \frac{\rho_0}{2}(\|\tilde{\epsilon}_u^{n+1}\|^2 - \|\epsilon_u^n\|^2 + \|\tilde{\epsilon}_u^{n+1} - \epsilon_u^n\|^2) \\
&+ \mu_0\delta t\|\nabla\tilde{\epsilon}_u^{n+1}\|^2 + \frac{\lambda\gamma\delta t}{2}\|\epsilon_w^{n+1}\|^2 + \frac{\lambda\gamma\delta t}{2}\|\Delta\epsilon_\phi^{n+1}\|^2 + \delta t(\nabla\psi^n, \tilde{\epsilon}_u^{n+1}) \\
=&\frac{S\lambda\gamma}{2\varepsilon^2}\delta t(\epsilon_\phi^{n+1} - \epsilon_\phi^n, \epsilon_w^{n+1}) + \lambda\delta t(\delta_t R_\phi^{n+1}, \epsilon_\phi^{n+1}) + \lambda\delta t(\tilde{R}_\phi^{n+1}, \epsilon_\phi^{n+1}) + \frac{\lambda\gamma\delta t}{2}(\delta_t G^n, \epsilon_w^{n+1}) \\
&+ \frac{\lambda\gamma\delta t}{2}(\delta_t R_w^{n+1}, \epsilon_w^{n+1}) + \delta t(\delta_t R_u^{n+1}, \tilde{\epsilon}_u^{n+1}) + \delta t(\tilde{R}_{u,u}^{n+1} + \tilde{R}_{u,\phi}^{n+1}, \tilde{\epsilon}_u^{n+1}) \\
&+ \frac{S\lambda\gamma}{2\varepsilon^2}\delta t(\epsilon_\phi^{n+1} - \epsilon_\phi^n, \Delta\epsilon_\phi^{n+1}) + \frac{\lambda\gamma\delta t}{2}(\delta_t G^n, \Delta\epsilon_\phi^{n+1}) + \frac{\lambda\gamma\delta t}{2}(\delta_t R_w^{n+1}, \Delta\epsilon_\phi^{n+1}).
\end{aligned} \tag{3.124}$$

All the terms on the RHS are easy to control except the third, fourth, seventh and ninth terms. Lemma 1 and 5, together with the assumptions on the exact solution imply the following estimates

$$\begin{aligned}
\|\delta_t\phi^n\|_1 \le \|\delta_t\phi(t^n)\|_1 + \|\delta_t e_\phi^n\|_1 \lesssim \delta t, \\
\|\phi^n\|_2 \le \|\phi(t^n)\|_2 + \|e_\phi^n\|_2 \le c, \\
\|\delta_t u^n\|_1 \le \|\delta_t u(t^n)\|_1 + \|\epsilon_u^n\|_1 \lesssim \delta t + \|\tilde{\epsilon}_u^n\|_1,
\end{aligned} \tag{3.125}$$

where we have also used the fact that $\epsilon_u^n = P_H(\tilde{\epsilon}_u^n)$ and that the projection P_H is stable in H^1. Using the property that f' is bounded and Lipschitz, we derive

$$\begin{aligned}
|\delta_t G^n| =& \left|\int_0^1 f'(s\phi(t^n) + (1-s)\phi^n)e_\phi^n\, ds - \int_0^1 f'(s\phi(t^{n-1}) + (1-s)\phi^{n-1})e_\phi^{n-1}\, ds\right| \\
=& \left|\int_0^1 f'(s\phi(t^n) + (1-s)\phi^n)\epsilon_\phi^n\, ds + e_\phi^{n-1}\int_0^1 \delta_t f'(s\phi(t^n) + (1-s)\phi^n))\, ds\right| \\
\le& \frac{c}{\varepsilon^2}\left(|\epsilon_\phi^n| + |e_\phi^{n-1}|\left(|\delta_t\phi(t^n)| + |\delta_t e_\phi^n|\right)\right).
\end{aligned}$$

Combining the above results with Lemma 1 and 5, we have

$$\begin{aligned}
&\frac{\lambda\delta t}{2}(\tilde{R}_\phi^{n+1}, \epsilon_\phi^{n+1}) \\
=&\frac{\lambda\delta t}{2}(\tilde{e}_u^n \cdot \nabla\delta_t\phi^n + \delta_t u(t^{n+1}) \cdot \nabla e_\phi^n + \tilde{\epsilon}_u^{n+1} \cdot \nabla\phi^n + u(t^n) \cdot \nabla\epsilon_\phi^n, \epsilon_\phi^{n+1}) \\
\le& c\delta t\big[\|\nabla\delta_t\phi^n\|\|\tilde{e}_u^n\|_1\|\epsilon_\phi^{n+1}\|_1 + \|\delta_t u(t^{n+1})\|_1\|\nabla e_\phi^n\|\|\epsilon_\phi^{n+1}\|_1 \\
&+ \|\nabla\phi^n\|\|\tilde{\epsilon}_u^{n+1}\|_1\|\epsilon_\phi^{n+1}\|_1 + \|u(t^n)\|_2\|\nabla\epsilon_\phi^n\|\|\epsilon_\phi^{n+1}\|\big] \\
\le& c\delta t\left(\delta t^2\|\tilde{e}_u^{n+1}\|_1^2 + \delta t^4 + \|\epsilon_\phi^{n+1}\|_1^2 + \|\epsilon_\phi^n\|_1^2\right) + \frac{\mu_0\delta t}{8}\|\nabla\tilde{\epsilon}_u^{n+1}\|^2.
\end{aligned}$$

$$
\begin{aligned}
\frac{\lambda\gamma\delta t}{2}(\delta_t G^n, \epsilon_w^{n+1}) \leq & c\delta t(\|\epsilon_\phi^n\|\|\epsilon_w^{n+1}\| + \|e_\phi^{n-1}\|_{L^6}\|\delta_t\phi(t^n)\|_{L^3}\|\epsilon_w^{n+1}\| + \|e_\phi^{n-1}\|_{L^6}\|\epsilon_\phi^n\|_{L^3}\|\epsilon_w^{n+1}\|) \\
\leq & c\delta t(\|\epsilon_\phi^n\|\|\epsilon_w^{n+1}\| + \|e_\phi^{n-1}\|_1\|\delta_t\phi(t^n)\|_1\|\epsilon_w^{n+1}\| + \|e_\phi^{n-1}\|_1\|\epsilon_\phi^n\|_1\|\epsilon_w^{n+1}\|) \\
\leq & c\delta t(\|\epsilon_\phi^n\|\|\epsilon_w^{n+1}\| + \delta t^2\|\epsilon_w^{n+1}\| + \delta t\|\epsilon_\phi^n\|_1\|\epsilon_w^{n+1}\|) \\
\leq & c\delta t(\delta t^4 + \|\epsilon_\phi^n\|_1^2) + \frac{\lambda\gamma\delta t}{8}\|\epsilon_w^{n+1}\|^2.
\end{aligned}
$$

For sufficiently small δt, we have

$$
\begin{aligned}
&\delta t(\tilde{R}_{u,u}^{n+1} + \tilde{R}_{u,\phi}^{n+1}, \tilde{\epsilon}_u^{n+1}) = \delta t(\tilde{R}_{u,u}^{n+1}, \tilde{\epsilon}_u^{n+1}) + (\tilde{R}_{u,\phi}^{n+1}, \tilde{\epsilon}_u^{n+1}) \\
\leq & c\delta t\big(\|\delta_t u^n\|_1\|\nabla\tilde{e}_u^n\|\,\|\tilde{\epsilon}_u^{n+1}\|_1 + \|e_u^n\|_1\|\nabla\delta_t u(t^{n+1})\|\|\tilde{\epsilon}_u^{n+1}\|_1 \\
&+ \|\epsilon_u^n\|\|u(t^n)\|_{W^{1,3}}\|\tilde{\epsilon}_u^{n+1}\|_1\big) + c\delta t\big(\|e_w^n\|_1\|\delta_t\phi^n\|_1\|\tilde{\epsilon}_u^{n+1}\|_1 + \|w(t^n)\|_{L^3}\|\nabla\epsilon_\phi^n\|\|\tilde{\epsilon}_u^{n+1}\|_1 \\
&+ \|\delta_t w(t^{n+1})\|_1\|\nabla e_\phi^n\|\|\tilde{\epsilon}_u^{n+1}\|_1 + \|\epsilon_w^{n+1}\|\|\phi^n\|_{W^{1,2d/(d-1)}}\|\tilde{\epsilon}_u^{n+1}\|_{L^{2d}}\big) \\
\leq & c\delta t\left((\delta t + \|\tilde{\epsilon}_u^n\|_1)\|\nabla\tilde{e}_u^n\|\|\tilde{\epsilon}_u^{n+1}\|_1 + \delta t\|e_u^n\|_1\|\tilde{\epsilon}_u^{n+1}\|_1 + \|\epsilon_u^n\|\|\tilde{\epsilon}_u^{n+1}\|_1\right) \\
&+ c\delta t\left(\delta t\|e_w^n\|\|\tilde{\epsilon}_u^{n+1}\|_1 + \delta t^2\|\tilde{\epsilon}_u^{n+1}\|_1 + \|\nabla\epsilon_\phi^n\|\|\tilde{\epsilon}_u^{n+1}\|_1 + \|\epsilon_w^{n+1}\|\|\tilde{\epsilon}_u^{n+1}\|_1^{1/2}\|\tilde{\epsilon}_u^{n+1}\|^{1/2}\right) \\
\leq & c\delta t\big(\delta t^2\|\tilde{\epsilon}_u^{n+1}\|_1^2 + \delta t^2\|\tilde{\epsilon}_u^n\|_1^2 + \delta t^2\|e_u^n\|_1^2 + \|\epsilon_u^n\|^2 + \|\tilde{\epsilon}_u^{n+1}\|^2 + \|\nabla\epsilon_\phi^n\|^2 \\
&+ \delta t^2\|e_w^n\|_1^2 + \delta t^4\big) + \frac{\lambda\gamma\delta t}{8}\|\epsilon_w^{n+1}\|^2 + \frac{\mu_0\delta t}{8}\|\nabla\tilde{\epsilon}_u^n\|^2 + \frac{\mu_0\delta t}{8}\|\nabla\tilde{\epsilon}_u^{n+1}\|^2.
\end{aligned}
$$

The ninth term in (3.124) can be controlled similarly as follows

$$
\begin{aligned}
\frac{\lambda\gamma\delta t}{2}(\delta_t G^n, \Delta\epsilon_\phi^{n+1}) \leq & c\delta t\big(\|\epsilon_\phi^n\|\|\Delta\epsilon_\phi^{n+1}\| + \|e_\phi^n\|_{L^6}\|\delta_t\phi(t^n)\|_{L^3}\|\Delta\epsilon_\phi^{n+1}\| \\
&+ \|e_\phi^n\|_{L^6}\|\epsilon_\phi^n\|_{L^3}\|\Delta\epsilon_\phi^{n+1}\|\big) \\
\leq & c\delta t(\delta t^4 + \|\epsilon_\phi^n\|_1^2) + \frac{\lambda\gamma\mu_0\delta t}{8}\|\Delta\epsilon_\phi^{n+1}\|^2.
\end{aligned}
$$

It remains to control $\delta t(\nabla\psi^n, \tilde{\epsilon}_u^{n+1})$ on the LHS of (3.126). We have the following result analogous to Lemma 1:

$$
\begin{aligned}
\delta t(\nabla\psi^n, \tilde{\epsilon}_u^{n+1}) = & \frac{\delta t^2}{2\rho_0}(\|\nabla\psi^{n+1}\|^2 - \|\nabla\psi^n\|^2) + \frac{\rho_0}{2}(\|\epsilon_u^{n+1}\|^2 - \|\tilde{\epsilon}_u^{n+1}\|^2) \\
&- \frac{\delta t^2}{\rho_0}(\delta_t R_p^{n+1}, \nabla\psi^n) - \delta t(\delta_t R_p^{n+1}, \tilde{\epsilon}_u^{n+1}),
\end{aligned}
$$

where the last two terms can be easily bounded by the Cauchy inequality.

Combining the above estimates into (3.124), we obtain

$$
\begin{aligned}
&\frac{\lambda}{2}(\|\epsilon_\phi^{n+1}\|^2 - \|\epsilon_\phi^n\|^2 + \|\epsilon_\phi^{n+1} - \epsilon_\phi^n\|^2) + \frac{\rho_0}{2}(\|\epsilon_u^{n+1}\|^2 - \|\epsilon_u^n\|^2 + \|\tilde{\epsilon}_u^{n+1} - \epsilon_u^n\|^2) \\
&+ \mu_0\delta t\|\nabla\tilde{\epsilon}_u^{n+1}\|^2 + \frac{\lambda\gamma\delta t}{2}\|\epsilon_w^{n+1}\|^2 + \frac{\lambda\gamma\delta t}{2}\|\Delta\epsilon_\phi^{n+1}\|^2 + + \frac{\delta t^2}{2\rho_0}(\|\nabla\psi^{n+1}\|^2 - \|\nabla\psi^n\|^2) \\
\le & c\delta t^3(\delta t^2 + \|\tilde{e}_u^{n+1}\|_1^2 + \|\tilde{e}_u^n\|_1^2 + \|e_u^n\|_1^2 + \|e_w^n\|_1^2) + c\delta t(\delta t^2\|\nabla\psi^n\|^2 + \|\tilde{\epsilon}_u^{n+1}\|^2 \\
&+ \|\epsilon_\phi^{n+1}\|^2 + \|\epsilon_\phi^n\|^2 + \|\epsilon_u^n\|^2) + \frac{\lambda\gamma\delta t}{4}\|\Delta\epsilon_\phi^{n+1}\|^2 + \frac{\lambda\gamma\delta t}{8}\|\Delta\epsilon_\phi^n\|^2 + \frac{\lambda\gamma\delta t}{4}\|\epsilon_w^{n+1}\|^2 \\
&+ \frac{\mu_0\delta t}{4}\|\nabla\tilde{\epsilon}_u^{n+1}\|^2 + \frac{\mu_0\delta t}{8}\|\nabla\tilde{\epsilon}_u^n\|^2,
\end{aligned}
\tag{3.126}
$$

where we have applied the inequality

$$
\|\nabla\epsilon_\phi^n\|^2 = -(\Delta\epsilon_\phi^n, \epsilon_\phi^n) \le \alpha\|\Delta\epsilon_\phi^n\|^2 + \frac{1}{4\alpha}\|\epsilon_\phi^n\|^2, \quad \forall\alpha > 0. \tag{3.127}
$$

Since $\|\tilde{\epsilon}_u^{n+1}\| \le \|\epsilon_u^n\| + \|\tilde{\epsilon}_u^{n+1} - \epsilon_u^n\|$, (3.126) would imply that for $n \ge 1$ and sufficiently small δt

$$
\begin{aligned}
&\frac{\lambda}{2}\|\epsilon_\phi^{n+1}\|^2 + \frac{\rho_0}{2}\|\epsilon_u^{n+1}\|^2 + \frac{\delta t^2}{2\rho_0}\|\nabla\psi^{n+1}\|^2 + \frac{1}{4}\sum_{k=2}^{n+1}\left(\lambda\|\epsilon_\phi^k - \epsilon_\phi^{k-1}\|^2 + \rho_0\|\tilde{\epsilon}_u^k - \epsilon_u^{k-1}\|^2\right) \\
&+ \sum_{k=2}^{n+1}\left(\frac{5\mu_0\delta t}{8}\|\nabla\tilde{\epsilon}_u^k\|^2 + \frac{\lambda\gamma\delta t}{8}\|\Delta\epsilon_\phi^k\|^2 + \frac{\lambda\gamma\delta t}{4}\|\epsilon_w^k\|^2\right) \\
\le & c\delta t^3\sum_{k=1}^{n+1}(\delta t^2 + \|\tilde{e}_u^k\|_1^2 + \|e_u^k\|_1^2 + \|e_w^k\|_1^2) + c\delta t\sum_{k=1}^{n+1}(\delta t^2\|\nabla\psi^k\|^2 + \|\epsilon_\phi^k\|^2 + \|\epsilon_\phi^k\|^2 + \|\epsilon_u^k\|^2) \\
&+ \frac{\lambda\gamma\delta t}{8}\|\Delta\epsilon_\phi^1\|^2 + \frac{\mu_0\delta t}{8}\|\nabla\tilde{\epsilon}_u^1\|^2 + \frac{\lambda}{2}\|\epsilon_\phi^1\|^2 + \frac{\rho_0}{2}\|\epsilon_u^1\|^2 + \frac{\delta t^2}{2\rho_0}\|\nabla\psi^1\|^2.
\end{aligned}
$$

Then, applying the Gronwall inequality to the above, and using Lemma 1 and the initial step estimate (3.119), we arrive at the desired result. ■

We are now ready to prove Theorem 3.

Proof [**Proof of Theorem 3**] We only show the case for the stabilized scheme (3.7)-(3.11), as the case for the convex splitting scheme (3.17)-(3.21) is very similar. Lemma 7 ensures that

$$
\|(\delta_t e_u)_{\delta t}\|_{l^\infty(L^2(\Omega)^d)} \lesssim \delta t^2, \qquad \|(\delta_t e_u)_{\delta t}\|_{l^2(L^2(\Omega)^d)} \lesssim \delta t^2. \tag{3.128}
$$

Using the similar arguments in Lemma 6, for the same Stokes system (3.93), we find

$$\|h_{\delta t}\|_{l^2(H^{-1}(\Omega)^d)} \lesssim \|\tilde{h}_{\delta t}\|_{l^2(H^{-1}(\Omega)^d)} + \frac{1}{\delta t}\|(\delta_t e_u)_{\delta t}\|_{l^2(L^2(\Omega)^d)} \lesssim \delta t. \tag{3.129}$$

Applying stand stability results for inhomogeneous Stokes system [14] to (3.93), there holds

$$\|\tilde{e}_u^{n+1}\|_1 + \|q^{n+1}\| \lesssim \|h^{n+1}\|_{-1} + \|g^{n+1}\|, \tag{3.130}$$

and we obtain

$$\|q_{\delta t}\|_{l^2(L^2(\Omega))} \lesssim \delta t. \tag{3.131}$$

The proof is complete. ■

3.4 Allen-Cahn Navier-Stokes equations

The Allen-Cahn phase-field model assumes the phase function ϕ follows the Allen-Cahn dynamics and the phase equations for ϕ can be written in a form similar to the Cahn-Hilliard Phase-field model (3.1)-(3.5). In detail, the Allen-Cahn Navier-Stokes equations are given by

$$\begin{cases} \phi_t + u \cdot \nabla\phi = \gamma(\Delta\phi - f(\phi)) \quad \text{in} \quad \Omega \subset \mathbb{R}^d, & (3.132) \\ \rho_0(u_t + (u\cdot\nabla)u) - \mu_0\Delta u + \nabla p - \lambda w \nabla\phi = 0, \quad \text{in} \quad \Omega \subset \mathbb{R}^d, & (3.133) \\ \nabla \cdot u = 0, \quad \text{in} \quad \Omega \subset \mathbb{R}^d, & (3.134) \\ u|_{\partial\Omega} = 0, \dfrac{\partial\phi}{\partial\mathbf{n}}|_{\partial\Omega} = 0, & (3.135) \end{cases}$$

with given initial data $u(0) = u_0$, $\phi(0) = \phi_0$. Note that the above system does not conserve the volume fraction, but this can be fixed by adding a Lagrange multiplier as in [29].

The above system satisfies the following energy law

$$\frac{d}{dt}\int_\Omega \left(\frac{1}{2}\rho_0|u|^2 + \frac{\lambda}{2}|\nabla\phi|^2 + \lambda F(\phi)\right) dx = -\int_\Omega \left(\mu_0|\nabla u|^2 + \frac{\lambda}{\gamma}|\phi_t + u\cdot\nabla\phi|^2\right) dx, \tag{3.136}$$

where $F(s) = \int_0^s f(\sigma)\, d\sigma$ and f are the same as those in the Cahn-Hilliard case.

The following stabilized scheme for (3.132)-(3.135) is introduced in [23]:

Given (u^n, ϕ^n, w^n, p^n), find $(\tilde{u}^{n+1}, \phi^{n+1})$ such that

$$\begin{cases} \frac{1}{\delta t}(\phi^{n+1} - \phi^n) + \tilde{u}^{n+1} \cdot \nabla\phi^n + \gamma w^{n+1} = 0, & (3.137) \\ w^{n+1} - \frac{S}{\gamma\varepsilon^2}(\phi^{n+1} - \phi^n) = -\Delta\phi^{n+1} + f(\phi^n), & (3.138) \\ \rho_0(\frac{\tilde{u}^{n+1} - u^n}{\delta t} + u^n \cdot \nabla\tilde{u}^{n+1}) - \mu_0\Delta\tilde{u}^{n+1} + \nabla p^n - \lambda w^{n+1}\nabla\phi^n = 0, & (3.139) \\ \frac{\partial\phi^{n+1}}{\partial\mathbf{n}}|_{\partial\Omega} = 0, \tilde{u}^{n+1}|_{\partial\Omega} = 0. & (3.140) \end{cases}$$

Given $(\tilde{u}^{n+1}, p^n)$, find (u^{n+1}, p^{n+1}) such that

$$\begin{cases} \rho_0\frac{u^{n+1} - \tilde{u}^{n+1}}{\delta t} + \nabla(p^{n+1} - p^n) = 0, & (3.141) \\ \text{div } u^{n+1} = 0, & (3.142) \\ u^{n+1} \cdot \mathbf{n}|_{\partial\Omega} = 0. & (3.143) \end{cases}$$

Note that we assume f satisfies the property (3.13) for the above scheme (3.137)-(3.143). The stability result is shown in [23]:

Theorem 4 *For $S \geq \frac{1}{2}$, the scheme (3.137)-(3.143) is unconditionally energy stable in the following sense:*

$$\begin{aligned} &\left[\frac{\lambda}{2}\|\nabla\phi^{n+1}\|^2 + \frac{\rho_0}{2}\|u^{n+1}\|^2 + \lambda\left(F(\phi^{n+1}), 1\right)\right] + \frac{\delta t^2}{2\rho_0}\|\nabla p^{n+1}\|^2 \\ &\quad +\mu_0\delta t\|\nabla\tilde{u}^{n+1}\|^2 + \lambda\gamma\delta t\|w^{n+1}\|^2 + \frac{\rho_0}{2}(\|\tilde{u}^{n+1} - u^n\|^2) \\ &\leq \left[\frac{\lambda}{2}\|\nabla\phi^n\|^2 + \frac{\rho_0}{2}\|u^n\|^2 + \lambda\left(F(\phi^n), 1\right)\right] + \frac{\delta t^2}{2\rho_0}\|\nabla p^n\|^2, \quad n \geq 0. \end{aligned}$$

The convex splitting scheme for the Allen-Cahn phase field model can be formulated similar to the Cahn-Hilliard case. From t_n to t_{n+1} ($n \geq 0$), we first compute the phase and intermediate velocity. Given $(u^n, w^n, \nabla p^n, \phi^n)$, find $(\tilde{u}^{n+1}, w^{n+1}, \phi^{n+1})$.

$$\begin{cases} \frac{\phi^{n+1} - \phi^n}{\delta t} + \tilde{u}^{n+1} \cdot \nabla\phi^n + \gamma w^{n+1} = 0, & (3.144) \\ w^{n+1} = -\Delta\phi^{n+1} + \left(f_c(\phi^{n+1}) - f_e(\phi^n)\right), & (3.145) \\ \rho_0(\frac{\tilde{u}^{n+1} - u^n}{\delta t} + u^n \cdot \nabla\tilde{u}^{n+1}) - \mu_0\Delta\tilde{u}^{n+1} + \nabla p^n - \lambda w^{n+1}\nabla\phi^n = 0, & (3.146) \\ \frac{\partial\phi^{n+1}}{\partial\mathbf{n}}|_{\partial\Omega} = 0, \tilde{u}^{n+1}|_{\partial\Omega} = 0. & (3.147) \end{cases}$$

Given $(\tilde{u}^{n+1}, \nabla p^n)$, find $(u^{n+1}, \nabla p^{n+1})$.

$$\begin{cases} \rho_0 \dfrac{u^{n+1} - \tilde{u}^{n+1}}{\delta t} + \nabla(p^{n+1} - p^n) = 0, & (3.148) \\ \text{div } u^{n+1} = 0, & (3.149) \\ u^{n+1} \cdot \mathbf{n}|_{\partial\Omega} = 0. & (3.150) \end{cases}$$

By using essentially the same arguments in [24], we have the following results:

Theorem 5 *The scheme (3.144)-(3.150) is stable in the following sense:*

$$\begin{aligned} &\left[\frac{\lambda}{2}\|\nabla\phi^{n+1}\|^2 + \frac{\rho_0}{2}\|u^{n+1}\|^2 + \lambda\left(F(\phi^{n+1}), 1\right)\right] + \frac{\delta t^2}{2\rho_0}\|\nabla p^{n+1}\|^2 \\ &\quad + \mu_0 \delta t \|\nabla \tilde{u}^{n+1}\|^2 + \lambda\gamma\delta t\|w^{n+1}\|^2 + \frac{\rho_0}{2}(\|\tilde{u}^{n+1} - u^n\|^2) \\ &\leq \left[\frac{\lambda}{2}\|\nabla\phi^n\|^2 + \frac{\rho_0}{2}\|u^n\|^2 + \lambda\left(F(\phi^n), 1\right)\right] + \frac{\delta t^2}{2\rho_0}\|\nabla p^n\|^2, \quad n \geq 0. \end{aligned}$$

We now define the error functions e_ϕ^n, e_w^n, e_u^n, $\tilde{e}_u^n$ and q^n the same way as in the Cahn-Hilliard-Navier-Stokes case. Then by using essentially the same arguments as in the proof of Theorem 3, we can prove the following results:

Theorem 6 *Under the assumption that solution (ϕ, w, u, p) is smooth enough, the numerical solution (u^n, p^n, ϕ^n, w^n) of the stabilized scheme* (3.137)-(3.143) *or the convex splitting scheme* (3.144)-(3.150) *satisfies the following error estimates for* $0 \leq n \leq [\frac{T}{\delta t}]$*:*

$$\begin{aligned} \|e_{\phi,\delta t}\|_{l^\infty(H^1(\Omega))} + \|e_{u,\delta t}\|_{l^\infty(L^2(\Omega)^d)} + \|\tilde{e}_{u,\delta t}\|_{l^\infty(L^2(\Omega)^d)} &\lesssim \delta t, \\ \|e_{w,\delta t}\|_{l^2(L^2(\Omega))} + \|e_{u,\delta t}\|_{l^2(H^1(\Omega)^d)} + \|\tilde{e}_{u,\delta t}\|_{l^2(H^1(\Omega)^d)}\| + \|q_{\delta t}\|_{l^2(L^2(\Omega))} &\lesssim \delta t, \\ \|e_{w,\delta t}\|_{l^\infty(L^2(\Omega))} + \|e_{u,\delta t}\|_{l^\infty(H^1(\Omega)^d)} + \|\tilde{e}_{u,\delta t}\|_{l^\infty(H^1(\Omega)^d)} + \|q_{\delta t}\|_{l^\infty(L^2(\Omega))} &\lesssim \delta t^{1/2}. \end{aligned}$$

3.5 Concluding remarks

We have rigorously derived error estimates for energy stable time discretizations of a phase-field model for two-phase incompressible flow, including both the Navier-Stokes-Cahn-Hilliard equations and the Navier-Stokes-Allen-Cahn equations. In the energy stable schemes, the Cahn-Hilliard/Allen-Cahn equation was discretized by the stabilized scheme or the convex splitting scheme and the Navier-Stokes equation was discretized by the projection method. The main difficulties of the analysis were the the splitting error in the projection step and the coupling between the phase function and velocity. We derived the optimal convergence rates for both phase functions and velocity in the H^1 norm and pressure in the L^2 norm. The analysis was presented for the Navier-Stokes-Cahn-Hilliard equations and could be easily extended to the Navier-Stokes-Allen-Cahn equations.

4. SPECTRAL METHOD FOR COMPLEX GEOMETRIES

In this chapter, we develop an efficient algorithm for the following problem:

$$\alpha u - \beta \Delta u = f, \quad \text{in } \Omega, \tag{4.1}$$
$$u|_{\partial\Omega} = g,$$

where Ω is a smooth bounded domain and $\alpha > 0, \beta > 0$ and f and g are smooth functions.

4.1 Introduction

Spectral method has been used to solve the PDEs after it was introduced by Orszag. Its main advantage over other numerical methods like finite element, finite difference is that the convergence rate is exponential, not algebraic. Hence its computational cost is less than other methods. However one of the main disadvantage is that the spectral method can be only applied to regular geometries (i.e. rectangle, cube, disk,...). There has been many attempts to apply the spectral method to complex geometries.

In [44], the author used a mapping to transform complex geometries into regular geometries so that traditional spectral method could be applied. The other approach is called a fictitious domain method. The fictitious domain method has been studied extensively in seventies [45, 46]. In the fictitious domain method, the original domain Ω is embedded into a larger and regular extended domain $\tilde{\Omega}$. And the problem is solved on the extended domain. Since the extended domain is regular, one may use structured meshes or currently existing solvers.

In this section, we introduce two algorithms of spectral method for complex geometries. In [47], the author develops a spectral method for complex geometries.

The author embed the original domain into a square. And the spectral collocation method without the boundary condition is applied to solve the problem. The spectral collocation method uses the Lagrange polynomial as a basis function and the matrix introduced by the spectral collocation method is full. In this section, we use Legendre polynomials as a basis and weak formulation. As a result, we could obtain ten digit accuracy for some test problems.

The second algorithm in this section is based on splitting of the equation (4.1). In [9], authors split the equation (4.1) into two parts and use a Fourier spectral method and the boundary integral equation solver. The main idea of the algorithm introduced in [9] is that a function need to be extended periodically over the extended domain. However it is known that a periodic expansion of a non-periodic function suffers from Gibbs phenomenon [48] so the convergence is slow. Hence we suggest a fictitious domain method which only requires a smooth extension of a given function.

Note that both algorithms suggested in this section assume we have a smooth extension of a given function. And obtaining a smooth extension would be a further research topic.

This chapter is organized as follows. In Section 4.2, we apply spectral method to a fictitious domain method suggested in [46] and observe that the boundary condition on the fictitious domain introduces derivative discontinuity. In Section 4.3, we introduce the first algorithm. In section 4.4, we introduce the second algorithm and the choice of the extended domain. The conclusion follows in Section 4.5.

4.2 Fictitious domain method with boundary condition

In this section, we summarize the fictitious domain method developed in [46] and apply it to the spectral method. In [46], the authors consider the following fictitious domain problem, given f, find $\tilde{u} \in H_0^1(\tilde{\Omega})$ such that

$$\begin{aligned} &\alpha\tilde{u} - \beta\Delta\tilde{u} = \tilde{f} \quad \text{in } \tilde{\Omega}, \\ &\tilde{u}|_{\partial\Omega} = g, \end{aligned} \tag{4.2}$$

where $\tilde{f}$ is a smooth extension of f on $\tilde{\Omega}$. It is obvious that $\tilde{u}|_{\Omega}$ solves problem (4.1). To solve (4.2), we consider the following variational form. Find $(\tilde{u}, \lambda) \in H_0^1 \times H^{-1/2}(\partial\Omega)$ such that

$$\begin{aligned} &\int_{\tilde{\Omega}} \alpha\tilde{u}v + \beta\nabla\tilde{u}\cdot\nabla v dx - \int_{\partial\Omega} \lambda v ds = \int_{\tilde{\Omega}} \tilde{f} v dx, \\ &\int_{\partial\Omega} \tilde{u}\mu ds = \int_{\partial\Omega} g\mu ds, \end{aligned} \tag{4.3}$$

for all $(v, \mu) \in H_0^1(\tilde{\Omega}) \times H^{-1/2}(\partial\Omega)$. In [46], a finite element was used to solve (4.3) as a test space and trial space. Here we use the Legendre-Galerkin method as follows.

$$X_N = \text{span}\{(L_i(x) - L_{i+2}(x))(L_j(y) - L_{j+2}(y))|0 \le i, j \le N-2\}, \tag{4.4}$$

where L_i is ith Legendre polynomial. It is known that X_N is a dense subset of $H_0^1(\tilde{\Omega})$. And it is shown in [21] that $\{(L_i(x)-L_{i+2}(x))(L_j(y)-L_{j+2}(y))\}$ forms a very efficient basis for the Legendre-Galerkin method. Let $\gamma : [0, 2\pi] \to \mathbb{R}^2$ be a parametrization of $\partial\Omega$. Given N_θ, let

$$\Lambda_{N_\theta} = \{f|f \text{ is a function from } \{\gamma(0), \gamma(\frac{2\pi}{N_\theta}), \cdots, \gamma(\frac{2(N_\theta - 1)\pi}{N_\theta})\} \text{ to } \mathbb{R}\}. \tag{4.5}$$

Using definition (4.4) and (4.5), we can define a discrete form of (4.3): find $(\tilde{u}_N, \lambda_{N_\theta}) \in X_N \times \Lambda_{N_\theta}$

$$\begin{aligned} &\int_{\tilde{\Omega}} \alpha\tilde{u}_N v_N + \beta\nabla\tilde{u}_N\cdot\nabla v_N dx + \int_{\partial\Omega} \lambda_{N_\theta} v ds = \int_{\tilde{\Omega}} I_N\tilde{f} v_N dx, \\ &\int_{\partial\Omega} \tilde{u}_N\mu_{N_\theta} ds = g\mu_{N_\theta} ds, \end{aligned} \tag{4.6}$$

for all $(v_N, \mu_{N_\theta}) \in X_N \times \Lambda_{N_\theta}$. In the following subsection, we observe the result of this numerical experiment of (4.6).

4.2.1 Numerical experiment

In this section, we perform a numerical test to observe the efficiency of the scheme (4.6). The domain Ω of the test problem is a circle which is centered at $(0, 0)$ with a radius $\frac{1}{2}$ and $\tilde{\Omega} = (-1, 1)^2$. Coefficients are $\alpha = \beta = 1$ and the exact solution is $u = sin(x + y)$.

Fig. 4.1 is a computed solution $\tilde{u}$ and an exact solution u on $\tilde{\Omega}$ and the error between two functions. Fig. 4.2 is a graph of the error as a function of N. Fig. 4.3 is a graph of the computed solution and its derivative along $y = 0$.

Fig. 4.1 shows that the computed solution $\tilde{u}_N$ and the exact solution u and their difference $u - \tilde{u}_N$ on $\tilde{\Omega}$ with $N = 60, N_\theta = 30$. The upper left graph is the computed solution $\tilde{u}_N$ and the upper right graph is the exact solution u. Since the exact solution does not belong to $H_0^1(\tilde{\Omega})$, two graphs are different. However, the lower left picture shows the error on Ω is negligible. In Fig. 4.2, the error on Ω as a function of N is shown to observe the order of accuracy. It is observed that the error decreases with the order $1/N$. Let h_N be the largest mesh size. Since $h_N \simeq 1/N$, (4.6) shows first-order accuracy, not spectral accuracy. The smoothness of $\tilde{u}_N$ gives us a better understanding of this situation.

The extended solution $\tilde{u}$ of (4.3) satisfies (4.1) and it also satisfies the following equation:

$$\begin{aligned} &\alpha w - \beta \Delta w = \tilde{f}, \qquad \text{in } \tilde{\Omega} \backslash \Omega, \\ &w|_{\partial\Omega} = g, \\ &w|_{\partial\tilde{\Omega}} = 0. \end{aligned} \tag{4.7}$$

Hence we have

$$\begin{aligned} \tilde{u}|_{\Omega} &= u, \\ \tilde{u}|_{\tilde{\Omega}\backslash\Omega} &= w. \end{aligned} \tag{4.8}$$

If the solution u of (4.1) and the solution w of (4.7) have different derivatives at $\partial\Omega$, $\tilde{u}$ cannot be smooth on $\tilde{\Omega}$. In this example, $\tilde{u}$ has a discontinuity of derivatives at $\partial\Omega$ and we can observe this in Fig. 4.3. Fig. 4.3 is the computed solution $\tilde{u}_N$ and its derivatives along $y = 0$. Ω is a circle which is centered at the origin and with radius $\frac{1}{2}$ and discontinuity of derivatives is expected at $x = \pm\frac{1}{2}$. The first graph shows the computed solution along $y = 0$. It is continuous but piecewise smooth. In the second and third graphs, we observe the discontinuity of derivatives of $\tilde{u}_N$ at $x = \pm\frac{1}{2}$ as expected.

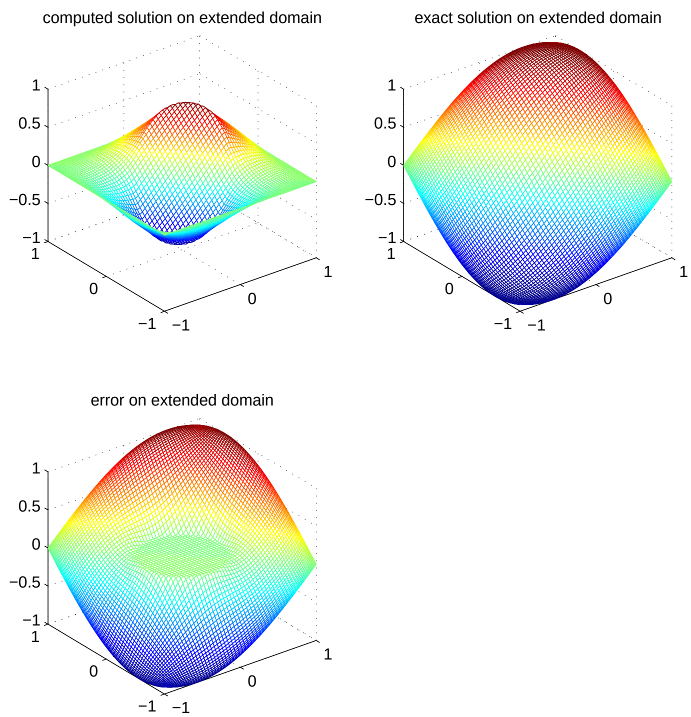

Figure 4.1. Exact and computed solutions and error on $\tilde{\Omega}$

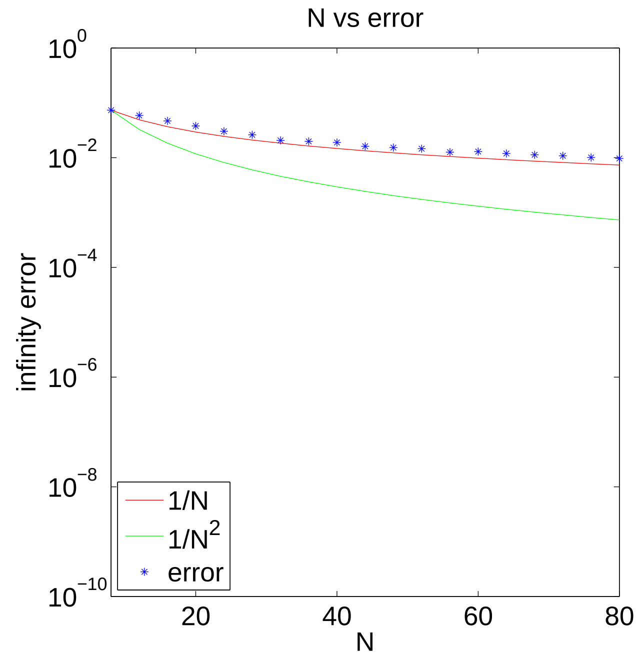

Figure 4.2. Error of computed solution

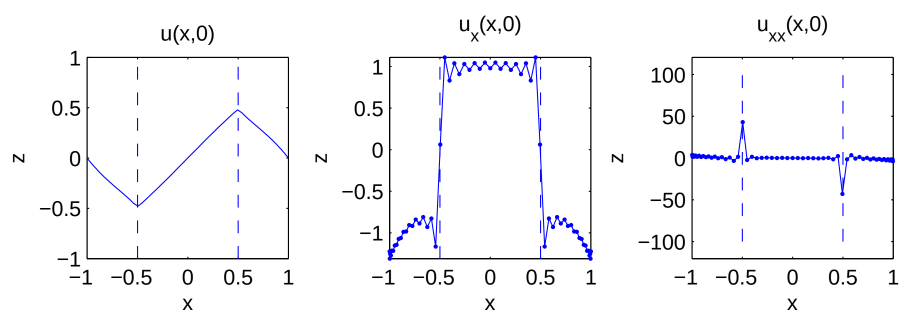

Figure 4.3. Derivatives of the computed solution

4.3 Fictitious domain methods without boundary condition

In Section 4.2, spectral accuracy was not obtained due to the discontinuity of derivatives. The main reason is boundary conditions are imposed on $\partial\Omega$ and $\partial\tilde{\Omega}$. In [47], the author applied a spectral collocation method with no boundary condition. In this section, we apply the Legendre-Galerkin method with no boundary condition on $\partial\tilde{\Omega}$. In this section, we consider the following formulation:

Given $\tilde{f}$, find $\tilde{u} \in H^2(\tilde{\Omega})$ such that

$$\begin{gathered} \int_{\tilde{\Omega}} (\alpha u - \beta\Delta\tilde{u}) v dx = \int_{\tilde{\Omega}} \tilde{f} v dx, \quad \text{ in } \tilde{\Omega}, \\ u|_{\partial\Omega} = g, \end{gathered} \tag{4.9}$$

for all $v \in L^2(\tilde{\Omega})$.

4.3.1 Description of method

To solve (4.9) numerically, we need to define the proper subspaces. We define the following discretized problem. Given $\tilde{f}$, find $u_N \in X_N$

$$\int_{\partial\Omega} (\alpha u_N - \beta\Delta u_N) v_N dx = \int_{\partial\Omega} I_N \tilde{f} v_N dx, \tag{4.10}$$

$$u_N(\gamma_k) = g(\gamma_k), \tag{4.11}$$

for all $v_N \in Y_N$ and $1 \leq k \leq M$ and γ_k is a discretization of $\partial\Omega$. The Legendre polynomial forms an orthogonal space with respect to the L^2-inner product. Hence the natural choice of X_N is

$$X_N = P_N \times P_N = \text{span}\{L_i(x)L_j(y) | 0 \leq i, j \leq N\}. \tag{4.12}$$

Note that choosing $Y_N = X_N$ would bring us overdetermined system. Hence we choose

$$Y_N = P_{N-2} \times P_{N-2} = \text{span}\{L_i(x)L_j(y) | 0 \leq i, j \leq N-2\}, \tag{4.13}$$

and we choose

$$M = 4N, \tag{4.14}$$

to make the system square. Hence the following system needs to be solved.

$$\begin{bmatrix} A \\ B \end{bmatrix} \bar{u} = \begin{bmatrix} y \\ z \end{bmatrix}, \tag{4.15}$$

where $A \in \mathbb{R}^{(N-1)^2 \times (N+1)^2}$ and $A\bar{u} = y$ correspond to (4.10) and $B \in \mathbb{R}^{4N \times (N+1)^2}$ and $B\bar{u} = z$ corresponds to (4.11).

The main purpose of this chapter is to observe the accuracy of the scheme. Hence we solve the system (4.15) by making $(N+1)^2$ by $(N+1)^2$ the full matrix and MATLAB backslash operator. The development of an efficient algorithm solving (4.15) could be future work.

4.3.2 Numerical experiment

In this section, we solve a test problem to observe the accuracy of the scheme. We solve the equation (4.10)-(4.11) on two different geometries. The domain is $\Omega_1 = \{(x,y)|(x/.6)^2 + (y/.9)^2 \leq 1\}$ and the exact solution is $u = sin(\pi/2(1 - (x/.6)^2 - (y/.9)^2))$. We call this an ellipse problem. The second geometry is $\Omega_2 = \{(x,y)|(x/.6)^2 + (y/.9)^2 \leq 1\}$ and the exact solution is $u = sin(\pi/2(1 - (x/.6)^2 - (y/.9)^2))$. We call this a circle problem. For both examples, we have $\alpha = 0, \beta = 1$.

Fig. 4.4 is the error on Ω_1 as a function of N. Fig. 4.5 is a few selected computed solutions of the ellipse problem on the extended domain $\tilde{\Omega}$ with various N. Fig. 4.6 is the error on Ω_2 as a function of N. Fig. 4.7 is a few selected computed solutions of the circle problem on $\tilde{\Omega}$ with various N.

In Fig. 4.4, we can observe two behaviors. When N is small, we can observe that the error decays very smoothly. As N gets larger, the error decays but behaves very unsmoothly. We can observe the behavior of the global solution to see why the error has such behavior. Fig. 4.5 shows the global solutions of the ellipse problem. It is observed that the computed solution converges to a smooth function when N is small, which reflects the smooth behavior of the error. However, it is observed that the global solution does not converge to a function. We conjecture that one reason is the well-posedness of the original problem (4.9). It is known that the solution exists and is unique on Ω, but its existence and uniqueness are not proved on $\tilde{\Omega}$. Hence if we could prove or disprove well-posedness of (4.9), we could have a clear understanding of this problem.

For the circle problem, we can observe spectral accuracy. This problem shows smooth error decay for small N in Fig. 4.6, but Fig. 4.7 shows that the computed global solution does not converge to a function for small N. Hence it is not always true that the global solution converges for small N.

In this section, we solved the problem (4.1) using the fictitious domain method. We discretized (4.9) and restricted the computed solution to Ω. For the example we tried, the global solution does not always converge to a function. And studying the well-posedness of (4.9) would give us better understanding of this behavior. In the following section, we a develop a fictitious method based on the splitting, which has well-posedness of PDEs.

4.4 Fictitious domain method using splitting

In [9], the authors developed a spectral embedding method to solve the advection-diffusion equation. The scheme is

1. Find smooth and periodic $\tilde{f}$ on $\tilde{\Omega}$ such that $\tilde{f}|_{\Omega} = f$.

2. Solve the following equation using the Fourier spectral method.

$$\begin{aligned} &\alpha\tilde{u} - \beta\Delta\tilde{u} = \tilde{f}, \text{ in } \tilde{\Omega}, \\ &\tilde{u}|_{\partial\tilde{\Omega}} = 0. \end{aligned} \tag{4.16}$$

3. In Ω, using a boundary element approach, find v such that

$$\begin{aligned} &\alpha v - \beta\Delta v = 0, \text{ in } \Omega \\ &v|_{\partial\Omega} = g - \tilde{u}. \end{aligned} \tag{4.17}$$

4. $u = \tilde{u} + v$.

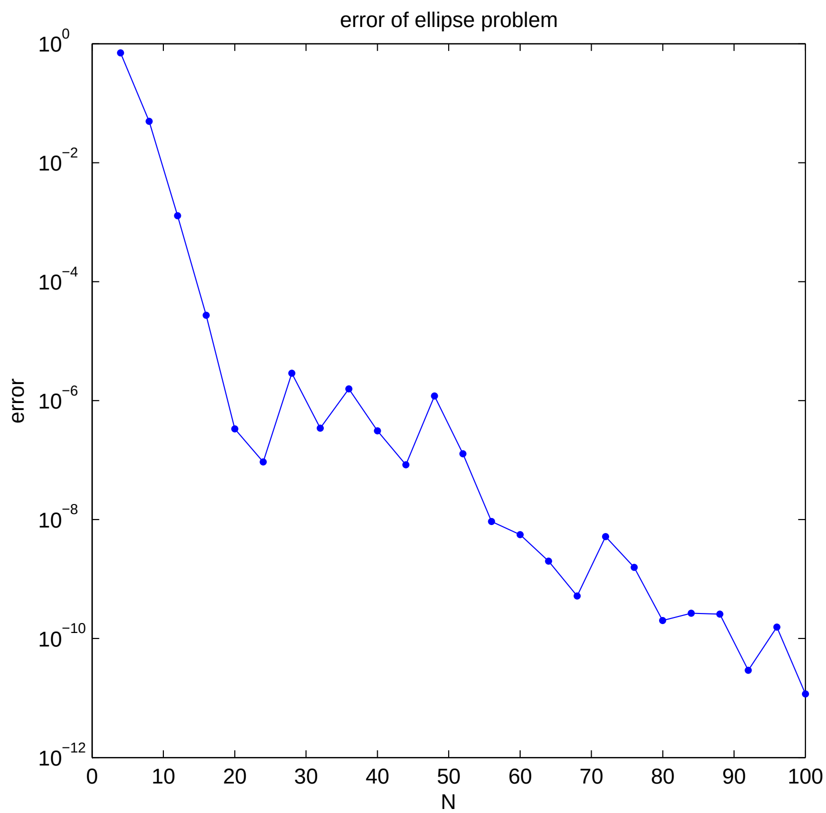

Figure 4.4. Error of the ellipse problem

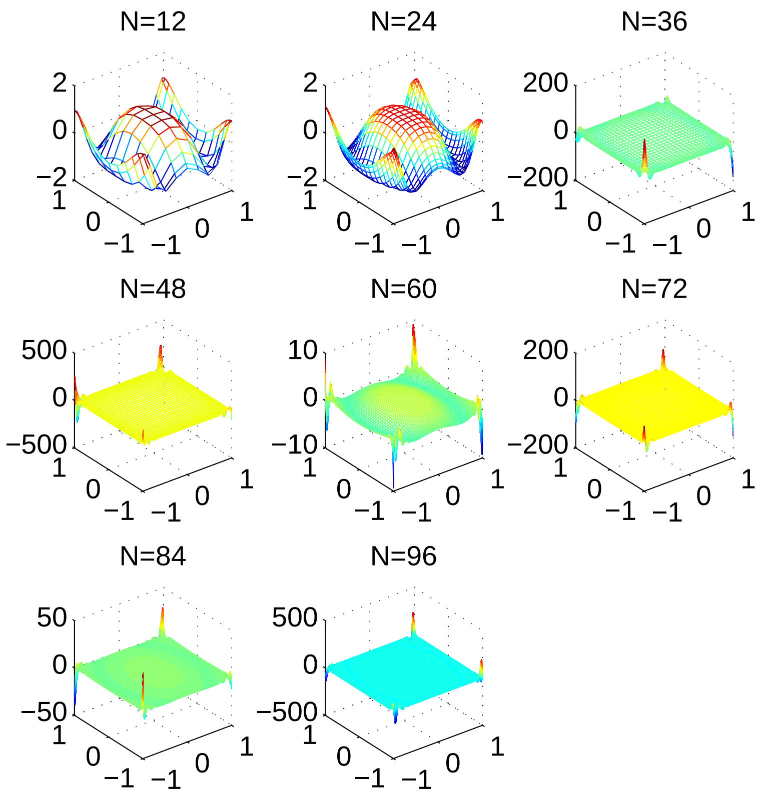

Figure 4.5. Global solutions of the ellipse problem

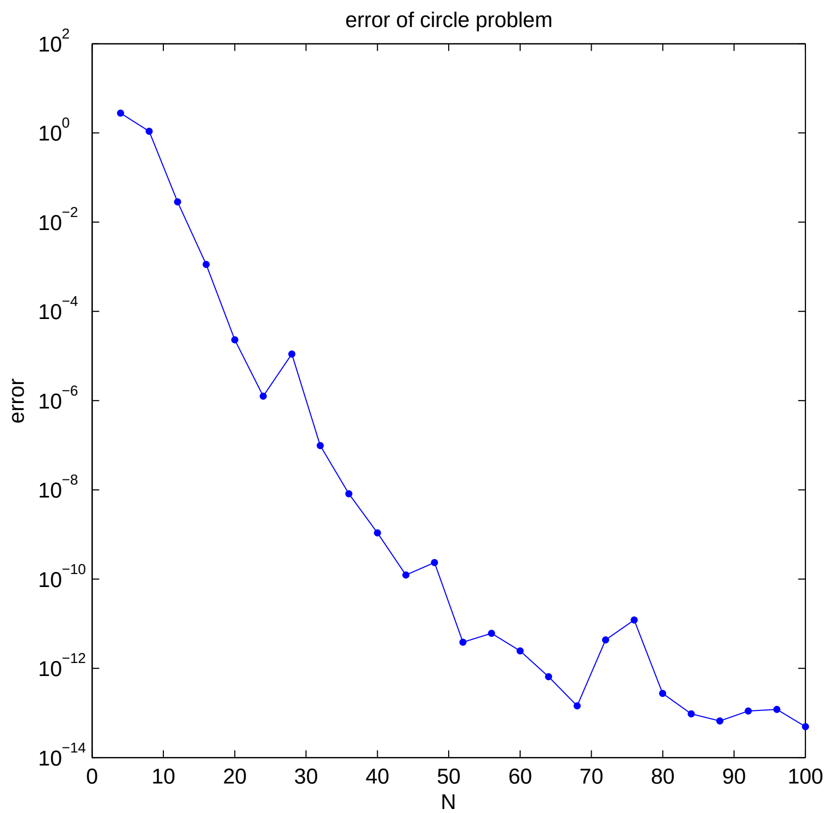

Figure 4.6. Error of the circle problem

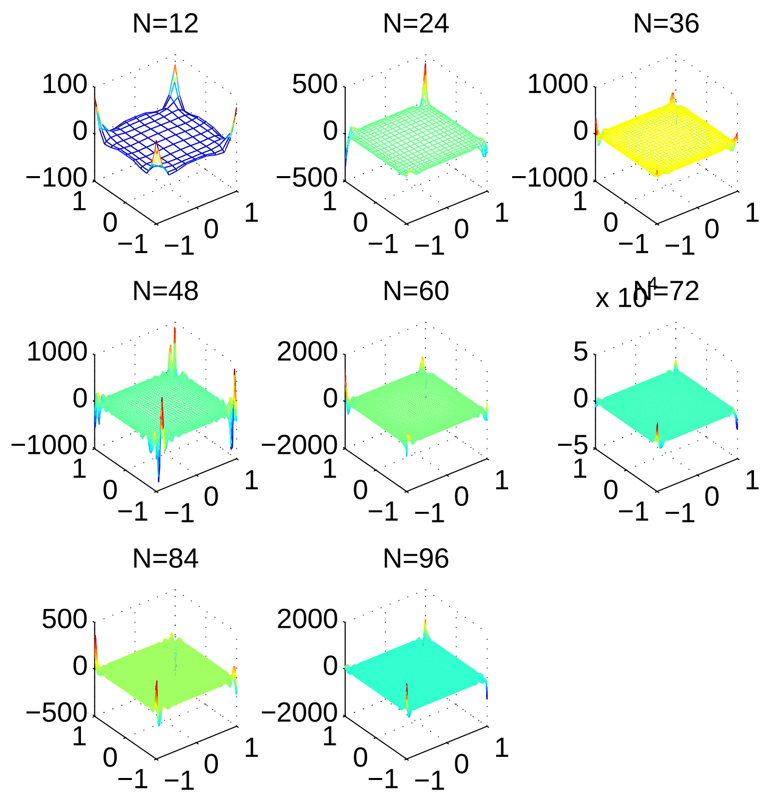

Figure 4.7. Global solutions of the circle problem

There are two main ideas. The first idea is that the authors obtain $\tilde{u}$ by the Fourier spectral method in step 2. The second idea is that the authors split the original equation into two equations (4.16) and (4.17) so that the boundary element solver could be used.

In this section we use a non-periodic solver in step 2 to avoid Gibss phenomenon. Hence the new scheme would be as follows.

1. Find smooth extension $\tilde{f}$ on $\tilde{\Omega}$ such that $\tilde{f}|_{\Omega} = f$.

2. Solve the following equation,

$$\begin{aligned} &\alpha\tilde{u} - \beta\Delta\tilde{u} = \tilde{f}, \text{ in } \tilde{\Omega}, \\ &\tilde{u}|_{\partial\bar{\Omega}} = 0. \end{aligned} \tag{4.18}$$

3. In Ω, using a boundary element approach, find v such that

$$\begin{aligned} &\alpha v - \beta\Delta v = 0, \text{ in } \Omega \\ &v|_{\partial\Omega} = g - \tilde{u}. \end{aligned} \tag{4.19}$$

4. $u = \tilde{u} + v$.

One of the main idea is the choice of the extended domain $\tilde{\Omega}$. In the following section, we show that the disk is a good choice rather than a square.

4.4.1 Choice of extended solver

In this subsection we discuss the choice of a proper extended domain $\tilde{\Omega}$. The Legendre-Galerkin method developed in [21] is very efficient in terms of complexity. Hence it is a good candidate among the existing spectral methods. However the domain is a d dimensional cube and this causes a singularity of the solution with smooth right-hand side function. We perform a numerical tests to observe this phenomenon. If $\tilde{f} = sin(\pi x)sin(\pi y)$, $u = \frac{1}{2\pi^2}sin(\pi x)sin(\pi y)$ solves (4.16) with $\alpha = 0$, $\beta = 1$. Since

u is smooth, we can expect spectral accuracy. However if $\tilde{f} = 1$, it is known in [21] that the following function,

$$u(x,y) = \frac{64}{\pi^2} \sum_{n,odd}^{\infty} \sum_{m,odd}^{\infty} (-1)^{\frac{1}{2}(n+m)} \frac{1}{nm(n^2+m^2)} cos(\frac{1}{2}n\pi x) cos(\frac{1}{2}m\pi y), \tag{4.20}$$

solves (4.16) with $\alpha = 0, \beta = 1, \bar{\Omega} = (-1,1)^2$. Fig. 4.8 shows the convergence rate of the numerical solution where the exact solution of $\tilde{f} = 1$ case is chosen for $N = 200$. If $\tilde{f} = sin(\pi x)sin(\pi y)$, we need twenty points to obtain round-off errors and we obtain spectral accuracy. However, we obtain algebraic accuracy for $\tilde{f} = 1$ case and we need $N = 100$ to obtain a ten-digit accuracy. Hence it is natural to try a different type of spectral method.

The main reason for algebraic accuracy is non-smoothness of the exact solution, and this was caused by corners of the domain. Since it is known in [49] that the solution is smooth at the boundary if the domain is smooth, we can avoid this problem by selecting a disk as an extended domain $\tilde{\Omega}$. We perform the same experiment on a disk. We use the pseudospectral method for disk essentially developed in [50] and [51]. The main idea is we use a polar coordinate with $-1 \le r \le 1$ and $0 \le \theta \le \pi$ to avoid singularity at $r = 0$, and we use a polynomial expansion in r direction and a Fourier expansion in θ direction as a basis function. The detailed implementation can be found in [51], chapter 11.

Fig. 4.9 shows a convergence rate of (4.16) with a different right-hand side $\alpha = 0, \beta = 1$. For both cases, we observe spectral accuracy. It is observed that we obtain round-off error for the case of $\tilde{f} = 1$ with only a few points. The polar coordinate form of (4.16) is

$$\begin{aligned} \alpha u - \beta(u_{rr} + \frac{1}{r}u_r + \frac{1}{r^2}u_{\theta\theta}) &= 1, \\ u|_{\tilde{\Omega}} &= 0. \end{aligned} \tag{4.21}$$

Because $u = \frac{1}{4}(1-r^2)$ solves (4.21) for $\alpha = 0$, $\beta = 1$ and u is a low degree polynomial. Hence we obtain round-off error with only a few points using the pseudospectral method.

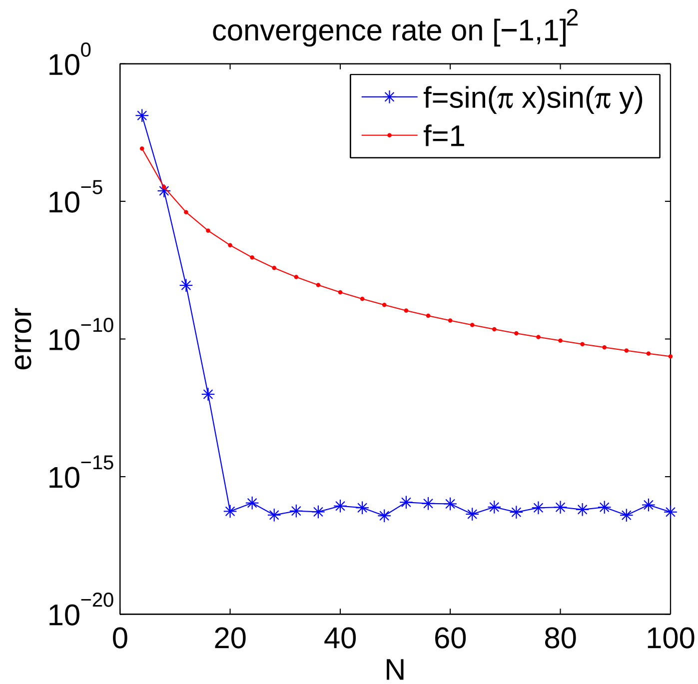

Figure 4.8. Convergence rate for $[-1, 1]^2$

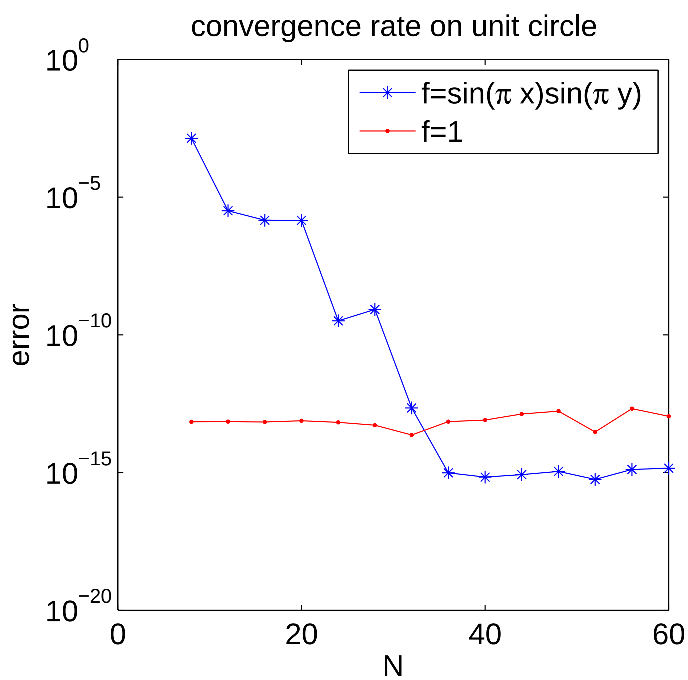

Figure 4.9. Convergence rate for disk

4.4.2 Boundary Integral Equation

The advantage of the method suggested in this section is that we can use existing solvers as a blackbox. In this subsection, we briefly explore one boundary integral equation solver introduced in [52]. We consider the following model problem:

$$\begin{aligned} -\Delta u &= 0, \quad \text{in } \Omega, \\ u|_{\partial\Omega} &= g. \end{aligned} \tag{4.22}$$

The double layer representation of (4.22) is

$$u(x) = \int_{\partial\Omega} \rho(s) \frac{\partial}{\partial n} log|x - s| ds, \quad x \in \Omega \tag{4.23}$$

where ρ is called a density function defined on $\partial\Omega$. And ρ is obtained by solving the following boundary integral equation:

$$-\pi\rho(t) + \int_{\partial\Omega} \rho(s) \frac{\partial}{\partial n} log|t - s| ds = g(t), \quad t \in \partial\Omega. \tag{4.24}$$

Hence the numerical scheme to solve (4.22) consists of two steps. We first obtain ρ from (4.24) and find $u(x)$ at $x \in \Omega$ by evaluating (4.23).

Let $r(t) = (\xi(t), \eta(t))$ be a parametrization of $\partial\Omega$ for $0 \leq t \leq 2\pi$. We can define K as follows:

$$K(t,s) = \begin{cases} \dfrac{\eta'(s)(\xi(t) - \xi(s)) - \xi'(s)(\eta(t) - \eta(s))}{(\xi(t) - \xi(s))^2 + (\eta(t) - \eta(s))^2}, & t \neq s, \\ \dfrac{\eta'(t)\xi''(t) - \xi'(t)\eta''(t)}{2(\xi('(t)^2 + \eta'(t)^2)}, & t = s. \end{cases} \tag{4.25}$$

And we can rewrite (4.24) as follows:

$$-\pi\rho(t) + \int_0^{2\pi} K(t,s)\rho(s) ds = g(t), \quad 0 \leq t \leq 2\pi. \tag{4.26}$$

Since K and functions g and ρ are periodic in $[0, 2\pi]$, the most efficient numerical method is a trapezoidal rule. For a given N, let $h = 2\pi/N$ and $t_j = jh$ for $j = 1, 2, \ldots,$. Then we can solve the system,

$$-\pi\rho_N(t_i) + h\sum_{j=1}^{N} K(t_i, t_j)\rho_N(t_j) = g(t_j), \quad j = 1, \ldots, N. \tag{4.27}$$

We can also use the trapezoidal rule to evaluate (4.22) if we have a density function ρ.

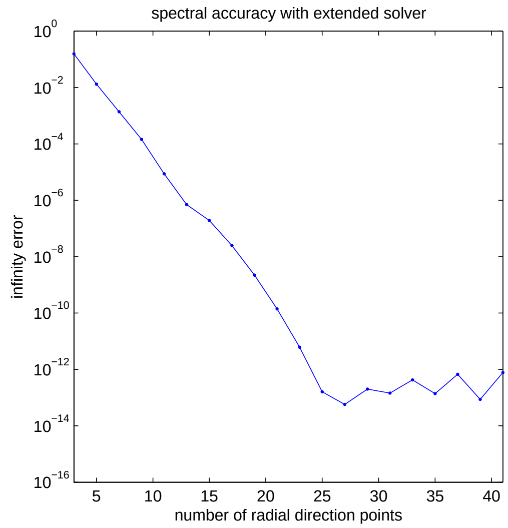

Figure 4.10. Convergence rate of extended domain solver

4.4.3 Numerical experiment

In this section we perform numerical tests to observe the order of accuracy. Since we have two different solvers, the extended domain solver and the BIE solver, we observe the convergence rate of each solver. The exact solution is $u(x, y) = e^{2x+y}$ and $\Omega = \{(x, y)|(x/a)^2 + (y/b)^2 = 1\}$ with $a = 7/8$ and $b = 1/2$ and the coefficients are $\alpha = 0, \beta = -1$ in (4.1).

Fig. 4.10 is the spectral accuracy of (4.18) part. In this example, we fix the accuracy of (4.19) solver to be the order of round-off error and increase the accuracy of (4.18). Fig. 4.11 is the spectral accuracy of (4.19). In this example, we fix accuracy of (4.18) solver to be the order of round-off error and increase the accuracy of (4.19). For both cases, we obtain spectral accuracy.

4.5 Concluding remarks

In this section, we developed two algorithms of spectral method for complex geometry. In Section 4.2, we applied a fictitious domain method with a homogeneous

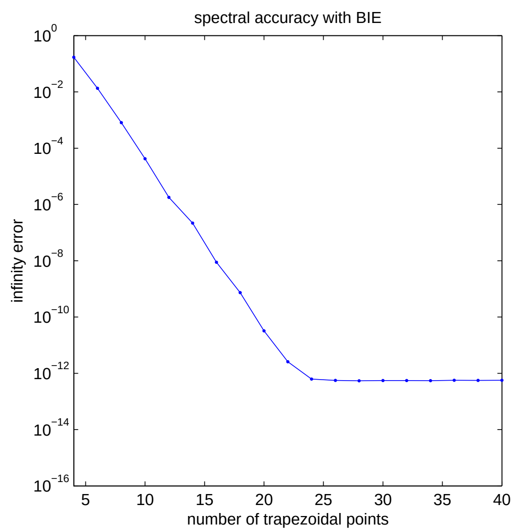

Figure 4.11. Convergence rate of BIE solver

boundary condition on $\tilde{\Omega}$ to spectral method. Since the exact solution is not smooth, we obtained the first-order accuracy. In Section 4.3, we did not impose a boundary condition on $\partial\Omega$. A numerical experiment shows that the solution of the extended solver did not converge to a function, but we could observe a high order of accuracy. In Section 4.4, we developed a non-periodic version of the spectral embedding method developed in [9]. We showed that the disk is a good choice for the extended domain $\tilde{\Omega}$. In that case, we obtained spectral accuracy.

In this chapter, we assumed we have smooth extension $\tilde{f}$ on $\tilde{\Omega}$. So the smooth extension $\tilde{f}$ to extended domain $\tilde{\Omega}$ could be a further research topic. The application of both methods suggested in Section 4.3 and Section 4.4 to a three-dimensional problems or real world problems also needs to be investigated.

5. SECOND-ORDER DEFECT CORRECTION

We present a solver for an ordinary differential equation based on spectral deferred correction (SDC). SDC uses Euler's method as a corrector and we use another kind of corrector. Our corrector is a high-order method in the sense that order of accuracy increases by two at each correction, while one in SDC. If the underlying problem is a constant coefficient linear problem, the method is the same as the collocation method. Hence it is A-stable. The distribution of quadrature points can be arbitrary; therefore Gauss type points can also be used.

5.1 Introduction

In this paper, we consider an efficient and accurate numerical scheme for solving the following ordinary differential equation,

$$y'(t) = f(t, y(t)), t \in (0, T], \tag{5.1}$$

$$y(0) = y_0, \tag{5.2}$$

where $f : (0, T) \times \mathbb{C}^n \to \mathbb{C}^n$, $y : (0, T) \to \mathbb{C}^n$ is a smooth function and $y_0 \in \mathbb{C}^n$. There has been extensive research to construct efficient and high-order numerical methods for ordinary differential equations. There are two types of strategies to obtain a high order of accuracy. One strategy is to construct a numerical scheme with a high order of consistency error (i.e. Runge-Kutta or multi-step method). For a non-stiff problem, the explicit Runge-Kutta method is efficient. For the stiff problem, the implicit Runge-Kutta method can be used to avoid an excessively small time step. The other direction to obtain a high order of accuracy is the defect correction type method.

The defect correction method is a generic term as described in [13]. Given an approximate solution, one finds the 'defect' which shows how good the approximate solution is. Then one finds the correction quantity using the defect. One obtains a better solution by adding the given approximate solution and correction quantity.

In [12], Dutt, Greengard and Rockhlin developed the spectral defect correction method (SDC), which is based on the classical defect correction method. They improved the classical defect correction method in two ways. They discretized the interval using Gauss points, rather than equidistance points, and they changed the differential equation into a Picard integral equation. In the classical defect correction method, a polynomial interpolation was performed on the equidistance points which lead to the well-known Runge's phenomenon. Obtaining high order was not easy, but SDC could avoid the Runge's phenomenon and obtain high order by using Gauss points.

Following SDC, there has been extensive research to extend the method. In [53], the authors showed that SDC is equivalent to a preconditioned Neumann series expansion if the original problem is linear. And they used an iterative solver, GMRES, to accelerate the convergence and improved the stability and the accuracy. In [54], the authors applied the Krylov Deferred Correction (KDC) method developed in [53] to a Differential Algebraic Equation (DAE). In their paper, KDC could obtain an accuracy similar to the existing DAE method, but it could take a much larger time step. In [55], the author developed the Semi-Implicit Spectral Deferred Correction (SISDC) which is a semi-implicit version of SDC. The SISDC is efficient when the equation of interest can be split into a stiff part and a non-stiff part. By treating the stiff term implicitly and the non-stiff term explicitly, the computational costs are reduced.

In this chapter, we suggest a different type of corrector. The new corrector improves the existing SDC in two directions. One is A-stability and the other is the order of accuracy. Stability of scheme is very important to avoid an excessively small time-step for stiff problems. One way to measure stability of the scheme is to look at the stability region. If the stability region covers the left-half plane, it is called A-stable [11]. However, SDC is not A-stable. Our new type of corrector solves collocation method after one correction if the underlying problem is a constant-coefficient, linear problem. Hence its stability function is the same as the collocation method. Therefore it is A-stable. In terms of order of accuracy, we also have an advantage. Using our corrector, the formal order of accuracy increases by two at each correction whereas backward Euler's method improves by one. And the extra cost we need to pay is $O(2k_1^2 n) = O(n)$ flops on top of the cost of the backward Euler corrector where n is the size of the problem and k_1+1 is the number of the nodes in subinterval. Since the problem size is nk_1 and k_1 is small (usually $k_1 < 10$), the extra cost is negligible.

The rest of the chapter is arranged as follows. In Section 5.2, we briefly review the SDC and introduce the algorithm. In Section 5.3, the new corrector is introduced. In Section 5.4, numerical experiments are performed. In the final section, we discuss our conclusions and further research possibilities.

5.2 Spectral deferred correction

We have two different kinds of mesh. The coarse mesh is

$$0 = t_0 < \cdots < t_P = T, \tag{5.3}$$

where we use equidistant points for simplicity.

$$t_{p+1} - t_p = h,$$

for $p = 0, \cdots, P-1$. And each mesh $[t_p, t_{p+1}]$ has fine mesh defined by $t_{p,k} = t_p + c_k h$ for $k = 0, \cdots, k_1$ where c_i is an arbitrary increasing sequences with $0 = c_0 < \cdots < c_{k_1} = 1$. Usually they are chosen to be equidistance points or Legendre-Gauss-Lobatto points. Let $t_p = t_{p,0}$ for simplicity. Since SDC is a one-step method, we will consider the problem (5.4) in a subinterval $[t_p, t_{p+1}]$ throughout the chapter. First we change the differential equation into an integral equation as follows:

$$y(t) = y_0 + \int_{t_p}^{t} f(s, y(s))ds. \tag{5.4}$$

Let $y(t)$ be the exact solution and $y^{[0]}(t)$ be an approximate solution. Then we can compute the residual $r(t)$ defined as

$$r(t) = \int_{t_p}^{t} f(s, y^{[0]}(s))ds - (y^{[0]}(t) - y_0). \tag{5.5}$$

If we call $\delta(t) = y(t) - y^{[0]}(t)$ an error, we can substitute $y(t)$ in equation (5.4), then

$$\begin{aligned} \delta(t) &+ y^{[0]}(t) = y_0 + \int_{t_p}^{t} f(s, \delta(s) + y^{[0]}(s))ds, \\ \delta(t) &= \int_{t_p}^{t} f(s, \delta(s) + y^{[0]}(s)) - f(s, y^{[0]}(s))ds + \int_{t_p}^{t} f(s, y^{[0]}(s))ds - (y^{[0]}(t) - y_0). \end{aligned}$$

Then we obtain the equation for the error $\delta(t)$:

$$\begin{cases} \delta(t) - \displaystyle\int_{t_p}^{t} f(s, y^{[0]}(s) + \delta(s)) - f(s, y^{[0]}(s))ds = r(t), \\ \delta(t_p) = 0. \end{cases} \tag{5.6}$$

The residual $r(t)$ is something we can compute and we have equation for the error $\delta(t)$. Hence we can compute the $\delta(t)$ in principle. To implement SDC, we need proper discretization.

We will be working in the space $\mathbb{C}^{nk_1}$ with the following notation:

$$\vec{\eta} = (\vec{\eta}_k)_{1 \le k \le k_1} = (\vec{\eta}_1, \cdots, \vec{\eta}_{k_1})^T \text{ where } \vec{\eta}_i \in \mathbb{C}^n \text{ for } 1 \le i \le k_1. \tag{5.7}$$

We can define the discrete residual operator $r_{k_1} : \mathbb{C}^{nk_1} \to \mathbb{C}^{nk_1}$. Suppose that we have initial value y_0. Then the discrete residual is

$$r_{k_1}\vec{\eta} = (\int_{t_p}^{t_{p,i}} (f(t_p, y_0)l_0(s) + \sum_{k=1}^{k_1} f(t_{p,k}, \vec{\eta}_k)l_k(s))ds - (\vec{\eta}_i - y_0))_{1 \le i \le k_1}, \tag{5.8}$$

where

$$l_k(t) = \frac{\Pi_{j \ne k}(t - t_{p,j})}{\Pi_{j \ne k}(t_{p,k} - t_{p,j})}.$$

Note that (5.8) is a Lagrange polynomial interpolation of (5.5) on $t_{p,k}$ for $1 \le k \le k_1$. Suppose we have an approximate solution $\vec{y}^{[j-1]} \in \mathbb{C}^{nk_1}$ where

$$\vec{y}_k^{[j-1]} \sim y(t_{p,k}),$$

for $1 \leq k \leq k_1$. Using (5.8), we can define the discrete residual $\vec{r}^{[j-1]} \in \mathbb{C}^{nk_1}$ as follows:

$$\vec{r}^{[j-1]} = r_{k_1}\vec{y}^{[j-1]}. \tag{5.9}$$

Then the vector of error $\vec{\delta} \in \mathbb{C}^{nk_1}$ where

$$\vec{\delta}_k^{[j]} \sim (y(t_{p,k}) - \vec{y}_k^{[j-1]}),$$

for $1 \leq k \leq k_1$ can be solved by the Euler's method. In this chapter, we present the backward Euler's method for simplicity. The first error $\vec{\delta}_1^{[j]}$ which is an approximate error at $t_{p,1}$ can found by solving the following equation:

$$\vec{\delta}_1^{[j]} - (t_{p,1} - t_{p_0})(f(t_{p,1}, \vec{y}_1^{[j-1]} + \vec{\delta}_1^{[j]}) - f(t_{p,1}, \vec{y}_1^{[j-1]})) = \vec{r}_1^{[j-1]}, \tag{5.10}$$

and the approximate errors $\vec{\delta}_k^{[j]}$ at $t_{p,k}$ for $k = 2, \ldots, k_1$ can be found by solving the following equation successively:

$$\vec{\delta}_k^{[j]} - \vec{\delta}_{k-1}^{[j]} - (t_{p,k} - t_{p,k-1})(f(t_{p,k}, \vec{y}_k^{[j-1]} + \vec{\delta}_k^{[j]}) - f(t_{p,k}, \vec{y}_k^{[j-1]})) = \vec{r}_k^{[j-1]} - \vec{r}_{k-1}^{[j-1]}. \tag{5.11}$$

We can describe the algorithm for SDC in the following subsection.

5.2.1 Algorithm for SDC

SDC is a one-step method and its algorithm is completely described by local behavior, advancing from t_p to t_{p+1}.

Comment: Fix h, k_1, J.

Comment: Assume $y(t_p) = y_0$ is given. We want to find $y(t_{p+1}) = y(t_p + h)$:

1. Find an initial guess $\vec{y}_k^{[0]} \sim y(t_{p,k})$ using Euler's method for $1 \leq k \leq k_1$ by solving (5.4).

   for $j = 1, \cdots, J$

   (a) Compute the residual $\vec{r}^{[j-1]} = r_{k_1}\vec{y}^{[j-1]}$ by (5.8);

   (b) Find $\vec{\delta}_k^{[j]}$ for $1 \leq k \leq k_1$ by solving (5.10) or (5.11).

   (c) Update the solution: $\vec{y}^{[j]} = \vec{y}^{[j-1]} + \vec{\delta}^{[j]}$.

   end for

5.3 New scheme

In this section we introduce a new scheme and how to obtain the solution.

5.3.1 Corrector

In this section we describe a new corrector besides the Euler's method. There are two main ideas. One is using a linearly implicit formulation described in [12] Section 6.3. The other one is using a high-order integration rule other than the Euler's method. We consider the following approximation:

$$f(t, y^{[j-1]}(t) + \delta(t)) - f(t, y^{[j-1]}(t)) \sim J(t, y^{[j-1]}(t))\delta(t). \tag{5.12}$$

We can apply (5.12) to our new scheme. Hence we solve the equation

$$\delta(t) - \int_{t_p}^{t} J(s, y^{[j-1]}(s))\delta(s)ds = r(t), \tag{5.13}$$

instead of (5.6). J is the Jacobian as

$$J(t, y) = \frac{\partial f(t, y)}{\partial y}. \tag{5.14}$$

The other key is the time discretization of (5.13). Consider the following operator $\hat{K}_{\vec{y}^{[j-1]}} : \mathbb{C}^{nk_1} \to \mathbb{C}^{nk_1}$ defined by

$$\hat{K}_{\vec{y}^{[j-1]}}\vec{\eta} = (\int_{t_p}^{t_{p,i}} \sum_{k=1}^{k_1} J(t_{p,k_*}, \vec{y}_{k_*}^{[j-1]})\vec{\eta}_k l_k(s)ds)_{1 \le i \le k_1}, \tag{5.15}$$

where $k_* = [\frac{k_1}{2}]$. k_* can be chosen as anything from 1 to k_1. Here we choose the middle one where $k_* = [\frac{k_1}{2}]$.

Then $id - \hat{K}_{\vec{y}^{[j]}}$ can be efficiently solved. The matrix notation of $id - \hat{K}_{\vec{y}^{[j-1]}}$ is

$$id - \hat{K}_{\vec{y}^{[j-1]}} = I_{k_1} \otimes I_n - hS_{k_1} \otimes J(t_{p,k_*}, \vec{y}_{k_*}^{[j-1]}), \tag{5.16}$$

where

$$(S_{k_1})_{i,j} = \int_0^{c_i} l_j(x)dx \quad \text{where } 1 \le i, j \le k_1. \tag{5.17}$$

Consider the Schur decomposition of the $S_{k_1} = Q_{k_1} R_{k_1} Q_{k_1}^T$ where R_{k_1} is an upper triangular matrix and Q_{k_1} is an orthogonal matrix. Then (5.16) is written as

$$(Q_{k_1} \otimes I_n)(I_{k_1} \otimes I_n - hR_{k_1} \otimes J(t_{p,k_*}, \vec{y}_{k_*}^{[j-1]}))(Q_{k_1}^T \otimes I_n). \tag{5.18}$$

Note that $I_{k_1} \otimes I_n - hR_{k_1} \otimes J(t_{p,k_*}, \vec{y}_{k_*}^{[j-1]})$ is an upper block triangular matrix by the definition of the tensor product. Even though the problem size is nk_1, we need to solve the problem of size n, k_1 times by using backsubstitution. Computational cost is the same as backward Euler's method except multiplying $Q_{k_1} \otimes I_n$ and $Q_{k_1}^T \otimes I_n$ which is $O(2k_1^2 n) = O(n)$.

5.3.2 Algorithm

new scheme is a one-step method and its algorithm is completely described by local behavior, advancing from t_p to t_{p+1}.

Comment: Fix h, k_1, J.

Comment: Assume $y(t_p)$ is given. We want to find $y(t_{p+1}) = y(t_p + h)$.

1. Find an initial guess $\vec{y}_k^{[0]} \sim y(t_{p,k})$ using FE or BE for $1 \le k \le k_1$ by solving (5.4).

   for $j = 1, \cdots, J$

   (a) Compute the residual $\vec{r}^{[j-1]} = r_{k_1} \vec{y}^{[j-1]}$ by (5.8);

   (b) Find $\vec{\delta}_k^{[j]} \sim \delta(t_{p,k})$ for $1 \le k \le k_1$ by solving $(id - \hat{K}_{\vec{y}^{[j-1]}})\vec{\delta}^{[j]} = \vec{r}^{[j-1]}$;

   (c) Update the solution: $\vec{y}^{[j]} = \vec{y}^{[j-1]} + \vec{\delta}^{[j]}$.

   end for

- Remark 1 . Diagonalization of S_{k_1} is numerically possible and would enable us to do a parallel computing. The process is numerically unstable even for the small k_1. A detailed discussion with numerical evidence is provided in Section 5.4.3.

- Remark 2. If the problem is a constant-coefficient, linear problem, we obtain the collocation solution of the problem. Hence extra iteration is unnecesary.

5.3.3 Stability of the scheme

Consider the following model problem:

$$\begin{aligned} y'(t) &= \lambda y(t), \quad t \in [0,1], \\ y(0) &= 1. \end{aligned} \tag{5.19}$$

If $\tilde{y}(1)$ is a numerical solution of (5.19) with $h = 1$, then the amplification factor or stability function $Am(\lambda)$ is defined by the formula

$$Am(\lambda) = \tilde{y}(1). \tag{5.20}$$

The set $\{\lambda \in \mathbb{C} : |Am(\lambda)| < 1\}$ is called a stability region, and if the stability region contains the left half plane, the method is called A-stable.

If the quadrature points $\{c_i\}$ are Gauss-Lobatto type points, the collocation method based on $\{c_i\}$ is called a Lobatto IIIA method. It is known that Lobatto IIIA method is A-stable [11]. The problem (5.19) is a constant-coefficient linear problem. From Remark 2 in Section 4.3.2, the stability function is the same as the collocation method. Hence the method suggested in this chapter is A-stable.

5.4 Numerical experiments

In the first subsection, we perform numerical experiments on an ordinary differential equations. We verify the order of accuracy and the relation between the new scheme and the collocation method.

The numerical scheme for the ordinary differential equation could be applied to the time-dependent partial differential equation. We choose our test example as the Allen-Cahn equation and verify the order of accuracy.

5.4.1 Linear ordinary differential equation

In this subsection we perform two numerical experiments. The purpose of the first example is to observe the order of the accuracy of the scheme. Also we observe the effect of the node distribution of the fine mesh. In the second example, we study the relation between new scheme and the collocation method.

Our first example is a variable-coefficient linear problem. In this problem, we observe the order of accuracy and the effect of choice of fine mesh. Consider the following problem:

$$y'(t) = -(2+t)(y - cos(2\pi t)) - 2\pi sin(2\pi t), \tag{5.21}$$

$$y(0) = 1. \tag{5.22}$$

$T = 1$ with the exact solution $cos(2\pi t)$. Fig. 5.1 is the result of the numerical scheme with Legendre-Gauss-Lobatto points and $k_1 = 4$. Note that each dotted lines are slope 1, 3, 5, 7, 8. The first thing we can observe is that we obtain two orders at each correction, which is proved in Section 5.3. We can also observe that maximum order achieved is $8 = 2k_1$ which is the order of the accuracy of the underlying collocation method when $\{c_i\}$ are chosen to be Legendre-Gauss-Lobatto points.

We perform the same experiment with different $\{c_i\}$ to observe the effect of fine mesh. We choose fine mesh to be $c_0 = 0, c_1 = 1/2, c_2 = 3/4, c_3 = 7/8, c_4 = 1$. Fig. 5.2 is the result of the experiment. Note that each dotted lines have slopes 1, 3, 4. We can observe at first correction the order of accuracy is 3. But further correction does not increase the order of accuracy and it is bounded by four. Same as the previous example, it is also the order of accuracy of the collocation method. So, Legendre-Gauss-Lobatto points can be chosen for the better performance.

To verify the relation to the collocation method we choose the following second example:

$$y'(t) = -2(y - cos(2\pi t)) - 2\pi sin(2\pi t), \tag{5.23}$$

$$y(0) = 1, \tag{5.24}$$

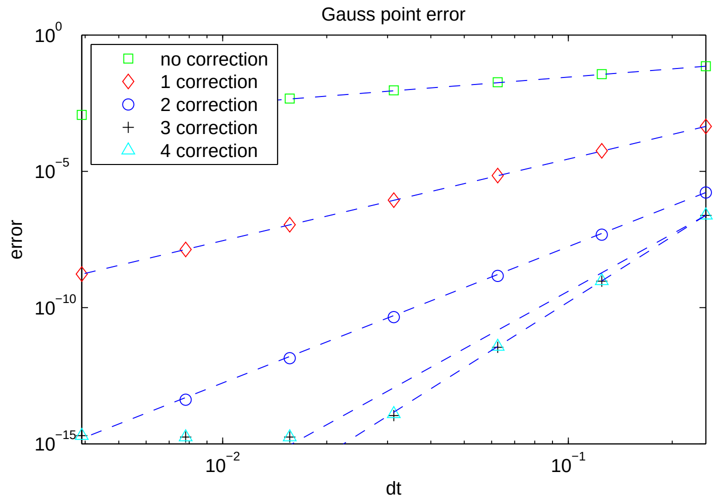

Figure 5.1. Error of the computed solution of (5.21) using LGL fine mesh

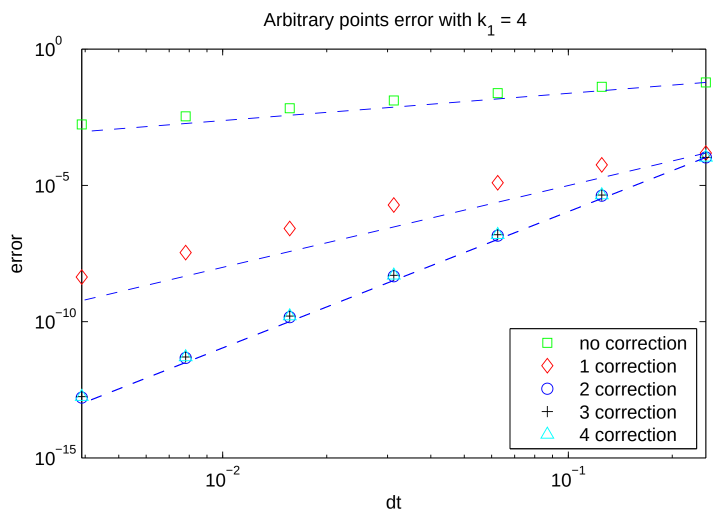

Figure 5.2. Error of the computed equation (5.21) using arbitrary fine mesh

where $T = 20$ with 5 Legendre-Gauss-Lobatto points ($k_1 = 4$) and the exact solution is $cos(2\pi t)$. The expected order of accuracy is 8. Each doted lines have slope one and eight. The 0 iteration is just the backward Euler's method; therefore we obtain the first order of accuracy. At one iteration we get almost the same solution as the

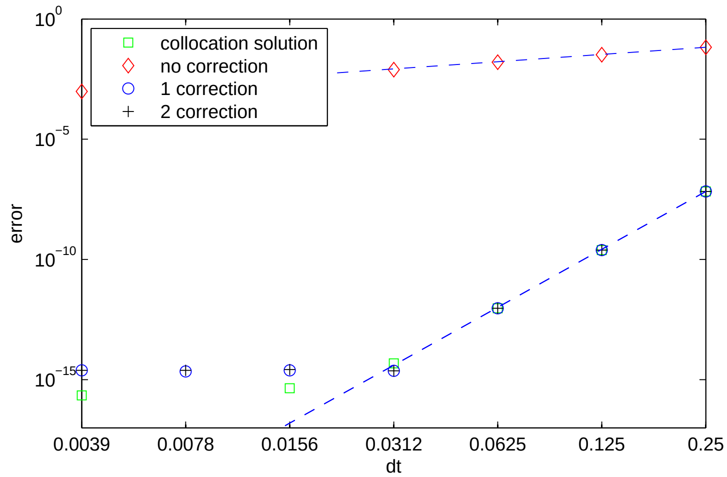

Figure 5.3. Error of the computed solution of (5.23)

collocation method. Since we have the collocation solution, the second iteration does not help.

5.4.2 2D Allen-Cahn equation

In this subsection, we apply the new scheme to the Allen-Cahn equation to observe the order of accuracy. The Allen-Cahn equation is

$$u_t = \gamma(\Delta u - \frac{1}{\epsilon^2}(u^3 - u)), \quad x \in \Omega, \tag{5.25}$$
$$\frac{\partial u}{\partial n}|_{\partial\Omega} = 0.$$

The Allen-Cahn equation was developed by Allen and Cahn to describe the behavior of anti-phase boundaries in crystalline solids. u is a concentration of each metallic component and ϵ is the width of the interface of two alloys.

We perform the order of accuracy test with the exact solution

$$u(t, x, y) = cos(t)cos(\pi x)cos(\pi y),$$

at $T = 1$ with $\gamma = \epsilon = 1$ with $\Omega = (-1, 1)^2$. The Legendre-Galerkin method was used for the spatial discretization. Fig. 5.4 is the result and dotted lines have slopes one, three and five. We can observe that the expected order of accuracy is obtained as the time step gets smaller. At each correction step, the following type of correction equation needs to be solved:

$$\frac{\gamma}{\epsilon^2}(3u^2 - 1)\delta - \gamma\Delta\delta = r, \qquad (5.26)$$
$$\frac{\partial\delta}{\partial n}|_{\partial\Omega} = 0.$$

(5.26) is a variable-coefficient equation so the GMRES iteration was used with a constant-coefficient equation solver as a preconditioner. For this particular problem, around four iterations were enough to achieve round-off error.

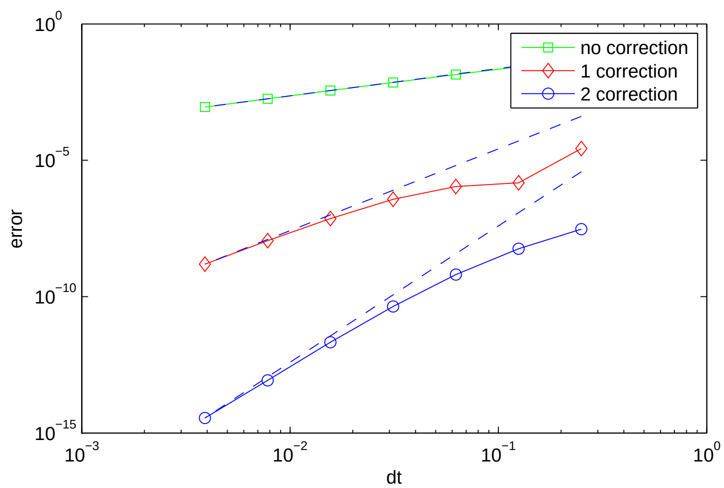

Figure 5.4. Order of accuracy for 2D Allen-Cahn equation (5.25)

5.4.3 Round-off error of eigenvalue decomposition

In new scheme the Schur decomposition $S_{k_1} = Q_{k_1} R_{k_1} Q_{k_1}^T$ was applied to solve the system. Since Schur decomposition results in triangular matrix R_{k_1}, the system $I_{k_1} \otimes I_n - dtR_{k_1} \otimes J$ needs to be solved in sequence just like backward substitution. But if we can do the eigenvalue decomposition to the integration matrix S_{k_1}, then every subproblem is completely decoupled, allowing for the parallel computing of the problem. The eigenvalues of S_{k_1} for $2 \leq k_1 \leq 50$ obtained from numerical computations are all distinct and S_{k_1} were diagonalizable. However, this approach should be restricted to small k_1 to prevent numerical instability. We can observe this from following the numerical experiment. Fig. 5.5 is the error plot as a function of k_1 for the problem:

$$y'(t) = -y(t),$$

$$y(0) = 1,$$

with $h = T_{max} = 1$. To see the effect of diagonalization of S_{k_1}, we solve the equation with two different methods in (5.16). One is with the Schur decomposition and the other one is with diagonalization of S_{k_1}. With both methods we observe the spectral accuracy until k_1 reaches 7. But as k_1 grows, the error of diagonalizing method grows, while the error of the Schur decomposition stays around the round-off error. One way to explain this situation is to observe the condition number of the diagonalization matrix V_{k_1} where $S_{k_1} = V_{k_1}\Lambda_{k_1}V_{k_1}^{-1}$ and Λ_{k_1} is a diagonal matrix. Since the eigenvalue decomposition matrix is not unique, we normalized V_{k_1} so that each column of V_{k_1} has L_2-norm one. Fig. 5.6 is the the condition number of V_{k_1} as a function of k_1. The condition number grows exponentially and if $k_1 = 8$ where V is an 8 by 8 matrix, the conditioner number is larger than 10^4 and it is ill conditioned. However the Schur decomposition gives us a unitary transform matrix which has a condition number 1 and the method is numerically stable.

5.5 Concluding remarks

In this chapter, we developed a defect correction type numerical scheme for an ordinary differential equation. The main property of the scheme is that we could recover order of accuracy by 2 at each correction and the cost is almost the same as the backward Euler's method. In terms of stability, the suggested scheme is the same as the collocation method if the underlying problem is a linear-constant coefficient problem. Since the Lobatto IIIA mehod is A-stable, the suggested method is A-stable.

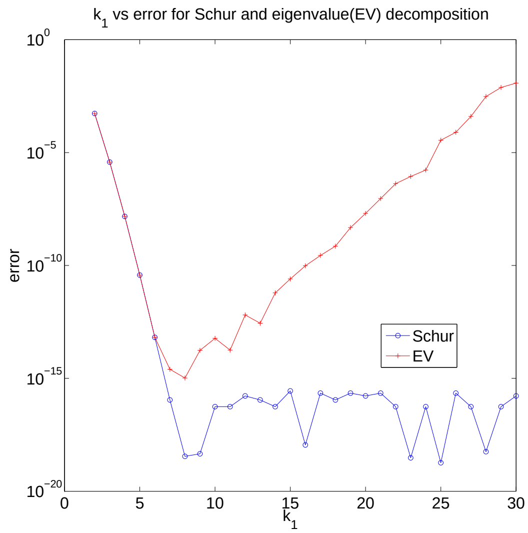

Figure 5.5. Error of the computed solution using Schur and eigenvalue decomposition

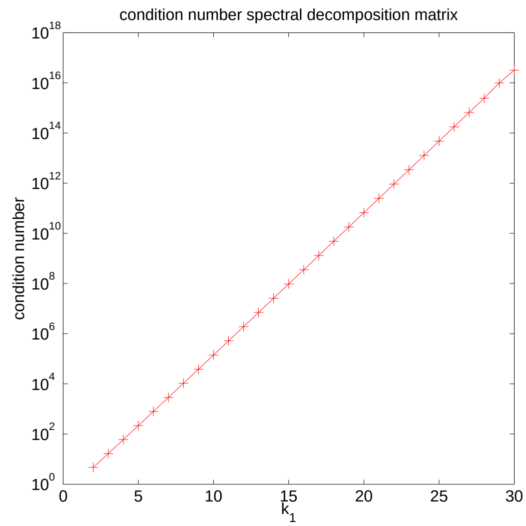

Figure 5.6. Condition number of eigenvalue decomposition matrix V_{k_1}

We verified the results by numerical experiment. This method can be also applied to PDEs and we could obtain a formal order of accuracy for the 2D Allen-Cahn equation.

However, it should be pointed out that the suggested method is based on Schur decomposition, hence, it is more complicated than other methods in terms of programming.

We have tested for the Allen-Cahn equation only but it could be applied to other time dependent PDEs.

APPENDIX

Appendix: A brief review of Legendre-Galerkin method

In this appendix, we review the Legendre-Galerkin method briefly on a model problem for simple geometries.

A.1 Legendre-Galerkin method for second-order two boundary value problem

In 1d case, we consider the following model problem:

$$\alpha u - \beta \Delta u = f, \qquad \text{in } \Omega, \tag{A.1}$$

$$a_{\pm} u(\pm 1) + b_{\pm} u'(\pm 1) = 0, \tag{A.2}$$

where α and β are positive real numbers and $\Omega = (-1, 1)$. The Galerkin method is given a function f, find $u_N \in X_N$ such that

$$\alpha \int_{\Omega} u_N v_N dx - \beta \int_{\Omega} u_N'' v_N dx = \int_{\Omega} I_N f v_N dx, \tag{A.3}$$

for all $v_N \in Y_N$ and $I_N f$ is an interpolation of f in X_N. Choosing a proper basis of X_N and Y_N is a crucial factor for building an efficient numerical scheme. In [21], the author developed a basis which consists of Legendre polynomials for simple geometries.

Hence it is natural to try the following function as a basis.

$$\phi_i(x) = L_i(x) + a_i L_{i+1}(x) + b_i L_{i+2}(x), \tag{A.4}$$

for $i \geq 0$. It is known that if we choose (a_i, b_i) to be the solution of the system

$$\begin{aligned} (a_+ + \frac{b_+}{2}(i+1)(i+2))a_i + (a_+ + \frac{b_+}{2}(i+2)(i+3))b_i &= -a_+ - \frac{b_+}{2} i(i+1), \\ -(a_- - \frac{b_-}{2}(i+1)(i+2))a_i + (a_- - \frac{b_-}{2}(i+2)(i+3))b_i &= -a_- + \frac{b_-}{2} i(i+1), \end{aligned} \tag{A.5}$$

and ϕ_i satisfies (A.2). And (A.5) can be solved uniquely under some mild condition. Details can be found in [56]. Note that in particular, if $a_{\pm} = 1$ and $b_{\pm} = 0$, we obtain $a_i = 0$ and $b_i = -1$. And if $a_\pm = 0$ and $b_\pm = 1$, we obtain, $a_i = 0$ and $b_i = -i(i+1)/((i+2)(i+3))$ for $i \geq 0$. We consider a homogeneous boundary condition for simplicity (i.e. $a_\pm = 1$ and $b_\pm = 0$). Given N, we can consider the following space:

$$X_N = P^N \cap H_0^1(\Omega) = \text{span}\{\phi_i(x) | 0 \leq i \leq N-2\}, \tag{A.6}$$

where P^N is a set of polynomials with degree equal or less than N. We can consider the following matrices:

$$S_{i,j} = \int_\Omega \phi_j(x)'\phi_i(x)'dx, \tag{A.7}$$

$$M_{i,j} = \int_\Omega \phi_j(x)\phi_i(x)dx, \tag{A.8}$$

$$\bar{f}_i = \int_\Omega I_N f \phi_i(x)dx, \tag{A.9}$$

for all $0 \leq i, j \leq N-2$. It is known that Legendre polynomials have the following properties.

$$\begin{aligned} \int_{-1}^{1} L_i(x)L_j(x)dx &= \frac{2}{2i+1}\delta_{ij}, \\ (2i+3)L_i(x) &= \frac{d}{dx}(L_{i+2}(x) - L_i(x)), \end{aligned} \tag{A.10}$$

for all $i, j \geq 0$ and δ_{ij} is a Kronecker delta. Using (A.10), we can find stiffness matrix S and mass matrix M. Let

$$S_{i,j} = \begin{cases} (4i+6), & \text{if } i = j, \\ 0 & \text{otherwise}, \end{cases} \tag{A.11}$$

and

$$M_{i,j} = M_{j,i} = \begin{cases} \dfrac{2}{2i+1} + \dfrac{2}{2i+5}, & \text{if } i = j, \\ \dfrac{2}{2i+5}, & \text{if } i = j+2, \\ 0, & \text{otherwise }, \end{cases} \tag{A.12}$$

for all $0 \leq i, j \leq N-2$. Then (A.3) has the following matrix form:

$$(\alpha M + \beta S)\bar{u} = \bar{f}, \tag{A.13}$$

where $\bar{u} = (u_0, \ldots, u_{N-2})^T$. Let $u_N(x) = \sum_{i=0}^{N-2} u_i\phi_i(x)$ and u be a solution of (A.1). Then it has following estimates

$$\|u - u_N\| + N\|u - u_N\| \simeq C(s)\|u\|_s, \tag{A.14}$$

where $u \in H^s$ and $\|\cdot\|_s$ is the norm on H^s.

A.2 Legendre-Galerkin method on second-order 2d boundary value problem

We consider the homogeneous boundary value problem:

$$\alpha u - \beta\Delta u = f, \quad \text{in } \Omega, \tag{A.15}$$

$$u|_{\partial\Omega} = 0. \tag{A.16}$$

The weak formulation of (A.15) is, find $u_N \in X_N$,

$$\alpha\int u_N v_N dxdy - \beta\int(\partial_{xx}u_N + \partial_{yy}u_N)v_N dxdy = \int I_N f v_N dxdy, \tag{A.17}$$

for all Y_N. We consider the following test space and trial space,

$$\begin{aligned} X_N &= (P_N \times P_N) \cap H_0^1(\Omega) = \text{span}\{\phi_i(x)\phi_j(y)|0 \leq i, j \leq N-2\}, \\ Y_N &= X_N. \end{aligned} \tag{A.18}$$

(A.17) has the following matrix form:

$$\alpha MUM^T + \beta(SUM^T + MUS^T) = F, \tag{A.19}$$

where

$$\bar{F}_{i,j} = \int_\Omega I_N f\phi_i(x)\phi_j(y)dxdy, \tag{A.20}$$

for all $0 \leq i, j \leq N-2$. Then

$$u_N(x, y) = \sum_{i=0}^{N-2}\sum_{j=0}^{N-2} U_{i,j}\phi_i(x)\phi_j(y), \tag{A.21}$$

solves (A.17).

Since Ω is separable and (A.15) is separable, we can develop an efficient method to solve (A.19) based on a separation of variables. Consider the following generalized eigenvalue problem:

$$ME = SE\Lambda, \tag{A.22}$$

where Λ is a diagonal matrix. We can put $W = E^{-1}U(E^T)^{-1}$, then we have

$$\alpha MEW(ME)^T + \beta(SEW(ME)^T + MEW(SE)^T) = F. \tag{A.23}$$

Substuting (A.22) to (A.23), we obtain

$$\alpha SE\Lambda W\Lambda^T(SE)^T + \beta(SEW\Lambda^T(SE)^T + SE\Lambda W(SE)^T) = F. \tag{A.24}$$

Multiplying $(SE)^{-1}$ to the left and $(SE)^{-T}$ to the right of (A.24), we obtain

$$\alpha\Lambda W\Lambda^T + \beta(W\Lambda^T + \Lambda^T W) = G, \tag{A.25}$$

where $G = (SE)^{-1}F(SE)^{-T}$. Note that (A.25) can be solved by the equation

$$W_{i,j} = G_{i,j}/(\alpha\Lambda_{ii}\Lambda_{jj} + \beta(\Lambda_{jj} + \Lambda_{ii})), \tag{A.26}$$

for all $0 \leq i, j \leq N-2$. And we find the solution by $U = EWE^T$. Hence the algorithm to solve the system (A.19) is as follows:

1. Precompute : Solve generalized eigenvalue problem (A.22) and compute $(SE)^{-1}$.

2. Find $G = (SE)^{-1}F(SE)^{-T}$.

3. Compute W by (A.26).

4. Obtain $U = EWE^T$.

REFERENCES

REFERENCES

[1] R. Courant, K. Friedrichs, and H. Lewy, "Über die partiellen Differenzengleichungen der mathematischen Physik," *Math. Ann.*, vol. 100, no. 1, pp. 32–74, 1928. [Online]. Available: http://dx.doi.org/10.1007/BF01448839

[2] P. D. Lax and R. D. Richtmyer, "Survey of the stability of linear finite difference equations," *Comm. Pure Appl. Math.*, vol. 9, pp. 267–293, 1956.

[3] J.-G. Liu and R. L. Pego, "Stable discretization of magnetohydrodynamics in bounded domains," *Commun. Math. Sci.*, vol. 8, no. 1, pp. 235–252, 2010. [Online]. Available: http://projecteuclid.org/euclid.cms/1266935021

[4] A. Prohl, "Convergent finite element discretizations of the nonstationary incompressible magnetohydrodynamics system," *M2AN Math. Model. Numer. Anal.*, vol. 42, no. 6, pp. 1065–1087, 2008. [Online]. Available: http://dx.doi.org/10.1051/m2an:2008034

[5] S. Badia, R. Planas, and J. V. Gutiérrez-Santacreu, "Unconditionally stable operator splitting algorithms for the incompressible magnetohydrodynamics system discretized by a stabilized finite element formulation based on projections," *Internat. J. Numer. Methods Engrg.*, vol. 93, no. 3, pp. 302–328, 2013. [Online]. Available: http://dx.doi.org/10.1002/nme.4392

[6] J. L. Guermond and J. Shen, "On the error estimates for the rotational pressure-correction projection methods," *Math. Comp.*, vol. 73, no. 248, pp. 1719–1737 (electronic), 2004. [Online]. Available: http://dx.doi.org/10.1090/S0025-5718-03-01621-1

[7] J. Shen and X. Yang, "Energy stable schemes for cahn-hilliard phase-field model of two phase incompressible flows equations," *Chinese Annals of Mathematics, Series B*, 2010, DOI: 10.1007/s11401-010-0599-y.

[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)*, ser. Mater. Res. Soc. Sympos. Proc. Warrendale, PA: MRS, 1998, vol. 529, pp. 39–46.

[9] M. Elghaoui and R. Pasquetti, "A spectral embedding method applied to the advection-diffusion equation," *J. Comput. Phys.*, vol. 125, no. 2, pp. 464–476, 1996. [Online]. Available: http://dx.doi.org/10.1006/jcph.1996.0108

[10] E. Hairer, S. P. Nørsett, and G. Wanner, *Solving ordinary differential equations. I*, 2nd ed., ser. Springer Series in Computational Mathematics. Berlin: Springer-Verlag, 1993, vol. 8, nonstiff problems.

[11] E. Hairer and G. Wanner, *Solving ordinary differential equations. II*, 2nd ed., ser. Springer Series in Computational Mathematics. Berlin: Springer-Verlag, 1996, vol. 14, stiff and differential-algebraic problems.

[12] A. Dutt, L. Greengard, and V. Rokhlin, "Spectral deferred correction methods for ordinary differential equations," *BIT*, vol. 40, no. 2, pp. 241–266, 2000. [Online]. Available: http://dx.doi.org/10.1023/A:1022338906936

[13] K. Böhmer, P. Hemker, and H. J. Stetter, "The defect correction approach," in *Defect correction methods (Oberwolfach, 1983)*, ser. Comput. Suppl. Springer, Vienna, 1984, vol. 5, pp. 1–32.

[14] R. Temam, *Navier-Stokes Equations: Theory and Numerical Analysis.* North-Holland, Amsterdam, 1984.

[15] R. Témam, "Sur l'approximation de la solution des équations de Navier-Stokes par la méthode des pas fractionnaires. II," *Arch. Rational Mech. Anal.*, vol. 33, pp. 377–385, 1969.

[16] A. J. Chorin, "Numerical solution of the Navier-Stokes equations," *Math. Comp.*, vol. 22, pp. 745–762, 1968.

[17] J. L. Guermond, P. Minev, and J. Shen, "An overview of projection methods for incompressible flows," *Comput. Methods Appl. Mech. Engrg.*, vol. 195, no. 44-47, pp. 6011–6045, 2006. [Online]. Available: http://dx.doi.org/10.1016/j.cma.2005.10.010

[18] J. L. Guermond and J. Shen, "A new class of truly consistent splitting schemes for incompressible flows," *J. Comput. Phys.*, vol. 192, no. 1, pp. 262–276, 2003. [Online]. Available: http://dx.doi.org/10.1016/j.jcp.2003.07.009

[19] M. Sermange and R. Temam, "Some mathematical questions related to the MHD equations," *Comm. Pure Appl. Math.*, vol. 36, no. 5, pp. 635–664, 1983. [Online]. Available: http://dx.doi.org/10.1002/cpa.3160360506

[20] R. Temam, *Navier-Stokes equations*, 3rd ed., ser. Studies in Mathematics and its Applications. North-Holland Publishing Co., Amsterdam, 1984, vol. 2, theory and numerical analysis, With an appendix by F. Thomasset.

[21] J. Shen, "Efficient spectral-Galerkin method. I. Direct solvers of second- and fourth-order equations using Legendre polynomials," *SIAM J. Sci. Comput.*, vol. 15, no. 6, pp. 1489–1505, 1994. [Online]. Available: http://dx.doi.org/10.1137/0915089

[22] Z. Zhang, Y. Ma, and Z. Qiao, "An adaptive time-stepping strategy for solving the phase field crystal model," *J. Comput. Phys.*, vol. 249, pp. 204–215, 2013. [Online]. Available: http://dx.doi.org/10.1016/j.jcp.2013.04.031

[23] J. Shen and X. Yang, "A phase-field model and its numerical approximation for two-phase incompressible flows with different densities and viscosities," *SIAM Journal on Scientific Computing*, vol. 32, no. 3, pp. 1159–1179, 2010.

[24] ——, "Energy stable schemes for cahn-hilliard phase-field model of two-phase incompressible flows," *Chinese Annals of Mathematics, Series B*, vol. 31, no. 5, pp. 743–758, 2010.

[25] M. E. Gurtin, D. Polignone, and J. Viñals, "Two-phase binary fluids and immiscible fluids described by an order parameter," *Math. Models Methods Appl. Sci.*, vol. 6, no. 6, pp. 815–831, 1996.

[26] D. Jacqmin, "Diffuse interface model for incompressible two-phase flows with large density ratios," *J. Comput. Phys.*, vol. 155, no. 1, pp. 96–127, 2007.

[27] D. M. Anderson, G. B. McFadden, and A. A. Wheeler, "Diffuse-interface methods in fluid mechanics," vol. 30, pp. 139–165, 1998.

[28] J. Lowengrub and L. Truskinovsky, "Quasi-incompressible Cahn-Hilliard fluids and topological transitions," *R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci.*, vol. 454, no. 1978, pp. 2617–2654, 1998.

[29] C. Liu and J. Shen, "A phase field model for the mixture of two incompressible fluids and its approximation by a Fourier-spectral method," *Physica D*, vol. 179, no. 3-4, pp. 211–228, 2003.

[30] P. Yue, J. J. Feng, C. Liu, and J. Shen, "A diffuse-interface method for simulating two-phase flows of complex fluids," *J. Fluid Mech*, vol. 515, pp. 293–317, 2004.

[31] J. Shen, "Modeling and numerical approximation of two-phase incompressible flows by a phase-field approach," in *Multiscale Modeling and Analysis for Materials Simulation, Lecture Note Series, Vol. 9. IMS, National University of Singapore, 2011, Edited by W. Bao and Q. Du*, pp. 147–196.

[32] R. Glowinski, "Finite element methods for incompressible viscous flow," in *Handbook of numerical analysis, Vol. IX*, ser. Handb. Numer. Anal., IX. Amsterdam: North-Holland, 2003, pp. 3–1176.

[33] J. Guermond, P. Minev, and J. Shen, "An overview of projection methods for incompressible flows," *Comput. Methods Appl. Mech. Engrg.*, vol. 195, pp. 6011–6045, 2006.

[34] X. Feng, "Fully discrete finite element approximations of the navier–stokes–cahn-hilliard diffuse interface model for two-phase fluid flows," *SIAM Journal on Numerical Analysis*, vol. 44, no. 3, pp. 1049–1072, 2006. [Online]. Available: http://dx.doi.org/10.1137/050638333

[35] X. Feng, Y. He, and C. Liu, "Analysis of finite element approximations of a phase field model for two-phase fluids," *Math. Comp.*, vol. 76, no. 258, pp. 539–571 (electronic), 2007.

[36] G. Grün, "On convergent schemes for diffuse interface models for two-phase flow of incompressible fluids with general mass densities," *SIAM Journal on Numerical Analysis*, vol. 51, no. 6, pp. 3036–3061, 2013. [Online]. Available: http://dx.doi.org/10.1137/130908208

[37] J. Shen and X. Yang, "Numerical approximations of allen-cahn and cahn-hilliard equations," *Discrete and Continuous Dynamical Systems. Series A*, 2010, to appear.

[38] C. Xu and T. Tang, "Stability analysis of large time-stepping methods for epitaxial growth models," *SIAM J. Numer. Anal.*, vol. 44, no. 4, pp. 1759–1779, 2006. [Online]. Available: http://dx.doi.org/10.1137/050628143

[39] C. M. Elliott and A. M. Stuart, "The global dynamics of discrete semilinear parabolic equations," *SIAM J. Numer. Anal.*, vol. 30, no. 6, pp. 1622–1663, 1993. [Online]. Available: http://dx.doi.org/10.1137/0730084

[40] X. Feng and A. Prohl, "Numerical analysis of the Allen-Cahn equation and approximation for mean curvature flows," *Numer. Math.*, vol. 94, no. 1, pp. 33–65, 2003. [Online]. Available: http://dx.doi.org/10.1007/s00211-002-0413-1

[41] ——, "Error analysis of a mixed finite element method for the Cahn-Hilliard equation," *Numer. Math.*, vol. 99, no. 1, pp. 47–84, 2004. [Online]. Available: http://dx.doi.org/10.1007/s00211-004-0546-5

[42] D. Kessler, R. H. Nochetto, and A. Schmidt, "A posteriori error control for the Allen-Cahn problem: circumventing Gronwall's inequality," *M2AN Math. Model. Numer. Anal.*, vol. 38, no. 1, pp. 129–142, 2004.

[43] J. Guermond and J. Shen, "On the error estimates of rotational pressure-correction projection methods," *Math. Comp*, vol. 73, pp. 1719–1737, 2004.

[44] S. A. Orszag, "Spectral methods for problems in complex geometries," *J. Comput. Phys.*, vol. 37, no. 1, pp. 70–92, 1980. [Online]. Available: http://dx.doi.org/10.1016/0021-9991(80)90005-4

[45] B. L. Buzbee, F. W. Dorr, J. A. George, and G. H. Golub, "The direct solution of the discrete Poisson equation on irregular regions," *SIAM J. Numer. Anal.*, vol. 8, pp. 722–736, 1971.

[46] R. Glowinski, T.-W. Pan, and J. Périaux, "A fictitious domain method for Dirichlet problem and applications," *Comput. Methods Appl. Mech. Engrg.*, vol. 111, no. 3-4, pp. 283–303, 1994. [Online]. Available: http://dx.doi.org/10.1016/0045-7825(94)90135-X

[47] S. H. Lui, "Spectral domain embedding for elliptic PDEs in complex domains," *J. Comput. Appl. Math.*, vol. 225, no. 2, pp. 541–557, 2009. [Online]. Available: http://dx.doi.org/10.1016/j.cam.2008.08.034

[48] D. Huybrechs, "On the Fourier extension of nonperiodic functions," *SIAM J. Numer. Anal.*, vol. 47, no. 6, pp. 4326–4355, 2010. [Online]. Available: http://dx.doi.org/10.1137/090752456

[49] L. C. Evans, *Partial differential equations*, ser. Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 1998, vol. 19.

[50] B. Fornberg, "A pseudospectral approach for polar and spherical geometries," *SIAM J. Sci. Comput.*, vol. 16, no. 5, pp. 1071–1081, 1995. [Online]. Available: http://dx.doi.org/10.1137/0916061

[51] L. N. Trefethen, *Spectral methods in MATLAB*, ser. Software, Environments, and Tools. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000, vol. 10. [Online]. Available: http://dx.doi.org/10.1137/1.9780898719598

[52] K. E. Atkinson, *The numerical solution of integral equations of the second kind*, ser. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, Cambridge, 1997, vol. 4. [Online]. Available: http://dx.doi.org/10.1017/CBO9780511626340

[53] J. Huang, J. Jia, and M. Minion, "Accelerating the convergence of spectral deferred correction methods," *J. Comput. Phys.*, vol. 214, no. 2, pp. 633–656, 2006. [Online]. Available: http://dx.doi.org/10.1016/j.jcp.2005.10.004

[54] ——, "Arbitrary order Krylov deferred correction methods for differential algebraic equations," *J. Comput. Phys.*, vol. 221, no. 2, pp. 739–760, 2007. [Online]. Available: http://dx.doi.org/10.1016/j.jcp.2006.06.040

[55] M. L. Minion, "Semi-implicit spectral deferred correction methods for ordinary differential equations," *Commun. Math. Sci.*, vol. 1, no. 3, pp. 471–500, 2003. [Online]. Available: http://projecteuclid.org/getRecord?id=euclid.cms/1250880097

[56] J. Shen, T. Tang, and L.-L. Wang, *Spectral methods*, ser. Springer Series in Computational Mathematics. Springer, Heidelberg, 2011, vol. 41, algorithms, analysis and applications. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-71041-7

VITA

VITA

Heejun Choi grew up in Busan, Korea. In 1999, he was admitted to the school of science at Yonsei University and graduated in 2005. He also attended graduate school at Yonsei University and graduated with M.S. degree in 2007. Later in the fall 2007, he attended Purdue University to pursue Ph.D. degree in the math department. He is now Ph.D. candidate in Mathematics.