Czech Technical University in Prague
Faculty of Nuclear Sciences and Physical Engineering

Doctoral Thesis

Variational Methods in Phase Transition Modelling

Prague 2022 Aleš Wodecki

This thesis is submitted to the Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, in partial fulfilment of the requirements for the degree of Doctor of Philosophy (Ph.D.) in Mathematical Engineering.

Copyright © 2022 Aleš Wodecki. All Rights Reserved.

Bibliografický záznam

Autor	Ing. Aleš Wodecki, České vysoké učení technické v Praze, Fakulta jaderná a fyzikálně inženýrská, Katedra matematiky
Název práce	Variační metody v modelování fázových přechodů
Studijní program	Aplikace přírodních věd
Studijní obor	Matematické inženýrství
Školitel	doc. Ing. Tomáš Oberhuber, Ph.D., České vysoké učení technické v Praze, Fakulta jaderná a fyzikálně inženýrská, Katedra matematiky
Školitel specialista	Ing. Pavel Strachota, Ph.D., České vysoké učení technické v Praze, Fakulta jaderná a fyzikálně inženýrská, Katedra matematiky
Akademický rok	2022
Počet stran	105
Klíčová slova	Metoda fázového pole, tuhnutí čisté směsi, anizotropická solidifikace, optimální řízení, numerická simulace

Bibliographic Entry

Author	Ing. Aleš Wodecki, Czech Technical University in Prague, Faculty of Nuclear Sciences and Physical Engineering, Department of Mathematics
Title of dissertation	Variational Methods in Phase Transition Modelling
Degree programme	Application of Natural Sciences
Field of study	Mathematical Engineering
Supervisor	doc. Ing. Tomáš Oberhuber, Ph.D., Czech Technical University in Prague, Faculty of Nuclear Sciences and Physical Engineering, Department of Mathematics
Specialist Supervisor	Ing. Pavel Strachota, Ph.D., Czech Technical University in Prague, Faculty of Nuclear Sciences and Physical Engineering, Department of Mathematics
Academic year	2022
Number of pages	105
Keywords	Phase field method, solidification of pure substances, anisotropic solidification, optimal control, numerical simulation

Abstrakt

Tato disertace se zabývá použitím variačních metod v modelování fázových přechodů. Nové varianty modelu fázového pole, vhodné pro modelování tuhnutí čistých látek, jsou navrženy a verifikovány s pomocí asymptotické analýzy a numerických experimentů. Je diskutováno optimální řízení Dirichletovy okrajové podmínky v problémů fázového pole. Existence optimálního řízení a podmínky optimality jsou odvozeny pro slabou formulaci zmíněného problému. S pomocí adjugované úlohy je pak odvozena metoda výpočtu gradientu, která je následně využita pro numerické simulace. V těchto simulacích je také diskutována vhodnost použití jednoho z nově navržených variant modelu.

Abstract

This thesis concerns itself with the use of variational methods in phase-field modeling. New variants of the phase-field model, applicable to the modeling of phase transitions, are proposed, contextualized and verified using asymptotic analysis and numerical experiments. Optimal control of the Dirichlet boundary condition in the phase-field problem is discussed. In its weak form the problem is shown to have a solution and the optimality conditions are derived. The aforementioned problem is then treated numerically, using the formal derivation of the adjoint problem that leads to an efficient means of gradient computation. In addition to this, it is shown that one of the newly proposed variants of the model has its benefits when applied in an optimal control setting.

Acknowledgements

I would like to thank everyone who gave me any kind of support during my Ph.D. studies. I appreciate all the assistance and understanding that my supervisors, doc. Ing. Tomáš Oberhuber, Ph.D. and Ing. Pavel Strachota, Ph.D. have provided throughout the entire duration of my studies. A lion share of my gratitude goes out to my family, my ever understanding and loving wife Nini, who has stood by me regardless of how much work needed to be put in at the moment. My children Marcus and Felix, who provide inspiration and daily motivation. My parents Jitka and Pavel, who have always been willing to give support whenever possible. I would like to further acknowledge the support of the following projects:

- OP RDE project no. CZ.02.1.01-0.0-0.0-16-019-0000765: Research Center for Informatics, Ministry of Education, Youth and Sports of the Czech Republic
- AZV project No. NV19-08-00071: Analysis of flow character and prediction of evolution in endovascular treated arteries by magnetic resonance imaging coupled with mathematical modeling, Ministry of Health of the Czech Republic
- OP RDE project no. CZ.02.1.01-0.0-0.0-16-019-0000778: Centre for Advanced Applied Sciences, group MATE, Operational Programme of Research, Development and Education, Ministry of Education, Youth and Sports of the Czech Republic

Author's declaration

I confirm having prepared the thesis on my own and having listed all used sources of information in the bibliography.

Prague, 8. December 2022 Aleš Wodecki

Contents

1. Introduction

Substance solidification is intriguing. When studying this phenomenon, we attempt to understand common occurrences like snow flake formation as well as more exotic ones like the solidification of helium, which is observable only in an experimental setting [54]. Influencing substance solidification is of interest in many industrial applications as well. In metal casting, for instance, controlled solidification is used to create a structure that can withstand particular types of strain [32]. Another example is monocrystalline silicone used in transistor manufacturing [48].

Models capable of predicting micro-structure evolution and the properties of the resulting solids vary, based on their applications. These models can be loosely divided into microscopic, mesoscopic and macroscopic. Atomistic simulations [39, 6], typically used to estimate material parameters, are an example of a microscopic model. These outputs are then used in other numerical simulations or compared with experimental values. Another significant class of models are macroscopic models, these use state equations to model the outcome of phase transitions [33, 24]. Each of the two aforementioned classes have their advantages and drawbacks. Microscopic simulations capture the nuances of the interactions between particles, but are very expensive computationally. On the other hand, macroscopic models are more efficient computationally, but do not capture microscopic scale phenomena.

Mesoscopic approaches focus on representing behavior that can be observed on a scale somewhere between the microscopic and macroscopic scales. In particular, pattern formation during solidification can be observed on the mesoscopic scale. To capture the evolution of these patterns, interface tracking models are often utilized. The Stefan Problem (SP) [23, 40, 57] describes the dynamics of an interface between the liquid and solid portion of a domain during solidification. In the SP, the interface is considered to be sharp and is tracked as a surface evolving in 3 dimensions. The Phase Field Model (PFM) [13, 22, 31, 71], on the other hand, introduces a scalar parameter that is used to track the interface as a level-set. These two approaches are related, since one can show that the sharp interface limit of the PFM identifies with the the SP [13].

The PFM became a viable way to simulate phase transitions in pure substances when the Cahn-Hilliard equation was formulated [4, 42]. Later on, the Allen-Cahn equation gave the phase field method interpretation applicable to alloy systems [2, 3]. More recently, this method has been developed to provide quantitatively correct results for physically realistic settings [49]. It has also found applications in various other fields like brittle crack propagation or tumor growth modeling [73, 60, 29] and even been integrated with machine learning [53].

In this contribution, modifications to the PFM governing pure substance solidification are proposed and justified using asymptotic analysis and numerical experiments [66]. A Dirichlet boundary condition optimization problem for controlling the extent of solidification governed by the PFM for pure substance solidification is formulated. This problem is then analyzed in its weak form and the well-posedness, Fréchet differentiability of the associated solution operator and optimality conditions are provided. For the purposes of numerical simulation, the strong form of the problem, along with the corresponding adjoint formulation is used to find a locally optimal solution using gradient descent.

The thesis is structured as follows. Part I addresses the fundamental aspects of phase field modeling applied to pure substance solidification. In particular, Chapter 2 is concerned with the derivation of a phase field model that governs the solidification of pure substances, using first principles. This brief introduction then leads into the next section, where novel reaction terms for this problem are proposed (Chapter 3). These reaction terms are validated with the help of asymptotic analysis and a number of numerical experiments, some of which are compared with experimental results.

Part II is dedicated to a PDE-constrained optimization problem, where the Dirichlet boundary condition for the heat equation, which leads to a particular phase field profile at final time, is found. Before

formulating the problem, we review the fundamentals of PDE-constrained optimization in Chapter 4. Chapter 5 is dedicated to the precise definition and analysis of the the optimization problem in question. Namely, the problem is posed in its weak form and the existence of a solution proven. Following this, the Fréchet differentiability of the solution operator and optimality conditions are provided. In Chapters 6 and 7, we provide a numerical treatment of the aforementioned optimization problem that focuses on highlighting the utility of one of the newly proposed reaction terms (Chapter 3). In these Chapters, the formal derivation of the adjoint problem is used to provide efficient means of gradient computation. Following this, several numerical examples in one and two spatial dimensions demonstrate the advantages of using the newly proposed reaction term for optimization purposes.

Listing of The Authors Contributions and Related Publications

This thesis covers the results of the following papers that have either been published, accepted or are under review:

- P. Strachota, A. Wodecki, and M. Beneš. Focusing the latent heat release in 3D phase field simulations of dendritic crystal growth. Modelling Simul. Mater. Sci. Eng., 29:065009, 2021.
  - Aleš Wodecki is a co-author, worked on the formulation of the new reaction terms and their respective free energy forms, assisted with editing the paper.
- A. Wodecki, P. Strachota, T. Oberhuber, K. Škardová, M. Balázsová, M. Bohatý. Numerical Optimization of the Dirichlet Boundary Condition in the Phase Field Model with an Application to Pure Substance Solidification. Under review in Computers & Mathematics with Applications.
  - Aleš Wodecki is the leading author, conceptualized the entire article, proposed the mathematical and numerical method, wrote the main body of the article.
- A. Wodecki, P. Strachota, T. Oberhuber, M. Balázsová. Existence of Optimal Control for Dirichlet Boundary Optimization in a Phase Field Problem. Under review in Journal of Applied Mathematics and Computing.
  - Aleš Wodecki is the leading author, conceptualized the entire article, authored the initial draft of the proof, wrote the main body of the article.

Besides the aforementioned publications, the author has contributed to the following related publications:

- P. Strachota and A. Wodecki. High Resolution 3D Phase Field Simulations of Single Crystal and Polycrystalline Solidification. Acta Physica Polonica A. 134:653-657, 2018.
  - Aleš Wodecki is a co-author, worked to develop the solver used in the article.
- P. Strachota, A. Wodecki, M. Beneš. Efficiency of A Hybrid Parallel Algorithm For Phase-Field Simulation of Polycrystalline Solidification in 3D. Proceedings Of The Conference Algoritmy, 131 - 140, 2020.
  - Aleš Wodecki is a co-author, worked to develop the solver used in the article.

Part I.

Phase Field Modeling of Pure Substance Solidification

2. Introduction to Phase Field Modeling

Consider a bounded domain $\Omega \subset \mathbb{R}^3$ filled with a pure substance divided into a solid and a liquid part. We are interested in describing the evolution of the phase interface over a time interval $(0, T)$. Let $u : \Omega \times (0, T) \to \mathbb{R}$ represent the temperature at point $(\boldsymbol{x}, t) \in (0, T) \times \Omega$. To track the evolution of the phase interface between the liquid and solid parts of the domain, we introduce a scalar function

$$p : \Omega \to [0, 1], \tag{2.1}$$

called the phase field function. More succinctly we define the solid subdomain $\Omega_s(t)$, the liquid subdomain $\Omega_l(t)$ and the phase interface $\Gamma(t)$ by

$$\Omega_s(t) = \left\{ \boldsymbol{x} \in \Omega; p(t, \boldsymbol{x}) > \frac{1}{2} \right\}, \tag{2.2}$$

$$\Omega_l(t) = \left\{ \boldsymbol{x} \in \Omega; p(t, \boldsymbol{x}) < \frac{1}{2} \right\}, \tag{2.3}$$

$$\Gamma(t) = \left\{ \boldsymbol{x} \in \Omega; p(t, \boldsymbol{x}) = \frac{1}{2} \right\}. \tag{2.4}$$

In the following sections, the differential equation that provides the evolution law for p is derived. It arises by minimizing the Cahn-Hilliard free energy functional [58]

$$F = \int_\Omega g(\nabla p(t, \boldsymbol{x})) + h(p(t, \boldsymbol{x}), u(t, \boldsymbol{x})) \, \mathrm{d}\boldsymbol{x}, \tag{2.5}$$

where $g : \mathbb{R}^3 \to \mathbb{R}$ represents the free energy contribution due to fluctuations at the interface and $h : \mathbb{R}^2 \to \mathbb{R}$ represents the bulk free energy contribution. The next section clarifies the origin of (2.5).

2.1. Derivation of Cahn-Hilliard Free Energy Using The Lattice Liquid Model

Cahn-Hilliard free energy may be derived using first principles. Since the derivation is known, the following sections only give a overview of the process at a level of detail necessary for understanding the content of the subsequent sections. More details on the derivation itself can be found in [58]. For further clarification of the definitions and concepts used in this section, refer to [8, 47, 36].

The function p (see (2.1)) is an example of a so called order parameter. These functions are used to distinguish (either directly or indirectly) between phases. When modeling pure substance solidification, one uses the order parameter p to describe the transition between the liquid and solid phases. In the solid phase, the molecules have a fixed position inside of a lattice, which corresponds to the maximal ordering $p = 1$. On the other hand, when the substance is liquid, the molecules lack any order and thus the order parameter $p = 0$ represents the disorder of the phase. In a binary mixture, relative concentration of the components is used as an additional order parameter that describes the system [58]. As we will see, the free energy expressions involving either of these can be derived using the same procedure. A lattice liquid model, where molecules labeled type A and type B are placed into fixed slots is used to provide clarification. Using this model, the expression for free energy involving the order parameter that describes relative concentration in a binary mixture is derived. The free energy expression for the pure substance case with respect to the appropriate order parameter then follows by just considering any slot occupied by a molecule of the second type as vacant.

To derive the expression for local free energy, a cube shaped subset $\hat{\Omega} \subset \Omega$ is considered. The subset $\hat{\Omega}$ is then covered by a rectangular lattice. Consider a binary substance consisting of molecules labeled type A and B. To represent this mixture, any position in the lattice is filled with a molecule of type A or type B. The subset $\hat{\Omega}$ is considered to have constant volume V and temperature u.

2.1.1. Lattice Liquid Model

The lattice liquid model is designed to capture the following two idealized properties that liquids typically exhibit:

- incompressibility of the liquid (the repulsion of molecules is modeled by the presence of the lattice),
- attraction of molecules.

Even though the results hold for three dimensions, the presented derivation is performed for $\hat{\Omega} \subset \mathbb{R}^2$ for the sake of clarity [58]. Let d_1 and d_2 be the number of positions in the horizontal and vertical directions ($d_1, d_2 \in \mathbb{N}^+$) and let $N \equiv d_1 d_2$. Then we define a lattice with dimensions d_1 and d_2 as the set of ordered pairs

$$L_2(d_1, d_2) \equiv \{(k, l) : k \in \{1, \ldots, d_1\}, l \in \{1, \ldots, d_2\}\}.$$

Additionally, we define a linear indexing $q : L_2(d_1, d_2) \to \{1, \ldots, N\}$ as

$$q(k, l) = k + d_1(l - 1).$$

Unless stated otherwise, we refer to the lattice positions using the linear indexing q of the lattice. Then any position in the lattice is described by an $i \in \{1, \ldots, N\}$. Let the energy contribution from the attraction between molecules with indices i and j be expressed by $E(i, j)$. Typically one assumes that [47, 36]

- $E(i, j) \geq 0$,
- $E(i, j) \to 0$ when $d(i, j) \to \infty$, where d is the euclidean distance between the two centers of the cubes represented by indices i and j.

The total interaction energy then becomes the sum of all contributions corresponding to the pairs $(i, j) \in L_2(d_1, d_2)$. Define the system configuration $\boldsymbol{p} = (p_i)_{i \in \{1, \ldots, N\}}$ as

$$p_i \equiv \begin{cases} 1 & \text{if the i-th partition contains molecule B} \\ 0 & \text{if the i-th partition contains molecule A.} \end{cases} \tag{2.6}$$

After briefly reviewing the grand canonical ensemble (Section 2.1.2), this setting is taken advantage of in Section 2.1.3 to define the Hamiltonian of an open system governed by the lattice liquid model.

2.1.2. The Grand-canonical Ensemble of the Lattice Liquid Model

The lattice liquid model described in Section 2.1.1 corresponds to a system with two components. Let μ_A and μ_B denote the chemical potentials of the components made out of molecules of type A and B respectively. Assuming that the chemical potentials of either component μ_A, μ_B do not change allows us to use grand-canonical statistics (GCS) [47] to describe this system. More specifically:

- The system is determined by the constant quantities volume V, temperature u and chemical potentials μ_A and μ_B.
- Since each particle occupies a volume of a given fixed size, the dependence on volume can be replaced with the total amount of molecules N.

- Though the total amount of molecules remains the same (thus modeling an incompressible fluid) the system is open and thus number of molecules of each component N_A, N_B can fluctuate ($N_A + N_B = N$).

Let $\boldsymbol{p} \equiv (p_i : i \in \{1, \ldots, N\})$ (see (2.6)) describe the configuration of the system. Call $\beta = \frac{1}{kT}$ the thermodynamic beta, where k is the Boltzmann constant. The partition function Z describes the characteristic properties of a thermodynamic system in equilibrium. The probability of finding the system in state $\boldsymbol{p}$ is [36]

$$\varrho_{gc}(\boldsymbol{p}) \equiv \frac{1}{Z} e^{-\beta H(\boldsymbol{p}) + \beta\mu_A N_A(\boldsymbol{p}) + \beta\mu_B N_B(\boldsymbol{p})},$$

where $N_A(\boldsymbol{p}), N_B(\boldsymbol{p})$ represent the number of molecules of type A and B respectively present in configuration $\boldsymbol{p}$ and $Z = Z(V, u, \mu_A, \mu_B)$ is the partition function of the grand canonical ensemble defined as

$$Z = \sum_{\boldsymbol{p} \in S_{\text{possible}}} e^{-\beta H(\boldsymbol{p}) + \beta\mu_A N_A(\boldsymbol{p}) + \beta\mu_B N_A(\boldsymbol{p})},$$

where S_{possible} is the set of all 2^n possible configurations of the system.

Another possible way of describing the grand canonical ensemble is by using the characteristic potential of the ensemble P. Let $\langle Q \rangle$ denote the average of a quantity $Q(\boldsymbol{p})$ taken over all $\boldsymbol{p} \in S_{\text{possible}}$. Using standard arguments [47], one can see that the characteristic potential of the ensemble P is related to the partition function by

$$F - \mu_A \langle N_A \rangle - \mu_B \langle N_B \rangle = U - Su - \mu_A \langle N_A \rangle - \mu_B \langle N_B \rangle \equiv P(V, u, \mu_A, \mu_B) = -ku \ln Z, \tag{2.7}$$

where F is Helmholtz free energy, U is the internal energy, S is the entropy of the system and k is Boltzmann's constant.

2.1.3. Derivation of Free Energy Using The Lattice Liquid Model

The Hamiltonian of a system described by the lattice liquid model with configuration $\boldsymbol{p}$ reads [36]

$$H(\boldsymbol{p}) \equiv -\sum_{i=1}^{N} \sum_{j \in v_i} \left[\varepsilon_{AA}(1 - p_i)(1 - p_j) + \varepsilon_{AB}(1 - p_i) p_j + \varepsilon_{AB} p_i (1 - p_j) + \varepsilon_{BB} p_i p_j\right], \tag{2.8}$$

where $\varepsilon_{AA}, \varepsilon_{AB}, \varepsilon_{BB}$ are the interaction energies between molecules of different types and v_i is the set of lattice positions directly adjacent to i. Using (2.8), the mean field approximation and relationship (2.7) gives the formula for free energy [58, 47]

$$F(N, u, \mu_A, \mu_B) = F_S + F_U,$$

where F_U^{discrete} and F_S^{discrete} are the internal and entropic components of free energy respectively, defined as

$$F_S^{\text{discrete}}(N, u, \langle N_B \rangle) \equiv -kT \ln\left(\frac{N!}{\langle N_B \rangle! (N - \langle N_B \rangle)!}\right), \; F_U^{\text{discrete}}(N, u, \langle N_B \rangle) \equiv \langle H \rangle.$$

Algebraic manipulation, the assumption of short range interactions and the application of the continuum limit leads to [58] the continuous variants

$$F_S^{\text{continuous}} = \int_{\hat{\Omega}} \varepsilon_1(x, u(\boldsymbol{x}))\left(p(\boldsymbol{x}) \ln p(\boldsymbol{x}) + (1 - p(\boldsymbol{x})) \ln(1 - p(\boldsymbol{x}))\right) \mathrm{d}\boldsymbol{x}, \tag{2.9}$$

$$F_U^{\text{continuous}} = \int_{\hat{\Omega}} \frac{1}{2} |W_0(\boldsymbol{x}) \nabla p(\boldsymbol{x})|^2 + \varepsilon_2(u^*) p(\boldsymbol{x})(1 - p(\boldsymbol{x})) \mathrm{d}\boldsymbol{x}, \tag{2.10}$$

where $\varepsilon_1(\boldsymbol{x}, u(\boldsymbol{x})), \varepsilon_2(u^*)$ are dependent on the temperature field and critical temperature respectively, $W_0(\boldsymbol{x})$ is the amplitude of the surface energy contribution and $p : \hat{\Omega} \to \mathbb{R}$ is the continuous equivalent of $\boldsymbol{p} = (p_i)_{i\in\{1,\dots,N\}}$.

Generalizing (2.9) and (2.10) to the entire domain and defining

$$\begin{aligned} h(p(\boldsymbol{x}), u(\boldsymbol{x})) &= \varepsilon_1(x, u(\boldsymbol{x}))\,(p(\boldsymbol{x}) \ln p(\boldsymbol{x}) + (1 - p(\boldsymbol{x})) \ln(1 - p(\boldsymbol{x}))) \\ &\quad + \varepsilon_2(u^*)\, p(\boldsymbol{x})\,(1 - p(\boldsymbol{x})), \\ g(\nabla p(\boldsymbol{x})) &= \frac{1}{2} |W_0(\boldsymbol{x})\, \nabla p(\boldsymbol{x})|^2 \end{aligned} \tag{2.11}$$

gives rise to

$$F = \int_\Omega g(\nabla p(\boldsymbol{x})) + h(p(\boldsymbol{x}), u(\boldsymbol{x}))\, \mathrm{d}\boldsymbol{x}. \tag{2.12}$$

For reasons described in the following section, $g : \mathbb{R}^3 \to \mathbb{R}$ and $h : \mathbb{R}^2 \to \mathbb{R}$ may be viewed as arbitrary functions such that g represents the free energy density generated by fluctuations at the interface and h represents the free energy density of the bulk phases. This leads to a view, where we do consider (2.12) as fundamental and we do not insist that the functions g and h are dictated by (2.11).

2.1.4. Generalization of the Free Energy Expression

The motivation behind the generalized form of free energy (2.12) is explained in this section.

While the derivation given in Sections 2.1.1-2.1.3 used a lot of approximations, it is still based on standard statistical physics that are considered reliable. Therefore, defining g and h using (2.11) seems like a sensible choice at first. When performing numerical simulations that capture the evolution of p governed by (2.12) the interface region (defined by the values $p \in (0, 1)$) is thin compared to the size of the domain (see Section 3.2, 3.4). However, the mesh resolution needs to be sufficient in order to capture the interface for the numerical simulation to be viable [66]. Using the physically accurate forms of g and h given by (2.11) causes the interface to be of microscopic dimensions for many physical substances [19, 20, 58] which results in a theoretical mesh resolution beyond the applicability limits of any numerical method. This motivates us to consider the general form of free energy (2.12).

The relationship (2.12), theoretically validated in the previous section, is often used as a starting point for phase field modeling. Using alternative forms of g and h allows us to control the thickness of the interface though a dimensionless parameter $\xi > 0$. Using a suitable choice of ξ, the interface can be made thick enough for numerical simulations to be possible. To legitimize this approach, the two following viewpoints can be taken.

1. The asymptotic behavior as $\xi \to 0$ can be investigated using matched asymptotic analysis [15, 14, 23] (see Section 3.2).

2. It is possible to compare numerical results and experimental (see Section 3.5). For instance, one may compare the dendrite structure formed at a particular time (number of dendrite tips, size, curve shapes) or compare growth dynamics (dendrite tip velocities) [56, 66]. Standalone numerical experiments may also be performed to demonstrate how robust a model is w.r.t. different parameter settings [51, 9, 24, 61].

2.2. Derivation of the Phase Field Model

The phase field model (PFM) can be used to describe the evolution of a thermodynamic system near critical point [2, 14, 45]. In particular, it can be used to describe the interface dynamics of a pure super-cooled melt. It consists of two partial differential equations: the heat and phase field equation.

Using the results of Sections 2.1.1-2.1.4 and the theory of dynamic critical phenomena [45], the phase field equation can be derived and coupled with the heat equation which results in a PFM that governs the solidification of a pure super-cooled melt.

2.2.1. Model A Equation

One of the results of the theory of dynamic critical phenomena is the model A equation [45]. It describes a thermodynamic system relaxing towards equilibrium in a non-oscillatory manner. Historically, the equation was first applied to explain the anomalous dampening of the sound in helium near the λ-point [54].

Let $\Omega \subset \mathbb{R}^3$, $p : \Omega \to [0,1]$ be the phase field parameter describing the solid liquid transition and $u : \Omega \to \mathbb{R}$ be the temperature of a super-cooled melt filling out the domain Ω. Furthermore, let F be the free energy functional given by (2.12). Using δ to denote the directional derivative (or formal Fréchet derivative), the model A equation reads [45]

$$\frac{\partial p\left(\boldsymbol{x},t\right)}{\partial t} = -c\delta F\left(p\right) + \theta\left(\boldsymbol{x},t\right) + cE\left(\boldsymbol{x},t\right), \tag{2.13}$$

where $\theta\left(\boldsymbol{x},t\right)$ is a Gaussian white noise satisfying certain conditions, $E\left(\boldsymbol{x},t\right)$ expresses the effect of external fields on the system. Removing the stochastic component by setting $\theta\left(\boldsymbol{x},t\right) = 0$ makes the model less accurate physically, but deterministic. Considering a system with no external force acting on it results in $E\left(\boldsymbol{x},t\right) = 0$. Finally, defining the relaxation parameter τ makes it possible to rewrite (2.13) as

$$\tau\frac{\partial p}{\partial t} = -\delta F\left(p\right). \tag{2.14}$$

The relaxation parameter τ determines the rate at which the order parameter p relaxes towards its minimum, given by the minimization of the free energy functional F. A more general form of the model A equation, featuring multiple order parameters, is discussed more thoroughly in [50].

2.2.2. Derivation of the Phase Field Equation

Recalling the free energy functional (2.12), the directional derivative of F at point p in direction δp reads

$$\delta F\left(p;\delta p\right) = \int_{\Omega}\sum_{j=1}^{3}\frac{\partial g\left(\nabla p\right)}{\partial y_j}\frac{\partial \delta p}{\partial x_j}\delta p + \frac{\partial h\left(p,u\right)}{\partial p}\delta p\mathrm{d}\boldsymbol{x},$$

where u is treated as a parameter and $g = g\left(\boldsymbol{y}\right)$. Applying Green's formula leads to

$$\delta F\left(p;\delta p\right) = \int_{\Omega}\frac{\partial h\left(p,u\right)}{\partial p}\delta p - \sum_{j,i=1}^{3}\frac{\partial^2 g\left(\nabla p\right)}{\partial y_j \partial y_i}\frac{\partial^2 p}{\partial x_j \partial x_i}\delta p\mathrm{d}\boldsymbol{x} + \overbrace{\int_{\partial\Omega}\sum_{j=1}^{3}\frac{\partial g\left(\nabla p;\xi\right)}{\partial y_j}\delta p n_j \mathrm{d}S}^{I.}. \tag{2.15}$$

Suppose that one of the two following boundary conditions are satisfied

$$\delta p\,|_{\partial\Omega} = 0\,, \tag{2.16}$$

$$\left.\frac{\partial g\left(\nabla p\right)}{\partial \boldsymbol{n}}\right|_{\partial\Omega} = 0. \tag{2.17}$$

The boundary conditions (2.16) and (2.17) cause the integral I. in (2.15) to vanish. Then the formal Fréchet derivative with respect to p becomes

$$\delta F(p) = -\sum_{j,i=1}^{3} \frac{\partial^2 g(\nabla p)}{\partial y_j \partial y_i} \frac{\partial^2 p}{\partial x_j \partial x_i} + \frac{\partial h(p,u)}{\partial p} = -\nabla \cdot \frac{\partial g(\nabla p)}{\partial (\nabla p)} + \frac{\partial h(p,u)}{\partial p}. \tag{2.18}$$

To determine the relaxation parameter τ, recall the Gibbs-Thomson relation which associates the interface velocity v_Γ with the mean curvature κ and super-cooling $\Delta^* u = u - u^*$ by

$$\Delta s \Delta^* u = -\sigma \kappa_\Gamma - \alpha \sigma v_\Gamma, \tag{2.19}$$

where σ is the surface tension, α is the coefficient of attachment kinetics at the boundary and Δs is the difference between entropy per unit volume between Ω_l and Ω_s. Introducing a dimensionless parameter $\xi > 0$ that is proportional to the thickness of the interface [15] and using matched asymptotic analysis [13, 15], the validity of (2.19) implies that the relaxation parameter is of the form

$$\tau(\xi) = \alpha \xi^2. \tag{2.20}$$

Combining (2.18) and (2.20) leads to the general form of the phase field equation

$$\alpha \xi^2 \frac{\partial p}{\partial t} = \nabla \cdot \frac{\partial g(\nabla p)}{\partial (\nabla p)} + f(u,p), \tag{2.21}$$

where

$$-\frac{\partial h(p,u)}{\partial p} \equiv f(u,p). \tag{2.22}$$

2.2.3. The General Phase Field Problem

The temperature gradients of u within the domain Ω drives the solidification process. Its evolution within the domain is captured by the heat equation

$$\varrho(u)\, c(u) \frac{\partial u}{\partial t} = \nabla \cdot (\lambda(u) \nabla u) + H \frac{\partial p}{\partial t}, \tag{2.23}$$

where ρ is material density, c is material specific heat capacity, H is the latent heat of fusion per unit volume and λ is the heat conductivity. All the material constants introduced thus far are summarized in Table 2.1. These can be used to define a dimensionless equivalents (denoted by a tilde) using the relationships from Table 2.2. Note that the relationships of Table 2.2 imply that $\tilde{u}^* = 1$, $\tilde{u}_{\text{ini}} = 0$ and that $\tilde{\beta}$ scales with the value of initial supercooling Δu_{ini} (for simplicity, we assume that the initial super-cooling is uniform). Additionally, it may be shown the Gibbs-Thomson relation (2.19) does not change form when transitioning to the dimensionless quantities (consider the definition of β) [66]

$$\tilde{\beta}(\tilde{u} - \tilde{u}^*) = -\tilde{\kappa}_\Gamma - \tilde{\alpha} \tilde{v}_\Gamma. \tag{2.24}$$

Using these along with (2.21), (2.23) and adding suitable boundary and initial conditions results in the formulation [16, 23, 66]

$$\frac{\partial u}{\partial t} = \Delta u + H \frac{\partial p}{\partial t} \qquad \text{in } (0,T) \times \Omega, \tag{2.25}$$

$$\alpha \xi^2 \frac{\partial p}{\partial t} = \nabla \cdot \frac{\partial g(\nabla p)}{\partial (\nabla p)} + f(u,p) \qquad \text{in } (0,T) \times \Omega, \tag{2.26}$$

$$u|_{t=0} = u_{\text{ini}}, p|_{t=0} = p_{\text{ini}} \qquad \text{in } \Omega, \tag{2.27}$$

$$u|_{\partial\Omega} = u_{\partial\Omega} \text{ or } \nabla u \cdot \boldsymbol{n} = 0 \qquad \text{on } (0,T) \times \partial\Omega, \tag{2.28}$$

$$p|_{\partial\Omega} = p_{\partial\Omega} \text{ or } \frac{\partial g(u,p)}{\partial (\nabla p)} \cdot \boldsymbol{n} = 0 \qquad \text{on } (0,T) \times \partial\Omega, \tag{2.29}$$

where the tildes are omitted for the sake of readability and some of the constants that are not visible directly in this formulation are contained in f (see section 2.2.5). We call (2.25)-(2.29) the phase field model (PFM) governing the solidification of pure substances.

Table 2.1.: Physical quantities in the Phase-field problem with surface tension

Quantity	SI Unit	Description
u	K	temperature
ρ	$\mathrm{kg} \cdot \mathrm{m}^{-3}$	density
c	$\mathrm{J} \cdot \mathrm{kg}^{-1}\mathrm{K}^{-1}$	specific heat capacity
λ	$\mathrm{W} \cdot \mathrm{m}^{-1} \cdot \mathrm{K}^{-1}$	heat conductivity
H	$\mathrm{J} \cdot \mathrm{m}^{-3}$	latent heat of fusion per unit vol.
u^*	K	melting point
Δu_{ini}	K	initial supercooling $\Delta u_{\mathrm{ini}} = u^* - u\,\vert_{t=0}$
σ	$\mathrm{J} \cdot \mathrm{m}^{-2}$	surface tension
Δs	$\mathrm{J} \cdot \mathrm{m}^{-3} \cdot \mathrm{K}$	entropy difference per unit volume
μ	$\mathrm{m} \cdot \mathrm{s}^{-1} \cdot \mathrm{K}^{-1}$	interface mobility (see [70, 71])
α	$\mathrm{m}^{-1} \cdot \mathrm{s}$	coefficient of attachment kinetics $\alpha = \frac{\Delta s}{\mu\sigma}$
β	$\mathrm{m}^{-1} \cdot \mathrm{K}^{-1}$	$\beta = \frac{\Delta s}{\sigma}$

Table 2.2.: Characteristic scales and dimensionless quantities

Quantity	Definition	SI Unit	Description
L_0	user-defined	m	length scale
t_0	$(\rho c/\lambda)\, L_0^2$	s	time scale
$\tilde{u}$	$(u - u^*)\,/\Delta u_{\mathrm{ini}} + 1$	1	temperature
$\tilde{H}$	$H/\left(\rho c \Delta u_{\mathrm{ini}}\right)$	1	latent heat
$\tilde{\alpha}$	$(\lambda/\left(\rho c\right))\, \alpha$	1	attach. kin. coef.
$\tilde{\beta}$	$\beta L_0 \Delta u_{\mathrm{ini}}$	1	
$\tilde{\boldsymbol{x}}$	$\boldsymbol{x}/L_0$	1	position
$\tilde{t}$	t/t_0	1	time

2.2.4. Isotropic and Anisotropic Phase Field Model

Introducing anisotropic surface energy into model (2.25)-(2.29) can be done by means of Finsler geometry [12]. This method of modeling anisotropy has been shown to be viable [18, 17, 63, 64]. In order to do this, we first introduce the Finsler dual metric $\phi^0 : \mathbb{R}^3 \to \mathbb{R}$. Let $\psi : \mathbb{R}^3 \to \mathbb{R}$ be the anisotropic surface energy (for examples see (2.43), (2.44)), then ϕ^0 is of the form

$$\phi^0(\boldsymbol{\eta}^*) = |\boldsymbol{\eta}^*|\, \psi\left(-\frac{\boldsymbol{\eta}^*}{|\boldsymbol{\eta}^*|}\right), \tag{2.30}$$

where $\boldsymbol{\eta}^* = \nabla p$ ensures that $-\frac{\boldsymbol{\eta}^*}{|\boldsymbol{\eta}^*|}$ is the outward pointing surface normal to the level set $p = \frac{1}{2}$. The anisotropic operator $T^0 : \mathbb{R}^3 \to \mathbb{R}^3$ is defined using the Finsler dual metric ϕ^0 as

$$T^0(\boldsymbol{\eta}) = \phi^0(\boldsymbol{\eta})\, \phi^0_{\boldsymbol{\eta}}(\boldsymbol{\eta}), \tag{2.31}$$

where

$$\phi^0_{\boldsymbol{\eta}}(\boldsymbol{\eta}) = \left(\frac{\partial}{\partial \eta_1^*}\phi^0(\boldsymbol{\eta}^*), \frac{\partial}{\partial \eta_2^*}\phi^0(\boldsymbol{\eta}^*), \frac{\partial}{\partial \eta_3^*}\phi^0(\boldsymbol{\eta}^*)\right).$$

Using the Finsler metric, we define the interface energy contribution in (2.12) as

$$g(\nabla p) = \frac{\xi^2}{2}\left(\phi^0(\nabla p)\right)^2. \tag{2.32}$$

Using (2.32) in (2.26) and (2.29) leads to the anisotropic phase field model which reads

$$\frac{\partial u}{\partial t} = \Delta u + H\frac{\partial p}{\partial t} \qquad \text{in } \mathcal{J}\times\Omega, \tag{2.33}$$

$$\alpha\xi^2\frac{\partial p}{\partial t} = \xi^2\nabla\cdot T^0(\nabla p) + f(u,p;\xi) \qquad \text{in } \mathcal{J}\times\Omega, \tag{2.34}$$

$$u\,|_{t=0} = u_{\text{ini}},\ p\,|_{t=0} = p_{\text{ini}} \qquad \text{in } \Omega, \tag{2.35}$$

$$b_c(u) = 0 \qquad \text{on } (0,T)\times\partial\Omega, \tag{2.36}$$

$$T^0(\nabla p)\cdot\boldsymbol{n} = 0 \qquad \text{on } (0,T)\times\partial\Omega, \tag{2.37}$$

considering $\psi(\boldsymbol{n}) = 1$ leads to the isotropic phase-field model, which reads

$$\frac{\partial u}{\partial t} = \Delta u + H\frac{\partial p}{\partial t} \qquad \text{in } \mathcal{J}\times\Omega, \tag{2.38}$$

$$\alpha\xi^2\frac{\partial p}{\partial t} = \xi^2\Delta p + f(u,p;\xi) \qquad \text{in } \mathcal{J}\times\Omega, \tag{2.39}$$

$$u\,|_{t=0} = u_{\text{ini}}, p\,|_{t=0} = p_{\text{ini}} \qquad \text{in } \Omega, \tag{2.40}$$

$$u\,|_{\partial\Omega} = u_{\partial\Omega} \text{ or } \nabla u\cdot\boldsymbol{n} = 0 \qquad \text{on } (0,T)\times\partial\Omega, \tag{2.41}$$

$$p\,|_{\partial\Omega} = p_{\partial\Omega} \text{ or } \nabla p\cdot\boldsymbol{n} = 0 \qquad \text{on } (0,T)\times\partial\Omega. \tag{2.42}$$

Examples of the anisotropic surface energy functions for 4-fold and 6-fold anisotropy are listed [59]. Assuming that the coordinate axes are aligned with the direction of anisotropy, they read

$$\psi(\boldsymbol{n}) = 1 + A_1\left(n_1^4 + n_2^4 + n_3^4 - 6\left(n_1^2n_2^2 + n_2^2n_3^2 + n_3^2n_1^2\right)\right), \tag{2.43}$$

$$\psi(\boldsymbol{n}) = 1 + (A_1 + 3A_2)\left(n_1^4 + n_2^4 + n_3^4\right) - \frac{3}{5}A_1 + A_2\left(66n_1^2n_2^2n_3^2 - \frac{17}{7}\right) \tag{2.44}$$

respectively, where A_1 and A_2 are the anisotropy parameters.

2.2.5. An Overview of Significant Reaction Terms

Different types of reaction terms f were proposed in [22, 13, 51]. The ones mentioned here do not cover all the known possibilities, but they only serve to illustrate some of the possible deficiencies that reaction terms may have. First, the isotropic reaction terms, used with (2.38)-(2.42) are discussed.

Historically, one of the first proposed forms of the reaction term was [22, 51, 66]

$$f(p,u;\xi) = ap(1-p)\left(p-\frac{1}{2}\right) - b\beta\xi(u-u^*), \tag{2.45}$$

where a, b and β are positive constants. The advantage of (2.45) is its simplicity (linearity in u), but its great disadvantage is that it loses its physical interpretation for values of u outside a particular range [66].

Another form of the reaction term is [51]

$$f(p,u;\xi) = ap(1-p)\left(p-\frac{1}{2} - b\beta\arctan(\gamma(u-u^*))\right), \tag{2.46}$$

where $\gamma > 0$. This terms flaw is that it causes the asymptotic expansion of the solution u of (2.25)-(2.29) in ξ to not converge to the Gibbs-Thomson relation even in the first order as $\xi \to 0$.

To remedy these deficiencies, Beneš [13] proposed a term of the form

$$f(p,\nabla p,u;\xi) = ap(1-p)\left(p-\frac{1}{2}\right) - \beta\xi^2|\nabla p|(u-u^*). \tag{2.47}$$

Note that f contains an additional dependency on ∇p. This term does not share the disadvantages of terms (2.45), (2.46), but numerical experiments have shown that in realistic simulations with large under-cooling, spontaneous nucleation [66] occurs. In addition to this, the addition of the $|\nabla p|$ could be considered non-physical in some sense, since the formulation of Cahn-Hilliard free energy (2.12) does not admit the dependence of h on the derivatives of p. In Chapter 3, a new reaction term f is proposed. This term maintains the advantages of (2.47), but is independent of ∇p and exhibits superior properties in numerical simulations [66].

The reaction term (2.47) can be augmented to include anisotropy [14] resulting in

$$f(u,p,\nabla p;\xi) = 2p(1-p)\left(p-\frac{1}{2}\right) + \xi^2 b\beta\phi^0(\nabla p)(u^*-u), \tag{2.48}$$

which is then used as the reaction term in (2.33)-(2.35).

3. New Forms of the Reaction Term for the PFM

We propose alternative reaction terms f for the anisotropic and isotropic variants of the PFM (2.25)-(2.29) [66]. As was detailed in Section 2.2.5, the reaction terms (2.47) and (2.48) are void of some of the limitations of (2.45), (2.46), but still posses some unwanted qualities. This motivates us to search for alternative reaction terms in hopes of improving the behavior of the resulting PFM.

Building on the discussion in Section 2.2.5, we focus on the following properties.

1. The asymptotic recovery (inner expansion) of the Gibbs-Thomson's relation (2.24) in the limit $\xi \to 0$ is possible. This is a desirable property that we would like to keep.

2. To allow the interface to form properly for arbitrary values of undercooling Δu_{ini}, we demand that the reaction term f has three roots: 0 and 1 and an additional root in $(0, 1)$ for any value of Δu_{ini}. This condition is necessary for the proper formation of the interface [8].

3. The latent heat release in the bulk phases should be minimal to prevent spontaneous nucleation.

4. Lastly, we aim to remove the dependence on ∇p, since the term should be independent of ∇p by the first principles derivation of Section 2.1. Removing the dependence of ∇p has the additional advantage of simplifying numerical schemes and analysis.

3.1. Decomposition of the Reaction Term

Consider the general form of the reaction term

$$f\left(u, p, \nabla p; \xi\right) = f_0\left(p\right) + \xi f_1\left(u, p, \nabla p; \xi\right) \tag{3.1}$$

where

$$f_0\left(p\right) = 2p\left(1-p\right)\left(p - \frac{1}{2}\right)$$

represents the derivative of the double-well potential [23]

$$\omega_0\left(p\right) = \frac{1}{2}\left(\left(p - \frac{1}{2}\right)^2 - \frac{1}{4}\right)^2. \tag{3.2}$$

Note that Landau theory [8] dictates the form of ω_0. The term f has roots at $0, \frac{1}{2}$ and 1 when $\xi = 0$. Non-zero $\xi > 0$ may displace the positions of the roots of f or even reduce the number of real roots. Significant displacement of the roots or reduction of their number is in conflict with Landau theory and prevents the proper formation of the interface given by (2.4).

3.2. Results of Matched Asymptotic Analysis

To discuss the relevant results of asymptotic analysis, consider an alternative coordinate system. Let $t \in (0, T)$, $\boldsymbol{x}_0 \in \Gamma\left(t\right)$ and z be a coordinate defined by

$$\bar{p}\left(z\right) = p\left(t, \boldsymbol{x}_0 + \xi z \boldsymbol{n}_\Gamma\right), \tag{3.3}$$

where $\boldsymbol{n}_\Gamma$ is the outward pointing unit normal vector to $\Gamma(t)$ at $\boldsymbol{x}_0$ with respect to $\Omega_s(t)$. We mention two important results of the asymptotic analysis shown in [15]. These will be used to justify some of the subsequent choices. The first result states that the function $\bar{p}$ satisfies

$$\bar{p}(z) = \bar{p}_{\text{asy}}(z) + o(\xi) \text{ where } \bar{p}_{\text{asy}}(z) = \frac{1}{2}\left[1 - \tanh\left(\frac{z}{2}\right)\right]. \tag{3.4}$$

The other important result is the asymptotic recovery of the Gibbs-Thomson relation as $\xi \to 0$. This relation has the form

$$\frac{\mathcal{I}_1}{\mathcal{I}_2} = -\kappa_\Gamma - \alpha v_\Gamma, \tag{3.5}$$

where

$$\mathcal{I}_1 = \int_{-\infty}^{+\infty} f_1\left(\bar{u}, \bar{p}, \frac{\mathrm{d}\bar{p}}{\mathrm{d}z}\right)\frac{\mathrm{d}\bar{p}}{\mathrm{d}z}\mathrm{d}z \tag{3.6}$$

and

$$\mathcal{I}_2 = \frac{1}{6}. \tag{3.7}$$

The relationship (3.7) holds due to (3.2). Considering (3.6) and (3.7) one can observe that f_1 needs to be of suitable form for (2.24) to hold.

3.3. Construction of the ΣP1-P Reaction Term

The term (2.47) is used as a starting point. Using (3.3) we may write

$$|\nabla p(\boldsymbol{x}_0)| = \left|\frac{\partial p}{\partial \boldsymbol{n}_\Gamma}(\boldsymbol{x}_0)\right| = \left|\frac{1}{\xi}\frac{\mathrm{d}\bar{p}}{\mathrm{d}z}(0)\right|. \tag{3.8}$$

To make the term independent of the gradient we use the fact that the asymptotic profile $\bar{p}_{\text{asy}}$ must satisfy the differential equation [15]

$$\frac{\mathrm{d}\bar{p}_{\text{asy}}}{\mathrm{d}z} = -2\bar{p}_{\text{asy}}(1 - \bar{p}_{\text{asy}}).$$

Using (3.4) motives us to write

$$|\nabla p(\boldsymbol{x}_0)| \approx \frac{2}{\xi} p(1-p).$$

Using this replacement, we formulate the resulting reaction term f as

$$f(u, p; \xi) = 2p(1-p)\left(p - \frac{1}{2} + \xi b\beta\frac{1}{2}(u^* - u)\right). \tag{3.9}$$

An equivalent formula is mentioned by Kobayashi in [51]. One can use (3.5)-(3.7) to find that the satisfaction of Gibbs-Thomson's relation (2.24) is ensured by setting $b = 1$ [66]. A major shortcoming of this formulation is that f only has a third real root between 0 and 1 (in p) if the condition

$$\xi\beta < 1 \tag{3.10}$$

holds. When $\xi > 0$ is fixed, β scales with initial supercooling Δu_{ini} and thus (3.10) translates into a bound on the supercooling. Violating this bound results in the model no longer having a physical interpretation. The existence of a third root between 0 and 1 ensures the proper formation of the phase interface. In order to preserve the third root, we propose the use of a differentiable sigmoid function

$$\Sigma\left(p;\varepsilon_1,\varepsilon_2\right) \equiv \begin{cases} 0 & p \leq \varepsilon_0 \\ 1 & p \geq \varepsilon_1 \\ \frac{3(p-\varepsilon_0)^2}{(\varepsilon_1-\varepsilon_0)^2} - \frac{2(p-\varepsilon_0)^3}{(\varepsilon_1-\varepsilon_0)^3} & p \in (\varepsilon_0,\varepsilon_1)\,, \end{cases} \tag{3.11}$$

where $0 < \varepsilon_0 < \varepsilon_1$ are small parameters. Consider a reformulation of the reaction term (3.9) using the sigmoid function that reads

$$f\left(p,u;\xi\right) = 2p\left(1-p\right)\left(p - \frac{1}{2} + \frac{1}{2}b\beta\xi\Sigma\left(p;\varepsilon_0,\varepsilon_1\right)\left(u-u^*\right)\right). \tag{3.12}$$

Computing the parameter b so that the Gibbs-Thomson relation (2.24) is satisfied leads to [66]

$$b = \frac{1}{6}\Big/\left(\frac{\varepsilon_0^3}{15} + \frac{\varepsilon_0^2\varepsilon_1}{10} - \frac{3\varepsilon_0^2}{20} - \frac{\varepsilon_0\varepsilon_1}{5} + \frac{\varepsilon_0\varepsilon_1^2}{10} - \frac{3\varepsilon_1^2}{20} + \frac{\varepsilon_1^3}{15} + \frac{1}{6}\right). \tag{3.13}$$

Note that in this case, the parameter b is dependent on $\varepsilon_0, \varepsilon_1$ and in general will differ from the one computed for the reaction term (3.9). The reaction term (3.12) will be refereed to as ΣP1-P.

3.4. Other Variants of Reaction Terms using the Σ Function

Allowing a dependence on ∇p, the function Σ can be incorporated into the reaction term (2.47). This modified term, which we call ΣGradP, reads

$$f\left(p,\nabla p,u;\xi\right) = ap\left(1-p\right)\left(p-\frac{1}{2}\right) - b\beta\xi^2\Sigma\left(p;\varepsilon_0,\varepsilon_1\right)\left|\nabla p\right|\left(u-u^*\right). \tag{3.14}$$

Additionally, anisotropy can be introduced by replacing $|\nabla p|$ with its anisotropic variant $\phi^0\left(\nabla p\right)$ resulting in

$$\hat{f}\left(u,p,\nabla p;\xi\right) = 2p\left(1-p\right)\left(p-\frac{1}{2}\right) + \xi^2 b\beta\Sigma\left(p;\varepsilon_0,\varepsilon_1\right)\phi^0\left(\nabla p\right)\left(u^*-u\right). \tag{3.15}$$

The term (3.15) is then used along with the anisotropic variant of the PFM (2.33)-(2.35). When a PFM uses the term (3.15), we call it a $\Sigma\phi^0$GradP model and when the term (2.48) is used we call it ϕ^0GradP. When it is clear that we deal with the anisotropic variant, we may omit the ϕ^0.

3.5. Validation of the Proposed Terms using Numerical Simulations

Apart from asymptotic analysis, the new variants of the reaction terms are additionally validated by numerical simulations [66]. We present some of these simulations, others may be found in the aforementioned article [66]. We focus in particular on the comparison of the original GradP model in its isotropic and anisotropic variant (2.47),(2.48) and the newly proposed reaction terms (3.12), (3.14) and (3.15). A summary of all the reaction terms involved can be found in Table 3.1.

All simulations were performed using an efficient hybrid OpenMP/MPI parallel implementation [64] of a numerical solver based on multi-point flux approximation finite volume scheme on a uniform rectangular mesh [65] and 4th order Runge-Kutta-Merson integrator with adaptive stepping in time [26].

3.5.1. Setup of the Numerical Experiments

In the following set of experiments, the domain Ω is a cube $\Omega \equiv (0,l)^3$. The discretization is uniform and we denote the number of nodes on one axis by N. The values of l and N along with the anisotropy

Table 3.1.: Overview of the

Term name	f is given by
GradP	(2.47)
ϕ^0GradP	(2.48)
ΣP1-P	(3.12)
ΣGradP	(3.14)
Σϕ^0GradP	(3.15)

strengths are listed with each simulation. Neumann boundary conditions are applied for both u and p. The initial condition for the phase field is given by

$$p_{\text{ini}}(\boldsymbol{x}) = \begin{cases} 1 & \boldsymbol{x} \in \Omega_s(0), \\ 0 & \boldsymbol{x} \in \Omega_l(0) = \Omega \backslash \Omega_s(0) \end{cases}$$

and a reference set of parameters $\alpha = 3$, $\beta = 300$, $L = 2$ and $\xi_0 = 0.011$ is given. To investigate the asymptotic behavior of the models $\frac{\xi_0}{2}$ or $\frac{\xi_0}{4}$ is sometimes considered in place of ξ_0. These parameters are used for all but the realistic simulations performed in Section 3.5.4.

3.5.2. Setting the Parameters ε_0 and ε_1

The behavior of the newly proposed terms (3.12), (3.14) and (3.15) is dependent on the choice of the parameters of $\Sigma(p; \varepsilon_0, \varepsilon_1)$. The values investigated read

$$\varepsilon_0 \in \{0, 0.05\}, \ \varepsilon_1 \in \{0.15, 0.20, 0.25, 0.30\}.$$

The function Σ as well as the profile of latent heat release around the interface $p(1-p)\, b\Sigma(p; \varepsilon_0, \varepsilon_1)$ are depicted in Figure 3.1. The results were obtained using $\ell = 6$, $N = 600$, and 4-fold anisotropy given by (2.43) with $A_1 = 0.02$. A nucleation site of diameter 0.05 at initial time $t = 0$ was placed at $\boldsymbol{x} = \boldsymbol{0}$ and the crystal evolution is assumed to be symmetric. All the final crystal shapes are then compared at time $t = 0.22$.

Figure 3.2 reveals that using the reference value $\xi = \xi_0$ leads to the three-dimensional crystal morphologies having significant differences depending on the values of $\varepsilon_0, \varepsilon_1$. This indicates that such a value of ξ is too large to experience the asymptotic behavior of the ΣP1-P model, which should be independent of $\varepsilon_0, \varepsilon_1$. This motivates repeating the simulations with a finer interface $\xi = \xi_0/2$ and $\xi = \xi_0/4$ in hopes to capture the predicted asymptotic behavior. At the same time, the mesh resolution was increased by keeping $N = 600$ and reducing the domain size from $\ell = 6$ to $\ell = 3$ and $\ell = 2$, respectively, which accommodates the growing crystal up to time $t = 0.22$ (hence the choice of t). In these experiments, we can observe much simpler and smaller dendritic structures, which makes it possible to visually represent all the combinations in Figure 3.3. Across all the combinations of $(\varepsilon_0, \varepsilon_1)$ we can observe that the decreasing ξ results in a convergence to a limit shape (as predicted by asymptotic analysis). Moreover, Figure 3.2 indicates that particular choices of $\varepsilon_0, \varepsilon_1$ could possibly increase the convergence rate to the asymptotic limit as $\xi \to 0$.

3.5.3. Comparison of Models

In this section, the ΣP1-P and ϕ^0GradP models are compared. Figure 3.4 compares the behavior of the ΣP1-P model and the ϕ^0GradP model for the values $\xi = \xi_0/2$ and $\xi = \xi_0/4$. The parameters $\varepsilon_0, \varepsilon_1$ of the ΣP1-P model are set to $\varepsilon_0 = 0.05$, $\varepsilon_1 = 0.2$, which is the setting for which the ϕ^0GradP and the ΣP1-P models produce results that are most similar in size and shape across all three values of ξ. With decreasing ξ, both models appear to converge to a common limit shape, even in the presence of anisotropy.

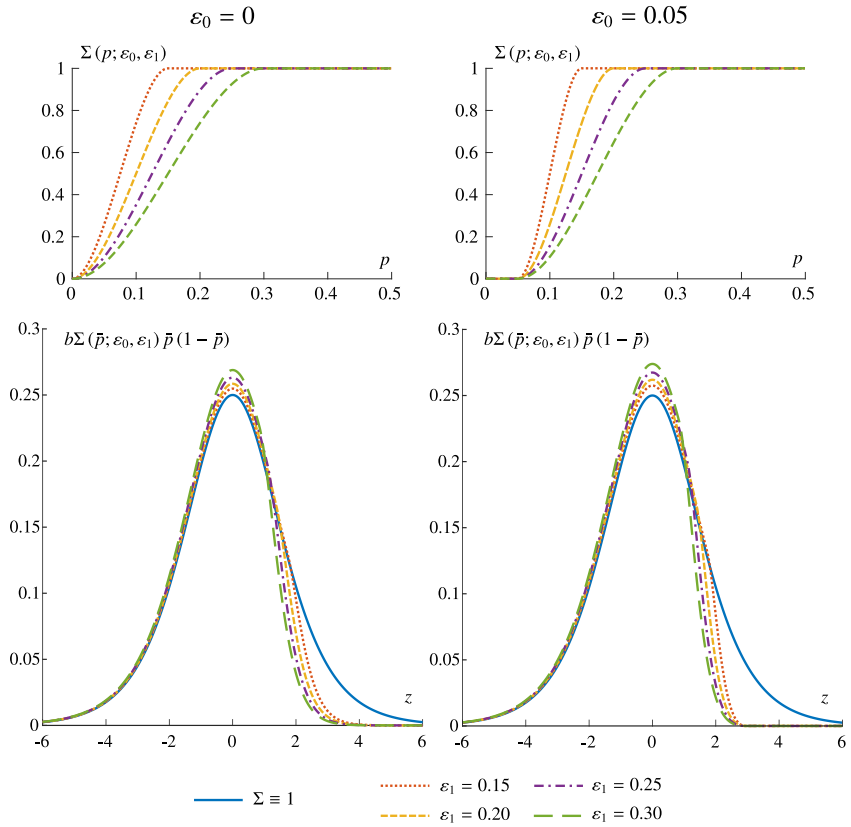

Figure 3.1.: Plots of $\Sigma(p;\varepsilon_0,\varepsilon_1)$ for different values of $\varepsilon_0,\varepsilon_1$ (top row) and of the term $b\Sigma(\bar{p}(z);\varepsilon_0,\varepsilon_1)\,\bar{p}(z)\,(1-\bar{p}(z))$ which is proportional to the latent heat release rate across the asymptotic profile of the diffuse interface (3.4) (bottom row). The value of b depends on $\varepsilon_0,\varepsilon_1$ as given by the satisfaction of the Gibbs-Thomson relation (2.24).

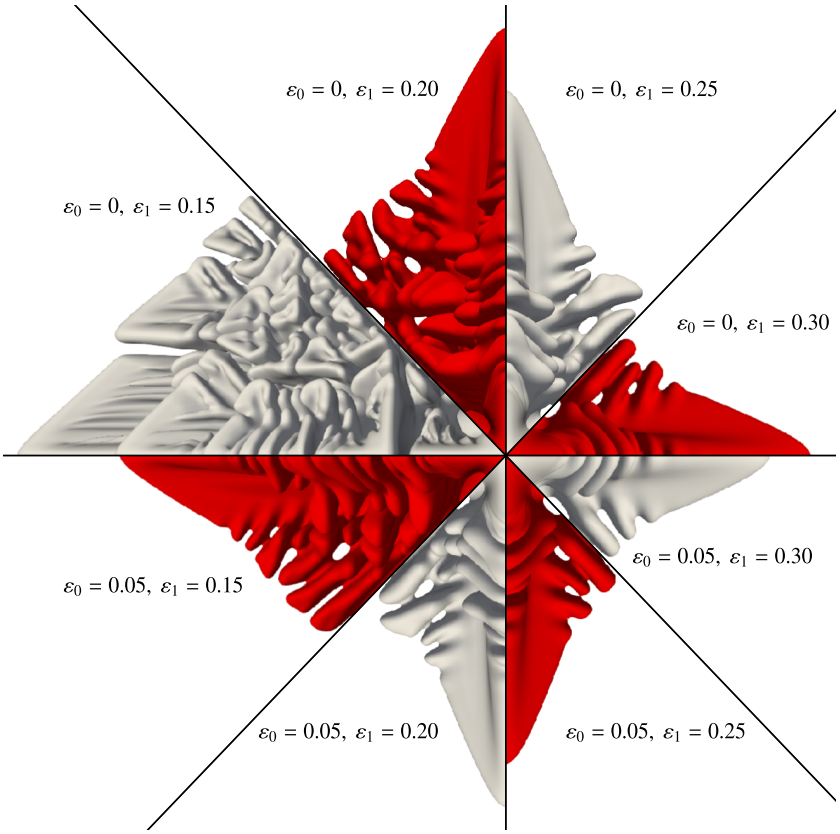

Figure 3.2.: Influence of $\varepsilon_0, \varepsilon_1$ on the crystal shape obtained by the ΣP1-P model with $\xi = \xi_0$ and b given by the adherence to (2.24).

The next set of results illustrates the properties of the ΣP1-P model in comparison with the original GradP and ϕ^0GradP models for a fixed interface thickness expressed by $\xi = \xi_0$. In addition, to see the practical importance of the correct value of b, we demonstrate the results of the ΣP1-P model in two versions: with b given by (3.13) and with $b = 1$.

Figure 3.5 demonstrates the solution of all the aforementioned models at time $t = 0.36$. Four-fold anisotropy given by (2.43) is applied and the parameters of the numerical simulation are $\ell = 6$, $N = 600$, $A_1 = 0.02$, and $\xi = \xi_0$. A nucleation site with radius 0.05 is located in the corner of the domain Ω. Inspired by the previous section, the ΣP1-P model parameters read $\varepsilon_0 = 0.05$, $\varepsilon_1 = 0.2$. For this choice, all four models produce qualitatively similar crystal shapes. In particular, a main dendrite with comparable amounts of side branching can be observed. Figures 3.5 and 3.2 show, among other things, that the ΣP1-P model is significantly more sensitive to changes in $\varepsilon_0, \varepsilon_1$ than b.

3.5.4. Rapid Solidification of Nickel

In this experiment, rapid dendritic solidification (large undercooling) of nickel is investigated. The results given by the novel models are compared with a number of experimental [56, 72, 41] and computational results [21, 46]. 6-fold anisotropy (2.44) with strengths $A_1 = 0.09, A_2 = -0.011$ is applied along with the dimensionless parameters that follow from applying the conversions from Table 2.2 to the material parameters for Nickel summarized in Table 3.2. The models involved in the study are ϕ^0GradP, Σϕ^0GradP, ΣP1-P without a noise term and ΣP1-P with temperature field perturbation. The temperature perturbation is realized using stochastic heat noise that replaces $f(u, p, \nabla p; \xi)$ with $f(u + \delta\hat{u}, p, \nabla p; \xi)$ and is applied to the ΣP1-P reaction term in some cases [66].

The main goal of the experiment is to compare the dendrite tip velocities for the following values of undercooling

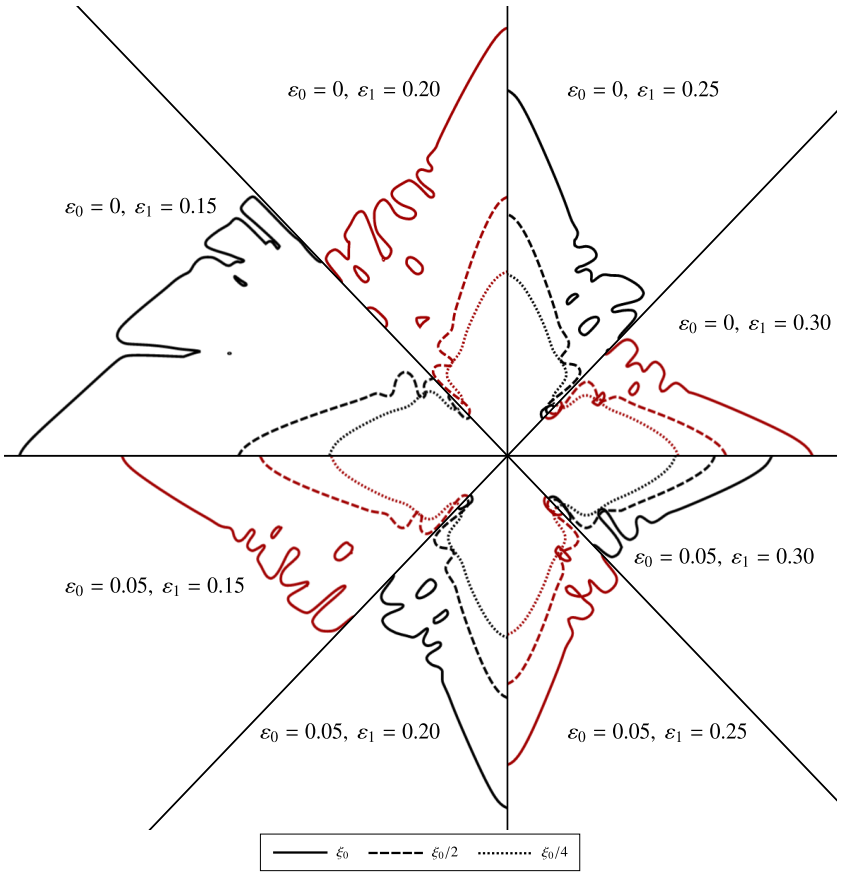

Figure 3.3.: Visual convergence study of the ΣP1-P model for $\xi \to 0$ with different settings of $\varepsilon_0, \varepsilon_1$.

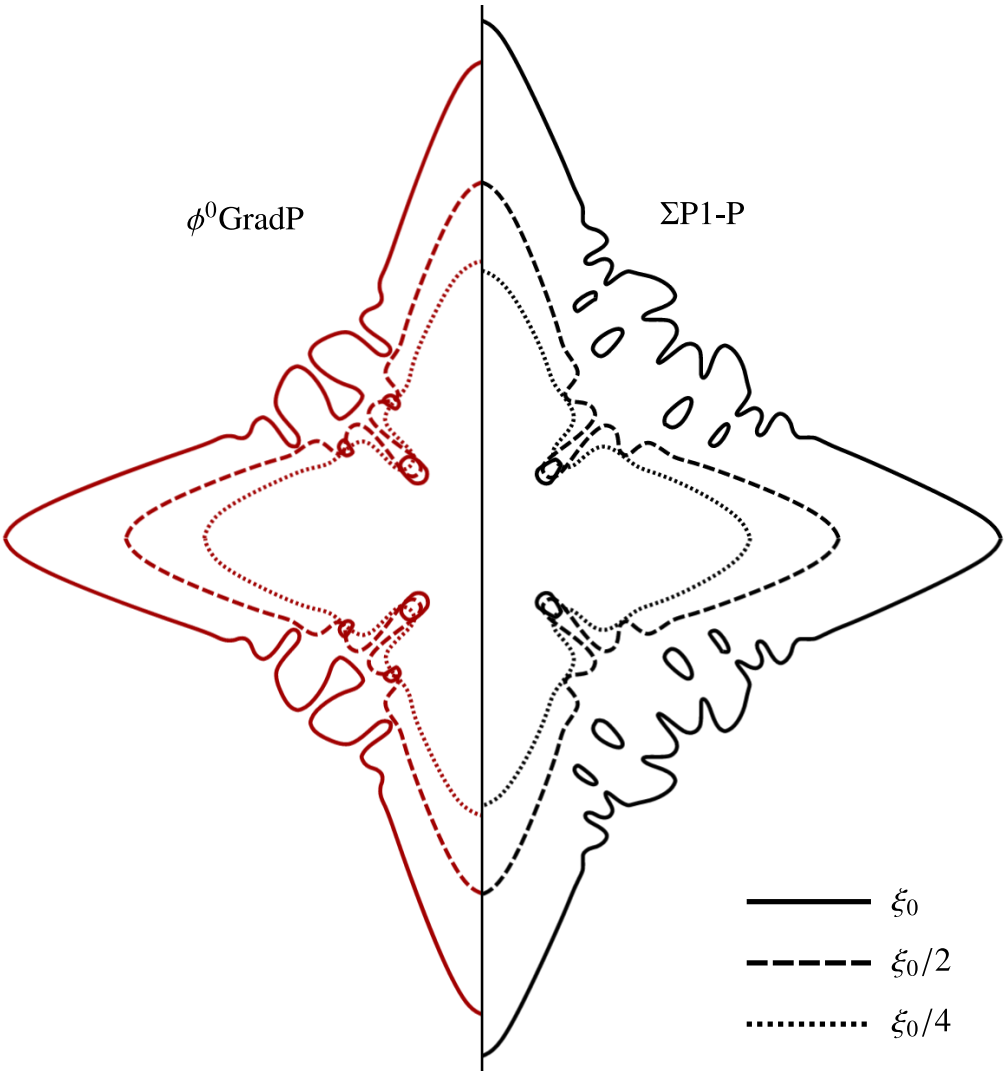

Figure 3.4.: Visual convergence study for $\xi \to 0$ comparing the ϕ^0GradP model and the ΣP1-P model with $\varepsilon_0 = 0.05, \varepsilon_1 = 0.2$.

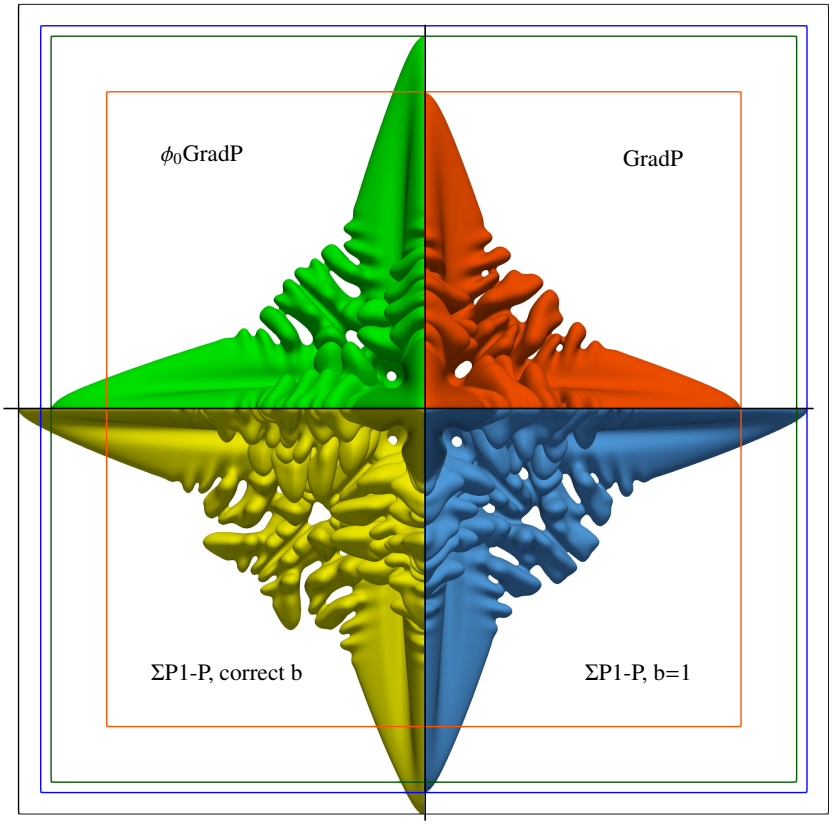

Figure 3.5.: Comparison of crystal shapes obtained by the four reaction term models. The squares indicate the bounding boxes for the individual crystals to aid in the visual comparison of average dendrite tip velocities.

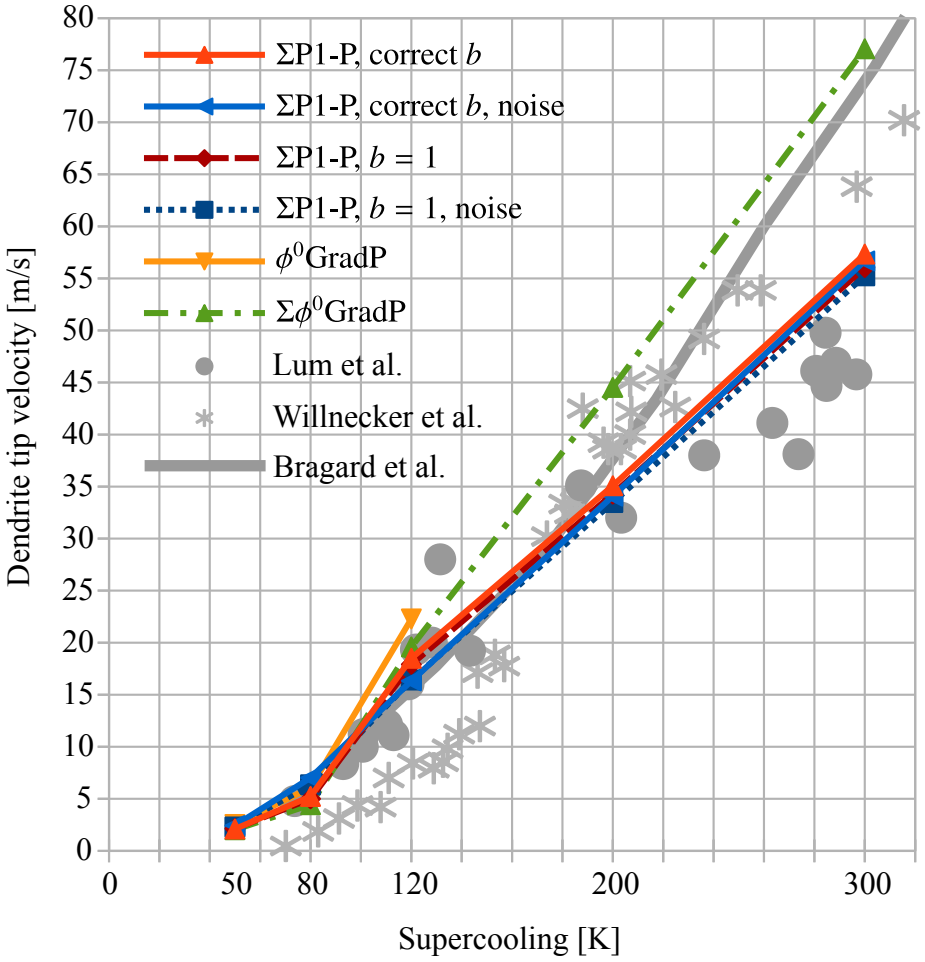

Figure 3.6.: Dendrite tip velocity in pure Ni solidification for large values of supercooling. Results obtained by the different models compared to experimental data by Lum et al. [56] and Willnecker et al. [72] and to phase field simulations by Bragard et al. [21].

$$\Delta u_{\mathrm{ini}} \in \{50\ \mathrm{K}, 80\ \mathrm{K}, 120\ \mathrm{K}, 200\ \mathrm{K}, 300\ \mathrm{K}\}$$

between numerical simulations utilizing various reaction terms (both our own and the ones found in literature) and experimental results. The results are summarized in Figure 3.6. The original ϕ^0GradP model fails after some time at $\Delta u_{\mathrm{ini}} = 120$ (see Figure 3.7) and fails immediately for $\Delta u_{\mathrm{ini}} \geq 200$ with spurious nucleation sites spreading all over the computational domain despite Δu_{ini} being far from the expected onset of homogeneous nucleation [35]. The $\Sigma\phi^0$GradP model gives the highest predictions of dendrite tip velocity. In all its variants, the ΣP1-P model yields very similar results that best agree with the measurements by Lum et al. [56]. In fact, all the variants of the ΣP1-P model indicate the transition from power-law to linear dependence described both by Lum et al. [56] and Willnecker et al. [72], which is not observed in the simulations of Bragard et al. [21]. Note that, switching between the two values of b for the ΣP1-P model has a negligible effect.

In conclusion, we have seen that the use of the terms including the Σ function prevent spontaneous nucleation for large values of undercooling. Furthermore, they have been shown to capture the transition from power-law to linear trend of dendrite tip velocity v_Γ dependent on undercooling Δu_{ini} rather well. Further numerical experiments discussing other aspects can be found in [66].

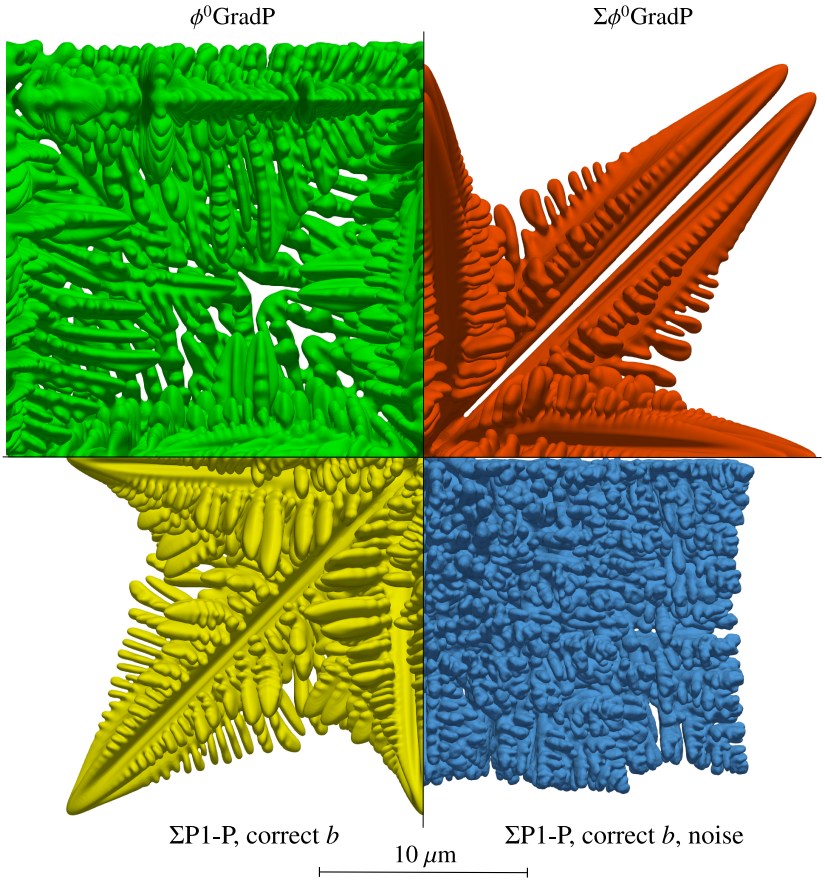

Figure 3.7.: Comparison of crystal shapes obtained by different models for rapid Ni solidification at time $t = 1.3$ μs and the initial supercooling $\Delta u_{\text{ini}} = 120$ K. The ϕ^0GradP model fails as spurious nucleation sites occur.

Table 3.2.: Physical and computational parameters of rapid solidification of pure Ni under very large supercooling. The relation $b = 1$ was used for the ϕ^0GradP and $\Sigma\phi^0$GradP models and both relations $b = 1$ and b consistent with the relation (3.5) were used for the ΣP1-P model.

Phys. qty	Value	Param.	Value
Δu_{ini}	$50 - 300$ K	L_0	10 μm
ρ	8900 kg $\cdot$ m^{-3}	t_0	10^{-5} s
c	609 J $\cdot$ kg^{-1}K^{-1}	ℓ	2
λ	54.2 W $\cdot$ m^{-1} $\cdot$ K^{-1}	N	480
L	2.35×10^9 J $\cdot$ m^{-3}	ξ	0.002
u^*	1728 K	ε_0	0.05
σ	0.37 J $\cdot$ m^{-2}	ε_1	0.2
Δs	1.36×10^6 J $\cdot$ m^{-3} $\cdot$ K	b	see caption
μ	1.99 m $\cdot$ s^{-1} $\cdot$ K^{-1}	δ	0.05

Part II.

Optimization on Normed Linear Spaces

4. General Optimization Theory

In this Chapter, standard optimization theory on Normed Linear Spaces is reviewed [43, 55]. Note that some of the concepts are introduced in a less general manner to better suit the purposes of this presentation.

4.1. Constrained Optimization Problem

Let X_S, X_C be reflexive Banach spaces and Z be a Hilbert space. The spaces X_S and X_C are called the solution and control space, respectively. Consider a set of admissible controls $W_{\text{ad}} \subset X_C$, where W_{ad} is a closed, convex and non-empty. Define the state equation as

$$e\left(y, \theta\right) = 0, \tag{4.1}$$

where $e : X_S \times X_C \to Z$ is called the state map. In a particular example, the state equation (4.1) could represent a system of ordinary differential equations (ODE) with initial conditions (IC) or a system of partial differential equations (PDE) with IC and boundary conditions (BC). In this setting, y represents the solution of the ODE or PDE and θ is part of the formulation of the differential equation. In particular, θ could represent the source term or be contained in the boundary condition for instance. Assume that there exists a map $S : X_C \to X_S$ such that

$$e\left(S\left(\theta\right), \theta\right) = 0.$$

We call S the solution operator. If e represents an ODE with IC (or PDE with IC and BC) the existence of the map S is precisely the statement that there exists a unique solution for any control $\theta \in X_C$.

Definition 1. (constrained minimization problem) Let $X_S, X_C, Z, W_{\text{ad}}$ and e be as described above and $J : X_S \times X_C \to \mathbb{R}$ be a functional. Then the constrained minimization problem (with possibly non-linear constraints) reads

$$\min_{\theta \in W_{\text{ad}}} J\left(y, \theta\right) \tag{4.2}$$

$$\text{s.t. } e\left(y, \theta\right) = 0 \text{ where } y \in X_C, \theta \in X_S. \tag{4.3}$$

We call $(y_0, \theta_0) \in X_S \times W_{\text{ad}}$ a local solution of (4.2)-(4.3) if $e\left(y_0, \theta_0\right) = 0$ and there exists an open neighborhood V in $X_S \times X_C$ such that

$$J\left(y, \theta\right) \geq J\left(y_0, \theta_0\right) \text{ for all } \left(y, \theta\right) \in X_S \times W_{\text{ad}} \cap V \text{ such that } e\left(y, \theta\right) = 0.$$

A local solution (y_0, θ_0) of (4.2)-(4.3) is called global if

$$J\left(y, \theta\right) \geq J\left(y_0, \theta_0\right) \text{ for all } \left(y, \theta\right) \in X_S \times U_{\text{ad}} \text{ such that } e\left(y, \theta\right) = 0.$$

4.2. Existence of Optimal Solution for the Constrained Optimization Problem

The existence of solution for the problem (4.2)-(4.3) can be addressed by imposing the following set of assumptions:

Definition 2. (regularity conditions for the constrained optimization problem (4.2)-(4.3)) If (4.2)-(4.3) are such that:

1. W_{ad} is a bounded set,
2. the solution operator $S : W_{\text{ad}} \to X_S$ of (4.3) exists and is bounded,
3. the map $e : X_S \times X_C \to Z$ in (4.3) is weak sequential continuous,
4. the map $J : X_S \times X_C \to \mathbb{R}$ is weak sequential lower semicontinuous,

then the optimization problem (4.2)-(4.3) is called regular.

Under the assumptions in this definition, the existence of a global solution may be proven.

Theorem 3. *Let (4.2)-(4.3) satisfy the assumptions of Definition 2. Then (4.2)-(4.3) has a global solution.*

Proof. Define

$$E \equiv \{(y, \theta) \in X_S \times W_{\text{ad}} : e(y, \theta) = 0\}$$

Since E is non-empty (due to assumption 2. of Definition 2) and $J \geq 0$, the set

$$\{J(y, \theta) : (y, \theta) \in E\} \subset [0, \infty)$$

is non-empty as well. Therefore $\inf_{(y,\theta)\in E} J(y, \theta) \equiv J_{\text{inf}}$ exists. Let $(y_n, \theta_n) \in E$ be such that

$$J(y_n, \theta_n) \overset{n\to\infty}{\to} J_{\text{inf}}.$$

Assumptions 1. and 2. of Definition 2 imply that (y_n, θ_n) is bounded in $X_S \times X_C$ and due to reflexivity there exists a sub-sequence (y_{k_n}, θ_{k_n}) and $(y_0, \theta_0) \in X_S \times X_C$ such that

$$(y_{k_n}, \theta_{k_n}) \rightharpoonup (y_0, \theta_0). \tag{4.4}$$

Since (4.4) implies that $\theta_{k_n} \rightharpoonup \theta_0$ in X_C and any convex strongly closed set is weakly closed $\theta_0 \in W_{\text{ad}}$ (weakly closed implies weakly sequentially closed). Due to point 3. of Definition 2 and (4.4) $e(y_0, \theta_0) = 0$ and so $(y_0, \theta_0) \in E$. To finalize the proof assumption 4. of Definition 2 is used to show that

$$J_{\text{inf}} = \lim_{n\to\infty} J(y_n, \theta_n) \geq J(y_0, \theta_0) \geq J_{\text{inf}},$$

which results in $J(y_0, \theta_0) = J_{\text{inf}}$. □

4.3. Basic Optimization Problem and Optimality Conditions

To derive optimality conditions (OC), a more elementary problem needs to be discussed first. Based on this theory and some additional assumptions, the OC for (4.2)-(4.3) may be formulated. The concepts of differentiability on normed linear space are used in this section. The notation used is standard and may be reviewed in the appendix.

Definition 4. (elementary minimization problem) Let X be a Banach space and U an open set in X. Suppose that $J : U \to \mathbb{R}$ and $C \subset X$ be non-empty, closed and convex such that $C \subset U$. Then the elementary minimization problem reads

$$\min_{x\in C} J(x). \tag{4.5}$$

We call a $x_0 \in C$ a local solution of (4.5) if there exists an open neighborhood V of x_0 in X such that

$$J(x) \geq J(x_0) \text{ for all } x \in C \cap V.$$

A local solution x_0 of (4.5) is called global if

$$J(x) \geq J(x_0) \text{ for all } x \in C.$$

Optimality conditions for this problem may be derived in a straightforward manner mirroring the techniques used in the finite dimensional case [55].

Theorem 5. *(optimality conditions for elementary problem) Consider the problem (4.5). Let $x_0 \in C$ be a local solution and let $J : U \to \mathbb{R}$ be Fréchet differentiable. Then the following variational inequality holds*

$$J'(x_0)[x - x_0] \geq 0 \text{ for any } x \in C. \tag{4.6}$$

Furthermore, if J is convex (4.6) is a necessary and sufficient condition for x_0 to be a global solution.

Proof. Suppose that $x_0 \in C$ is a local optimum for (4.5). Then for any $x \in C$ we have $x_0 + t(x - x_0) \in C$ if $t \in [0, 1]$. Local optimality implies that there exists a $\epsilon > 0$ such that

$$J(x_0 + t(x - x_0)) - J(x_0) \geq 0 \text{ for any } t \in [0, \epsilon).$$

Dividing by t and taking the limit $t \to 0$ along with the assumed Fréchet differentiability gives (4.6) immediately. This argument would also obviously hold for a global solution.

Next, we prove that if J is convex and (4.6) holds for $x_0 \in C$, then x_0 is a global solution. By convexity for any $x \in C$ we get

$$J(x_0 + t(x - x_0)) \leq (1 - t) J(x_0) + tJ(x) \text{ for all } t \in (0, 1],$$

then

$$J(x) - J(x_0) \geq \frac{J(x_0 + t(x - x_0)) - J(x_0)}{t}.$$

Since the right-hand side tends to $J'(x_0)[x - x_0]$ as $t \to 0$ and the optimality condition (4.6) holds, we get

$$J(x) - J(x_0) \geq 0 \text{ for any } x \in C.$$

□

Even though the preceding theorem assumes Fréchet differentiability, this requirement is not mandatory to get a necessary condition for $\delta J(x_0, x - x_0)$. Repeating the argument of the first part of the proof under weaker assumption gives the following lemma.

Lemma 6. *(optimality conditions for elementary problem) Consider the problem (4.5). Let $x_0 \in C$ be a local solution. Then the following variational inequality holds*

$$\delta J(x_0, x - x_0) \geq 0 \text{ for any } x \in C \tag{4.7}$$

if $\delta J(x_0, x - x_0)$ exists.

Next, the results of this section are applied to the constrained minimization problem (4.2)-(4.3).

4.4. Optimality Conditions

In order to provide optimality conditions, it is useful to impose additional regularity conditions.

Definition 7. (strong regularity of the constrained optimization problem) If (4.2)-(4.3) is such that

1. J and e are continuously Fréchet differentiable,
2. for any $\theta \in X_C$ the state equation $e(y, \theta) = 0$ has a unique solution, i.e. the solution operator $S : X_C \to X_S$ is well defined,
3. $e_y(S(\theta), \theta) \in B(X_S, Z)$ is an isomorphism between Banach spaces for any $\theta \in X_C$,

then the optimization problem (4.2)-(4.3) is called strongly regular.

Applying the implicit function theorem under these conditions implies that the solution operator S is locally defined and Fréchet differentiable for any $\theta \in X_C$ (and C^1). Then the reduced problem may be formulated. Define the reduced functional as

$$\widehat{J}(\theta) \equiv J(y(\theta), \theta),$$

where $y(\theta)$ is used instead of $S(\theta)$ to denote the solution y of the state equation (4.1) for a given $\theta \in X_C$. Then the minimization of (4.2), (4.3) can be reformulated as

$$\min_{\theta \in W_{\text{ad}}} \widehat{J}(\theta). \tag{4.8}$$

The regularity conditions in Definition 7 allow us to formulate optimality conditions for (4.8).

Theorem 8. *(optimality conditions for constrained optimization) If $\theta_0 \in W_{ad}$ is a local solution of (4.8), then*

$$\widehat{J}'(\theta_0)[\theta - \theta_0] \geq 0 \text{ for any } \theta \in W_{ad}. \tag{4.9}$$

Proof. Since W_{ad} is convex and closed and $\widehat{J}$ is Fréchet differentiable (thanks to the regularity conditions of Definition 7), we can directly apply Theorem 5 to (4.8). □

To formulate the optimality conditions in their final form, we discuss the different ways the Fréchet derivative $\widehat{J}'$ may be evaluated.

4.4.1. Sensitivity Analysis Approach to Derivative Computation

One way of resolving the derivative of the reduced cost functional $\hat{J}$ in (4.8) is to directly compute it. The regularity conditions from Definition 7 and the chain rule for Fréchet derivatives gives

$$\hat{J}'(\theta)[\vartheta] = J_y(y(\theta), \theta)\left[y'(\theta)\vartheta\right] + J_\theta(y(\theta), \theta)[\vartheta], \tag{4.10}$$

where $\theta, \vartheta \in X_C$. Taking the derivative of the state equation (4.1) results in

$$e_y(y(\theta), \theta)\, y'(\theta)\vartheta = -e_\theta(y(\theta), \theta)[\vartheta]. \tag{4.11}$$

The equation (4.11) is called the linearized state equation.

When working out the derivative of $\hat{J}$ using the sensitivity approach, one first computes (4.11) to find $y'(\theta)\vartheta$. Then (4.10) can be computed to get the Fréchet derivative of $\hat{J}$.

4.4.2. Adjoint Approach to Derivative Computation

For simplicity, let X_C and X_S be Hilbert spaces. Define the Lagrange function $L : Y \times U \times Z \to \mathbb{R}$ as

$$L(y, \theta, \lambda) \equiv J(\theta, u) + \langle \lambda, e(y, \theta) \rangle_Z , \tag{4.12}$$

where $\lambda \in Z$ is called the adjoint variable. Notice that $e(y(\theta), \theta) = 0$ for all $\theta \in X_C$ implies that

$$L(y(\theta), \theta, \lambda) = \hat{J}(\theta) . \tag{4.13}$$

Using (4.13) the Fréchet derivative of the reduced functional $\hat{J}$ can be calculated using the Lagrangian function (4.12). The Fréchet derivative of $\hat{J}$ at $\theta \in X_C$ in direction $\vartheta \in X_C$ becomes

$$\begin{aligned} \hat{J}'(\theta)\vartheta &= L_y(y(\theta), \theta, \lambda) y'(\theta)\vartheta + L_\theta(y(\theta), \theta, \lambda)\vartheta \\ &= J_y(y(\theta), \theta) y'(\theta)\vartheta + J_\theta(y(\theta), \theta)\vartheta \\ &+ \left\langle \lambda, e_y(y(\theta), \theta) y'(\theta)\vartheta + e_\theta(y(\theta), \theta)\vartheta \right\rangle_Z . \end{aligned}$$

Let $\hat{J}_y \in X_S$ be the representative of the functional $J_y(y(\theta), \theta) : X_S \to \mathbb{R}$. Using this, the expression is parsed in a particular way to give

$$\begin{aligned} L_\theta(y(\theta), \theta, \lambda) s = &\left\langle \hat{J}_y, y'(\theta)\vartheta \right\rangle_Y + \left\langle e_y(y(\theta), \theta)^* \lambda, y'(\theta)\vartheta \right\rangle_Y \\ &+ J_\theta(y(\theta), \theta)\vartheta + \langle \lambda, e_\theta(y(\theta), \theta)\vartheta \rangle_Z \\ = &\left\langle \hat{J}_y + e_y(y(\theta), \theta)^* \lambda, y'(\theta)\vartheta \right\rangle \\ &+ J_\theta(y(\theta), \theta)\vartheta + \langle \lambda, e_\theta(y(\theta), \theta)\vartheta \rangle_Z . \end{aligned}$$

If we can find a $\lambda_0 \in Z$ such that

$$\hat{J}_y + e_y(y(\theta), \theta)^* \lambda_0 = 0. \tag{4.14}$$

The expression for the Fréchet derivative computation then becomes

$$L_\theta(y(\theta), \theta, \lambda_0)\vartheta = J_\theta(y(\theta), \theta)\vartheta + \langle \lambda_0, e_\theta(y(\theta), \theta)\vartheta \rangle_Z . \tag{4.15}$$

We call (4.14) the adjoint equation and λ_0 the solution to the adjoint equation. Note that condition (4.14) is equivalent to

$$L_y(y(\theta), \theta, \lambda_0) = 0. \tag{4.16}$$

Replacing the inner products with duality brackets, the derivation above holds, when X_S, X_C are Banach spaces also [43, 55].

4.4.3. Formulation of Optimality Conditions For the Constrained Minimization Problem

Utilizing the approach detailed in Section 4.4.2, the optimality conditions for problem (4.8) may be formulated. Considering strong regularity conditions (Definition 7), we combine Theorem 8 with (4.14) - (4.16) to formulate the optimality conditions.

Theorem 9. *(Optimality conditions) Let (y_0, θ_0) be a local solution ($y_0 = y(\theta_0)$) to (4.2), (4.3) and let the strong regularity conditions hold. Then there exists an adjoint state $\lambda_0 \in Z$ such that*

$$\begin{aligned} e(y_0, \theta_0) &= 0, \\ e_y(y_0, \theta_0)\lambda_0 &= -J_y(y_0, \theta_0), \\ (J_\theta(y_0, \theta_0) + e_\theta(y_0, \theta_0)^* \lambda_0)[\theta - \theta_0] &\geq 0 \text{ for any } \theta \in W_{ad}. \end{aligned}$$

Using Lagrange formalism these can be rewritten as

$$L_p\left(y_0, \theta_0, \lambda_0\right) = e\left(y_0, \theta_0\right) = 0, \tag{4.17}$$

$$L_y\left(y_0, \theta_0, \lambda_0\right) = 0, \tag{4.18}$$

$$L_\theta\left(y_0, \theta_0, \lambda_0\right)\left[\theta - \theta_0\right] \geq 0 \text{ for any } \theta \in W_{ad}. \tag{4.19}$$

Note that strong regularity conditions were considered for the sake of simplicity, they are sufficient to derive the optimality conditions, but not necessary (Section 5.7). In a numerical setting, we might not always be able to guarantee that the strong regularity conditions of Definition 7 hold. In this case, it is possible to use the strong form of the problem and formally compute all the derivatives (Chapter 6).

5. Weak Formulation of the Dirichlet Boundary Condition Optimization for the Phase Field problem

Constrained optimization problems where the state equation (4.3) is a PFM have been studied thoroughly since the 90's. Distributed control (DC) of the PFM in which the source term of the heat equation is optimized, was addressed first [44]. Many developments have taken place since. On the theoretical level, the existence and optimality conditions for the DC of the PFM have been shown for many different variants of the reaction term (2.22), some of which come from singular potentials [27, 29]. In addition to the theoretical advancements in the DC of the PFM, many publications document interesting applications including tumor growth modeling [62], inductive heating [75] and isothermal alloy solidification [7].

Another branch of control problems in which the state equation (4.3) is a PFM are problems, where the Neumann or Robin Control boundary conditions (NoR) [28, 74] are sought after. Since DC problems are typically supplemented with homogeneous Neumann (natural) or Dirichlet boundary conditions, both the DC and NoR control problems naturally result in state equations of variational type [67, 69, 52], which lend themselves to straightforward analysis. In the case of Dirichlet boundary condition control, a variational formulation does not naturally arise. This leads to a number of possible formulations that will be discussed in Section 5.1.

Next, we describe the problem in question. Consider a PFM that governs the solidification of a pure supercooled melt (2.25)-(2.29) as the state equation (4.3). The Dirichlet boundary condition of the heat equation (2.28) that results in a particular phase field profile at final time is sought after. The full description of the problem is laid out in Section 5.2. The state equation is shown to have a unique solution for any control and the necessary auxiliary results, such as the boundedness of the solution operator with respect to the control and problem specific embedding statements are provided. The aforementioned results are then used to prove that a solution of the optimal control problem exists. In addition to this, the Fréchet differentiability of the associated solution operator and optimality conditions are provided.

The precise formulation of the problem (Section 5.2) is motivated by considering the possible approaches to deriving a state equation for non-homogeneous Dirichlet boundary conditions. These are reviewed in the following section. To keep things as clear as possible, the Laplace equation is used as an illustrative example. In an effort to improve readability, the differential symbols $\mathrm{d}x$ and $\mathrm{d}S$ are left out in spatial integrals over volume or surface respectively in Sections 5.1-5.7.

5.1. Overview of the Weak and Very Weak Formulation of the Laplace Equation with Non-homogeneous Dirichlet Boundary Conditions

Suppose that $\Omega \subset \mathbb{R}^n$ is a bounded domain and consider

$$-\Delta y = 0 \qquad \text{in } \Omega, \tag{5.1}$$

$$y = \theta \qquad \text{on } \partial\Omega, \tag{5.2}$$

where $y : \Omega \to \mathbb{R}$ is the unknown function and $\theta : \partial\Omega \to \mathbb{R}$ such that $\theta \neq 0$ is given.

5. Weak Formulation of the Dirichlet Boundary Condition Optimization for the Phase Field problem

We mention three different formulations of (5.1)-(5.2) leading to different analytical treatments. The first of these is the Dirichlet lift, where the solution is split into the zero and non-zero trace part. The second is the approximation by Robin boundary conditions, which leads to the need to set a small parameter $\delta > 0$ appropriately to get a good enough approximation. The third option is to use the very weak formulation of the problem, which arises by applying Green's formula twice.

When using the technique of Dirichlet lift (or solution split) [67, 69] to formulate (5.1)-(5.2), one assumes that $\theta \in H^{\frac{1}{2}}(\partial\Omega)$. This implies that there is a $\eta \in H^1(\Omega)$ such that $\mathrm{Tr}(\eta) = \theta$. Viewing (5.1) as an equality between distributions and (5.2) using the trace operator yields

$$a(y,\varphi) \equiv \int_\Omega \nabla y \cdot \nabla \varphi = 0 \qquad \text{for all } \varphi \in C_0^\infty(\Omega), \tag{5.3}$$

$$\mathrm{Tr}(y) = \mathrm{Tr}(\eta) = \theta, \tag{5.4}$$

where $y \in H^1(\Omega)$. The condition (5.4) allows us to define

$$\hat{y} \equiv y - \eta \in H_0^1(\Omega). \tag{5.5}$$

The equation (5.3) may be reformulated by using (5.5), which results in

$$a(\hat{y},\varphi) = -a(\eta,\varphi) \qquad \text{for all } \varphi \in C_0^\infty(\Omega), \tag{5.6}$$

where $\hat{y} \in H_0^1(\Omega)$. The density of $C_0^\infty(\Omega)$ in $H_0^1(\Omega)$ shows that (5.6) is indeed a variational form and can be treated by conventional methods. Since $a : H_0^1(\Omega) \times H_0^1(\Omega) \to \mathbb{R}$ is a bounded bilinear form and $a(\eta,\cdot) \in H_0^1(\Omega)^*$, the relationship (5.6) can be seen as an equality in $H_0^1(\Omega)^*$ (find $\hat{y} \in H_0^1(\Omega)$ such that $a(\hat{y},\cdot) = -a(\eta,\cdot)$ in $H_0^1(\Omega)^*$).

The Robin boundary condition can be used to reformulate (5.1)-(5.2) by considering $\theta \in L_2(\partial\Omega)$ and replacing (5.2) with the Robin boundary condition [5, 11]

$$\delta \frac{\partial y}{\partial \overrightarrow{n}} + y = \theta \qquad \text{on } \partial\Omega, \tag{5.7}$$

where $\delta > 0$. Observe that (5.7) gives (5.2) when $\delta \to 0^+$. Viewing (5.1) in the sense of distributions yields

$$a(y,\varphi) = \frac{1}{\delta}\int_{\partial\Omega} (\theta - y)\,\varphi \qquad \text{for all } \varphi \in H^1(\Omega), \tag{5.8}$$

where $y \in H^1(\Omega)$.

Lastly, the very weak formulation is discussed [38, 37]. To recast (5.1)-(5.2) into the very weak form, consider $\theta \in L_2(\partial\Omega)$. Formally using Green's formula twice and applying the same techniques as in the derivation of (5.8) leads to

$$\int_\Omega y \Delta\varphi = -\int_{\partial\Omega} \theta \frac{\partial\varphi}{\partial\overrightarrow{n}} \qquad \text{for all } \varphi \in H^2(\Omega) \cap H_0^1(\Omega), \tag{5.9}$$

where $y \in L_2(\Omega)$.

Among other things, the formulations (5.6), (5.8), and (5.9) impose different regularity requirements on the derivatives of the solution y (or $\hat{y}$) and Dirichlet boundary condition θ. These differences are summarized in Table 5.1. It is clear that the Dirichlet lift formulation (5.6) requires the strongest regularity for both the solution and the boundary condition (use $H^{\frac{1}{2}}(\partial\Omega) \subset L_2(\partial\Omega)$). The Robin boundary condition approximation (5.8) eases the requirement for the boundary condition from $H^{\frac{1}{2}}(\partial\Omega)$ to

Formulation	Solution regularity	Dirichlet boundary condition regularity
(5.6)	$H^1(\Omega)$	$H^{\frac{1}{2}}(\partial\Omega)$
(5.8)	$H^1(\Omega)$	$L_2(\partial\Omega)$
(5.9)	$L_2(\Omega)$	$L_2(\partial\Omega)$

Table 5.1.: Comparison of regularity between the formulations (5.6), (5.8) and (5.9).

$L_2(\partial\Omega)$. Lastly, the very weak formulation (5.9) relaxes the regularity requirement for the solution from $H^1(\Omega)$ to $L_2(\Omega)$, resulting in the weakest regularity requirement of $(y, \theta) \in L_2(\Omega) \times L_2(\partial\Omega)$ overall.

Because of these considerations, either (5.8) or (5.9) might seem like the superior choice when formulating state equations that contain Dirichlet boundary conditions. There are, however, drawbacks that come with these approaches. For instance, the correspondence of problem (5.8) to the original one will be subject to the choice of $\delta > 0$ [5, 11]. Furthermore, the very weak formulation of a state equation [38, 37, 52] can be harder to analyze as there are fewer analytical tools available.

5.2. Dirichlet Control for the PFM Using the Dirichlet Lift

The Dirichlet boundary condition control for parabolic differential equations has been described only scarcely in literature. In the particular case of a single linear parabolic PDE, the approximation by Robin boundary conditions [10] or the very weak formulation [38, 52] may be taken advantage of. The technique of Dirichlet lift has only been applied to optimization problems constrained by an Elliptic PDE [25].

In this contribution, the possibilities of the Dirichlet lift technique are explored for an optimization problem constrained by the PFM that governs the solidification of a pure melt (2.25)-(2.29). Note that this system features a non-linearity, so the full analysis of the system must be provided. By using the Dirichlet lift, we avoid the difficulties of having to choose a suitable $\delta > 0$ for a Robin boundary approximation (see 5.8) and working with the very weak formulation (see (5.9)). The drawback of this choice is the relatively strong regularity demand imposed on the control which follows from the analysis.

As mentioned before, the state equation (2.25)-(2.29) with Dirichlet boundary conditions is considered. This model is non-dimensional and described fully in Section 2.2.3. The goal of the optimization is to achieve a certain shape of the solid body (described by the phase field p) at final time. This is captured by the form of the functional J.

Let $n \in \{1, 2, 3\}$, $T > 0$ and $\Omega \subset \mathbb{R}^n$ be a bounded domain with a Lipschitz boundary. Let $u : \overline{\Omega} \times [0, T] \to \mathbb{R}$ and $p : \overline{\Omega} \times [0, T] \to \mathbb{R}$ represent the temperature and phase field respectively. Then the problem in question reads

$$\min_{((u,p),\theta)\in X_S\times W_{\mathrm{ad}}} J(u,p,\theta) \equiv \frac{1}{2}\int_\Omega |p(T)-p_{\mathrm{f}}|^2 + \frac{\gamma}{2}\int_0^T\int_{\partial\Omega} |\theta|^2 \,\mathrm{d}t \tag{5.10}$$

$$\begin{aligned}
\text{s.t.} \qquad u_t &= \Delta u + H p_t && \text{on } (0,T)\times\Omega, && (5.11)\\
u|_{\partial\Omega} &= \theta && \text{on } [0,T]\times\partial\Omega, && (5.12)\\
u|_{t=0} &= u_{\mathrm{ini}} && \text{on } \Omega, && (5.13)\\
\alpha\xi^2 p_t &= \xi^2\Delta p + f(u,p,\xi), && && (5.14)\\
f(u,p,\xi) &= p(1-p)\left(p-\frac{1}{2}\right) - b\beta\xi u, && && (5.15)\\
p|_{\partial\Omega} &= 0 && \text{on } [0,T]\times\partial\Omega, && (5.16)\\
p|_{t=0} &= p_{\mathrm{ini}} && \text{on } \Omega, && (5.17)
\end{aligned}$$

where α, ξ, b and H are dimensionless material constants detailed in Section 2.2.3, $\gamma > 0$ is the regularization parameter, $u_{\mathrm{ini}} \in H^1(\Omega)$, $p_{\mathrm{ini}} \in H^1(\Omega)$, $\theta : [0,T]\times\partial\Omega \to \mathbb{R}$ is the boundary control and $p_{\mathrm{f}} \in L_2(\Omega)$ is the target phase field profile at final time. The profile p_{f} represents the optimal state, which is to be achieved. To this end, the Dirichlet boundary control θ, such that $J(u,p,\theta)$ is minimized, is sought after.

To achieve a Dirichlet lift formulation (see Section 5.1), define the control space as

$$X_C \equiv H^4(\Omega\times(0,T)). \tag{5.18}$$

The reason for imposing this strong regularity will become apparent in the following analysis. Let $\theta \in \mathrm{Tr}(X_C)$, i.e. there exists $\eta \in X_C$ for which

$$\mathrm{Tr}(\eta) = \theta. \tag{5.19}$$

Define the solution of the state equation as a (u,p) such that

$$(u,p) \in X_S \equiv W\left(0,T;2,2;H^1(\Omega),L_2(\Omega)\right) \times W\left(0,T;2,2;H_0^1(\Omega),L_2(\Omega)\right)$$

and (5.10)-(5.17) is satisfied in the distributional sense. Following the principles of the Dirichlet lift (see (5.3)-(5.6)), define

$$\hat{u} \equiv u - \eta \in W\left(0,T;2,2;H_0^1(\Omega),L_2(\Omega)\right). \tag{5.20}$$

Let $W_{\mathrm{ad}} \subset X_C$ be closed, convex and bounded, then the problem in question reads

$$\min_{((\hat{u},\hat{p}),\eta)} J(\hat{u},\hat{p},\eta) \equiv \frac{1}{2}\int_\Omega |\hat{p}(T)-p_{\mathrm{f}}|^2 + \frac{\gamma}{2}\int_0^T\int_{\partial\Omega} |\mathrm{Tr}\ (\eta)|^2 \,\mathrm{d}t \tag{5.21}$$

$$\begin{aligned}
\text{s.t.} \qquad \hat{u}_t - \Delta\hat{u} &= H\hat{p}_t - \eta_t + \Delta\eta && \text{on } (0,T)\times\Omega, && (5.22)\\
\hat{u}|_{\partial\Omega} &= 0 && \text{on } [0,T]\times\partial\Omega, && (5.23)\\
\hat{u}|_{t=0} &= u_{\mathrm{ini}} - \eta|_{t=0} && \text{on } \Omega, && (5.24)\\
\alpha\xi^2\hat{p}_t &= \xi^2\Delta\hat{p} + f(\hat{u},\hat{p},\eta,\xi), && && (5.25)\\
f(\hat{u},\hat{p},\xi) &= \hat{p}(1-\hat{p})\left(\hat{p}-\frac{1}{2}\right) - b\beta\xi(\hat{u}+\eta), && && (5.26)\\
\hat{p}|_{\partial\Omega} &= 0 && \text{on } [0,T]\times\partial\Omega, && (5.27)\\
\hat{p}|_{t=0} &= p_{\mathrm{ini}} && \text{on } \Omega, && (5.28)
\end{aligned}$$

where the equations (5.22) and (5.25) are posed in the sense of distributions. An ordered pair

$$(\hat{u},\hat{p}) \in \hat{X}_S \equiv W\left(0,T;2,2;H_0^1(\Omega),L_2(\Omega)\right) \times W\left(0,T;8,2;H_0^1(\Omega),L_2(\Omega)\right) \tag{5.29}$$

that satisfies (5.22)-(5.28) is called a solution of the state equation (5.22)-(5.28). The reason for the higher regularity of the space in which $\hat{p}$ lies will become apparent in Section 5.5. It follows from the principles of the Dirichlet lift that a solution of (5.22)-(5.28) is also a solution to (5.11)-(5.17) when (5.19) holds.

Theorem 10. *Let the control space X_C and solution space $\hat{X}_S$ be given by (5.18) and (5.29). respectively. Let $W_{ad} \subset X_C$ be closed, convex, and bounded. Then problem (5.21)-(5.28) has a (not necessarily unique) solution.*

The following sections are dedicated to the proof of the theorem above. In Section 5.3, problem specific embedding results are provided. These are then used in Section 5.4 to show that there exists a unique solution of the state equation (5.22)-(5.28) for any $\eta \in X_C$. The results of Sections 5.3, 5.4 are then used to show the existence of optimal control for (5.21)-(5.28) in Section 5.5, which concludes the proof of Theorem 10.

5.3. Specific Embedding Based Results

Two results based on embedding statements are derived in this section. They are then used to prove that the solution operator of (5.22)-(5.28) exists and further utilized to prove the existence of optimal control for the entire problem (5.21)-(5.28).

Lemma 11. *Let $\Omega \subset \mathbb{R}^n$ be a bounded domain with a Lipschitz boundary for $n \in \{1,2,3\}$ and $T > 0$. Let $f \in H^4(\Omega \times (0,T))$. Then*

$$\|f(t)\|_{H^1(\Omega)} < +\infty \text{ for any } t \in [0,T]. \tag{5.30}$$

Proof. Since Ω has a Lipschitz boundary, so does $\Omega \times (0,T)$. Using Theorem 20 and setting $m = 4$, $p = 2$ and $k = 1$ leads to

$$H^4(\Omega \times (0,T)) \hookrightarrow C^{1,\beta}\left(\overline{\Omega} \times [0,T]\right) \text{ for all } \beta \in (0,1].$$

This means that there exists a Hölder continuous function $g \in C^{1,\beta}\left(\overline{\Omega} \times [0,T]\right)$ such that $f = g$ in $H^4(\Omega \times (0,T))$ and specifically $g(t) \in C^{1,\beta}\left(\overline{\Omega}\right)$ for any $t \in [0,T]$. Since $\|g(t)\|_{H^1(\Omega)} = \|f(t)\|_{H^1(\Omega)}$, the conclusion (5.30) holds. □

Lemma 12. *Let $\Omega \subset \mathbb{R}^n$ be a bounded domain (with an arbitrary boundary) for $n \in \{1,2,3\}$ and $T > 0$. Let $f_n \rightharpoonup f$ in $W\left(0,T;8,2;H_0^1(\Omega),L_2(\Omega)\right)$. Then*

$$f_n^3 \to f^3 \text{ in } L_2\left(0,T;H_0^1(\Omega)^*\right), \tag{5.31}$$
$$f_n^2 \to f^2 \text{ in } L_2\left(0,T;H_0^1(\Omega)^*\right). \tag{5.32}$$

Proof. Using Theorem 21 and setting $m = 1$, $k = 0$, $p = 2$ and $q = 5$, the Gelfand triple

$$H_0^1(\Omega) \hookrightarrow\hookrightarrow L_5(\Omega) \hookrightarrow L_2(\Omega)$$

can be constructed. Applying Theorem 22 (Aubin-Lions) gives

$$f_n \to f \text{ in } L_8(0,T;L_5(\Omega)).$$

This suffices to prove strong convergence of $f_n^3 \to f^3$ in $L_2\left(0,T;L_{\frac{5}{3}}(\Omega)\right)$, since the reflexivity of $L_2\left(0,T;L_{\frac{5}{3}}(\Omega)\right)$ and the embedding

$$H_0^1(\Omega) \hookrightarrow L_{\frac{5}{2}}(\Omega) \tag{5.33}$$

give rise to

$$L_2\left(0,T;L_{\frac{5}{3}}(\Omega)\right) \cong L_2\left(0,T;L_{\frac{5}{3}}(\Omega)\right)^{**} \cong L_2\left(0,T;L_{\frac{5}{2}}(\Omega)^*\right) \hookrightarrow L_2\left(0,T;H_0^1(\Omega)^*\right). \tag{5.34}$$

More precisely, (5.33) may be used to estimate the $H_0^1(\Omega)^*$-norm of $\int\limits_\Omega w\cdot \in H_0^1(\Omega)^*$, when $w \in L_{\frac{5}{3}}(\Omega)$ as follows

$$\begin{aligned}\left\|\int_\Omega w\cdot\right\|_{H_0^1(\Omega)^*} &= \sup_{\substack{u\in H_0^1(\Omega)\\ \|u\|_{H_0^1(\Omega)}=1}} \left|\int_\Omega wu\right| \le \sup_{\substack{u\in H_0^1(\Omega)\\ \|u\|_{H_0^1(\Omega)}=1}} \|w\|_{L_{\frac{5}{3}}(\Omega)}\|u\|_{L_{\frac{5}{2}}(\Omega)}\\ &\le \sup_{\substack{u\in H_0^1(\Omega)\\ \|u\|_{H_0^1(\Omega)}=1}} c\|w\|_{L_{\frac{5}{3}}(\Omega)}\|u\|_{H_0^1(\Omega)} = c\|w\|_{L_{\frac{5}{3}}(\Omega)}.\end{aligned} \tag{5.35}$$

The estimate (5.35) can be used to show that

$$\|w\|^2_{L_2\left(0,T;H_0^1(\Omega)^*\right)} = \int_0^T \left\|\int_\Omega w\cdot\right\|^2_{H_0^1(\Omega)^*} \mathrm{d}t \le \int_0^T c^2\|w\|^2_{L_{\frac{5}{3}}(\Omega)}\,dt = c^2\|w\|^2_{L_2\left(0,T;L_{\frac{5}{3}}(\Omega)\right)}, \tag{5.36}$$

which is just (5.34).

For any two numbers $a, b \in \mathbb{R}$, the estimate

$$\left|b^3 - a^3\right| \le 3\left(a^2+b^2\right)|b-a| \tag{5.37}$$

may be derived. Using (5.37), the estimate

$$\begin{aligned}\left\|f_n^3 - f^3\right\|_{L_{\frac{5}{3}}(\Omega)} &\le 3\left(\int_\Omega \left(f_n^2(x)+f^2(x)\right)^{\frac{5}{3}}|f_n(x)-f(x)|^{\frac{5}{3}}\right)^{\frac{3}{5}}\\ &\le 3\left\|f_n^2(f_n-f)\right\|_{L_{\frac{5}{3}}(\Omega)} + 3\left\|f^2(f_n-f)\right\|_{L_{\frac{5}{3}}(\Omega)} \text{ for all } t\in[0,T]\end{aligned} \tag{5.38}$$

can be constructed. For any $v, w \in L^5(\Omega)$, Hölder's inequality can be used with exponents $\frac{3}{2}$ and 3 to arrive at

$$\left\|w^2v\right\|_{L_{\frac{5}{3}}(\Omega)} = \left(\int_\Omega w^{\frac{10}{3}}v^{\frac{5}{3}}\right)^{\frac{3}{5}} \le \left(\int_\Omega w^5\right)^{\frac{2}{5}}\left(\int_\Omega v^5\right)^{\frac{1}{5}} = \|w\|^2_{L_5(\Omega)}\|v\|_{L_5(\Omega)}. \tag{5.39}$$

Combining the estimates (5.38) and (5.39) yields

$$\left\|f_n^3 - f^3\right\|_{L_{\frac{5}{3}}(\Omega)} \le 3\|f_n\|^2_{L_5(\Omega)}\|f_n-f\|_{L_5(\Omega)} + 3\|f\|^2_{L_5(\Omega)}\|f_n-f\|_{L_5(\Omega)}.$$

Squaring, integrating over $[0,T]$ and using Young's inequality then yields

$$\left\|f_n^3 - f^3\right\|^2_{L_2\left(0,T;L_{\frac{5}{3}}(\Omega)\right)} \leq 6 \int_0^T \|f_n(t)\|^4_{L_5(\Omega)} \|f_n(t) - f(t)\|^2_{L_5(\Omega)} \,\mathrm{d}t + 6 \int_0^T \|f(t)\|^4_{L_5(\Omega)} \|f_n(t) - f(t)\|^2_{L_5(\Omega)} \,\mathrm{d}t. \tag{5.40}$$

Applying Hölder's inequality to the right-hand side of (5.40) gives

$$\left\|f_n^3 - f^3\right\|^2_{L_2\left(0,T;L_{\frac{5}{3}}(\Omega)\right)} \leq 6 \left(\int_0^T \|f_n(t)\|^8_{L_5(\Omega)} \,\mathrm{d}t\right)^{\frac{1}{2}} \left(\int_0^T \|f_n(t) - f(t)\|^4_{L_5(\Omega)} \,\mathrm{d}t\right)^{\frac{1}{2}} + 6 \left(\int_0^T \|f(t)\|^8_{L_5(\Omega)} \,\mathrm{d}t\right)^{\frac{1}{2}} \left(\int_0^T \|f_n(t) - f(t)\|^4_{L_5(\Omega)} \,\mathrm{d}t\right)^{\frac{1}{2}}.$$

Note that $f_n \to f$ in $L_8(0,T;L_5(\Omega))$, $\|f_n\|^4_{L_8(0,T;L_5(\Omega))}$ is bounded and $\|f\|^4_{L_8(0,T;L_5(\Omega))}$ is finite. Furthermore

$$f_n \to f \text{ in } L_4(0,T;L_5(\Omega)),$$
$$f_n \to f \text{ in } L_2(0,T;L_5(\Omega)),$$

since $L_8(0,T;L_5(\Omega)) \hookrightarrow L_4(0,T;L_5(\Omega))$. Altogether,

$$f_n^3 \to f^3 \text{ in } L_2\left(0,T;L_{\frac{5}{3}}(\Omega)\right) \Rightarrow f_n^3 \to f^3 \text{ in } L_2\left(0,T;H_0^1(\Omega)^*\right),$$

which completes the proof of (5.31). To get the result (5.32), replace the estimate (5.37) with

$$\left|b^2 - a^2\right| \leq (|a| + |b|)\,|b - a|$$

and use

$$\|w\|_{L_{\frac{5}{2}}(\Omega)} \leq c\,\|w\|_{L_5(\Omega)}$$

in the estimate analogous to (5.39). □

5.4. Analysis of the State Equation

The existence and boundedness of the solution operator of the system (5.22)-(5.28) is proven in this section. The result is utilized in Section 5.5 to prove that (5.21)-(5.28) has a solution. First, the weak formulation of the problem (5.22)-(5.28) is stated. To simplify the notation, $\hat{u}$ and $\hat{p}$ are replaced by u and p, respectively.

By standard methods, the weak formulation of (5.22)-(5.28) reads

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega u\left(t\right)v+\int_\Omega \nabla u\left(t\right)\cdot\nabla v+\int_\Omega \nabla\eta\left(t\right)\cdot\nabla v=H\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega p\left(t\right)v-\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega \eta\left(t\right)v$$
$$\text{for all } v\in H_0^1\left(\Omega\right) \text{ a.e. in } \left[0,T\right], \tag{5.41}$$
$$\alpha\xi^2\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega p\left(t\right)w+\xi^2\int_\Omega \nabla p\left(t\right)\cdot\nabla w=\int_\Omega f_0\left(p\left(t\right)\right)w-b\beta\xi\int_\Omega\left(u\left(t\right)+\eta\left(t\right)\right)w$$
$$\text{for all } w\in H_0^1\left(\Omega\right) \text{ a.e. in } \left[0,T\right], \tag{5.42}$$
$$u\left(0\right)=u_{\mathrm{ini}}-\eta\left|_{t=0}\right., \tag{5.43}$$
$$p\left(0\right)=p_{\mathrm{ini}}, \tag{5.44}$$

where $f_0:\mathbb{R}\to\mathbb{R}$ is defined as $f_0\left(y\right)\equiv y\left(1-y\right)\left(y-\frac{1}{2}\right)$.

Theorem 13. *Let the control space X_C and solution space X_S (the hat is omitted for the sake of readability) be given by (5.18) and (5.29) respectively and let $\eta\in X_C$. Then (5.41)-(5.44) has a unique solution in X_S.*

Proof. The steps of the proof are as follows. First, the m-th Galerkin approximation of the problem (5.41)-(5.44) is defined (part A)). Following this, the key energy estimate is derived (part B)). Lastly, the existence and uniqueness of the solution is proven (parts C and D)).

A) Consider a countable set of functions $w_i:\Omega\to\mathbb{R}$ such that

$$\left(w_i\right)_{i\in\mathbb{N}} \text{ form an orthonormal basis of } L_2\left(\Omega\right), \tag{5.45}$$
$$\left(w_i\right)_{i\in\mathbb{N}} \text{ form an orthogonal basis of } H_0^1\left(\Omega\right). \tag{5.46}$$

It can be shown that such a set exists constructively. For instance, the eigenfunctions of the Laplacian on $H_0^1\left(\Omega\right)$ satisfy (5.45)-(5.46) [34]. For any $m\in\mathbb{N}$, let $u_m,p_m:\left[0,T\right]\to H_0^1\left(\Omega\right)$ be two vector valued functions such that

$$u_m\left(t\right)\equiv\sum_{k=1}^m \alpha_m^k\left(t\right)w_k,\quad p_m\left(t\right)\equiv\sum_{k=1}^m \beta_m^k\left(t\right)w_k. \tag{5.47}$$

Additionally, let (5.47) be such that the problem

$$\int_\Omega \dot{u}_m\left(t\right)w_k+\int_\Omega \nabla u_m\left(t\right)\cdot\nabla w_k+\int_\Omega \nabla\eta\left(t\right)\cdot\nabla w_k=H\int_\Omega \dot{p}_m\left(t\right)w_k-\int_\Omega \dot{\eta}\left(t\right)w_k$$
$$\text{for}\quad k=1,\ldots,m, \text{ a.e. in } \left[0,T\right], \tag{5.48}$$
$$\alpha\xi^2\int_\Omega \dot{p}_m\left(t\right)w_k+\xi^2\int_\Omega \nabla p_m\left(t\right)\cdot\nabla w_k=\int_\Omega f_0\left(p_m\left(t\right)\right)w_k-b\beta\xi\int_\Omega\left(u_m\left(t\right)+\eta\left(t\right)\right)w_k$$
$$\text{for}\quad k=1,\ldots,m, \text{ a.e. in } \left[0,T\right], \tag{5.49}$$

with the initial conditions

$$\alpha_m^k\left(0\right)=\int_\Omega\left(u_{\mathrm{ini}}-\eta\left(0\right)\right)w_k\quad\text{for}\quad k=1,\ldots,m, \tag{5.50}$$
$$\beta_m^k\left(0\right)=\int_\Omega p_{\mathrm{ini}}w_k\quad\text{for}\quad k=1,\ldots,m, \tag{5.51}$$

is satisfied for all $m \in \mathbb{N}$. Standard ODE theory can be applied to show that (5.48)-(5.51) has a unique solution [34, 15]. This implies that the Galerkin approximations u_m, p_m are well defined. We call the problem (5.48)-(5.51) the m-th Galerkin approximation of (5.41)-(5.44).

B) Many of the techniques featured in this section have been inspired by previous works on PDE analysis [34, 15].

In some situations, the notation

$$(f,g) \equiv \int_\Omega fg, \tag{5.52}$$

$$\langle f,g \rangle \equiv \int_\Omega fg + \int_\Omega \nabla f \cdot \nabla g, \tag{5.53}$$

$$B[f,g] \equiv \int_\Omega \nabla f \cdot \nabla g, \tag{5.54}$$

where $f, g \in H^1(\Omega)$ is used. Let the time derivative of the vector valued function $f : [0,T] \to H^1(\Omega)$ be denoted by $\dot{f}$. The function evaluation at a point is also omitted, i.e. instead of $f(t) = g(t)$ a.e. in $[0,T]$, we write $f = g$ a.e. in $[0,T]$. In the following, the symbols $c_1, c_2, \ldots$ denote positive constants.

Consider a fixed $m \in \mathbb{N}$. Multiply (5.48) by $\dot{\alpha}_m^k$ and (5.49) by $\dot{\beta}_m^k$ for each $k = 1, \ldots, m$. Adding up the equations multiplied by $\dot{\alpha}_m^k$ and $\dot{\beta}_m^k$ separately gives

$$(\dot{u}_m, \dot{u}_m) + B[u_m, \dot{u}_m] + B[\eta, \dot{u}_m] = H(\dot{p}_m, \dot{u}_m) - (\dot{\eta}, \dot{u}_m) \text{ a.e. in } [0,T], \tag{5.55}$$

$$\alpha\xi^2 (\dot{p}_m, \dot{p}_m) + \xi^2 B[p_m, \dot{p}_m] = (f_0(p_m), \dot{p}_m) - b\beta\xi (u_m + \eta, \dot{p}_m) \text{ a.e. in } [0,T]. \tag{5.56}$$

Using (5.54) to rewrite some of the terms in (5.56) leads to

$$B[p_m, \dot{p}_m] = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t} \|\nabla p_m\|^2_{L_2(\Omega)}, \tag{5.57}$$

$$f_0(p_m)\dot{p}_m = -\frac{\mathrm{d}}{\mathrm{d}t}\omega_0(p_m), \tag{5.58}$$

where $\omega_0(y) = \frac{1}{4}\left(\left(y - \frac{1}{2}\right)^2 - \frac{1}{4}\right)^2$ for any $y \in \mathbb{R}$. Using (5.57) and (5.58), the system (5.55)-(5.56) may be reformulated as

$$\|\dot{u}_m\|^2_{L_2(\Omega)} + B[u_m, \dot{u}_m] + B[\eta, \dot{u}_m] = H(\dot{p}_m, \dot{u}_m) - (\dot{\eta}, \dot{u}_m) \text{ a.e. in } [0,T], \tag{5.59}$$

$$\alpha\xi^2 \|\dot{p}_m\|^2_{L_2(\Omega)} + \xi^2 \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t} \|\nabla p_m\|^2_{L_2(\Omega)} + \frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega \omega_0(p_m) = -b\beta\xi (\dot{u}_m + \eta, \dot{p}_m) \text{ a.e. in } [0,T]. \tag{5.60}$$

The bilinearity of $B[\cdot,\cdot]$ gives rise to the identity

$$B[u_m + \eta, \dot{u}_m] = B[u_m + \eta, \dot{u}_m + \dot{\eta}] - B[u_m + \eta, \dot{\eta}]. \tag{5.61}$$

Using (5.61), we can rewrite (5.59) as

$$\|\dot{u}_m\|^2_{L_2(\Omega)} + B[u_m + \eta, \dot{u}_m + \dot{\eta}] = H(\dot{p}_m, \dot{u}_m) - (\dot{\eta}, \dot{u}_m) + B[u_m + \eta, \dot{\eta}] \text{ a.e. in } [0,T]. \tag{5.62}$$

Applying the Young's and Schwarz inequalities to the right-hand side of (5.62) leads to the estimates

$$
\begin{aligned}
|H(\dot{p}_m, \dot{u}_m)| &\leq \frac{\delta_1}{2} \|\dot{u}_m\|^2_{L_2(\Omega)} + \frac{H^2}{2\delta_1} \|\dot{p}_m\|^2_{L_2(\Omega)} \text{ a.e. in } [0,T], \\
|-(\dot{\eta}, \dot{u}_m)| &\leq \frac{1}{2\delta_2} \|\dot{\eta}\|^2_{L_2(\Omega)} + \frac{\delta_2}{2} \|\dot{u}_m\|^2_{L_2(\Omega)} \text{ a.e. in } [0,T], \\
|B[u_m + \eta, \dot{\eta}]| &\leq \frac{c_1\delta_3}{2} \|u_m + \eta\|^2_{H^1(\Omega)} + \frac{c_2}{2\delta_3} \|\dot{\eta}\|^2_{H^1(\Omega)} \text{ a.e. in } [0,T],
\end{aligned}
$$

where $\delta_1, \delta_2 > 0$. Setting $\delta_1 = \delta_2 = \frac{1}{2}$ yields

$$
\begin{aligned}
\frac{1}{2} \|\dot{u}_m\|^2_{L_2(\Omega)} + \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t} \|\nabla(u_m + \eta)\|^2_{L_2(\Omega)} \leq H^2 \|\dot{p}_m\|^2_{L_2(\Omega)} + \|\dot{\eta}\|^2_{L_2(\Omega)} + \frac{c_1\delta_3}{2} \|u_m + \eta\|^2_{H^1(\Omega)} \\
+ \frac{c_2}{2\delta_3} \|\dot{\eta}\|^2_{H^1(\Omega)} \text{ a.e. in } [0,T]. \qquad (5.63)
\end{aligned}
$$

Estimating the right-hand side of (5.60) using analogous methods with $\delta_4 > 0$ gives rise to

$$
|-b\beta\xi(u_m + \eta, \dot{p}_m)| \leq b\beta\xi \left(\frac{\delta_4}{2} \|u_m + \eta\|^2_{L_2(\Omega)} + \frac{1}{2\delta_4} \|\dot{p}_m\|^2_{L_2(\Omega)} \right) \text{ a.e. in } [0,T].
$$

Setting $\delta_4 = \frac{b\beta\xi}{\alpha\xi^2}$ leads to the estimate (see (5.60))

$$
\frac{\alpha\xi^2}{2} \|\dot{p}_m\|^2_{L_2(\Omega)} + \xi^2 \frac{\mathrm{d}}{\mathrm{d}t} \|\nabla p_m\|^2_{L_2(\Omega)} + \frac{\mathrm{d}}{\mathrm{d}t} \int_\Omega \omega_0(p_m) \leq \frac{(b\beta)^2}{\alpha} \|u_m + \eta\|^2_{L_2(\Omega)} \text{ a.e. in } [0,T]. \qquad (5.64)
$$

Multiplying (5.63) by $\frac{\alpha\xi^2}{4H^2}$ and adding the result to (5.64) gives

$$
\begin{aligned}
&\frac{\alpha\xi^2}{8H^2} \|\dot{u}_m\|^2_{L_2(\Omega)} + \frac{\alpha\xi^2}{4} \|\dot{p}_m\|^2_{L_2(\Omega)} + \frac{\alpha\xi^2}{8H^2}\frac{\mathrm{d}}{\mathrm{d}t} \|\nabla(u_m + \eta)\|^2_{L_2(\Omega)} + \xi^2 \frac{\mathrm{d}}{\mathrm{d}t} \|\nabla p_m\|_{L_2(\Omega)} + \frac{\mathrm{d}}{\mathrm{d}t} \int_\Omega \omega_0(p_m) \\
\leq & c_3 \left(\delta_3 \|u_m + \eta\|^2_{H^1(\Omega)} + \|u_m + \eta\|^2_{L_2(\Omega)} + \|\dot{\eta}\|^2_{L_2(\Omega)} + \frac{1}{\delta_3} \|\dot{\eta}\|^2_{H^1(\Omega)} \right) \text{ a.e. in } [0,T]. \qquad (5.65)
\end{aligned}
$$

Using the continuity of the trace operator $\mathrm{Tr} : H^1(\Omega) \to L_2(\partial\Omega)$, the generalized Poincaré's inequality and $u_m \in H^1_0(\Omega)$ to overestimate the right-hand side of (5.65) results in

$$
\begin{aligned}
& c_3 \left(\delta_3 \|u_m + \eta\|^2_{H^1(\Omega)} + \|u_m + \eta\|^2_{L_2(\Omega)} + \|\dot{\eta}\|^2_{L_2(\Omega)} + \frac{1}{\delta_3} \|\dot{\eta}\|^2_{H^1(\Omega)} \right) \\
\leq & c_4 \|\nabla(u_m + \eta)\|^2_{L_2(\Omega)} + c_5 \int_{\partial\Omega} |\mathrm{Tr}(u_m + \eta)|^2 + c_6 \|\dot{\eta}\|^2_{H^1(\Omega)} \\
\leq & c_7 \left(\|\nabla(u_m + \eta)\|^2_{L_2(\Omega)} + \|\eta\|^2_{H^1(\Omega)} + \|\dot{\eta}\|^2_{H^1(\Omega)} \right) \text{ a.e. in } [0,T]. \qquad (5.66)
\end{aligned}
$$

Note that $\|\eta\|^2_{H^1(\Omega)}$ and $\|\dot{\eta}\|^2_{H^1(\Omega)}$ must be finite almost everywhere in $(0,T)$ since $\eta \in H^4(\Omega \times (0,T))$. Dropping the positive terms $\frac{\alpha\xi^2}{8H^2} \|\dot{u}_m\|^2_{L_2(\Omega)}, \frac{\alpha\xi^2}{4} \|\dot{p}_m\|^2_{L_2(\Omega)}$ from the left-hand side of (5.65) and applying (5.66) leads to

$$
\begin{aligned}
& \frac{\alpha\xi^2}{8H^2}\frac{\mathrm{d}}{\mathrm{d}t} \|\nabla(u_m + \eta)\|^2_{L_2(\Omega)} + \xi^2 \frac{\mathrm{d}}{\mathrm{d}t} \left(\|\nabla p_m\|^2_{L_2(\Omega)} \right) + \frac{\mathrm{d}}{\mathrm{d}t} \int_\Omega \omega_0(p_m) \\
\leq & c_7 \left(\|\nabla(u_m + \eta)\|^2_{L_2(\Omega)} + \|\eta\|^2_{H^1(\Omega)} + \|\dot{\eta}\|^2_{H^1(\Omega)} \right) \text{ a.e. in } [0,T]. \qquad (5.67)
\end{aligned}
$$

Note that both $\|\nabla p_m\|^2_{L_2(\Omega)}$ and $\int\limits_\Omega \omega_0(p_m)$ are non-negative terms. Adding them to the right-hand side of (5.67) and adjusting the constants yields

$$\begin{aligned}
&\frac{\alpha\xi^2}{8H^2}\frac{\mathrm{d}}{\mathrm{d}t}\|\nabla(u_m+\eta)\|^2_{L_2(\Omega)}+\xi^2\frac{\mathrm{d}}{\mathrm{d}t}\left(\|\nabla p_m\|^2_{L_2(\Omega)}\right)+\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega\omega_0(p_m)\\
\leq& c_8\left(\frac{\alpha\xi^2}{8H^2}\|\nabla(u_m+\eta)\|^2_{L_2(\Omega)}+\xi^2\|\nabla p_m\|^2_{L_2(\Omega)}+\int_\Omega\omega_0(p_m)\right)\\
&+c_7\|\eta\|^2_{H^1(\Omega)}+c_7\|\dot\eta\|^2_{H^1(\Omega)}\ \text{ a.e. in }\ [0,T]\,.
\end{aligned}\tag{5.68}$$

Using Grönwall's lemma (Lemma 18) with the setting

$$\begin{aligned}
\varrho(t)&\equiv\frac{\alpha\xi^2}{8H^2}\|\nabla(u_m(t)+\eta(t))\|^2_{L_2(\Omega)}+\xi^2\|\nabla p_m(t)\|^2_{L_2(\Omega)}+\int_\Omega\omega_0(p_m(t))\,,\\
\phi(t)&\equiv c_8,\\
\psi(t)&\equiv c_7\|\eta(t)\|^2_{H^1(\Omega)}+c_7\|\dot\eta(t)\|^2_{H^1(\Omega)}\,,
\end{aligned}\tag{5.69}$$

leads to

$$\begin{aligned}
&\frac{\alpha\xi^2}{8H^2}\|\nabla(u_m(t)+\eta(t))\|^2_{L_2(\Omega)}+\xi^2\|\nabla p_m(t)\|^2_{L_2(\Omega)}+\int_\Omega\omega_0(p_m(t))\\
\leq& e^{tc_8}\left(\frac{\alpha\xi^2}{8H^2}\|\nabla(u_m(0)+\eta(0))\|^2_{L_2(\Omega)}+\xi^2\|\nabla p_m(0)\|^2_{L_2(\Omega)}+\int_\Omega\omega_0(p_m(0))\right)\\
&+e^{tc_8}\int_0^t c_7\left(\|\eta(r)\|^2_{H^1(\Omega)}+\|\dot\eta(r)\|^2_{H^1(\Omega)}\right)\mathrm{d}r,
\end{aligned}\tag{5.70}$$

where $t\in[0,T]$. Taking $t=T$ on the right-hand side of (5.70) gives an overestimation for any $t\in[0,T]$ and leads to

$$e^{Tc_8}\int_0^T c_7\left(\|\eta(r)\|^2_{H^1(\Omega)}+\|\dot\eta(r)\|^2_{H^1(\Omega)}\right)\mathrm{d}r\leq c_9\|\eta\|^2_{H^2(\Omega\times(0,T))}\leq c_{10}\|\eta\|^2_{H^4(\Omega\times(0,T))}\,.\tag{5.71}$$

The initial condition for the Galerkin approximation (5.50) gives

$$\begin{aligned}
\|\nabla(u_m(0)+\eta(0))\|^2_{L_2(\Omega)}&\leq\|u_m(0)+\eta(0)\|^2_{H^1(\Omega)}\\
&=\left\|\sum_{k=1}^m(u_{\mathrm{ini}}-\eta(0),w_k)w_k+\sum_{k=1}^\infty(\eta(0),w_k)w_k\right\|^2_{H^1(\Omega)}\\
&=\left\|\sum_{k=1}^m(u_{\mathrm{ini}},w_k)w_k+\sum_{k=m+1}^\infty(\eta(0),w_k)w_k\right\|^2_{H^1(\Omega)}\\
&\leq 2\left\|\sum_{k=1}^\infty(u_{\mathrm{ini}},w_k)w_k\right\|^2_{H^1(\Omega)}+2\left\|\sum_{k=1}^\infty(\eta(0),w_k)w_k\right\|^2_{H^1(\Omega)}\\
&=2\|u_{\mathrm{ini}}\|^2_{H^1(\Omega)}+2\|\eta(0)\|^2_{H^1(\Omega)}\,.
\end{aligned}\tag{5.72}$$

Taking $\eta \in H^4(\Omega \times (0,T))$ into account, (5.72) may be refined further. From Lemma 11, it follows that $\|\eta(0)\|^2_{H^1(\Omega)} < \infty$. Using Lemma 11, one obtains an estimate in the the $H^4(\Omega \times (0,T))$-norm that reads

$$\|\eta(0)\|^2_{H^1(\Omega)} \leq c_{11} \|\eta\|^2_{H^4(\Omega\times(0,T))}. \tag{5.73}$$

The term $\|\nabla p_m(0)\|^2_{L_2(\Omega)}$ on the right-hand side of (5.70) can be estimated using (5.51) as

$$\|\nabla p_m(0)\|^2_{L_2(\Omega)} \leq \|p_{\text{ini}}\|^2_{H^1_0(\Omega)}. \tag{5.74}$$

Lastly, the term $\int_\Omega \omega_0(p_m(0))$ in (5.70) is estimated. Applying Young's inequality, constants $c_{12}, c_{13}, c_{14}, c_{15}$ can be found such that

$$c_{12}y^4 - c_{13} \leq \frac{1}{a}\omega_0(y) \leq c_{14}y^4 + c_{15}$$

holds. It follows that

$$\left| \int_\Omega \omega_0(p_m(0)) \right| \leq c_{14} \int_\Omega p_m(0)^4 + c_{15}\mu(\Omega). \tag{5.75}$$

Considering Theorem 21 for $n = 1, 2, 3$ results in

$$H^1_0(\Omega) \hookrightarrow L_4(\Omega).$$

Using this embedding along with (5.74) gives rise to the estimate

$$\|p_m(0)\|^4_{L_4(\Omega)} \leq c_{16} \|p_m(0)\|^4_{H^1_0(\Omega)} \leq c_{17} \|p_{\text{ini}}\|^4_{H^1(\Omega)}. \tag{5.76}$$

Applying (5.71), (5.72), (5.73), (5.75), and (5.76) leads to

$$\begin{aligned} &\frac{\alpha\xi^2}{8H^2} \|\nabla(u_m(t) + \eta(t))\|^2_{L_2(\Omega)} + \xi^2 \|\nabla p_m(t)\|^2_{L_2(\Omega)} + \int_\Omega \omega_0(p_m(t)) \\ \leq& c_{18} \left(\|u_{\text{ini}}\|^2_{H^1(\Omega)} + \|p_{\text{ini}}\|^2_{H^1_0(\Omega)} + \|p_{\text{ini}}\|^4_{H^1_0(\Omega)} + \|\eta\|^2_{H^4(\Omega\times(0,T))} + 1 \right) \text{ a.e. in } [0,T]. \end{aligned} \tag{5.77}$$

Since

$$\begin{aligned} u_{\text{ini}} &\in H^1(\Omega), \\ p_{\text{ini}} &\in H^1_0(\Omega), \\ \eta &\in H^4(\Omega \times (0,T)), \end{aligned}$$

and ω_0 is positive (see (5.58)), taking the essential supremum of (5.77) shows that

$$\begin{aligned} (\nabla(u_m + \eta))_{m\in\mathbb{N}} &\text{ is bounded in } L_\infty(0,T;L_2(\Omega)), \\ (\nabla p_m)_{m\in\mathbb{N}} &\text{ is bounded in } L_\infty(0,T;L_2(\Omega)). \end{aligned}$$

The Poincare and Young inequalities further provide

$$\begin{aligned} \|u_m\|^2_{L_\infty(0,T;H^1_0(\Omega))} &\leq c_{19} \|\nabla u_m\|^2_{L_\infty(0,T;L_2(\Omega))} \\ &\leq c_{20} \|\nabla(u_m + \eta)\|^2_{L_\infty(0,T;L_2(\Omega))} + c_{21} \|\nabla\eta\|^2_{L_\infty(0,T;L_2(\Omega))} \\ &\leq c_{22} \|\nabla(u_m + \eta)\|^2_{L_\infty(0,T;L_2(\Omega))} + c_{23} \|\eta\|^2_{H^4(\Omega\times(0,T))}. \end{aligned} \tag{5.78}$$

Similarly,

$$\|p_m\|^2_{L_\infty\left(0,T;H^1_0(\Omega)\right)} \leq \|\nabla p_m\|^2_{L_\infty(0,T;L_2(\Omega))}\,. \tag{5.79}$$

The estimates (5.78), (5.79) and (5.77) imply that

$$(u_m)_{m\in\mathbb{N}} \text{ is bounded in } L_\infty\left(0,T;H^1_0\left(\Omega\right)\right), \tag{5.80}$$
$$(p_m)_{m\in\mathbb{N}} \text{ is bounded in } L_\infty\left(0,T;H^1_0\left(\Omega\right)\right). \tag{5.81}$$

In addition to this result, a bound on the time derivatives is needed in order to conclude the proof of existence. Letting $\varrho\left(t\right)$ be as in (5.69) and using a similar estimate to (5.71), the inequality (5.70) can be viewed as

$$\varrho\left(t\right) \leq e^{tc_8}\left(\varrho\left(0\right) + c_{24}\left\|\eta\right\|^2_{H^4(\Omega\times(0,T))}\right) \text{ for all } t\in[0,T]\,. \tag{5.82}$$

Integrating (5.82) with respect to t over $[0,T]$ gives rise to

$$\int_0^T \varrho\left(r\right)\mathrm{d}r \leq \frac{1}{c_8}\left(\varrho\left(0\right) + c_{24}\left\|\eta\right\|^2_{H^4(\Omega\times(0,T))}\right)\left(e^{Tc_8}-1\right). \tag{5.83}$$

Returning to the inequality (5.65), applying (5.66) and retaining all the terms on the left-hand side (as opposed to the estimate 5.77), one arrives at

$$\begin{aligned}
&\frac{\alpha\xi^2}{8H^2}\left\|\dot{u}_m\right\|^2_{L_2(\Omega)} + \frac{\alpha\xi^2}{4}\left\|\dot{p}_m\right\|^2_{L_2(\Omega)} \\
&\quad + \frac{\alpha\xi^2}{8H^2}\frac{\mathrm{d}}{\mathrm{d}t}\left\|\nabla\left(u_m+\eta\right)\right\|^2_{L_2(\Omega)} + \xi^2\frac{\mathrm{d}}{\mathrm{d}t}\left\|\nabla p_m\right\|^2_{L_2(\Omega)} + \frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega \omega_0\left(p_m\right) \\
\leq& c_7\left(\left\|\nabla\left(u_m+\eta\right)\right\|^2_{L_2(\Omega)} + \left\|\eta\right\|^2_{H^1(\Omega)} + \left\|\dot{\eta}\right\|^2_{H^1(\Omega)}\right) \text{ a.e. in } [0,T]\,.
\end{aligned} \tag{5.84}$$

The terms $\frac{8H^2c_7}{\alpha}\left\|\nabla p_m\right\|^2_{L_2(\Omega)}$ and $\frac{8H^2c_7}{\alpha\xi^2}\int_\Omega \omega_0\left(p_m\right)$ are non-negative. Adding them to the right-hand side of (5.84), adjusting the constants and integrating with respect to t over $[0,T]$ yields

$$\begin{aligned}
&\frac{\alpha\xi^2}{8H^2}\int_0^T \left\|\dot{u}_m\left(t\right)\right\|^2_{L_2(\Omega)}\mathrm{d}t + \frac{\alpha\xi^2}{4}\int_0^T \left\|\dot{p}_m\left(t\right)\right\|^2_{L_2(\Omega)}\mathrm{d}t + \varrho\left(T\right) - \varrho\left(0\right) \\
\leq& \frac{1}{c_8}\frac{8H^2c_7}{\alpha\xi^2}\left(\varrho\left(0\right) + c_{24}\left\|\eta\right\|^2_{H^4(\Omega\times(0,T))}\right)\left(e^{Tc_8}-1\right) \\
&+ c_7\int_0^T \left\|\eta\left(t\right)\right\|^2_{H^1(\Omega)}\mathrm{d}t + c_7\int_0^T \left\|\dot{\eta}\left(t\right)\right\|^2_{H^1(\Omega)}\mathrm{d}t,
\end{aligned} \tag{5.85}$$

where (5.83) was used. Utilizing the fact that $\left(e^{Tc_8}-1\right) > 0$ allows us to overestimate some of the terms on the right-hand side of (5.85) using

$$\begin{aligned}
&\frac{1}{c_8}\frac{8H^2c_7}{\alpha\xi^2}\left\|\eta\right\|^2_{H^4(\Omega\times(0,T))}\left(e^{Tc_8}-1\right) + c_7\int_0^T \left\|\eta\left(t\right)\right\|^2_{H^1(\Omega)}\mathrm{d}t + c_7\int_0^T \left\|\dot{\eta}\left(t\right)\right\|^2_{H^1(\Omega)}\mathrm{d}t \\
\leq& c_{25}\left\|\eta\right\|^2_{H^4(\Omega\times(0,T))}\,.
\end{aligned}$$

Since c_7 can be chosen such that $\frac{1}{c_8}\frac{8H^2c_7}{\alpha\xi^2} > 1$ (see (5.68)), adding $\varrho(0)$ to both sides of (5.85) and dropping the positive term $\varrho(T)$ leads to

$$\begin{aligned}&\frac{\alpha\xi^2}{8H^2}\int_0^T \|\dot{u}_m(t)\|^2_{L_2(\Omega)}\,\mathrm{d}t + \frac{\alpha\xi^2}{4}\int_0^T \|\dot{p}_m(t)\|^2_{L_2(\Omega)}\,\mathrm{d}t\\ &\leq c_{26}\varrho(0) + c_{25}\|\eta\|^2_{H^4(\Omega\times(0,T))}.\end{aligned} \tag{5.86}$$

The inequality (5.86) can be rewritten as

$$\begin{aligned}&\frac{\alpha\xi^2}{8H^2}\|\dot{u}_m\|^2_{L_2(0,T;L_2(\Omega))} + \frac{\alpha\xi^2}{4}\|\dot{p}_m\|^2_{L_2(0,T;L_2(\Omega))}\\ \leq& c_{27}\left(\|u_{\mathrm{ini}}\|^2_{H^1(\Omega)} + \|p_{\mathrm{ini}}\|^2_{H^1_0(\Omega)} + \|p_{\mathrm{ini}}\|^4_{H^1_0(\Omega)} + \|\eta\|^2_{H^4(\Omega\times(0,T))} + 1\right),\end{aligned} \tag{5.87}$$

where the estimates (5.73), (5.74) (5.75) and (5.76) were used once again to estimate $\varrho(0)$.

In conclusion, the estimate (5.87) is used to see that

$$(\dot{u}_m)_{m\in\mathbb{N}}, (\dot{p}_m)_{m\in\mathbb{N}} \text{ are bounded in } L_2(0,T;L_2(\Omega)) \tag{5.88}$$

and combining (5.80), (5.81) with (5.87) shows that

$$(u_m)_{m\in\mathbb{N}}, (p_m)_{m\in\mathbb{N}} \text{ are bounded in } W\left(0,T;q,2;H^1_0(\Omega),L_2(\Omega)\right), \tag{5.89}$$

where $q\in\mathbb{N}$ is arbitrary. The freedom in the choice of q will be used in the following section to provide the needed strong convergence.

C) The space $W\left(0,T;q,2;H^1_0(\Omega),L_2(\Omega)\right)$ is reflexive for any $q\in\mathbb{N}$, this means that any bounded sequence has a weakly convergent sub-sequence. Applying this to (5.89) implies that there exist sub-sequences such that

$$u_{h_m} \rightharpoonup u^{(q)} \text{ in } L_q\left(0,T;H^1_0(\Omega)\right), \tag{5.90}$$
$$p_{h_m} \rightharpoonup p^{(q)} \text{ in } L_q\left(0,T;H^1_0(\Omega)\right), \tag{5.91}$$
$$\dot{u}_{h_m} \rightharpoonup \dot{u}^{(q)} \text{ in } L_2(0,T;L_2(\Omega)), \tag{5.92}$$
$$\dot{p}_{h_m} \rightharpoonup \dot{p}^{(q)} \text{ in } L_2(0,T;L_2(\Omega)). \tag{5.93}$$

First, choose sub-sequences $(u_{h_m}), (p_{h_m})$ that converge weakly in $W\left(0,T;8,2;H^1_0(\Omega),L_2(\Omega)\right)$ to $u^{(8)}, p^{(8)}$, respectively. Next, find sub-sequences of $(u_{h_m}), (p_{h_m})$ that converge weakly in $W\left(0,T;2,2;H^1_0(\Omega),L_2(\Omega)\right)$ to $u^{(2)}, p^{(2)}$, respectively. This results in sequences $(u_m), (p_m)$ (the sub-sequence notation is dropped) with the following properties

$$(u_m) \rightharpoonup u^{(8)}, (p_m) \rightharpoonup p^{(8)} \text{ in } W\left(0,T;8,2;H^1_0(\Omega),L_2(\Omega)\right), \tag{5.94}$$
$$(u_m) \rightharpoonup u^{(2)}, (p_m) \rightharpoonup p^{(2)} \text{ in } W\left(0,T;2,2;H^1_0(\Omega),L_2(\Omega)\right). \tag{5.95}$$

Applying Theorem 22 along with the Gelfand triple $H^1_0(\Omega) \hookrightarrow\hookrightarrow L_2(\Omega) \hookrightarrow H^1_0(\Omega)^*$, we find that (5.94), (5.95) lead to $u^{(2)} = u^{(8)} \equiv u$ and $p^{(2)} = p^{(8)} \equiv p$.

To show that the pair (u,p) is a solution to (5.41)-(5.44), consider two functions of the form

$$\varphi(t) \equiv \sum_{i=1}^m a^i(t)\,w_i,\ \ \psi(t) \equiv \sum_{i=1}^m b^i(t)\,w_i, \tag{5.96}$$

where $a^i, b^i : [0,T] \to \mathbb{R}$ are smooth coefficients. It is a standard result that functions of the form (5.96) are dense in $L_2\left(0,T;H_0^1(\Omega)\right)$ [34]. Multiplying (5.48) by $a^i(t)$ and (5.49) by $b^i(t)$ for all $i \in \{1,\dots,m\}$ and summing each of them separately leads to

$$\int_\Omega \dot{u}_m\varphi + \int_\Omega \nabla u_m \cdot \nabla\varphi + \int_\Omega \nabla\eta\cdot\nabla\varphi = H\int_\Omega \dot{p}_m\varphi - \int_\Omega \dot{\eta}\varphi \text{ a.e. in } [0,T], \tag{5.97}$$

$$\alpha\xi^2\int_\Omega \dot{p}_m\psi + \xi^2\int_\Omega \nabla p_m\cdot\nabla\psi = \int_\Omega f_0(p_m)\,\psi - b\beta\xi\int_\Omega (u_m+\eta)\,\psi \text{ a.e. in } [0,T]. \tag{5.98}$$

Integrating (5.97)-(5.98) with respect to t over $[0,T]$ gives rise to

$$\int_0^T\int_\Omega \dot{u}_m\varphi\mathrm{d}t + \int_0^T\int_\Omega \nabla u_m\cdot\nabla\varphi\mathrm{d}t + \int_0^T\int_\Omega \nabla\eta\cdot\nabla\varphi\mathrm{d}t = H\int_0^T\int_\Omega \dot{p}_m\varphi\mathrm{d}t - \int_0^T\int_\Omega \dot{\eta}\varphi\mathrm{d}t, \tag{5.99}$$

$$\alpha\xi^2\int_0^T\int_\Omega \dot{p}_m\psi\mathrm{d}t + \xi^2\int_0^T\int_\Omega \nabla p_m\cdot\nabla\psi\mathrm{d}t = \int_0^T\int_\Omega f_0(p_m)\,\psi\mathrm{d}t - b\beta\xi\int_0^T\int_\Omega (u_m+\eta)\,\psi\mathrm{d}t. \tag{5.100}$$

To show that the solution exists, the convergence of all the terms in (5.99)-(5.100) needs to be investigated. The terms

$$\int_0^T\int_\Omega \dot{u}_m\varphi\mathrm{d}t, \int_0^T\int_\Omega \dot{p}_m\varphi\mathrm{d}t \text{ and } \int_0^T\int_\Omega \dot{p}_m\psi\mathrm{d}t$$

converge, since $\varphi,\psi \in L_2\left(0,T;H_0^1(\Omega)^*\right)$ (using the canonical pairing). Moreover, (5.94) and (5.95) ensure that

$$\int_0^T\int_\Omega \dot{u}_m\varphi\mathrm{d}t \to \int_0^T\int_\Omega \dot{u}\varphi\mathrm{d}t, \quad \int_0^T\int_\Omega \dot{p}_m\varphi\mathrm{d}t \to \int_0^T\int_\Omega \dot{p}\varphi\mathrm{d}t, \quad \int_0^T\int_\Omega \dot{p}_m\psi\mathrm{d}t \to \int_0^T\int_\Omega \dot{p}\psi\mathrm{d}t.$$

The convergence

$$\int_0^T\int_\Omega u_m\psi\mathrm{d}t \to \int_0^T\int_\Omega u\psi\mathrm{d}t$$

follows from $\psi \in L_2\left(0,T;H_0^1(\Omega)^*\right)$ and (5.95). The treatment of the remaining linear terms $\int_0^T\int_\Omega \nabla u_m\cdot\nabla\varphi\mathrm{d}t, \int_0^T\int_\Omega \nabla p_m\cdot\nabla\psi\mathrm{d}t$ is similar. It is clear that [34]

$$\int_0^T B[\cdot,\varphi]\,\mathrm{d}t \in L_2\left(0,T;H_0^1(\Omega)^*\right),$$

which results in

$$\int_0^T\int_\Omega \nabla u_m\cdot\nabla\varphi\mathrm{d}t \to \int_0^T\int_\Omega \nabla u\cdot\nabla\varphi\mathrm{d}t, \quad \int_0^T\int_\Omega \nabla p_m\cdot\nabla\psi\mathrm{d}t \to \int_0^T\int_\Omega \nabla p\cdot\nabla\psi\mathrm{d}t,$$

where (5.94)-(5.95) was used.

Since

$$p_m \rightharpoonup p \text{ in } W\left(0,T;8,2;H_0^1(\Omega),L_2(\Omega)\right),$$

Lemma 12 can be applied. This leads to the convergence of the term $\int_0^T \int_\Omega f_0(p_m)\,\psi \mathrm{d}t$. More specifically, the cubic and quadratic terms converge:

$$\begin{aligned}(p_m)^3 &\to p^3 \text{ in } L_2\left(0,T;H_0^1(\Omega)^*\right),\\ (p_m)^2 &\to p^2 \text{ in } L_2\left(0,T;H_0^1(\Omega)^*\right),\end{aligned} \tag{5.101}$$

which are the only non-linearities contained in the term f_0. The relations (5.101) imply that

$$\begin{aligned}\left|\left((p_m)^3-p^3\right)(\psi)\right| &\le \left\|(p_m)^3-p^3\right\|_{L_2\left(0,T;H_0^1(\Omega)^*\right)}\|\psi\|_{L_2\left(0,T;H_0^1(\Omega)\right)}\to 0,\\ \left|\left((p_m)^2-p^2\right)(\psi)\right| &\le \left\|(p_m)^2-p^2\right\|_{L_2\left(0,T;H_0^1(\Omega)^*\right)}\|\psi\|_{L_2\left(0,T;H_0^1(\Omega)\right)}\to 0,\end{aligned}$$

which concludes the proof of convergence of all the terms in (5.99)-(5.100). Applying integration by parts to the time derivatives in (5.41)-(5.42), equations (5.99)-(5.100) and taking the limit shows that the initial conditions are satisfied by the limit functions as well. Finally, we may conclude that (5.41)-(5.44) has a solution for an arbitrary $\eta \in H^4(\Omega\times(0,T))$.

D) To prove that the solution provided in part C) is unique, suppose that (u_1,p_1) and (u_2,p_2) are both solutions of (5.41)-5.44). Define $u_{12}\equiv u_1-u_2$ and $p_{12}\equiv p_1-p_2$. Since the pairs (u_1,p_1) and (u_2,p_2) are both characterized by being a solution to (5.41)-(5.44), subtracting the two systems results in

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega u_{12}(t)\,v+\int_\Omega \nabla u_{12}(t)\cdot\nabla v = H\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega p_{12}(t)\,v$$
$$\text{for all } v\in H_0^1(\Omega) \text{ a.e. in } [0,T], \tag{5.102}$$
$$\alpha\xi^2\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega p_{12}(t)\,w+\xi^2\int_\Omega \nabla p_{12}(t)\cdot\nabla w = \int_\Omega \left[f_0(p_1(t))-f_0(p_2(t))\right]w - b\beta\xi\int_\Omega u_{12}(t)\,w$$
$$\text{for all } w\in H_0^1(\Omega) \text{ a.e. in } [0,T], \tag{5.103}$$
$$u_{12}|_{t=0}=0,\quad p_{12}|_{t=0}=0. \tag{5.104}$$

Setting $v=u_{12}(t)-Hp_{12}(t)$ and $w=p_{12}(t)$ in (5.102) and (5.103), respectively, and using the notation (5.52)-(5.54) leads to

$$(\dot u_{12}-H\dot p_{12},u_{12}-Hp_{12})+B[u_{12}-Hp_{12},u_{12}-Hp_{12}]+HB[p_{12},u_{12}-Hp_{12}]=0, \tag{5.105}$$
$$\alpha\xi^2(\dot p_{12},p_{12})+\xi^2B[p_{12},p_{12}]=(f_0(p_1)-f_0(p_2),p_{12})-b\beta\xi(u_{12},p_{12}) \text{ a.e. in } [0,T]. \tag{5.106}$$

The regularity properties of the solution guarantee that there exists a constant $C_{f_0}>0$ such that

$$\left\|\frac{f_0(p_1)-f_0(p_2)}{p_1-p_2}\right\|_{L_2(\Omega)}\le C_{f_0},$$

where f_0 is given by (5.58). Using the Young and Schwarz inequalities, the following estimates can be made:

$$|-HB\left[p_{12}, u_{12} - Hp_{12}\right]| \leq \frac{H^2}{2} \left\|\nabla p_{12}\right\|_{L_2(\Omega)}^2 + \frac{1}{2} \left\|\nabla\left(u_{12} - Hp_{12}\right)\right\|_{L_2(\Omega)}^2, \tag{5.107}$$

$$\begin{aligned} |\left(f_0\left(p_1\right) - f_0\left(p_2\right), p_{12}\right)| &\leq C_{f_0} \left\|p_{12}\right\|_{L_4(\Omega)}^2 \leq C_{f_0} \left\|p_{12}\right\|_{L_2(\Omega)}^{\frac{1}{2}} \left\|p_{12}\right\|_{L_6(\Omega)}^{\frac{6}{4}} \\ &\leq C_{f_0} C_{\mathrm{emb}}^{\frac{6}{4}} \left\|p_{12}\right\|_{L_2(\Omega)}^{\frac{1}{2}} \left\|p_{12}\right\|_{H_0^1(\Omega)}^{\frac{6}{4}} \\ &\leq c_{28} \left[\frac{\varepsilon^4 \left\|p_{12}\right\|_{L_2(\Omega)}^2}{4} + \frac{\left\|\nabla p_{12}\right\|_{L_2(\Omega)}^2}{\frac{4}{3}\varepsilon^{\frac{4}{3}}}\right], \end{aligned} \tag{5.108}$$

$$|-b\beta\xi\left(u_{12}, p_{12}\right)| \leq \frac{c_{29}}{2}\left[\left\|u_{12}\right\|_{L_2(\Omega)}^2 + \left\|p_{12}\right\|_{L_2(\Omega)}^2\right]. \tag{5.109}$$

In (5.108), $C_{\mathrm{emb}} > 0$ is the constant of the continuous embedding $H_0^1\left(\Omega\right) \hookrightarrow L_6\left(\Omega\right)$. Combining (5.105) with (5.107) results in

$$\frac{\mathrm{d}}{\mathrm{d}t} \left\|u_{12} - Hp_{12}\right\|_{L_2(\Omega)}^2 + \left\|\nabla\left(u_{12} - Hp_{12}\right)\right\|_{L_2(\Omega)}^2 \leq H^2 \left\|\nabla p_{12}\right\|_{L_2(\Omega)}^2. \tag{5.110}$$

Similarly, the estimate

$$\frac{\alpha\xi^2}{2} \frac{\mathrm{d}}{\mathrm{d}t} \left\|p_{12}\right\|_{L_2(\Omega)}^2 + \frac{3\xi^2}{4} \left\|\nabla p_{12}\right\|_{L_2(\Omega)}^2 \leq c_{30} \left\|p_{12}\right\|_{L_2(\Omega)}^2 + c_{31} \left\|u_{12}\right\|_{L_2(\Omega)}^2 \tag{5.111}$$

arises by combining (5.109), (5.108) with (5.106), where $\varepsilon > 0$ in (5.108) is set so that

$$\frac{c_{28}}{\frac{4}{3}\varepsilon^{\frac{4}{3}}} = \frac{\xi^2}{4}.$$

Adding the $\frac{\xi^2}{2H^2}$ multiple of (5.110) to (5.111) along with using the estimate

$$\begin{aligned} \left\|u_{12}\right\|_{L_2(\Omega)}^2 &\leq \left(\left\|u_{12} - Hp_{12}\right\|_{L_2(\Omega)} + H \left\|p_{12}\right\|_{L_2(\Omega)}\right)^2 \\ &\leq 2 \left\|u_{12} - Hp_{12}\right\|_{L_2(\Omega)}^2 + 2H^2 \left\|p_{12}\right\|_{L_2(\Omega)}^2, \end{aligned}$$

gives rise to

$$\begin{aligned} &\frac{\xi^2}{2H^2} \left\|\nabla\left(u_{12} - Hp_{12}\right)\right\|_{L_2(\Omega)}^2 + \frac{\xi^2}{4} \left\|\nabla p_{12}\right\|_{L_2(\Omega)}^2 \\ &+ \frac{\alpha\xi^2}{2} \frac{\mathrm{d}}{\mathrm{d}t} \left\|p_{12}\right\|_{L_2(\Omega)}^2 + \frac{\xi^2}{2H^2} \frac{\mathrm{d}}{\mathrm{d}t} \left\|u_{12} - Hp_{12}\right\|_{L_2(\Omega)}^2 \\ &\leq c_{32} \left\|u_{12} - Hp_{12}\right\|_{L_2(\Omega)}^2 + c_{33} \left\|p_{12}\right\|_{L_2(\Omega)}^2 \text{ a.e. in } [0, T]. \end{aligned} \tag{5.112}$$

Dropping the positive terms on the left-hand side of (5.112) and increasing the constants on the right-hand side so that

$$\begin{aligned} &\frac{\alpha\xi^2}{2} \frac{\mathrm{d}}{\mathrm{d}t} \left\|p_{12}\right\|_{L_2(\Omega)}^2 + \frac{\xi^2}{2H^2} \frac{\mathrm{d}}{\mathrm{d}t} \left\|u_{12} - Hp_{12}\right\|_{L_2(\Omega)}^2 \\ &\leq c_{34} \left(\frac{\alpha\xi^2}{2} \left\|u_{12} - Hp_{12}\right\|_{L_2(\Omega)}^2 + \frac{\xi^2}{2H^2} \left\|p_{12}\right\|_{L_2(\Omega)}^2\right) \text{ a.e. in } [0, T], \end{aligned}$$

for some constant $c_{34} > 0$ sets the stage for using Grönwall's lemma (Lemma 18). Applying said lemma yields

$$\frac{\alpha\xi^2}{2} \left\|p_{12}\left(t\right)\right\|_{L_2(\Omega)}^2 + \frac{\xi^2}{2H^2} \left\|u_{12}\left(t\right) - Hp_{12}\left(t\right)\right\|_{L_2(\Omega)}^2 \tag{5.113}$$

$$\leq e^{c_{34}T} \left(\frac{\alpha\xi^2}{2} \left\|u_{12}\left(0\right) - Hp_{12}\left(0\right)\right\|_{L_2(\Omega)}^2 + \frac{\xi^2}{2H^2} \left\|p_{12}\left(0\right)\right\|_{L_2(\Omega)}^2\right) \text{ a.e. in } [0, T]. \tag{5.114}$$

Integrating (5.113) over $[0,T]$ and using the initial conditions of the problem (5.102)-(5.103) results in

$$\begin{aligned}\|p_{12}\|^2_{L_2(0,T;L_2(\Omega))} = 0 \Rightarrow p_1 = p_2,\\ \|u_{12} - Hp_{12}\|^2_{L_2(0,T;L_2(\Omega))} = 0 \Rightarrow u_1 = u_2.\end{aligned}$$

□

5.5. Existence of Optimal Solution

The notation (5.18), (5.29) is applied in this section. Suppose that $W_{\text{ad}} \subset X_C$ is a closed, convex and bounded set. Next, the proof of Theorem 10 is provided.

Proof. The following sufficient conditions (see Definition 2 and Theorem 3) need to be shown to hold:
(c1) The solution operator of the system $S : W_{\text{ad}} \to \hat{X}_S$ is a bounded operator.
(c2) The state equation operator

$$e : \hat{X}_S \times W_{\text{ad}} \to L_2\left(0,T; H_0^1(\Omega)^*\right)^2 \times L_2(\Omega)^2$$

is weakly sequentially continuous.
(c3) J (defined by (5.21)) is convex and continuous.

Proving condition (c1) relies on the boundedness of W_{ad}. Taking the supremum of (5.77) while taking (5.78) and (5.79) into account yields

$$\begin{aligned}&c_{35}\|u_m\|^2_{L_2\left(0,T;H_0^1(\Omega)\right)} + c_{36}\|p_m\|^2_{L_2\left(0,T;H_0^1(\Omega)\right)}\\ \leq& c_{39}\left(\|u_{\text{ini}}\|^2_{H^1(\Omega)} + \|p_{\text{ini}}\|^2_{H_0^1(\Omega)} + \|p_{\text{ini}}\|^4_{H_0^1(\Omega)} + 1 + M^2\right),\end{aligned}$$

where $M > 0$ is the constant bounding W_{ad} in X_C. From this it follows that there exists a constant $M_1 > 0$ dependent only on the Sobolev norms of the initial conditions $u_{\text{ini}}, p_{\text{ini}}$, final time T, and the constant M such that

$$c_{35}\|u_m\|^2_{L_2\left(0,T;H_0^1(\Omega)\right)} + c_{36}\|p_m\|^2_{L_2\left(0,T;H_0^1(\Omega)\right)} \leq M_1\left(u_{\text{ini}}, p_{\text{ini}}, T, M\right). \tag{5.115}$$

Analogously, the derivatives can be estimated (see (5.87)) by

$$c_{37}\|\dot{u}_m\|^2_{L_2(0,T;L_2(\Omega))} + c_{38}\|\dot{p}_m\|^2_{L_2(0,T;L_2(\Omega))} \leq M_2\left(u_{\text{ini}}, p_{\text{ini}}, T, M\right). \tag{5.116}$$

Taking the limit inferior of the estimates (5.115), (5.116), and adding them results in

$$\|(u,p)\|_{W\left(0,T;L_2(\Omega),H_0^1(\Omega)\right)^2} \leq M_3\left(u_{\text{ini}}, p_{\text{ini}}, T, M\right).$$

This concludes the proof of boundedness of the solution operator.

The convergence of the non-linear terms needs to be checked in order to guarantee that (c2) holds. Note that the direct application of Lemma 12 provides the necessary convergence.

To prove (c3), consider the embedding given by Theorem 23

$$W\left(0,T;2,2;H_0^1(\Omega), L_2(\Omega)\right) \hookrightarrow C\left([0,T]; L_2(\Omega)\right). \tag{5.117}$$

Take

$$(u_1,p_1),(u_2,p_2) \in W\left(0,T;2,2;H_0^1(\Omega), L_2(\Omega)\right) \times W\left(0,T;8,2;H_0^1(\Omega), L_2(\Omega)\right) = \hat{X}_S$$

and $\eta_1, \eta_2 \in H^4(\Omega \times (0,T)) = X_C$. With the help of Hölder's inequality, we make the estimate

$$\begin{aligned}
&|J(u_1,p_1,\eta_1) - J(u_2,p_2,\eta_2)| \\
\leq &\int_\Omega \left[|p_f| \, |p_1(T) - p_2(T)| + \frac{1}{2} |p_1(T) + p_2(T)| \, |p_1(T) - p_2(T)| \right] \\
&+ \frac{\gamma}{2} \int_0^T \int_{\partial\Omega} \left[\mathrm{Tr}\ (\eta_1) \left(\mathrm{Tr}\ (\eta_1 - \eta_2) \right) + \mathrm{Tr}\ (\eta_2) \left(\mathrm{Tr}\ (\eta_1 - \eta_2) \right) \right] \\
&\leq \left(\int_\Omega |p_f|^2 \right)^{\frac{1}{2}} \left(\int_\Omega |p_1(T) - p_2(T)|^2 \right)^{\frac{1}{2}} \\
&+ \left(\int_\Omega |p_1(T) + p_2(T)|^2 \right)^{\frac{1}{2}} \left(\int_\Omega |p_1(T) - p_2(T)|^2 \right)^{\frac{1}{2}} \\
&+ \frac{\gamma}{2} \left(\int_0^T \int_{\partial\Omega} \mathrm{Tr}\ (\eta_1)^2 \right)^{\frac{1}{2}} \left(\int_0^T \int_{\partial\Omega} \mathrm{Tr}\ (\eta_1 - \eta_2)^2 \right)^{\frac{1}{2}} \\
&+ \frac{\gamma}{2} \left(\int_0^T \int_{\partial\Omega} \mathrm{Tr}\ (\eta_2)^2 \right)^{\frac{1}{2}} \left(\int_0^T \int_{\partial\Omega} \mathrm{Tr}\ (\eta_1 - \eta_2)^2 \right)^{\frac{1}{2}}.
\end{aligned} \tag{5.118}$$

Considering $(u_1,p_1,\eta_1) \in \hat{X}_S \times X_C$ such that $\|(u_1,p_1,\eta_1) - (u_2,p_2,\eta_2)\|_{\hat{X}_S \times X_C} < 1$ makes the terms

$$\left(\int_\Omega |p_1(T) + p_2(T)|^2 \right)^{\frac{1}{2}}, \frac{\gamma}{2} \left(\int_0^T \int_{\partial\Omega} \mathrm{Tr}\ (\eta_1)^2 \right)^{\frac{1}{2}} \text{ and } \frac{\gamma}{2} \left(\int_0^T \int_{\partial\Omega} \mathrm{Tr}\ (\eta_2)^2 \right)^{\frac{1}{2}}$$

bounded, where the embedding statement (5.117) was used to ensure the boundedness of the first term. Using the boundedness of the trace operator and the same embedding statement as before, it is possible to make

$$\left(\int_\Omega |p_1(T) - p_2(T)|^2 \right)^{\frac{1}{2}}, \left(\int_\Omega |p_1(T) - p_2(T)|^2 \right)^{\frac{1}{2}}, \quad \text{and} \quad \left(\int_0^T \int_{\partial\Omega} \mathrm{Tr}\ (\eta_1 - \eta_2)^2 \right)^{\frac{1}{2}}$$

arbitrarily small by choosing an appropriate $0 < \delta < 1$ such that

$$\|(u_1,p_1,\eta_1) - (u_2,p_2,\eta_2)\|_{\hat{X}_S \times X_C} < \delta.$$

This concludes the proof and thus J is continuous. Since J is convex due to the linearity of the integral and convexity of the square function, (c3) holds. □

5.6. Fréchet Differentiability of the Solution Operator

In this section, the solution operator of (5.41)-(5.44) is shown to be Fréchet differentiable. Since the analysis provided in this section is separate from the proofs of Sections (5.4)-(5.5), the constant numbering is reset and the pair (u,p) no longer represents the solution of the original system. Conforming

to the standard notation commonly used in Fréchet differentiability proofs of (PDE) solution operators [55, 43, 68], we adhere to the following notation. Let $S : X_C \to X_S$ be the solution operator of (5.41)-(5.44), $\bar{\eta}, \eta \in X_C$, $(u,p) = S'(\bar{\eta})\,\eta$ and

$$S(\bar{\eta}) = (\bar{u}, \bar{p}), \tag{5.119}$$

then the formal Fréchet derivative (u,p) at point $\bar{\eta}$ in direction η satisfies the system

$$u_t - \Delta u = Hp_t - \eta_t + \Delta\eta, \tag{5.120}$$

$$\alpha\xi^2 p_t - \xi^2 \Delta p = -3\bar{p}^2 p + 3\bar{p}p - \frac{1}{2}p - b\beta\xi\,(u+\eta), \tag{5.121}$$

$$u|_{t=0} = -\eta|_{t=0}\,, p|_{t=0} = 0, u|_{\partial\Omega} = 0, p|_{\partial\Omega} = 0. \tag{5.122}$$

Next, let $S(\bar{\eta}+\eta) = (\tilde{u}, \tilde{p})$ be the solution to the perturbed system i.e.

$$\tilde{u}_t - \Delta\tilde{u} = H\tilde{p}_t - \bar{\eta}_t + \Delta\bar{\eta} - \eta_t + \Delta\eta, \tag{5.123}$$

$$\alpha\xi^2\tilde{p}_t - \xi^2\Delta\tilde{p} = -\tilde{p}^3 + \frac{3}{2}\tilde{p}^2 - \frac{1}{2}\tilde{p} - b\beta\xi\,(u+\bar{\eta}+\eta), \tag{5.124}$$

$$\tilde{u}|_{t=0} = u_{\text{ini}} - (\bar{\eta}+\eta)|_{t=0}\,, \tilde{p}|_{t=0} = p_{\text{ini}}, \tilde{u}|_{\partial\Omega} = 0, \tilde{p}|_{\partial\Omega} = 0. \tag{5.125}$$

Using the notation induced by (5.119) $(\bar{u}, \bar{p})$ satisfy

$$\bar{u}_t - \Delta\bar{u} = H\bar{p}_t - \bar{\eta}_t + \Delta\bar{\eta}, \tag{5.126}$$

$$\alpha\xi^2\bar{p}_t - \xi^2\Delta\bar{p} = -\bar{p}^3 + \frac{3}{2}\bar{p}^2 - \frac{1}{2}\bar{p} - b\beta\xi\,(\bar{u}+\bar{\eta}), \tag{5.127}$$

$$\bar{u}|_{t=0} = u_{\text{ini}} - \bar{\eta}|_{t=0}\,, \bar{p}|_{t=0} = p_{\text{ini}}, \bar{u}|_{\partial\Omega} = 0, \bar{p}|_{\partial\Omega} = 0. \tag{5.128}$$

Subtracting (5.126)-(5.128) from (5.123)-(5.125) leads to

$$(\tilde{u}-\bar{u})_t - \Delta(\tilde{u}-\bar{u}) = H(\tilde{p}-\bar{p})_t - \eta_t + \Delta\eta, \tag{5.129}$$

$$\alpha\xi^2(\tilde{p}-\bar{p})_t - \xi^2\Delta(\tilde{p}-\bar{p}) = f_0(\tilde{p}) - f_0(\bar{p}) - b\beta\xi\,(\tilde{u}-\bar{u}+\eta), \tag{5.130}$$

$$(\tilde{u}-\bar{u})|_{t=0} = -\eta|_{t=0}\,, (\tilde{p}-\bar{p})|_{t=0} = 0, (\tilde{u}-\bar{u})|_{\partial\Omega} = 0, (\tilde{p}-\bar{p})|_{\partial\Omega} = 0. \tag{5.131}$$

Next, it is shown that function f defined as

$$f : W\left(0,T;8,2;H_0^1(\Omega), L_2(\Omega)\right) \to L_2(0,T;L_2(\Omega)) \cong L_2(\Omega\times(0,T))$$

is Fréchet differentiable. Suppose that $\bar{y}, y \in W\left(0,T;8,2;H_0^1(\Omega), L_2(\Omega)\right)$, then

$$\begin{aligned}
&\left\|f_0(\bar{y}+y) - f_0(\bar{y}) - f'(\bar{y})\,y\right\|_{L_2(\Omega\times(0,T))} \\
&= \left\|-(\bar{y}+y)^3 + \frac{3}{2}(\bar{y}+y)^2 - \frac{1}{2}(\bar{y}+y) + (\bar{y})^3 - \frac{3}{2}(\bar{y})^2 + \frac{1}{2}(\bar{y}) + 3\bar{y}^2 y - 3\bar{y}y + \frac{1}{2}y\right\|_{L_2(\Omega\times(0,T))} \\
&= \left\|-3\bar{y}y^2 - y^3 + \frac{3}{2}y^2\right\|_{L_2(\Omega\times(0,T))} \leq 3\left\|\bar{y}y^2\right\|_{L_2(\Omega\times(0,T))} + \left\|y^3\right\|_{L_2(\Omega\times(0,T))} + \frac{3}{2}\left\|y^2\right\|_{L_2(\Omega\times(0,T))} \\
&\leq 3\|\bar{y}\|_{L_6(\Omega\times(0,T))}\|y\|^2_{L_6(\Omega\times(0,T))} + \|y\|^3_{L_6(\Omega\times(0,T))} + c_1\|y\|^2_{L_6(\Omega\times(0,T))}.
\end{aligned} \tag{5.132}$$

Taking the continuous embeddings

$$W\left(0,T;8,2;H_0^1(\Omega), L_2(\Omega)\right) \hookrightarrow L_6\left(0,T;H_0^1(\Omega)\right) \hookrightarrow L_6(0,T;L_6(\Omega))$$

into consideration shows that f is Fréchet differentiable. Let

$$r_{f_0}(\tilde{p},\bar{p}) \equiv f_0(\tilde{p}) - f_0(\bar{p}) - f'(\bar{p})(\tilde{p}-\bar{p})$$

be the remainder associated with the Fréchet derivative of f_0 at point $\bar{p}$ in direction $\tilde{p}-\bar{p}$ and define an auxiliary problem

$$(u^\varrho)_t - \Delta u^\varrho = H(p^\varrho)_t, \tag{5.133}$$

$$\alpha\xi^2(p^\varrho)_t - \xi^2\Delta p^\varrho = -3\bar{p}^2p^\varrho + 3\bar{p}p^\varrho - \frac{1}{2}p^\varrho - b\beta\xi p^\varrho + r_{f_0}(\tilde{p},\bar{p}), \tag{5.134}$$

$$u^\varrho|_{t=0} = 0, p^\varrho|_{t=0} = 0, u^\varrho|_{\partial\Omega} = 0, p^\varrho|_{\partial\Omega} = 0. \tag{5.135}$$

One can show that if $\tilde{p},\bar{p}$ are the solutions of (5.123)-(5.122) and (5.126)-(5.128) respectively, the auxiliary problem (5.133)-(5.135) has a unique solution and that there exists a constant $C_2 > 0$ such that (see Theorem 14)

$$\|(u^\varrho,p^\varrho)\|_{W\left(0,T;8,2;H_0^1(\Omega),L_2(\Omega)\right)^2} \le C_1\|r_{f_0}(\tilde{p},\bar{p})\|_{L_2(0,T;L_2(\Omega))}. \tag{5.136}$$

Note that (5.132) implies that $r_{f_0}(\tilde{p},\bar{p}) \in L_2(0,T;L_2(\Omega))$ (since $\tilde{p},\bar{p} \in L_6(0,T;L_6(\Omega))$). Subtracting (5.120)-(5.122) and (5.133)-(5.135) from (5.129)-(5.131) gives rise to the problem

$$(\tilde{u}-\bar{u}-u^\varrho-u)_t - \Delta(\tilde{u}-\bar{u}-u^\varrho-u) = H(\tilde{p}-\bar{p}-p^\varrho-p)_t \tag{5.137}$$

$$\alpha\xi^2(\tilde{p}-\bar{p}-p^\varrho-p)_t - \xi^2\Delta(\tilde{p}-\bar{p}-p^\varrho-p) = f_0'(\bar{p})(\tilde{p}-\bar{p}-p^\varrho-p) - b\beta\xi(\tilde{u}-\bar{u}-u^\varrho-u) \tag{5.138}$$

$$(\tilde{u}-\bar{u}-u^\varrho-u)|_{t=0} = 0, (\tilde{p}-\bar{p}-p^\varrho-p)|_{t=0} = 0, \tag{5.139}$$

$$(\tilde{u}-\bar{u}-u^\varrho-u)|_{\partial\Omega} = 0, (\tilde{p}-\bar{p}-p^\varrho-p)|_{\partial\Omega} = 0. \tag{5.140}$$

Notice that (5.137)-(5.140) is just a special case of problem (5.133)-(5.135) and thus has a unique solution and

$$\tilde{u}-\bar{u}-u^\varrho-u = 0, \tilde{p}-\bar{p}-p^\varrho-p = 0, \tag{5.141}$$

since $0 \in L_2(0,T;L_2(\Omega))$ replaces the remainder.

In addition to this, it is possible to derive a stability estimate for problem (5.129)-(5.131) (see Theorem 15). This estimate reads

$$\left\|\begin{pmatrix}\tilde{u}\\ \tilde{p}\end{pmatrix} - \begin{pmatrix}\bar{u}\\ \bar{p}\end{pmatrix}\right\|_{W\left(0,T;8,2;H_0^1(\Omega),L_2(\Omega)\right)^2} \le C_2\|\eta\|_{H^4(\Omega\times(0,T))}. \tag{5.142}$$

To show the Fréchet differentiability of the solution operator, one makes the estimate

$$\begin{aligned}
&\frac{\|(u^\varrho,p^\varrho)\|_{W\left(0,T;8,2;H_0^1(\Omega),L_2(\Omega)\right)^2}}{\|\eta\|_{H^4(\Omega\times(0,T))}} \\
&\le \frac{\|r_{f_0}(\tilde{p},\bar{p})\|_{L_2(0,T;L_2(\Omega))}}{\|\eta\|_{H^4(\Omega\times(0,T))}} \\
&= \frac{\|r_{f_0}(\tilde{p},\bar{p})\|_{L_2(0,T;L_2(\Omega))}}{\|\tilde{p}-\bar{p}\|_{W\left(0,T;8,2;H_0^1(\Omega),L_2(\Omega)\right)}} \frac{\|\tilde{p}-\bar{p}\|_{W\left(0,T;8,2;H_0^1(\Omega),L_2(\Omega)\right)}}{\|\eta\|_{H^4(\Omega\times(0,T))}} \\
&\le C_2\frac{\|r_{f_0}(\tilde{p},\bar{p})\|_{L_2(0,T;L_2(\Omega))}}{\|\tilde{p}-\bar{p}\|_{W\left(0,T;8,2;H_0^1(\Omega),L_2(\Omega)\right)}},
\end{aligned} \tag{5.143}$$

where (5.136) and (5.142) were used. Using (5.142) once again, along with the Fréchet differentiability of f_0 shows that

$$\frac{\|r_{f_0}(\tilde{p},\bar{p})\|_{L_2(0,T;L_2(\Omega))}}{\left\|\begin{pmatrix}\tilde{u}\\ \tilde{p}\end{pmatrix}-\begin{pmatrix}\bar{u}\\ \bar{p}\end{pmatrix}\right\|_{W\left(0,T;8,2;H_0^1(\Omega),L_2(\Omega)\right)^2}}\to 0 \text{ as } \eta\to 0 \text{ in } H^4(\Omega\times(0,T))$$

and consequently

$$\frac{\|(u^\varrho,p^\varrho)\|_{W\left(0,T;8,2;H_0^1(\Omega),L_2(\Omega)\right)^2}}{\|\eta\|_{H^4(\Omega\times(0,T))}}\to 0 \text{ as } \eta\to 0 \text{ in } H^4(\Omega\times(0,T)),$$

which concludes the proof of Fréchet differentiability. Next, the stability estimates (5.136) and (5.142) are proven.

Theorem 14. *Let $\tilde{p},\bar{p}$ be the solutions to (5.123)-(5.125) and (5.126)-(5.128) respectively, then there exists a unique solution of the problem (5.133)-(5.135) and the stability estimate*

$$\|(u^\varrho,p^\varrho)\|_{W\left(0,T;8,2;H_0^1(\Omega),L_2(\Omega)\right)^2}\leq C_1\|r_{f_0}(\tilde{p},\bar{p})\|_{L_2(0,T;L_2(\Omega))}. \tag{5.144}$$

holds.

Proof. Let $(w_k)_{k\in\mathbb{N}}$ be such that (5.45)-(5.46) hold. For each $m\in\mathbb{N}$ let there be two vector valued functions $u_m^\varrho,p_m^\varrho:[0,T]\to H_0^1(\Omega)$ such that

$$u_m^\varrho(t)\equiv\sum_{k=1}^m\alpha_m^k(t)w_k,\quad p_m^\varrho(t)\equiv\sum_{k=1}^m\beta_m^k(t)w_k. \tag{5.145}$$

We consider the approximate problem

$$\begin{aligned}\int_\Omega\dot{u}_m^\varrho(t)w_k+\int_\Omega\nabla u_m^\varrho(t)\cdot\nabla w_k=&H\int_\Omega\dot{p}_m^\varrho(t)w_k\\ &k=1,\dots,m,\ \text{a.e. in } [0,T], \end{aligned}\tag{5.146}$$

$$\begin{aligned}\alpha\xi^2\int_\Omega\dot{p}_m^\varrho(t)w_k+\xi^2\int_\Omega\nabla\dot{p}_m^\varrho(t)\cdot\nabla w_k=&-3\int_\Omega\bar{p}(t)^2p_m^\varrho(t)w_k+3\int_\Omega\bar{p}(t)p_m^\varrho(t)w_k\\ &-\frac{1}{2}\int_\Omega p_m^\varrho(t)w_k-b\beta\xi\int_\Omega u_m^\varrho(t)w_k\\ &+\int_\Omega r_{f_0}(\tilde{p},\bar{p})(t)w_k\\ &k=1,\dots,m,\ \text{a.e. in } [0,T]\end{aligned}\tag{5.147}$$

with the initial conditions

$$\alpha_m^k(0)=0\quad k=1,\dots,m, \tag{5.148}$$

$$\beta_m^k(0)=0\quad k=1,\dots,m. \tag{5.149}$$

Multiplying (5.146) by $\dot{\alpha}_m^k$ and (5.147) by $\dot{\beta}_m^k$ for each $k = 1, \dots, m$ and adding up the equations multiplied by $\dot{\alpha}_m^k$ and $\dot{\beta}_m^k$ separately gives

$$\left\|\dot{u}_m^\varrho\right\|_{L_2(\Omega)}^2 + \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left\|\nabla u_m^\varrho\right\|_{L_2(\Omega)}^2 = H\left(\dot{p}_m^\varrho, \dot{u}_m^\varrho\right) \tag{5.150}$$

$$\begin{aligned}\alpha\xi^2\left\|\dot{p}_m^\varrho\right\|_{L_2(\Omega)}^2 + \xi^2\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left\|\nabla p_m\right\|_{L_2(\Omega)}^2 &= 3\left(-\bar{p}^2 p_m^\varrho, \dot{p}_m^\varrho\right) + 3\left(\bar{p}p_m^\varrho, \dot{p}_m^\varrho\right) \\ &\quad - \frac{1}{2}\left(p_m^\varrho, \dot{p}_m^\varrho\right) - b\beta\xi\left(u_m^\varrho, \dot{p}_m^\varrho\right) \\ &\quad + \left(r_{f_0}\left(\tilde{p}, \bar{p}\right), \dot{p}_m^\varrho\right).\end{aligned} \tag{5.151}$$

Applying the Hölder and Young inequalities leads to the following estimates

$$\left|H\left(\dot{p}_m^\varrho, \dot{u}_m^\varrho\right)\right| \le \frac{H\delta_1}{2}\left\|\dot{p}_m^\varrho\right\|_{L_2(\Omega)}^2 + \frac{H}{2\delta_1}\left\|\dot{u}_m^\varrho\right\|_{L_2(\Omega)}^2, \tag{5.152}$$

$$\begin{aligned}\left|3\left(-\bar{p}^2 p_m^\varrho, \dot{p}_m^\varrho\right)\right| &\le 3\left(\int_\Omega \bar{p}^4\left(p_m^\varrho\right)^2\right)^{\frac{1}{2}}\left(\int_\Omega\left(\dot{p}_m^\varrho\right)^2\right)^{\frac{1}{2}} \\ &\le 3\left(\int_\Omega \bar{p}^6\right)^{\frac{1}{3}}\left(\int_\Omega\left(p_m^\varrho\right)^6\right)^{\frac{1}{6}}\left(\int_\Omega\left(\dot{p}_m^\varrho\right)^2\right)^{\frac{1}{2}} \\ &= 3\left\|\bar{p}\right\|_{L_6(\Omega)}^2\left\|p_m^\varrho\right\|_{L_6(\Omega)}\left\|\dot{p}_m^\varrho\right\|_{L_2(\Omega)} \\ &\le \frac{3\delta_2}{2}\left\|\dot{p}_m^\varrho\right\|_{L_2(\Omega)}^2 + \frac{3}{2\delta_2}\left\|\bar{p}\right\|_{L_6(\Omega)}^4\left\|p_m^\varrho\right\|_{L_6(\Omega)}^2,\end{aligned}$$

$$\begin{aligned}\left|3\left(\bar{p}p_m^\varrho, \dot{p}_m^\varrho\right)\right| &\le 3\left(\int_\Omega \bar{p}^4\right)^{\frac{1}{4}}\left(\int_\Omega\left(p_m^\varrho\right)^4\right)^{\frac{1}{4}}\left(\int_\Omega\left(\dot{p}_m^\varrho\right)^2\right)^{\frac{1}{2}} \\ &\le \frac{3\delta_3}{2}\left\|\dot{p}_m^\varrho\right\|_{L_2(\Omega)}^2 + \frac{3}{2\delta_3}\left\|\bar{p}\right\|_{L_4(\Omega)}\left\|p_m^\varrho\right\|_{L_4(\Omega)},\end{aligned}$$

$$\left|-\frac{1}{2}\left(p_m^\varrho, \dot{p}_m^\varrho\right)\right| \le \frac{\delta_4}{4}\left\|p_m^\varrho\right\|_{L_2(\Omega)}^2 + \frac{1}{4\delta_4}\left\|\dot{p}_m^\varrho\right\|_{L_2(\Omega)}^2,$$

$$\left|-b\beta\xi\left(u_m^\varrho, \dot{p}_m^\varrho\right)\right| \le \frac{\delta_5 b\beta\xi}{2}\left\|\dot{p}_m^\varrho\right\|_{L_2(\Omega)}^2 + \frac{b\beta\xi}{2\delta_5}\left\|u_m^\varrho\right\|_{L_2(\Omega)}^2,$$

$$\left|\left(r_{f_0}\left(\tilde{p}, \bar{p}\right), \dot{p}_m^\varrho\right)\right| \le \frac{\delta_6}{2}\left\|\dot{p}_m^\varrho\right\|_{L_2(\Omega)}^2 + \frac{1}{2\delta_6}\left\|r_{f_0}\left(\tilde{p}, \bar{p}\right)\right\|_{L_2(\Omega)}^2.$$

The constants δ_2, δ_3, δ_4, δ_5 and δ_6 can be set so that (5.151) gives rise to

$$\begin{aligned}\frac{\alpha\xi^2}{2}\left\|\dot{p}_m^\varrho\right\|_{L_2(\Omega)}^2 + \xi^2\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left\|\nabla p_m\right\|_{L_2(\Omega)}^2 &\le c_1\left\|p_m^\varrho\right\|_{L_6(\Omega)}^2 + c_2\left\|p_m^\varrho\right\|_{L_4(\Omega)} \\ &\quad + c_3\left\|p_m^\varrho\right\|_{L_2(\Omega)}^2 + c_4\left\|u_m^\varrho\right\|_{L_2(\Omega)}^2 + c_5\left\|r_{f_0}\left(\tilde{p}, \bar{p}\right)\right\|_{L_2(\Omega)}^2,\end{aligned} \tag{5.153}$$

where $\left\|\bar{p}\left(t\right)\right\|_{L^6(\Omega)}^4$, $\left\|\bar{p}\left(t\right)\right\|_{L^4(\Omega)}$ could be overestimated for any $t \in [0, t]$ since $H_0^1\left(\Omega\right) \hookrightarrow L^q\left(\Omega\right)$ for $q \in [2, 6]$, and $\bar{p} \in L_\infty\left(0, T; H_0^1\left(\Omega\right)\right)$. Setting δ_1 in estimate (5.152) appropriately and applying it to (5.150) yields

$$\frac{1}{2}\left\|\dot{u}_m^\varrho\right\|_{L_2(\Omega)}^2 + \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left\|\nabla u_m^\varrho\right\|_{L_2(\Omega)}^2 \le \frac{H^2}{2}\left\|\dot{p}_m^\varrho\right\|_{L^2(\Omega)}^2. \tag{5.154}$$

Multiplying (5.154) by $\frac{\alpha\xi^2}{2H^2}$ and adding it to (5.153) results in

$$\frac{\alpha\xi^2}{4}\left\|\dot{p}_m^\varrho\right\|_{L_2(\Omega)}^2+\xi^2\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left\|\nabla p_m\right\|_{L^2(\Omega)}^2+\frac{\alpha\xi^2}{4H^2}\left\|\dot{u}_m^\varrho\right\|_{L_2(\Omega)}^2+\frac{\alpha\xi^2}{4H^2}\frac{\mathrm{d}}{\mathrm{d}t}\left\|\nabla u_m^\varrho\right\|_{L_2(\Omega)}^2$$
$$\leq c_1\left\|p_m^\varrho\right\|_{L^6(\Omega)}^2+c_2\left\|p_m^\varrho\right\|_{L^4(\Omega)}+c_3\left\|p_m^\varrho\right\|_{L^2(\Omega)}^2+c_4\left\|u_m^\varrho\right\|_{L^2(\Omega)}^2+c_5\left\|r_{f_0}\left(\tilde{p},\bar{p}\right)\right\|_{L^2(\Omega)}^2. \quad (5.155)$$

Using Poincare's inequality and the continuous embedding $H_0^1(\Omega)\hookrightarrow L^q(\Omega)$ for $q\in[2,6]$ the estimate (5.155) may be rewritten as

$$\frac{\alpha\xi^2}{4}\left\|\dot{p}_m^\varrho\right\|_{L_2(\Omega)}^2+\xi^2\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left\|\nabla p_m^\varrho\right\|_{L_2(\Omega)}^2+\frac{\alpha\xi^2}{4H^2}\left\|\dot{u}_m^\varrho\right\|_{L_2(\Omega)}^2+\frac{\alpha\xi^2}{4H^2}\frac{\mathrm{d}}{\mathrm{d}t}\left\|\nabla u_m^\varrho\right\|_{L_2(\Omega)}^2$$
$$\leq c_6\left\|\nabla p_m^\varrho\right\|_{L_2(\Omega)}^2+c_7\left\|\nabla p_m^\varrho\right\|_{L_2(\Omega)}^2+c_8\left\|\nabla p_m^\varrho\right\|_{L_2(\Omega)}^2+c_9\left\|\nabla u_m^\varrho\right\|_{L_2(\Omega)}^2+c_5\left\|r_{f_0}\left(\tilde{p},\bar{p}\right)\right\|_{L_2(\Omega)}^2. \quad (5.156)$$

Dropping the positive terms $\frac{\alpha\xi^2}{4}\left\|\dot{p}_m^\varrho\right\|_{L_2(\Omega)}^2$ and $\frac{\alpha\xi^2}{4H^2}\left\|\dot{u}_m^\varrho\right\|_{L_2(\Omega)}^2$ from the left hand side and adjusting the constants gives

$$+\frac{\mathrm{d}}{\mathrm{d}t}\left\|\nabla p_m^\varrho\right\|_{L_2(\Omega)}^2+c_{10}\frac{\mathrm{d}}{\mathrm{d}t}\left\|\nabla u_m^\varrho\right\|_{L_2(\Omega)}^2$$
$$\leq c_{11}\left(\left\|\nabla p_m^\varrho\right\|_{L_2(\Omega)}^2+c_{10}\left\|\nabla p_m^\varrho\right\|_{L_2(\Omega)}+\left\|r_{f_0}\left(\tilde{p},\bar{p}\right)\right\|_{L_2(\Omega)}^2\right). \quad (5.157)$$

Applying Grönwall's lemma with the setting

$$\begin{aligned}\varrho(t)&=\left\|p_m^\varrho(t)\right\|_{L_2(\Omega)}^2+c_{10}\left\|\nabla u_m^\varrho(t)\right\|_{L_2(\Omega)}^2\\ \phi(t)&\equiv c_{11},\\ \psi(t)&\equiv c_{11}\left\|r_{f_0}\left(\tilde{p},\bar{p}\right)\right\|_{L_2(\Omega)}^2,\end{aligned} \quad (5.158)$$

gives rise to

$$\left\|p_m^\varrho(t)\right\|_{L_2(\Omega)}^2+c_{10}\left\|\nabla u_m^\varrho(t)\right\|_{L_2(\Omega)}^2$$
$$\leq e^{c_{11}T}\left(\left\|\nabla p_m^\varrho(0)\right\|_{L_2(\Omega)}^2+c_{10}\left\|\nabla u_m^\varrho(0)\right\|_{L_2(\Omega)}^2\right)+e^{c_{11}T}c_{11}\left\|r_{f_0}\left(\tilde{p},\bar{p}\right)\right\|_{L_2(0,T;L_2(\Omega))}^2. \quad (5.159)$$

Taking the initial conditions (5.148)-(5.149) into account yields

$$\left\|p_m^\varrho(t)\right\|_{L_\infty\left(0,T;H_0^1(\Omega)\right)}+c_{10}\left\|u_m^\varrho(t)\right\|_{L_\infty\left(0,T;H_0^1(\Omega)\right)}\leq c_{13}\left\|r_{f_0}\left(\tilde{p},\bar{p}\right)\right\|_{L_2(0,T;L_2(\Omega))}^2. \quad (5.160)$$

To get an estimate for the derivatives, apply Grönwall's lemma to (5.157) again and integrate with respect to t:

$$\int_0^T\varrho(t)\,\mathrm{d}t\leq\left(e^{c_{11}t}-1\right)\left\|r_{f_0}\left(\tilde{p},\bar{p}\right)\right\|_{L_2(0,T;L_2(\Omega))}^2. \quad (5.161)$$

Integrating (5.156) while taking (5.161) into account gives

$$c_{14}\left\|\dot{p}_m^\varrho\right\|_{L_2(0,T;L_2(\Omega))}^2+c_{15}\left\|\dot{u}_m^\varrho\right\|_{L_2(\Omega)}^2+\varrho(T)-\varrho(0)$$
$$\leq c_{16}\int_0^T\varrho(t)\,\mathrm{d}t+e^{c_{11}T}c_{11}\left\|r_{f_0}\left(\tilde{p},\bar{p}\right)\right\|_{L_2(0,T;L_2(\Omega))}^2$$
$$\leq c_{17}\left\|r_{f_0}\left(\tilde{p},\bar{p}\right)\right\|_{L_2(0,T;L_2(\Omega))}^2. \quad (5.162)$$

Combining (5.160) and (5.162) shows that (5.144) holds. The existence of a solution is then proved with the help of arguments analogous to the ones used in the proof of Theorem 13. The uniqueness of the solution is a direct consequence of the linearity of the equations and estimate (5.136). □

Theorem 15. *(Stability estimate) Let $S(\bar{\eta}) = (\bar{u}, \bar{p})$ and $S(\bar{\eta} + \eta) = (\tilde{u}, \tilde{p})$, then the estimate*

$$\left\| \begin{pmatrix} \tilde{u} \\ \tilde{p} \end{pmatrix} - \begin{pmatrix} \bar{u} \\ \bar{p} \end{pmatrix} \right\|_{W\left(0,T;8,2;H_0^1(\Omega),L_2(\Omega)\right)^2} \leq C_2 \left\| \eta \right\|_{H^4(\Omega\times(0,T))}$$

holds.

Proof. Let $(w_k)_{k\in\mathbb{N}}$ be such that (5.45)-(5.46) hold. For the sake of readability, denote $S(\bar{\eta}) = (\bar{u}, \bar{p})$, $S(\bar{\eta} + \eta) = (\tilde{u}, \tilde{p})$ and let (u_1^m, p_1^m) (u_2^m, p_2^m) be the Galerkin approximations that solve the finite dimensional problems corresponding to $(\bar{u}, \bar{p})$ and $(\tilde{u}, \tilde{p})$ respectively. Then taking the difference of the two approximative problems gives rise to

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega u_{12}^m(t)\, w_k + \int_\Omega \nabla u_{12}^m(t) \cdot \nabla w_k + \int_\Omega \nabla \eta(t) \cdot \nabla w_k = H \frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega p_{12}^m(t)\, w_k - \frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega \eta(t)\, w_k$$
$$\text{for all } k = 1, \dots m \text{ a.e. in } [0,T], \tag{5.163}$$

$$\alpha\xi^2 \frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega p_{12}^m(t)\, w_k + \xi^2 \int_\Omega \nabla p_{12}^m(t) \cdot \nabla w_k = \int_\Omega \left(f_0\left(p_1^m(t)\right) - f_0\left(p_2^m(t)\right)\right) w_k \tag{5.164}$$
$$- b\beta\xi \int_\Omega \left(u_1^m(t) + \eta(t)\right) w_k \tag{5.165}$$
$$\text{for all } k = 1, \dots m \text{ a.e. in } [0,T], \tag{5.166}$$

$$\alpha_m^k(0) = \int_\Omega -\eta(0)\, w_k \quad k = 1, \dots, m, \tag{5.167}$$

$$\beta_m^k(0) = 0 \quad k = 1, \dots, m, \tag{5.168}$$

where $u_{12}^m \equiv u_1^m - u_2^m$, $p_{12}^m \equiv p_1^m - p_2^m$ and α_m^k, β_m^k are the coefficients of the Galerkin approximations u_{12}^m, p_{12}^m. Multiplying equations (5.163) and (5.166) by α_m^k, and β_m^k, respectively for each $k = 1, \dots, m$ and adding them up separately yields

$$\left\|\dot{u}_{12}^m\right\|_{L_2(\Omega)}^2 + \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left\|\nabla\left(u_{12}^m + \eta\right)\right\|_{L_2(\Omega)}^2 = B\left[u_{12}^m + \eta, \dot{\eta}\right]$$
$$+ H\left(\dot{p}_{12}^m, \dot{u}_{12}^m\right) - \left(\dot{\eta}, \dot{u}_{12}^m\right), \tag{5.169}$$

$$\alpha\xi^2 \left\|\dot{p}_{12}^m\right\|_{L_2(\Omega)}^2 + \frac{\xi^2}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left\|\nabla p_{12}^m\right\|_{L_2(\Omega)}^2 = \left(f_0\left(p_1^m\right) - f_0\left(p_2^m\right), \dot{p}_{12}^m\right) - b\beta\xi\left(u_{12} + \eta, \dot{p}_{12}\right) \text{ a.e. in } [0,T]. \tag{5.170}$$

The non-linear term can be estimated in the following manner

$$\begin{aligned}
\left\| \frac{f_0\left(p_1^m\right) - f_0\left(p_2^m\right)}{p_1^m - p_2^m} \right\|_{L_3(\Omega)} &\leq \left\| -\frac{1}{2} \right\|_{L_3(\Omega)} + \left\| \frac{\left(p_1^m - p_2^m\right)\left(p_1^m + p_2^m\right)}{p_1^m - p_2^m} \right\|_{L_3(\Omega)} \\
&+ \left\| \frac{\left(p_1^m - p_2^m\right)\left(\left(p_1^m\right)^2 + p_1^m p_2^m + \left(p_2^m\right)^2\right)}{p_1^m - p_2^m} \right\|_{L_3(\Omega)} \\
&\leq c_{18} + \left\|p_1^m\right\|_{L_3(\Omega)} + \left\|p_2^m\right\|_{L_3(\Omega)} \\
&+ \left\|p_1^m\right\|_{L_6(\Omega)}^2 + \left\|p_1^m\right\|_{L_6(\Omega)} \left\|p_2^m\right\|_{L_6(\Omega)} + \left\|p_2^m\right\|_{L_6(\Omega)}^2 \\
&\equiv c\left(p_1^m, p_2^m\right).
\end{aligned}$$

This estimate can be applied to $\left(f_0\left(p_1^m\right) - f_0\left(p_2^m\right), \dot{p}_{12}^m\right)$ in the following way

$$\begin{aligned}|(f_0(p_1^m)-f_0(p_2^m),\dot{p}_{12}^m)| &= \left|\left(\frac{f_0(p_1^m)-f_0(p_2^m)}{p_{12}^m}, p_{12}^m\dot{p}_{12}^m\right)\right| \\ &\leq \frac{1}{2\delta_1}c(p_1^m,p_2^m)^2\|p_{12}^m\|_{L_6(\Omega)}^2+\frac{\delta_1}{2}\|\dot{p}_{12}^m\|_{L_2(\Omega)}^2 \\ &\leq \frac{c_{19}}{2\delta_1}c(p_1^m,p_2^m)^2\|\nabla p_{12}^m\|_{L_2(\Omega)}^2+\frac{\delta_1}{2}\|\dot{p}_{12}^m\|_{L_2(\Omega)}^2, \end{aligned} \tag{5.171}$$

where Hölder's inequality was used twice along with the continuous embedding $H_0^1(\Omega)\hookrightarrow L_6(\Omega)$. If there exists a $M_1>0$ such that

$$\|\bar{\eta}\|_{X_c}, \|\bar{\eta}+\eta\|_{X_c}\leq M_1, \tag{5.172}$$

one arrives at

$$c(p_1^m,p_2^m)^2\leq M_2,$$

due to the estimate (5.77).

The following estimates are made

$$\begin{aligned}|B[u_{12}^m+\eta,\dot{\eta}]| &\leq \frac{1}{2\delta_2}\|\dot{\eta}\|_{H^1(\Omega)}^2+\frac{\delta_2}{2}\|\nabla(u_{12}^m+\eta)\|_{L_2(\Omega)}^2, \\ |H(\dot{p}_{12}^m,\dot{u}_{12}^m)| &\leq \frac{H^2}{2\delta_3}\|\dot{p}_{12}^m\|_{L_2(\Omega)}^2+\frac{\delta_3}{2}\|\dot{u}_{12}^m\|_{L_2(\Omega)}^2, \\ |-(\dot{\eta},\dot{u}_{12}^m)| &\leq \frac{1}{2\delta_4}\|\dot{\eta}\|_{L_2(\Omega)}^2+\frac{\delta_4}{2}\|\dot{u}_{12}^m\|_{L_2(\Omega)}^2, \\ |-b\beta\xi(u_{12}+\eta,\dot{p}_{12})| &\leq \frac{c_{20}\delta_4}{2}\|\nabla(u_{12}^m+\eta)\|_{L_2(\Omega)}^2+\frac{c_{21}\delta_4}{2}\|\dot{p}_{12}\|_{L_2(\Omega)}^2. \end{aligned}$$

Using these estimates along with (5.171) and combining (5.169) with (5.170) gives rise to

$$\begin{aligned}c_{22}\|\dot{u}_{12}^m\|_{L_2(\Omega)}^2+c_{23}\frac{\mathrm{d}}{\mathrm{d}t}\|\nabla(u_{12}^m+\eta)\|_{L_2(\Omega)}^2+c_{24}\|\dot{p}_{12}^m\|_{L_2(\Omega)}^2+c_{25}\frac{\mathrm{d}}{\mathrm{d}t}\|\nabla p_{12}^m\|_{L_2(\Omega)}^2 \\ \leq c_{26}\left(\|\nabla(u_{12}^m+\eta)\|_{L_2(\Omega)}^2+\|\nabla p_{12}^m\|_{L_2(\Omega)}^2+\|\eta\|_{X_c}^2\right). \end{aligned} \tag{5.173}$$

Dropping the terms $c_{22}\|\dot{u}_{12}^m\|_{L_2(\Omega)}^2$, $c_{24}\|\dot{p}_{12}^m\|_{L_2(\Omega)}^2$ and adjusting the constants on the right-hand side gives

$$\begin{aligned}c_{23}\frac{\mathrm{d}}{\mathrm{d}t}\|\nabla(u_{12}^m+\eta)\|_{L_2(\Omega)}^2+c_{25}\frac{\mathrm{d}}{\mathrm{d}t}\|\nabla p_{12}^m\|_{L_2(\Omega)}^2 \\ \leq c_{27}\left(c_{23}\|\nabla(u_{12}^m+\eta)\|_{L_2(\Omega)}^2+c_{25}\|\nabla p_{12}^m\|_{L_2(\Omega)}^2+\|\eta\|_{X_c}^2\right). \end{aligned}$$

Using Grönwall's lemma yields

$$\begin{aligned}c_{23}\|\nabla(u_{12}^m+\eta)(t)\|_{L_2(\Omega)}^2+c_{25}\|\nabla p_{12}^m(t)\|_{L_2(\Omega)}^2 \\ \leq e^{tc_{27}}\left(c_{23}\|\nabla(u_{12}^m+\eta)(0)\|_{L_2(\Omega)}^2+c_{25}\|\nabla p_{12}^m(0)\|_{L_2(\Omega)}^2+t\|\eta\|_{X_c}^2\right). \end{aligned} \tag{5.174}$$

Using the initial conditions and over-estimating t by T gives

$$c_{23}\|\nabla(u_{12}^m+\eta)(t)\|_{L_2(\Omega)}^2+c_{25}\|\nabla p_{12}^m(t)\|_{L_2(\Omega)}^2\leq c_{28}\|\eta\|_{X_c}^2,$$

which shows that

$$\|u_{12}^m\|_{L_\infty\left(0,T;H_0^1(\Omega)\right)} + \|p_{12}^m\|_{L_\infty\left(0,T;H_0^1(\Omega)\right)} \leq c_{29} \|\eta\|_{X_c}.$$

Considering the continuous embedding $L_\infty\left(0,T;H_0^1(\Omega)\right) \hookrightarrow L_8\left(0,T;H_0^1(\Omega)\right)$ and taking the limes inferior of the relationship gives

$$\|\tilde{u}-\bar{u}\|_{L_8\left(0,T;H_0^1(\Omega)\right)} + \|\tilde{p}-\bar{p}\|_{L_8\left(0,T;H_0^1(\Omega)\right)} \leq c_{29} \|\eta\|_{X_c}, \tag{5.175}$$

since $(u_{12}^m, p_{12}^m) \rightharpoonup (\tilde{u}-\bar{u}, \tilde{p}-\bar{p})$ in $L_8\left(0,T;H_0^1(\Omega)\right)^2$.

To complete the proof, the stability estimate for the derivatives is shown. Letting

$$\varrho(t) = c_{23} \|\nabla\left(u_{12}^m+\eta\right)(t)\|_{L_2(\Omega)}^2 + c_{25} \|\nabla p_{12}^m(t)\|_{L_2(\Omega)}^2$$

and integrating (5.174) over $[0,T]$ yields

$$\int_0^T \varrho(t)\,\mathrm{d}t \leq \frac{1}{c_{27}}\left(e^{Tc_{27}}-1\right)\left(\varrho(0) + c_{30}\|\eta\|_{X_c}^2\right). \tag{5.176}$$

Returning to (5.173) and integrating over $[0,T]$ while keeping (5.176) in mind leads to

$$\begin{aligned} c_{22}\|\dot{u}_{12}^m\|_{L_2(0,T;L_2(\Omega))}^2 + c_{24}\|\dot{p}_{12}^m\|_{L_2(0,T;L_2(\Omega))}^2 + \varrho(T) - \varrho(0) \\ \leq \frac{1}{c_{27}}\left(e^{Tc_{27}}-1\right)\varrho(0) + c_{31}\|\eta\|_{X_c}^2. \end{aligned}$$

Adding $\varrho(0)$ to both sides and dropping $\varrho(T)$ since it's positive gives the final estimate for derivatives:

$$\begin{aligned} c_{22}\|\dot{u}_{12}^m\|_{L_2(0,T;L_2(\Omega))}^2 + c_{24}\|\dot{p}_{12}^m\|_{L_2(0,T;L_2(\Omega))}^2 \\ \leq c_{32}\varrho(0) + c_{31}\|\eta\|_{X_c}^2 \leq c_{32}\|\eta\|_{X_c}^2. \end{aligned} \tag{5.177}$$

Using $(\dot{u}_{12}^m, \dot{p}_{12}^m) \rightharpoonup \left(\frac{\mathrm{d}}{\mathrm{d}t}(\tilde{u}-\bar{u}), \frac{\mathrm{d}}{\mathrm{d}t}(\tilde{p}-\bar{p})\right)$ in $L_2\left(0,T;L_2(\Omega)\right)^2$ and taking the limes inferior of (5.177) yields

$$c_{22}\left\|\frac{\mathrm{d}}{\mathrm{d}t}(\tilde{u}-\bar{u})\right\|_{L_2(0,T;L_2(\Omega))}^2 + c_{24}\left\|\frac{\mathrm{d}}{\mathrm{d}t}(\tilde{p}-\bar{p})\right\|_{L_2(0,T;L_2(\Omega))}^2 \leq c_{32}\|\eta\|_{X_c}^2.$$

Combining this estimate with (5.175) shows that

$$\|(\tilde{u},\tilde{p})-(\bar{u},\bar{p})\|_{W\left(0,T;8,2;H_0^1(\Omega),L_2(\Omega)\right)} \leq c_{33}\|\eta\|_{X_c}^2$$

for some $c_{33} > 0$ if (5.172) holds. □

This means that the solution operator is locally Lipschitz continuous on X_c and Lipschitz continuous on any bounded subset of X_c.

5.7. Optimality Conditions

In this section, the optimality conditions for problem (5.21)-(5.28) are provided.

Theorem 16. *(First order optimality conditions) Let $\bar{\eta} \in X_c$ be the optimal control of problem (5.21)-(5.28) and $u(\bar{\eta}) = \bar{u}$, $p(\bar{\eta}) = \bar{p}$ be the solution to (5.22)-(5.28). Then there exist $\lambda_u, \lambda_p \in L_2\left(0,T; H_0^1(\Omega)\right)$ such that*

$$\begin{aligned}
(\lambda_u)_t + \Delta\lambda_u &= b\beta\xi\lambda_p && \text{in } (0,T)\times\Omega,\\
\lambda_u|_{\partial\Omega} &= 0 && \text{on } \partial\Omega\times[0,T],\\
\lambda_u|_{t=T} &= 0 && \text{in } \Omega,\\
\gamma\xi^2(\lambda_p)_t + \xi^2\Delta\lambda_p &= H(\lambda_u)_t + 3\bar{p}^2\lambda_p - 3\bar{p}\lambda_p + \frac{1}{2}\lambda_p && \text{in } (0,T)\times\Omega,\\
\lambda_p|_{\partial\Omega} &= 0 && \text{on } \partial\Omega\times[0,T],\\
\lambda_p|_{t=T} &= \frac{1}{\alpha\xi^2}\left(p_f - \bar{p}|_{t=T}\right) && \text{in } \Omega,
\end{aligned} \tag{5.178}$$

is satisfied in its weak formulation. Furthermore

$$\int_0^T\int_\Omega \frac{\mathrm{d}}{\mathrm{d}t}(\eta-\bar{\eta})\lambda_u \mathrm{d}t - \int_0^T\int_\Omega \Delta(\eta-\bar{\eta})\lambda_u \mathrm{d}t + \int_0^T\int_\Omega b\beta\xi(\eta-\bar{\eta})\lambda_p \mathrm{d}t + \int_\Omega (\eta-\bar{\eta})|_{t=0}\lambda_u|_{t=0}$$
$$+\gamma\int_0^T\int_{\partial\Omega} Tr\ (\bar{\eta})\ Tr\ (\eta)\,\mathrm{d}t \geq 0 \tag{5.179}$$

$$\text{for all } \eta \in W_{ad}. \tag{5.180}$$

Proof. Using Lagrangian formalism, we define

$$L : X_s \times X_c \times L_2\left(0,T; H_0^1(\Omega)^{**}\right) \times \left(L_2(\Omega)^*\right)^2 \to \mathbb{R}$$

as

$$\begin{aligned}
L((u,p),\eta,(\lambda_u,\lambda_p,\lambda_1,\lambda_2)) \equiv{}& \frac{1}{2}\int_\Omega |p(T)-p_{\mathrm{f}}|^2 + \frac{\gamma}{2}\int_0^T\int_{\partial\Omega} |\mathrm{Tr}\ (\eta)|^2\,\mathrm{d}t\\
&+\int_0^T\int_\Omega (u_t - \Delta u - Hp_t + \eta_t - \Delta\eta)\lambda_u \mathrm{d}t\\
&+\int_0^T\int_\Omega \left(\gamma\xi^2 p_t - \xi^2\Delta p - p(1-p)\left(p-\frac{1}{2}\right) + b\beta\xi(u+\eta)\right)\lambda_p \mathrm{d}t\\
&+\int_\Omega (u|_{t=0} - u_{\mathrm{ini}} + \eta|_{t=0})\lambda_1 + \int_\Omega (p|_{t=0} - p_{\mathrm{ini}})\lambda_2,
\end{aligned} \tag{5.181}$$

where the reflexivity of the space $L_2\left(0,T; H_0^1(\Omega)^{**}\right) \cong L_2\left(0,T; H_0^1(\Omega)\right)$ was used as well as the Riesz representation $L_2(\Omega)^* \cong L_2(\Omega)$. Since the solution operator is Fréchet differentiable, it is possible to differentiate $L(\bar{u},\bar{p},\bar{\eta},(\lambda_u,\lambda_p,\lambda_1,\lambda_2))$ with respect to η and get

$$\begin{aligned}
&D_\eta L(u(\bar{\eta}),p(\bar{\eta}),\bar{\eta},(\lambda_u,\lambda_p,\lambda_1,\lambda_2))[\eta]\\
&= L_{(u,p)}(S(\bar{\eta}),\bar{\eta},(\lambda_u,\lambda_p,\lambda_1,\lambda_2))S'(\bar{\eta})\eta + L_\eta(S(\bar{\eta}),\bar{\eta},(\lambda_u,\lambda_p,\lambda_1,\lambda_2))\eta\\
&\qquad \text{for any } \eta \in X_c,
\end{aligned}$$

where D_η denotes the Fréchet derivative with respect to η. Finding $(\lambda_u, \lambda_p, \lambda_1, \lambda_2)$ such that

$$L_{(u,p)}\left(S\left(\bar{\eta}\right), \bar{\eta}, \left(\lambda_u, \lambda_p, \lambda_1, \lambda_2\right)\right) = 0 \tag{5.182}$$

Defining

$$\lambda_u\left|_{t=0}\right. = \lambda_1, \lambda_p\left|_{t=0}\right. - \lambda_u\left|_{t=0}\right. = \lambda_2 \tag{5.183}$$

leads to the adjoint equation 5.178. Using an analogous procedure as in Theorem 14, the existence and uniqueness of the solution for problem (5.178) may be argued. Finally, using the solution of (5.178) while considering (5.183) gives rise to the optimality condition (5.180). □

6. Numerical Optimization of the Dirichlet Boundary Conditions in the Phase Field Problem

The Dirichlet boundary condition optimization in a PFM governing the solidification of a pure substance (2.25)-(2.29) is addressed numerically. The problem in question is identical to the one discussed in Chapter 5 (see (5.10)-(5.17)). Opposed to Chapter 5, the strong formulation is applied. Adjoint techniques are used to ultimately give an efficient method of gradient computation. Gradient descent is then used to iteratively improve an initial guess and eventually arrive at a local minimum, where the first order optimality conditions are satisfied. Two different reaction terms ((2.45) and (3.12)) are considered in the experiments and the advantages of using the newly proposed reaction term (3.12) are discussed. Unlike Chapter 5, the optimization problem is considered to be unconstrained, i.e. $W_{\text{ad}} = X_C$.

6.1. The Numerical Advantage of Using Adjoint Methods

Following the theory and notation listed in Chapter 4, the adjoint and direct methods of gradient computation are compared. Consider a PDE constrained optimization problem of the form (4.2)-(4.3). Assuming strong regularity (Definition 7), the results of Sections 4.4.1 and 4.4.2 can be directly compared from a numerical perspective.

When following the sensitivity approach to gradient computation, we resolve

$$\hat{J}'(\theta)[\vartheta] = J_y(y(\theta),\theta)\left[y'(\theta)\vartheta\right] + J_\theta(y(\theta),\theta)[\vartheta],$$

where $J(y(\theta),\theta) = \hat{J}(\theta)$, by directly computing $y'(\theta)\vartheta$ using the linearized state equation (4.11). Assume that the control space is discretized by $N \in \mathbb{N}$ functions labeled $(\theta_i)_{i\in\{1,\dots,N\}}$ and the control guess is θ_0. Then solving (4.11) requires the solution for each of the pairs $(\theta_0,\theta_i)_{i\in\{1,\dots,N\}}$. The solutions $S(\theta_0+\delta\theta_i)$ and $S(\theta_0)$ are then utilized to get an approximation for $y'(\theta_0)\theta_i$ using

$$y'(\theta_0)\theta_i \approx \frac{S(\theta_0+\delta\theta_i) - S(\theta_0)}{\delta} \tag{6.1}$$

for each $i \in \{1,\dots,N\}$ and some $\delta > 0$. In the case of boundary condition control, which is characteristic by the control space X_C being approximated by a large amount of parameters (basis functions $(\theta_i)_{i\in\{1,\dots,N\}}$), this approach becomes unfeasible. In addition to this, the approximation (6.1) requires a suitable choice of $\delta > 0$, which is a separate issue altogether.

The mentioned issues motivate the use of adjoint techniques. Following the results of Section 4.4.2, we summarize that

$$\hat{J}'(\theta)[\vartheta] = L_u(y(\theta),\theta,\lambda)\vartheta = L_y(y(\theta),\theta,\lambda)y'(\theta)\vartheta + L_\theta(y(\theta),\theta,\lambda)\vartheta. \tag{6.2}$$

Furthermore, by finding a λ_0 that solves the adjoint problem

$$L_y(y(\theta),\theta,\lambda_0) = 0, \tag{6.3}$$

we may reformulate the computation (6.2) as

$$\hat{J}'(\theta)[\vartheta] = L_\theta(y(\theta),\theta,\lambda_0)\vartheta, \tag{6.4}$$

where most notably the direct resolution of the operator $y'(\theta)$ is avoided. The adjoint problem (6.3) typically takes on the form of another PDE system (with initial and boundary conditions). This was seen for the weak form of the problem in Section 5.7. Analogous results for the strong formulation are derived in the following section.

6.2. Adjoint Problem for Optimizing the Solution of the Phase Field Problem

Using strong formalism, the procedure shown in Section 6.1 is applied to the optimization problem (5.10)-(5.17). Let $\Omega \subset \mathbb{R}^n$ be a bounded domain and let $T > 0$. The course of solidification (or melting) of a pure material in Ω is described by the evolution of the phase field p with values between 0 and 1. Denote the solid and liquid subdomains as $\Omega_s(t)$ and $\Omega_l(t)$, respectively, and the phase interface by $\Gamma(t)$ (as defined by (2.2)-(2.4)). As in Chapter 5, the aim is to obtain a Dirichlet boundary condition for the heat equation θ that results in a particular phase-field profile p at a given time T.

More succinctly, consider the problem

$$\min_{\theta} J(p,u,\theta) \equiv \frac{1}{2}\int_{\Omega} |p(T,\boldsymbol{x}) - p_{\mathrm{f}}(\boldsymbol{x})|^2 \,\mathrm{d}\boldsymbol{x} + \frac{\gamma}{2}\int_0^T \int_{\partial\Omega} |\theta(t,\boldsymbol{x})|^2 \,\mathrm{d}S\mathrm{d}t \tag{6.5}$$

s.t.

$$\begin{aligned}
u_t &= \Delta u + H p_t, && \text{in } (0,T)\times\Omega, && (6.6)\\
u|_{\partial\Omega} &= \theta && \text{on } [0,T]\times\partial\Omega, && (6.7)\\
u|_{t=0} &= u_{\mathrm{ini}} && \text{in } \Omega, && (6.8)\\
\alpha\xi^2 p_t &= \xi^2\Delta p + f(u,p;\xi) && \text{in } (0,T)\times\Omega, && (6.9)\\
p|_{\partial\Omega} &= p_{\mathrm{bc}} && \text{on } [0,T]\times\partial\Omega, && (6.10)\\
p|_{t=0} &= p_{\mathrm{ini}} && \text{in } \Omega, && (6.11)
\end{aligned}$$

where $p_{\mathrm{f}} \in \mathrm{L}^2(\Omega)$ in (6.5) is the phase field target profile, γ denotes the strength of the regularization and

$$f(u,p;\xi) = p(1-p)\left(p - \frac{1}{2}\right) - \beta\xi(u - u^*). \tag{6.12}$$

For the sake of clarity, only the reaction term (6.12) is used in the computations that follow. The final result for the derivation process when f is equal to (3.12) is mentioned at the end of the section. The model parameters α, β, ξ and H are described in Section 2.2.3.

As detailed in Chapter 3 and [66], the term f is subject to a condition that ensures proper formation of the phase interface. This condition reads

$$\beta\xi(u - u^*) \in \left(-\frac{\sqrt{3}}{36}, \frac{\sqrt{3}}{36}\right). \tag{6.13}$$

When (6.13) is violated, the results given by (6.6)-(6.11) lose their physical interpretation. In Section 7, both results that adhere to (6.13) and the ones that violate this bound are discussed. Note that the alternative reaction term (3.12) is not subject to a restriction of this kind, which gives it a profound advantage over (6.12).

The adjoint problem is derived as follows. Consider a setting, which corresponds to the strong form of the problem:

- $X_S \equiv \mathrm{C}^2\left([0,T]\times\overline{\Omega}\right)^2$,
- $X_C \equiv \mathrm{C}(\partial\Omega\times[0,T])$,

- $Z \equiv \left(\mathrm{L}^2\left([0,T]\times\overline{\Omega}\right)\times\mathrm{L}^2\left(\Omega\right)\times\mathrm{L}^2\left(\partial\Omega\times[0,T]\right)\right)^2$.

Define the state equation operator component-wise as

$$\begin{aligned}
e^1\left(u,p\right) =& u_t-\Delta u-Hp_t,\\
e^2\left(u,\theta\right) =& \theta-u\left|_{\partial\Omega}\right.,\\
e^3\left(u\right) =& u_{\mathrm{ini}}-u\left|_{t=0}\right.,\\
e^4\left(u,p\right) =& \alpha\xi^2p_t-\xi^2\Delta p-p\left(1-p\right)\left(p-\frac{1}{2}\right)+\beta\xi\left(u-u^*\right),\\
e^5\left(p\right) =& p_{\mathrm{bc}}-p\left|_{\partial\Omega}\right.,\\
e^6\left(p\right) =& p_{\mathrm{ini}}-p\left|_{t=0}\right..
\end{aligned} \tag{6.14}$$

Let $\lambda\equiv\left(l_1,l_2,l_3,q_1,q_2,q_3\right)\in Z$, then the Lagrangian for the problem (6.5)-(6.11) reads

$$\begin{aligned}
L\left(u,p,\theta,\lambda\right) =& J\left(u,p,\theta\right)+\lambda\left(e\left(u,p,\theta\right)\right)\\
=& \frac{1}{2}\int\limits_{\Omega}\left|p\left|_{t=T}\right.-p_{\mathrm{f}}\right|^2\mathrm{d}\boldsymbol{x}+\frac{\gamma}{2}\int\limits_0^T\int\limits_{\partial\Omega}\left|\theta\right|^2\mathrm{d}S\mathrm{d}t\\
&+\int\limits_0^T\int\limits_{\Omega}\left(u_t-\Delta u-Hp_t\right)l_1\mathrm{d}\boldsymbol{x}\mathrm{d}t+\int\limits_{\Omega}\left(u\left|_{t=0}\right.-u_{\mathrm{ini}}\right)l_2\mathrm{d}\boldsymbol{x}+\int\limits_0^T\int\limits_{\partial\Omega}\left(u\left|_{\partial\Omega}\right.-\theta\right)l_3\mathrm{d}S\mathrm{d}t\\
&+\int\limits_0^T\int\limits_{\Omega}\left(\alpha\xi^2p_t-\xi^2\Delta p-p\left(1-p\right)\left(p-\frac{1}{2}\right)+\beta\xi\left(u-u^*\right)\right)q_1\mathrm{d}\boldsymbol{x}\mathrm{d}t\\
&+\int\limits_{\Omega}\left(p\left|_{t=0}\right.-p_{\mathrm{ini}}\right)q_2\mathrm{d}\boldsymbol{x}+\int\limits_0^T\int\limits_{\partial\Omega}\left(p\left|_{\partial\Omega}\right.-p_{\mathrm{bc}}\right)q_3\mathrm{d}S\mathrm{d}t.
\end{aligned} \tag{6.15}$$

Compare (6.15) to (5.181) and notice that the Lagrangians for the strong and weak formulation differ due to the lift technique being applied and the inclusion of some of the relationships in functional spaces. To find the appropriate λ_0 such that the reduced gradient computation (6.4) can be used, (6.3) is solved. Consider $(\hat{u},\hat{p})\in Y$, then

$$\begin{aligned}
L_{(u,p)}\left(u,p,\theta,\lambda\right)\left[(\hat{u},\hat{p})\right] =& \int\limits_{\Omega}\left(p\left|_{t=T}\right.-p_f\right)\hat{p}\left|_{t=T}\right.\mathrm{d}\boldsymbol{x}+\overbrace{\int\limits_0^T\int\limits_{\Omega}\left(\hat{u}_t-\Delta\hat{u}-H\hat{p}_t\right)l_1\mathrm{d}\boldsymbol{x}\mathrm{d}t}^{I.} \qquad (6.16)\\
&+\int\limits_{\Omega}\hat{u}\left|_{t=0}\right.l_2\mathrm{d}\boldsymbol{x}+\int\limits_0^T\int\limits_{\partial\Omega}\hat{u}\left|_{\partial\Omega}\right.l_3\mathrm{d}S\mathrm{d}t\\
&+\overbrace{\int\limits_0^T\int\limits_{\Omega}\left(\alpha\xi^2\hat{p}_t-\xi^2PFM\Delta\hat{p}+3p^2\hat{p}-3p\hat{p}+\frac{1}{2}\hat{p}+\beta\xi\hat{u}\right)q_1\mathrm{d}\boldsymbol{x}\mathrm{d}t}^{II.}\\
&+\int\limits_{\Omega}\hat{p}\left|_{t=0}\right.q_2\mathrm{d}\boldsymbol{x}+\int\limits_0^T\int\limits_{\partial\Omega}\hat{p}\left|_{\partial\Omega}\right.q_3\mathrm{d}S\mathrm{d}t. \qquad (6.17)
\end{aligned}$$

Condition (4.16) is satisfied if and only if

$$L_{(u,p)}\left(u,p,\theta,\lambda\right)\left[\hat{u},\hat{p}\right]=0 \text{ for all } (\hat{u},\hat{p})\in Y.$$

To this end, we use Green's formula to offload the derivatives in expressions $I.$ and $II.$ of (6.17). Expression $I.$ becomes

$$\begin{aligned}
I.=&-\int_0^T\int_\Omega \left(l_1\right)_t\hat{u}\mathrm{d}\boldsymbol{x}\mathrm{d}t+\int_\Omega \left(l_1\hat{u}\right)|_{t=T}-\left(l_1\hat{u}\right)|_{t=0}\,\mathrm{d}\boldsymbol{x}+\int_0^T\int_\Omega \nabla l_1\cdot\nabla\hat{u}\mathrm{d}\boldsymbol{x}\mathrm{d}t\\
&-\int_0^T\int_{\partial\Omega} l_1\nabla\hat{u}\cdot\boldsymbol{n}\mathrm{d}S\mathrm{d}t+\int_0^T\int_\Omega H\left(l_1\right)_t\hat{p}\mathrm{d}\boldsymbol{x}\mathrm{d}t+\int_\Omega -H\left(l_1\hat{p}\right)|_{t=T}+H\left(l_1\hat{p}\right)|_{t=0}\,\mathrm{d}\boldsymbol{x}\\
=&-\int_0^T\int_\Omega \left(l_1\right)_t\hat{u}\mathrm{d}\boldsymbol{x}\mathrm{d}t+\int_\Omega \left(l_1\hat{u}\right)|_{t=T}-\left(l_1\hat{u}\right)|_{t=0}\,\mathrm{d}\boldsymbol{x}-\int_0^T\int_\Omega \Delta l_1\hat{u}\mathrm{d}\boldsymbol{x}\mathrm{d}t+\int_0^T\int_{\partial\Omega}\hat{u}\nabla l_1\cdot\boldsymbol{n}\mathrm{d}S\mathrm{d}t\\
&-\int_0^T\int_{\partial\Omega} l_1\nabla\hat{u}\cdot\boldsymbol{n}\mathrm{d}S\mathrm{d}t+\int_0^T\int_\Omega H\left(l_1\right)_t\hat{p}\mathrm{d}\boldsymbol{x}\mathrm{d}t+\int_\Omega -H\left(l_1\hat{p}\right)|_{t=T}+H\left(l_1\hat{p}\right)|_{t=0}\,\mathrm{d}\boldsymbol{x}.
\end{aligned}\tag{6.18}$$

Analogously, expression $II.$ can be rewritten as

$$\begin{aligned}
II.=&-\int_0^T\int_\Omega \alpha\xi^2\left(q_1\right)_t\hat{p}\mathrm{d}\boldsymbol{x}\mathrm{d}t+\int_\Omega \alpha\xi^2\left(q_1\hat{p}\right)|_{t=T}-\alpha\xi^2\left(q_1\hat{p}\right)|_{t=0}\,\mathrm{d}\boldsymbol{x}\\
&+\xi^2\left[\int_0^T\int_\Omega -\Delta q_1\hat{p}-\int_0^T\int_{\partial\Omega} q_1\nabla\hat{p}\cdot\boldsymbol{n}\mathrm{d}S\mathrm{d}t+\int_0^T\int_{\partial\Omega}\hat{p}\nabla q_1\cdot\boldsymbol{n}\mathrm{d}S\mathrm{d}t\right]\\
&+\int_0^T\int_\Omega\left(3p^2q_1-3pq_1+\frac{1}{2}q_1\right)\hat{p}\mathrm{d}\boldsymbol{x}\mathrm{d}t+\int_0^T\int_\Omega \beta\xi q_1\hat{u}\mathrm{d}\boldsymbol{x}\mathrm{d}t.
\end{aligned}\tag{6.19}$$

Using (6.18), (6.19) along with (6.17) results in

$$
\begin{aligned}
L_{(u,p)}\left(u,p,\theta,\lambda\right)\left[\hat{u},\hat{p}\right] =& \int_{\Omega}\left(p\left|_{t=T}\right. - p_f - Hl_1\left|_{t=T}\right. + \alpha\xi^2 q_1\left|_{t=T}\right.\right)\hat{p}\left|_{t=T}\right.\mathrm{d}\boldsymbol{x} \\
&+ \int_0^T\int_{\Omega}\left(-\left(l_1\right)_t - \Delta l_1 + \beta\xi q_1\right)\hat{u}\mathrm{d}\boldsymbol{x}\mathrm{d}t + \int_{\Omega}\left(l_2 - l_1\right)\left|_{t=0}\right.\hat{u}\left|_{t=0}\right.\mathrm{d}\boldsymbol{x} \\
&+ \int_0^T\int_{\partial\Omega}\hat{u}\left(\nabla l_1\cdot\boldsymbol{n} + l_3\right)\mathrm{d}S\mathrm{d}t - \int_0^T\int_{\partial\Omega} l_1\nabla\hat{u}\cdot\boldsymbol{n}\mathrm{d}S\mathrm{d}t + \int_{\Omega}\left(l_1\hat{u}\right)\left|_{t=T}\right.\mathrm{d}\boldsymbol{x} \\
&+ \int_0^T\int_{\Omega}\left(-\alpha\xi^2\left(q_1\right)_t - \xi^2\Delta q_1 + H\left(l_1\right)_t + 3p^2q_1 - 3pq_1 + \frac{1}{2}q_1\right)\hat{p}\mathrm{d}\boldsymbol{x}\mathrm{d}t \\
&+ \int_{\Omega}\left(Hl_1 - \alpha\xi^2 q_1 + q_2\right)\hat{p}\left|_{t=0}\right.\mathrm{d}\boldsymbol{x} - \int_0^T\int_{\partial\Omega}\xi^2 q_1\nabla\hat{p}\cdot\boldsymbol{n}\mathrm{d}S\mathrm{d}t \\
&+ \int_0^T\int_{\partial\Omega}\left(q_3 + \xi^2\nabla q_1\cdot\boldsymbol{n}\right)\hat{p}\left|_{\partial\Omega}\right.\mathrm{d}S\mathrm{d}t. \qquad (6.20)
\end{aligned}
$$

From (6.20), we see that by providing a l_1 that solves

$$
\begin{aligned}
\left(l_1\right)_t + \Delta l_1 =& \beta\xi q_1 && \text{in } (0,T)\times\Omega, \\
l_1\left|_{\partial\Omega}\right. =& 0 && \text{on } \partial\Omega\times[0,T], \\
l_1\left|_{t=T}\right. =& 0 && \text{in } \Omega \qquad (6.21)
\end{aligned}
$$

and a q_1 that solves

$$
\begin{aligned}
\alpha\xi^2\left(q_1\right)_t + \xi^2\Delta q_1 =& H\left(l_1\right)_t + 3\tilde{y}^2 q_1 - 3\tilde{y}q_1 + \frac{1}{2}q_1 && \text{in } (0,T)\times\Omega, \\
q_1\left|_{\partial\Omega}\right. =& 0 && \text{on } \partial\Omega\times[0,T], \\
q_1\left|_{t=T}\right. =& \frac{1}{\alpha\xi^2}\left(p_{\mathrm{f}} - p\left|_{t=T}\right.\right) && \text{in } \Omega \qquad (6.22)
\end{aligned}
$$

(in the weak sense) causes (6.20) to reduce to

$$
\begin{aligned}
L_{(u,p)}\left(u,p,\theta,\lambda\right)\left[\hat{u},\hat{p}\right] =& \int_{\Omega}\left(l_2 - l_1\right)\left|_{t=0}\right.\hat{u}\left|_{t=0}\right.\mathrm{d}\boldsymbol{x} + \int_0^T\int_{\partial\Omega}\hat{u}\left(\nabla l_1\cdot\boldsymbol{n} + l_3\right)\mathrm{d}S\mathrm{d}t \qquad (6.23) \\
&+ \int_{\Omega}\left(Hl_1 - \alpha\xi^2 q_1 + q_2\right)\hat{p}\left|_{t=0}\right.\mathrm{d}\boldsymbol{x} + \int_0^T\int_{\partial\Omega}\left(q_3 + \xi^2\nabla q_1\cdot\boldsymbol{n}\right)\hat{p}\left|_{\partial\Omega}\right.\mathrm{d}S\mathrm{d}t. \qquad (6.24)
\end{aligned}
$$

Lastly, we set

$$
\begin{aligned}
l_2 =& l_1\left|_{t=0}\right., \\
l_3 =& -\nabla l_1\cdot\boldsymbol{n}\left|_{\partial\Omega}\right., \\
q_2 =& \left(\alpha\xi^2 q_1 - Hl_1\right)\left|_{t=0}\right., \\
q_3 =& -\xi^2\nabla q_1\cdot\boldsymbol{n}\left|_{\partial\Omega}\right. \qquad (6.25)
\end{aligned}
$$

so that (6.24) becomes the zero operator. Consider a backwards time transformation $t \to T - t$ of (6.21) and (6.22).

More succinctly, define

$$\begin{aligned} l(t) =& l_1(T-t), \\ q(t) =& q_1(T-t), \\ z(t) =& p(T-t). \end{aligned} \tag{6.26}$$

Using (6.26) allows us to write

$$\begin{aligned} l_t = \Delta l - \beta\xi q \quad & \text{in } (0,T)\times\Omega, \\ l|_{\partial\Omega} = 0 \quad & \text{on } \partial\Omega\times[0,T], \\ l|_{t=0} = 0 \quad & \text{in } \Omega, \end{aligned} \tag{6.27}$$

$$\begin{aligned} \alpha\xi^2 q_t = \xi^2\Delta q + H l_t - 3z^2 q + 3zq - \frac{1}{2}q \quad & \text{in } (0,T)\times\Omega, \\ q|_{\partial\Omega} = 0 \quad & \text{on } \partial\Omega\times[0,T], \\ q|_{t=0} = \frac{1}{\alpha\xi^2}\left(p_f - p|_{t=T}\right) \quad & \text{in } \Omega, \end{aligned} \tag{6.28}$$

where the equations for l, q resemble the heat equation (6.6) and the Allen-Cahn equation (6.9), respectively. The system of equations (6.27), (6.28) differs substantially from the primary one (6.6)-(6.11) due to the reaction term of the form $-3z^2q + 3zq - \frac{1}{2}q$ in (6.28) and the source term $-\beta\xi q$ in (6.27). Once the system (6.27)-(6.28) is solved and the additional adjoint variables are set according to (6.25), the Fréchet derivative of $\hat{J}$ can be evaluated using (6.4). In particular, let $\vartheta \in U$ be a functional variation, then the formal Fréchet derivative at point $\theta \in U$ in direction s reads

$$\hat{J}'(\theta)[\vartheta] = \gamma \int_0^T \int_{\partial\Omega} \theta\vartheta \mathrm{d}S\mathrm{d}t - \int_0^T \int_{\partial\Omega} l_3\vartheta \mathrm{d}S\mathrm{d}t. \tag{6.29}$$

Some of the experiments in Section 7 show that considering only the linear reaction term (6.12) is insufficient and may lead to non-physical results ((6.13) is violated). For this reason, the alternative reaction (3.12) proposed in Section 3.3 is considered. Assume that $b = 1$, then the reaction term (3.12) reads

$$f(u,p;\xi) = 2p(1-p)\left(p - \frac{1}{2} + \xi\beta\frac{1}{2}\Sigma(p;\varepsilon_0,\varepsilon_1)(u^* - u)\right), \tag{6.30}$$

where $\Sigma(p;\varepsilon_0,\varepsilon_1)$ is a differentiable sigmoid limiter function defined by (3.11). Since (6.30) does not lose its physical interpretation for any value of undercooling (Chapter 3), replacing (6.12) with (6.30) makes it possible to perform numerical experiments for virtually any value of undercooling (see Section 7.2).

The derivation procedure of the adjoint equation is technically identical to the one performed for the linear term (6.12). Because of this, only the final result will be listed. The adjoint system reads

$$\begin{aligned} l_t = \Delta l - \beta\xi q\left(z\Sigma(z;\varepsilon_0,\varepsilon_1) - z^2\Sigma(z;\varepsilon_0,\varepsilon_1)\right) \text{ in } (0,T)\times\Omega, \quad & \text{in } (0,T)\times\Omega, \\ l|_{\partial\Omega} = 0, \quad & \text{on } \partial\Omega\times[0,T], \\ l|_{t=0} = 0, \quad & \text{in } \Omega, \end{aligned} \tag{6.31}$$

$$\begin{aligned} \alpha\xi^2 q_t &= \xi^2 \Delta q + H l_t - 6z^2 q + 6zq - q - q\Xi(z,w) && \text{in } (0,T)\times\Omega, \\ q\,|_{\partial\Omega} &= 0 && \text{on } \partial\Omega \times [0,T], \\ q\,|_{t=0} &= \frac{1}{\alpha\xi^2}\left(p_{\mathrm{f}} - p\,|_{t=T}\right) && \text{in } \Omega, \end{aligned} \tag{6.32}$$

where $w(t) = u(T-t)$,

$$\Xi(z,w) = \beta(w-u^*)\left(\Sigma(z;\varepsilon_0,\varepsilon_1) + z\Sigma'(z;\varepsilon_0,\varepsilon_1) - 2z\Sigma(z;\varepsilon_0,\varepsilon_1) - z^2\Sigma'(z;\varepsilon_0,\varepsilon_1)\right), \tag{6.33}$$

l, q and z are defined as in (6.26) and the auxiliary relationships (6.25) hold. As a result, the gradient is still resolved using (6.29). Notice that Σ' is present in (6.33). This highlights the importance of proposing Σ as a differentiable function (Chapter 3), since lower regularity would lead to the inability to derive the adjoint problem.

6.3. General Numerical Framework

The numerical method that solves (6.5)-(6.11) is described. The First-Optimize-Then-Discretize method is applied [43, 55]. Following this approach, the optimality conditions are first derived and then numerically treated (discretized). The resulting first order optimality conditions are then solved using gradient descent.

6.3.1. Numerical Treatment Using the First-Optimize-Then-Discretize Paradigm

First, the optimality conditions of type (4.17)-(4.19) are stated for an unconstrained optimization problem. Let $W_{\mathrm{ad}} = X_C$, let $(y_0, \theta_0) \in X_S \times X_C$ be the solution to (4.2)-(4.3) and λ_0 be the solution to the respective adjoint problem (4.16). Then the optimality conditions read

$$L_\lambda(y_0,\theta_0,\lambda_0) = e(y_0,\theta_0) = 0, \tag{6.34}$$

$$L_y(y_0,\theta_0,\lambda_0) = e_y(y_0,\theta_0)^*\lambda_0 + J_y(y_0,\theta_0) = 0, \tag{6.35}$$

$$L_\theta(y_0,\theta_0,\lambda_0)\,\vartheta = J_\theta(y_0,\theta_0)\,\vartheta + \langle \lambda_0, e_\theta(y_0,\theta_0)\,\vartheta\rangle_Z \geq 0 \;\forall \vartheta \in X_C. \tag{6.36}$$

The conditions (6.34)-(6.36) state that $(y_0, \theta_0, \lambda_0)$ satisfy the state equation, the adjoint problem, and the necessary condition for a local minimum (compare to (4.17)-(4.19)).

The equations (6.34) and (6.35) are then discretized using a suitable numerical method. For example, supposing that (6.34) and (6.35) is a system of PDEs with the requisite (initial and/or boundary) conditions, the finite difference method (FDM) or the finite element method (FEM) can be applied. The condition (6.36) can then be approximated by suitably interpolating the numerical solutions of (6.34), (6.35).

Even though the following discussion easily generalizes to virtually any PDE-constrained problem, we focus on the problem at hand for the sake of clarity. More succinctly, suppose that (6.34) is given by (6.6)-(6.11), the adjoint problem (6.35) assumes the form (6.27)-(6.28) (or (6.31)-(6.32)).

Both of these problems are solved numerically using the FDM. The spatial mesh is uniform and the time step, denoted Δt fixed. Let the numerical solution of the state equation y^h, the control θ^h, and the adjoint variable λ^h all be mesh functions. A mesh function is only defined on a discrete set (or subset) of mesh points. Let X_S^h, X_C^h, and Z^h be the discrete solution, control and adjoint variable spaces respectively. With this in mind, the discrete version of (6.34)-(6.36) can then be solved as follows:

1. Start with an initial guess $\theta_h := \theta_{h,0} \in X_C^h$.
2. Use the numerical solver for the primary problem (FDM) to compute $y^h = S\left(\theta^h\right) \in X_S^h$.
3. Run the adjoint problem solver (FDM) to get $\lambda^h\left(y^h\right) \in Z^h$.

4. Compute all components of the gradient $\nabla \hat{J}^h$. This is the discrete analogue of the Fréchet derivative of $\hat{J}$ and involves computing the variation with respect to all the basis vectors of X_C^h. Note that this step uses the adjoint variable λ^h obtained in the previous step.

5. Perform one step of gradient descent to update the control θ^h using the rule $\theta^h := \theta^h - \varepsilon \nabla \hat{J}^h$, where $\varepsilon > 0$ is a given step size.

6. Go to step 2 unless a suitable stop criterion is satisfied.

The stop criteria applied can vary. One option is, for instance, to run the described algorithm a fixed number of times, another is to stop it when the cost function J or $\left|\nabla \hat{J}_h\right|$ falls below a certain threshold [43].

6.3.2. Details of the Numerical Method

In this section, a detailed numerical treatment of the primary problem (6.6)-(6.11), the adjoint problem (6.27)-(6.28) (or (6.31)-(6.32)) and gradient computation (6.29) is laid out.

Section 7 addresses simulations in one or two spatial dimensions. For this reason, we assume two spatial dimensions since the results for a one dimensional spatial domain are easy to obtain from the following.

Let $\Omega = (0, L_{x_1}) \times (0, L_{x_2})$ and $T > 0$. Both the adjoint and primary problems are solved using the finite difference method on a uniform spatial mesh. The time step and the spatial mesh resolution are

$$\Delta t = \frac{T}{N_t - 1},\ \Delta x_1 = \frac{L_{x_1}}{N_{x_1} - 1},\ \Delta x_2 = \frac{L_{x_2}}{N_{x_2} - 1}$$

where N_t denotes the number of time layers and N_{x_1}, N_{x_2} the number of mesh points in the x_1 and x_2 directions respectively. The explicit Euler scheme is used for time stepping. Let f be a real-valued function defined on $(0, T) \times \Omega$ and f^h be its approximation (f^h is defined using functional evaluation of f at points of the mesh) by the respective mesh function. Then the time derivative replacement of f reads

$$f_t\left(k\Delta t, i\Delta x_1, j\Delta x_2\right) \approx f_t^h\left(k\Delta t, i\Delta x_1, j\Delta x_2\right) \equiv \frac{f^h\left(\left(k+1\right)\Delta t, i\Delta x_1, j\Delta x_2\right) - f^h\left(k\Delta t, i\Delta x_1, j\Delta x_2\right)}{\Delta t}. \tag{6.37}$$

The Laplacian of f is replaced using the central difference quotient

$$\Delta f\left(k\Delta t, i\Delta x_1, j\Delta x_2\right) \approx \Delta_h f^h\left(k\Delta t, i\Delta x_1, j\Delta x_2\right) \equiv \tag{6.38}$$

$$\frac{f^h\left(k\Delta t, \left(i+1\right)\Delta x_1, j\Delta x_2\right) - 2f^h\left(k\Delta t, i\Delta x_1, j\Delta x_2\right) + f^h\left(k\Delta t, \left(i-1\right)\Delta x_1, j\Delta x_2\right)}{\left(\Delta x_1\right)^2} \tag{6.39}$$

$$+\frac{f^h\left(k\Delta t, i\Delta x_1, \left(j+1\right)\Delta x_2\right) - 2f^h\left(k\Delta t, i\Delta x_1, j\Delta x_2\right) + f^h\left(k\Delta t, i\Delta x_1, \left(j-1\right)\Delta x_2\right)}{\left(\Delta x_2\right)^2}. \tag{6.40}$$

The integral in (6.29) is evaluated using a piece-wise constant interpolation as follows. Let $\mathit{\Pi}$ be a subset of the time-space mesh representing the boundary ($[0, T] \times \partial\Omega$) and let $f^h : \mathit{\Pi} \to \mathbb{R}$, $\theta^h : \mathit{\Pi} \to \mathbb{R}$ be mesh functions. To increase the readability of the following text, the subscript h is dropped for both f^h and θ^h. Let $S_{\mathit{\Pi}}$ be an interpolation operator that maps a mesh function on $\mathit{\Pi}$ to a piecewise constant function on $[0, T] \times \partial\Omega$. More precisely, the definition of $S_{\mathit{\Pi}} f$ on the subset of $[0, T] \times \partial\Omega$, where $x_2 = 0$ reads

$$S_{\mathit{\Pi}} f\left(t, x_1, 0\right) = f_{k,i,0} \text{ for } \left(t, x_1, 0\right) \in M_{k,i,0},, \tag{6.41}$$

where

$$M_{k,i,0} = \left(\max\left(0,(k-1)\,\Delta t + \frac{\Delta t}{2}\right), \min\left(T, k\Delta t + \frac{\Delta t}{2}\right)\right) \times \left(\max\left(0,(i-1)\,\Delta x_1 + \frac{\Delta x_1}{2}\right), \min\left(L_{x_1}, i\Delta x_1 + \frac{\Delta x_1}{2}\right)\right) \times \{0\}, \tag{6.42}$$

k denotes the time step and i represents the x_1 coordinate of the spatial mesh. The set $M_{k,i,0}$ can be described as a rectangle in $(0,T)\times\partial\Omega$ centered at the point $(k\Delta t, i\Delta x_1, 0)$. Note that the rectangle $M_{k,i,0}$ is "cut off" in the extremal cases $k = 0, N_t$ and $i = 0, N_{x_1}$. Using (6.41)-(6.42), the definition for S_{Π} can be easily obtained for the other edges. The numerical scheme implied by (6.37)-(6.38) does not use the points in the corners of the spatial domain Ω. This motivates leaving them out of the approximation of the integral (6.29), giving rise to

$$\begin{aligned}\int_0^T \int_{\partial\Omega} \left(S_{\Pi} f\right)\left(S_{\Pi} s\right) \mathrm{d}S\mathrm{d}t \approx & \sum_{k=0}^{N_t-1}\sum_{i=1}^{N_{x_1}-1} \left[f_{k,i,0}s_{k,i,0} + f_{k,i,N_{x_2}}s_{k,i,N_{x_2}}\right]\Delta t\Delta x_1 \\ & + \sum_{k=0}^{N_t-1}\sum_{j=1}^{N_{x_2}-1} \left[f_{k,0,j}s_{k,0,j} + f_{k,N_{x_1},j}s_{k,N_{x_1},j}\right]\Delta t\Delta x_2,\end{aligned} \tag{6.43}$$

where j represents the x_2 coordinate position in the mesh and $\gamma = 0$ is chosen for simplicity. Notice that omitting the corners of the spatial domain does not affect the convergence of (6.43) as $N_t, N_{x_1}, N_{x_2} \to +\infty$.

The approximation (6.43) and the finite difference approximation l_3^h of l_3 in (6.25) are used to provide the following computation rule for the k,i,j-th component of the gradient

$$\int_0^T \int_{\partial\Omega} \left(S_{\Pi} l_{3,h}\right)\left(S_{\Pi} e_{k,i,j}\right) \mathrm{d}S\mathrm{d}t = \begin{cases} \left(l_{3,h}\right)_{k,i,j}\Delta t\Delta_{x_1} & \text{if } j = \{1, N_{x_2}-1\} \\ \left(l_{3,h}\right)_{k,i,j}\Delta t\Delta x_2 & \text{if } i = \{1, N_{x_1}-1\}, \end{cases} \tag{6.44}$$

where $e_{k,i,j}$ is the characteristic function of $M_{k,i,j}$. The adherence of l to the homogeneous Dirichlet boundary condition (see (6.27)) implies that l_3^h in gradient computation (6.44) reduces to the interior values of l_1 (see (6.26), (6.25)) nearest to the boundary $\partial\Omega$.

7. Numerical Results

In the following, the problem (6.5)-(6.11) is solved numerically in one and two spatial dimensions. To this end, the general numerical framework described in Section 6.3 is used. The simulations in one spatial dimension make use only of the linear reaction term (6.12) (Sections 7.1.1-7.1.3). In these simulations, the effects of regularization, changes in final time and different initial guesses are discussed. Additionally, some of there simulations violate the bound (6.13), which results in a non-physical formation of the interface. When this happens, we call a simulation and the associated optimal control non-realistic.

As anticipated, two dimensional simulations featuring the linear reaction term (6.12) can become non-realistic also. A direct head to head comparison between the behaviors obtained using the linear reaction term (6.12) and the term ΣP1-P (6.30) is described in Section 7.2.1. In these experiments, a crystal is moved from "north" to "south". This pair of simulations lets us fully appreciate the effect of the reaction term (6.30). Lastly, the experiment of Section 7.2.2 shows how the term (6.30) may be used to find an optimal control that separates a solid body in two.

In order to comment on the experiments, recall the definitions (2.2)-(2.4) and the related terminology of Section 6.2. The terms "crystal" and "solid subdomain" are used interchangeably to refer to $\Omega_{\mathrm{s}}(t)$. When $\xi > 0$ in (6.5)-(6.11) is small enough, the value of p inside $\Omega_{\mathrm{s}}(t)$ and $\Omega_{\mathrm{l}}(t)$ is virtually 1 and 0, respectively [66, 16, 22], excluding a thin transition layer between the phases centered around $\Gamma(t)$. For both of the models considered, this transition layer attains a characteristic hyperbolic tangent profile as described in Section 3.2. Additionally, denote the numerical approximation of u as u^h, the numerical approximation of p as p^h and let $P^h : X_S \to X_S^h$ be the projection operator onto the space time mesh.

7.1. Dirichlet Boundary Condition Control for the Phase Field Problem in 1D

The physical parameters for all the simulations performed in one spatial dimension are set according to the values in Table 7.1. The settings do not correspond to any particular material necessarily. Considering the reaction term f defined by (6.12), the influence of regularization γ, changes in final time T and initial guess for the control are discussed.

Table 7.1.: Parameter settings for the phase field simulations in Section 7.1.

Param.	Value	Physical Meaning
α	1	coefficient of attachment kinetics
β	2	dimensionless representation of supercooling
ξ	0.005	interface thickness scaling
u^*	0.5	melting temperature
H	1	latent heat
L_x	1	spatial dimension in the x direction

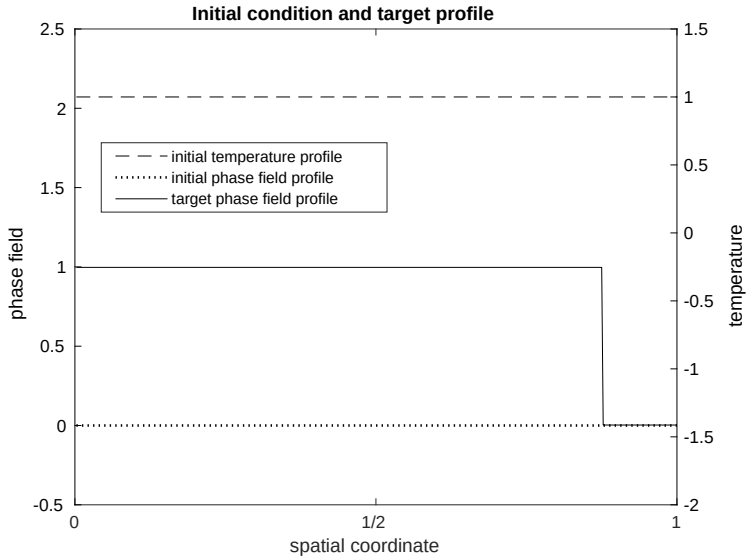

Figure 7.1.: The initial temperature and phase field spatial profiles $u_{\text{ini}}, p_{\text{ini}}$ along with the target profile p_{f} for experiments 1, 2, 3. The values of the boundary condition p_{bc} are given by (7.1).

7.1.1. Controlling the Extent of Crystal Growth

A control that produces a crystal of prescribed length, described by the final phase field profile p_{f}, inside the spatial domain Ω at the fixed final time $T > 0$ is found. The phase field boundary condition is set to

$$\begin{aligned} p_{\text{bc}}(t,0) &= 1 && \forall t \in [0,T)\,, \\ p_{\text{bc}}(t,1) &= 0 && \forall t \in [0,T)\,. \end{aligned} \tag{7.1}$$

The initial conditions $u_{\text{ini}}, p_{\text{ini}}$ and the target profile p_{f} are depicted in Figure 7.1. The initial setup given by these can be intuitively viewed as the domain Ω having a single nucleation site at $x = 0$. In this case, p_{f} is merely the characteristic function of the target crystal shape. Alternatively, p_{f} could be chosen as a continuous function which includes the characteristic shape of the interface given by (3.4). The simulations in Sections 7.1.2 and 7.1.3 make use of this characteristic profile in the definition of p_{f}. The initial guess for the control is

$$\begin{aligned} \theta_0(t,0) &= 0 && \forall t \in [0,T)\,, \\ \theta_0(t,1) &= 1 && \forall t \in [0,T)\,. \end{aligned} \tag{7.2}$$

The three following numerical experiments adhere to the aforementioned setting. The parameters including the spatial mesh resolution, the number of time steps as well as the difference of the final solution from the prescribed phase field profile are summarized in Table 7.2.

The resulting temperature u^h and phase field p^h spatial profiles at final time T are depicted in Figure 7.2. Figure 7.3 shows the respective temporal control profiles of the Dirichlet boundary condition.

Two final times are considered. In experiment 1, the final time T is set to 0.1. It is observed that the values of u^h do not violate the bound (6.13) and the phase field p^h at final time T matches the prescribed profile $P^h p_{\text{f}}$. Furthermore, adding regularization has no effect on the solution, which was investigated by incrementally adding regularization γ up to the value of $\gamma = 10^{-7}$.

In experiments 2 and 3 the final time T is halved to 0.05. This shortened time frame "forces" the optimized control to use higher temperatures and the bound (6.13) is violated, causing the experiment

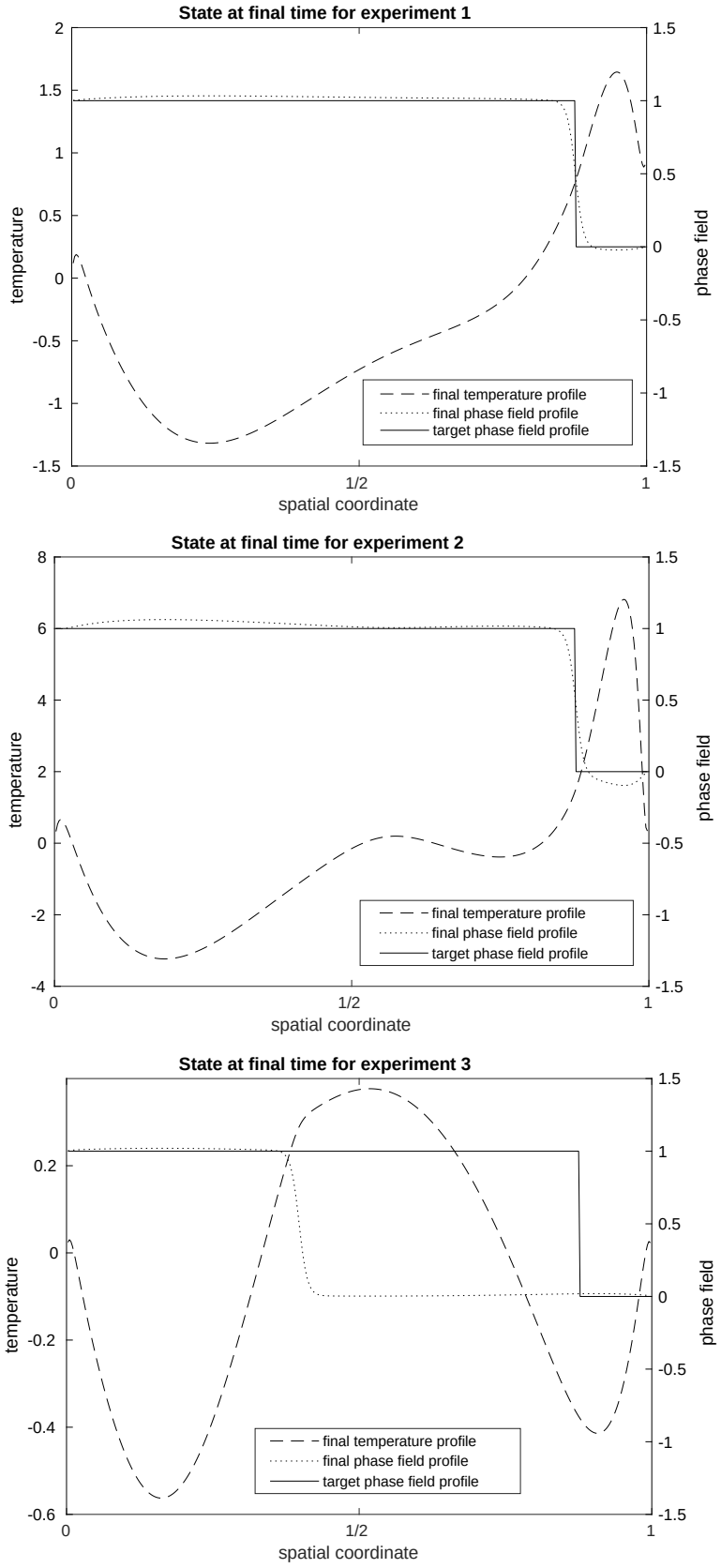

Figure 7.2.: Final temperature and phase field spatial profiles of experiments 1, 2, 3. We observe that p_h reaches the target $P^h p_\mathrm{f}$ in experiments 1 and 2. In experiment 3, the interface of p_h does not reach its target $P^h p_\mathrm{f}$ because sufficient regularization is added to prevent non-realistic behavior.

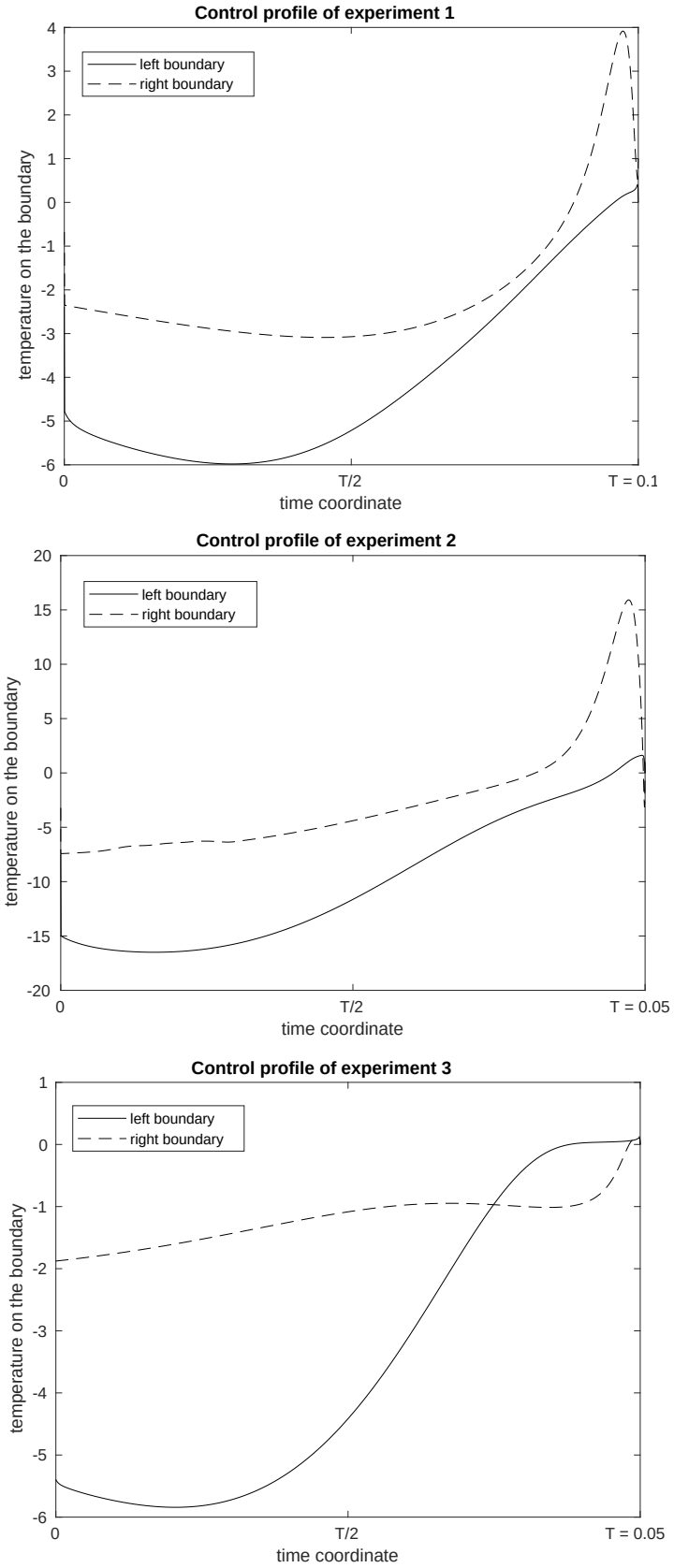

Figure 7.3.: Optimized temporal control profiles of experiments 1, 2, 3. In experiment 2, the values of the temperature violate the bound (6.13) due to the short final time T. The regularization applied in experiment 3 fixes this issue at the cost of not attaining the target profile (see Figure 7.2).

Table 7.2.: Settings for experiments 1, 2, 3 and the respective values of the difference (error) from the prescribed profile.

Parameter	Simulation number 1	2	3
number of time steps N_t	$4 \cdot 10^5$	$4 \cdot 10^5$	$4 \cdot 10^5$
number of grid points N_x	400	400	400
initial control given by	(7.2)	(7.2)	(7.2)
final time T	0.1	0.05	0.05
regularization parameter γ	0	0	$5 \cdot 10^{-11}$
gradient descent step size ε	$3 \cdot 10^{15}$	$2 \cdot 10^{16}$	$2 \cdot 10^{16}$
number of iterations	100	100	100
$\left\|p^h - P^h p_f\right\|_2$ at $t = T$	1.117846	1.276584	13.64326

to be non-physical. The settings of experiment 3 mirror those of experiment 2, with the exception of added regularization $\gamma = 5 \cdot 10^{-11}$. This is sufficient to prevent the non-physicality of the experiment at the cost of not fully achieving the prescribed profile p_f.

This first set of experiments shows that the linear term (6.12) is inadequate, when the control θ^h is likely to induce temperatures that cause rapid solidification. The use of a more advanced model, that is not limited in this way is discussed in Sections 7.2.1 and 7.2.2.

7.1.2. Keeping Crystal Separation

Different initial guesses have an effect on the result of numerical minimization. A handful of examples, in which, among other things, different initial guesses produce different results are shown. This is because the numerical minimization solves the necessary conditions for a local minimum (6.34)-(6.36) and so global minima might not be found using this technique. Instead, the initial guesses can be manipulated in an effort to get a different or more suitable local minimum.

The setting for all of these is as follows. Initially (at $t = 0$), two symmetrically placed crystals in the interior of the domain are separated by a gap. The final profile p_f demands that they grow towards the edges of the domain, but not towards each other. The target profile p_f and initial conditions u_ini, p_ini are depicted in Figure 7.4. In contrast with Section 7.1.1, p_f is constructed as a continuous function with the characteristic shape (3.4) across the interface. The boundary condition for p reads

$$p_\mathrm{bc}(t, x) = 1 \qquad \forall x \in \{0, 1\}, \ \forall t \in [0, T). \tag{7.3}$$

The first four experiments (labeled 4-7) are performed with the non-symmetric initial control guess (7.2) and two different final times $T = 0.05$, $T = 0.4$. Regularization $\gamma > 0$ is applied in experiments 6 and 7. The last experiment (labeled 8) makes use of the symmetric control guess

$$\theta_0(x, t) = 0 \qquad \forall x \in \{0, 1\}, \ \forall t \in [0, T). \tag{7.4}$$

The settings for all of these experiments can be reviewed in Table 7.3.

The value of $\left\|p^h - P^h p_\mathrm{f}\right\|_2$ (see Table 7.3) is small for all the experiments as the target profile p_f is achieved (see Figure 7.5). In addition to this, all the experiments in this section are physical since u is kept within the admissible bounds given by (6.13).

Lastly, some key differences between the experiments and observations are listed (see Figure 7.6).

- Experiments 4 and 5 demonstrate how the choice of different final times T ($T = 0.05$ and $T = 0.4$) affects the optimized control θ^h when no regularization is applied ($\gamma = 0$). Comparing the control

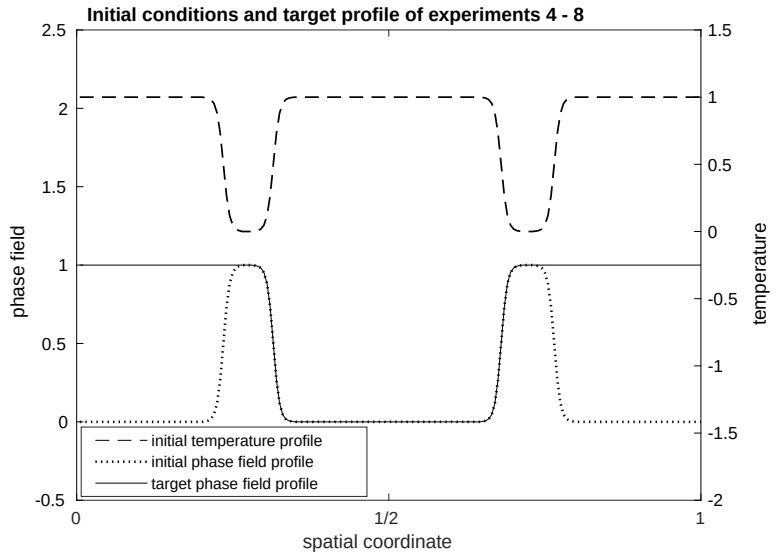

Figure 7.4.: The initial temperature and phase field profiles $p_{\text{ini}}, u_{\text{ini}}$ along with the target profile p_{f} for experiments 4 through 8. The values of the boundary condition p_{bc} are given by (7.3).

Table 7.3.: The settings for experiments 4 through 8 and the respective values of the difference (error) from the prescribed profile.

	Simulation Number				
Parameter	4	5	6	7	8
number of time steps N_t	10^5	10^5	10^5	10^5	10^5
number of grid points N_x	200	200	200	200	200
initial control given by	(7.2)	(7.2)	(7.2)	(7.2)	(7.4)
T (final time)	0.05	0.4	0.4	0.4	0.4
regularization parameter γ	0	0	$5 \cdot 10^{-10}$	10^{-9}	0
gradient descent step size ε	$2 \cdot 10^{14}$	$3 \cdot 10^{13}$	10^{13}	10^{13}	$3 \cdot 10^{14}$
number of iterations	150	100	100	125	100
$\left\|p^h - P^h p_{\text{f}}\right\|_2$	0.3359370	0.3392653	0.4725518	0.8469198	0.3293709

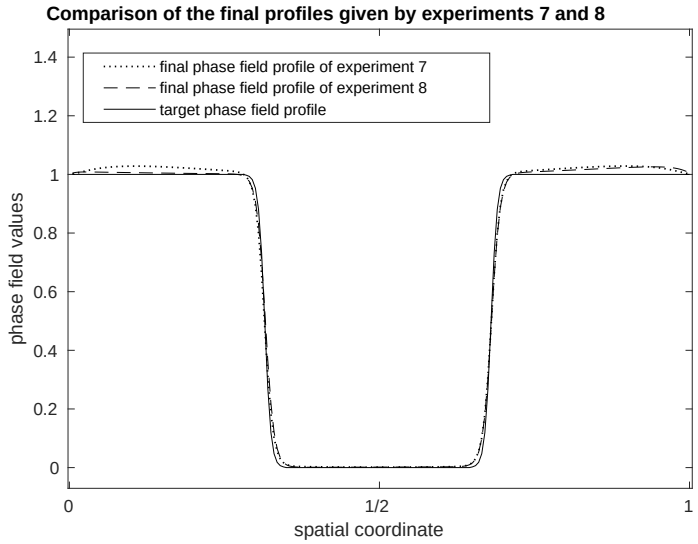

Figure 7.5.: Comparison of the best and the worst (in terms of the error $\left\|p^h - P^h p_{\text{f}}\right\|_2$) phase field profile estimations for experiments 4 through 8.

Table 7.4.: Settings for experiment 9 and the respective value of the difference (error) from the prescribed profile.

Parameter	Value
number of time steps N_t	10^5
number of grid points N_x	200
initial control given by	(7.4)
T (final time)	0.1
regularization parameter γ	0
gradient descent step size ε	$\varepsilon = 10^{16}$, $\varepsilon' = 5 \cdot 10^{15}$
number of iterations	225 with step ε, 25 with ε'
$\left\|p^h - P^h P_\mathrm{f}\right\|_2$	0.5932381

profiles of these experiments shows that the action of the control is delayed when the time interval is elongated.

- Experiments 5, 6 and 7 demonstrate the effects of different values of regularization $\gamma > 0$ with a fixed final time $T = 0.4$. Comparing experiments 5 and 6 shows that adding regularization $\gamma > 0$ decreases the magnitude of the optimal control θ^h (by one order), while simultaneously distributing its action over a larger time interval. Applying stronger regularization (compare experiments 6 and 7) yields an almost symmetric temporal control profile.

- Experiments 5 and 8 show the effect of the initial guess θ_0. The only difference between experiments 5 and 8 is that experiment 5 uses the non-symmetric initial guess (7.2) and experiment 8 features the symmetric initial guess (7.4). Unsurprisingly, the final temporal control profile of experiment 8 is symmetric and completely different from the temporal control profile of experiment 5. We conclude that two local minima of comparable quality (see Table 7.3) were found.

- Experiments 5, 6, and 7 also indicate that the strength of regularization $\gamma > 0$ is linked to the symmetry of the optimal control temporal profiles θ^h.

- Lastly, it is observed, that the bound (6.13) was not violated in any of the experiments. This shows that the linear reaction term (6.12) is sufficient to provide physical results in this particular case.

7.1.3. Moving a Gap Between Crystals

Experiment 9 showcases how even a highly non-trivial control θ^h can be obtained. Consider two crystals occupying most of the domain Ω with a small liquid gap between them. The final profile p_f expresses a desired final state, where the gap between the two crystals is moved to a different location. This is reflected by setting u_ini, p_ini and the target profile p_f as shown in Figure 7.7 and using the boundary condition (7.3). The full setup of the experiment is summarized in Table 7.4. To ensure the convergence of the method, two different gradient descent parameters ε and ε' were used.

The target profile p_f is achieved by a surprising process. The optimal control θ^h first makes sure the entire domain is solidified. Then, by applying high temperatures at the right boundary of the spatial domain, θ^h recreates the gap between the crystals and moves it into the position prescribed by p_f.

The resulting temporal control profile θ^h is displayed in Figure 7.8. In addition to this, six snapshots of the evolution of $\left(u^h\left(\theta^h\right), p^h\left(\theta^h\right)\right)$ close to final time T are depicted in Figure 7.10, with the corresponding times marked in Figure 7.9. The condition (6.13) is violated during the simulation, making it non-realistic.

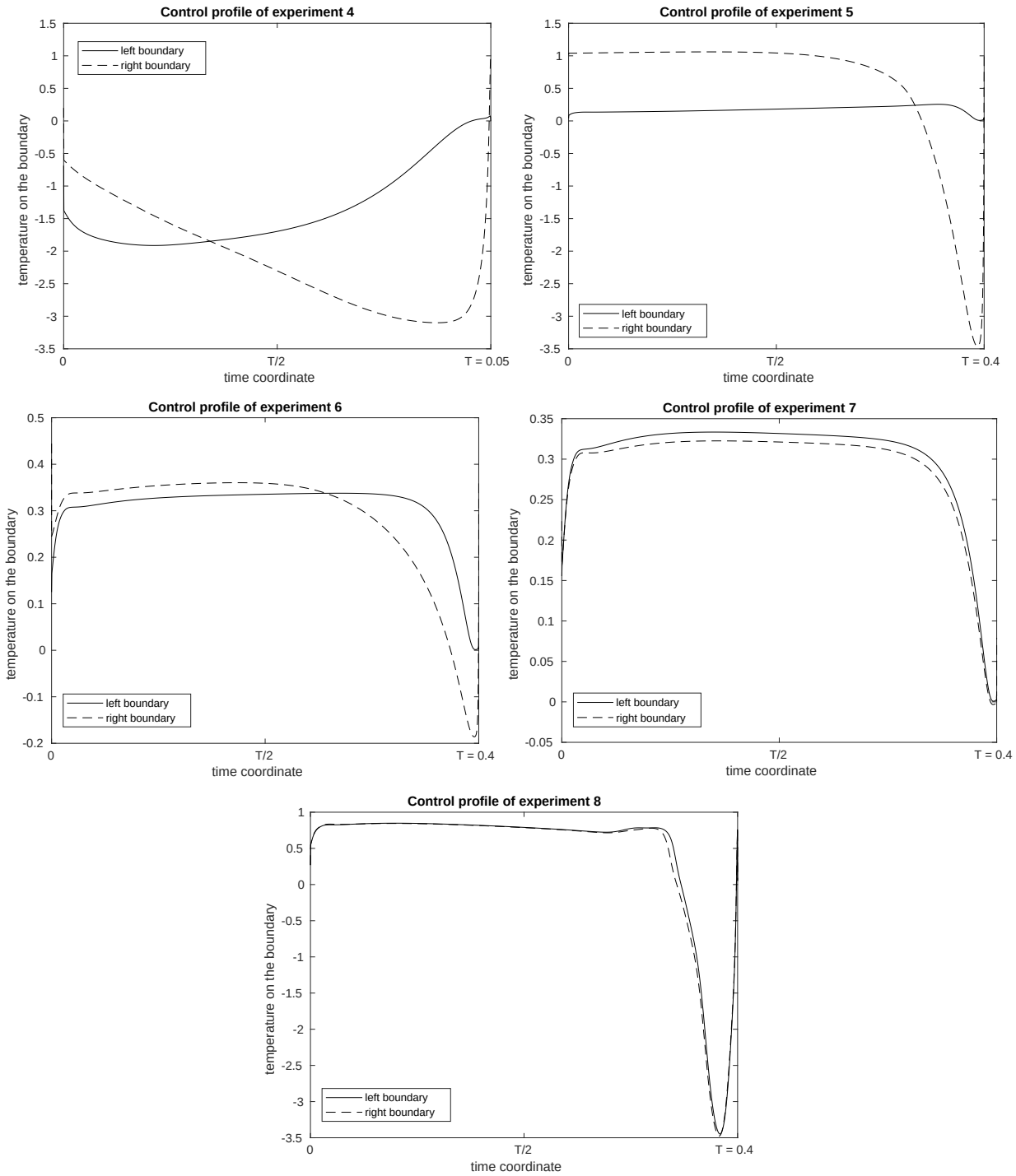

Figure 7.6.: Optimized temporal control profiles of experiments 4 through 8. Experiments 4 and 5 show how an optimization without regularization responds to an increase of final time T. Regularization is then added in experiments 6 and 7. The control profiles flatten and the control becomes more evenly distributed. Lastly, experiment 8 shows a different non-regularized control given by the symmetric initial condition (7.4).

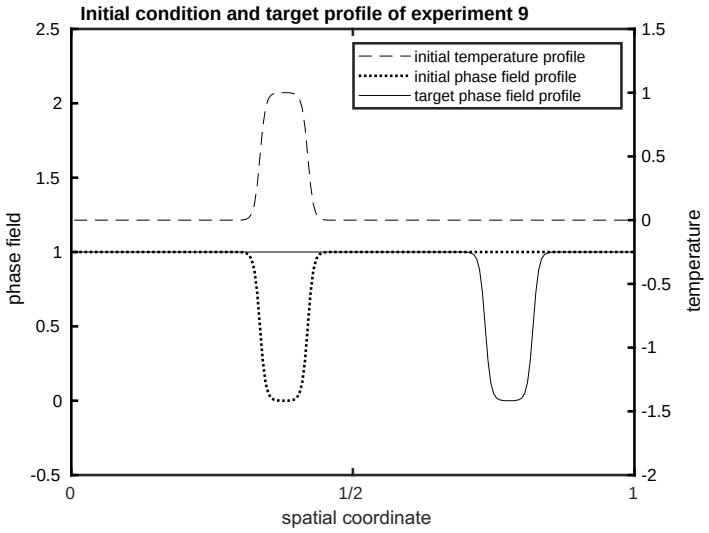

Figure 7.7.: The initial temperature and phase field profiles $p_{\text{ini}}, u_{\text{ini}}$ along with the target profile p_{f} for experiment 9. The values of the boundary condition p_{bc} are given by (7.3).

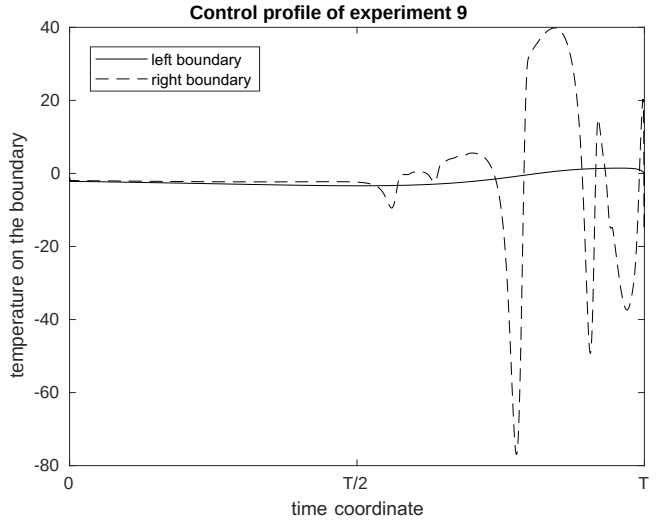

Figure 7.8.: Temporal control profiles for experiment 9.

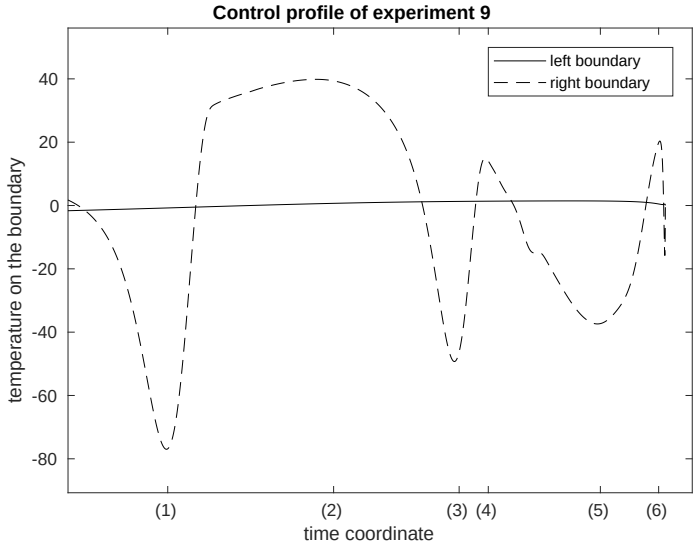

Figure 7.9.: A detailed look at the latter part of the optimal temporal control profile of experiment 9.

Table 7.5.: Parameter settings for the phase field simulations in Section 7.2.

Param.	Value	Physical Meaning
α	3.0	coefficient of attachment kinetics
β	300	dimensionless representation of supercooling
ξ	0.0101	interface thickness scaling
u^*	1.0	melting temperature
H	2.0	latent heat
ϵ_0	0	parameter of the sigmoid function (a limiter)
ϵ_1	0.2	parameter of the sigmoid function (a limiter)
L_{x_1}	0.6	spatial dimension in the x_1 direction
L_{x_2}	1.0	spatial dimension in the x_2 direction

7.2. Dirichlet Boundary Condition Control for the Phase Field Problem in 2D

Sections 7.1-7.1.3 featured a range of experiments, in which the linear reaction term (6.12) found utility as well as a couple of scenarios, where its shortcomings became apparent. In addition to this, the effects of simulation specific settings (changes in initial guess, applying regularization, changes in final time T, etc.) were discussed.

In this section, simulations performed in two spatial dimensions are discussed. The focus shifts to highlighting the difference between the linear reaction term (6.12) and the more advanced ΣP1-P reaction term (6.30). Section 7.2.1 focuses on a head to head comparison of the two reaction terms and Section 7.2.2 shows how the non-trivial objective of splitting a crystal into two can be achieved.

Parameters that are common to all the experiments detailed in this section are listed in Table 7.5.

7.2.1. Moving a Crystal from North to South with Different Reaction Terms

In this pair of experiments, a control that moves a crystal from one position in the domain to another is sought after. It is also required that the crystal shape at final time matches the prescribed shape p_{f}. The first of the two experiments (labeled 10) makes use of the linear reaction term (6.12) and the second (labeled 11) uses the ΣP1-P reaction term (6.30).

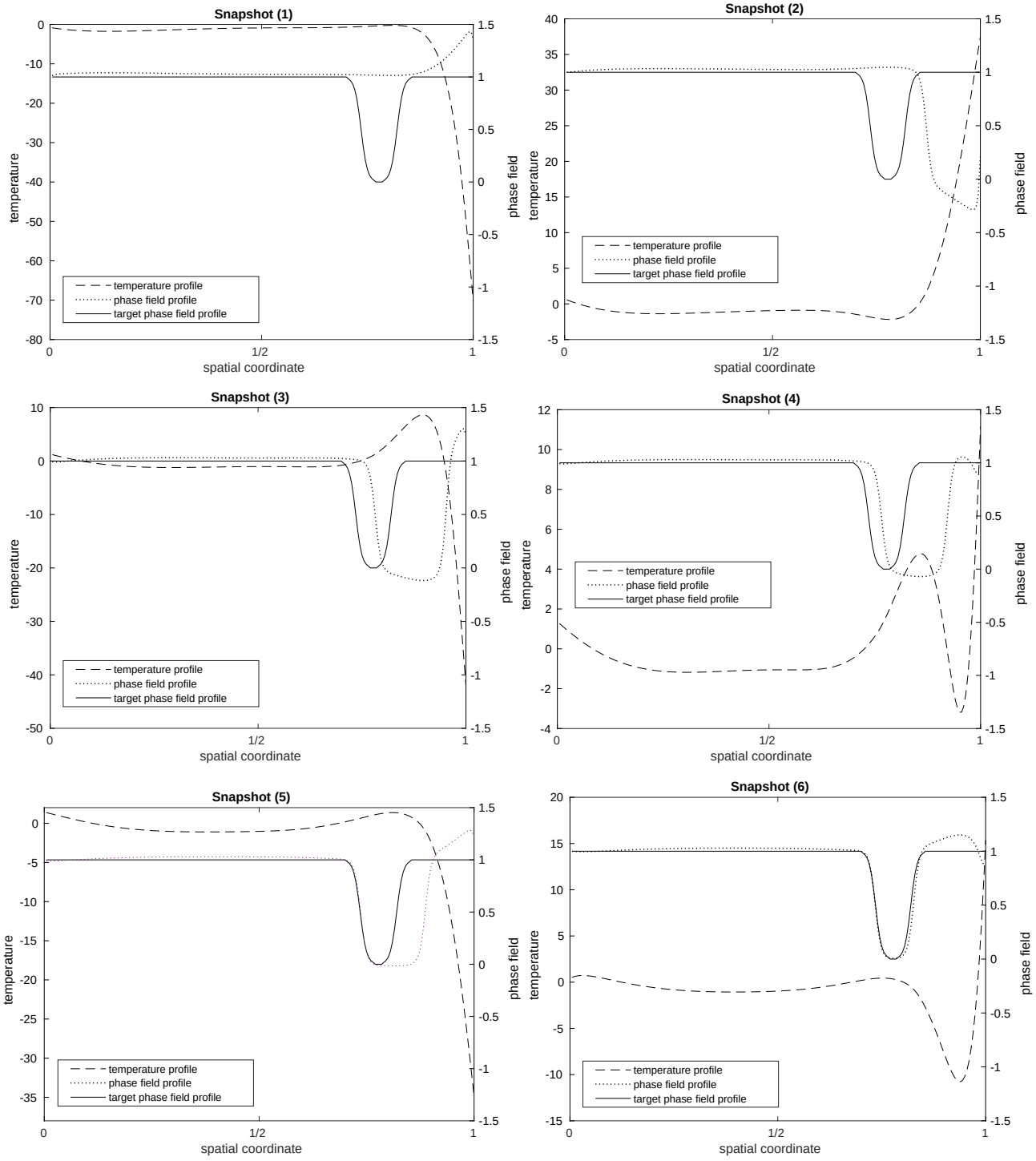

Figure 7.10.: The spatial profiles of the phase field and temperature at the significant times (1)–(6) marked in Figure 7.9.

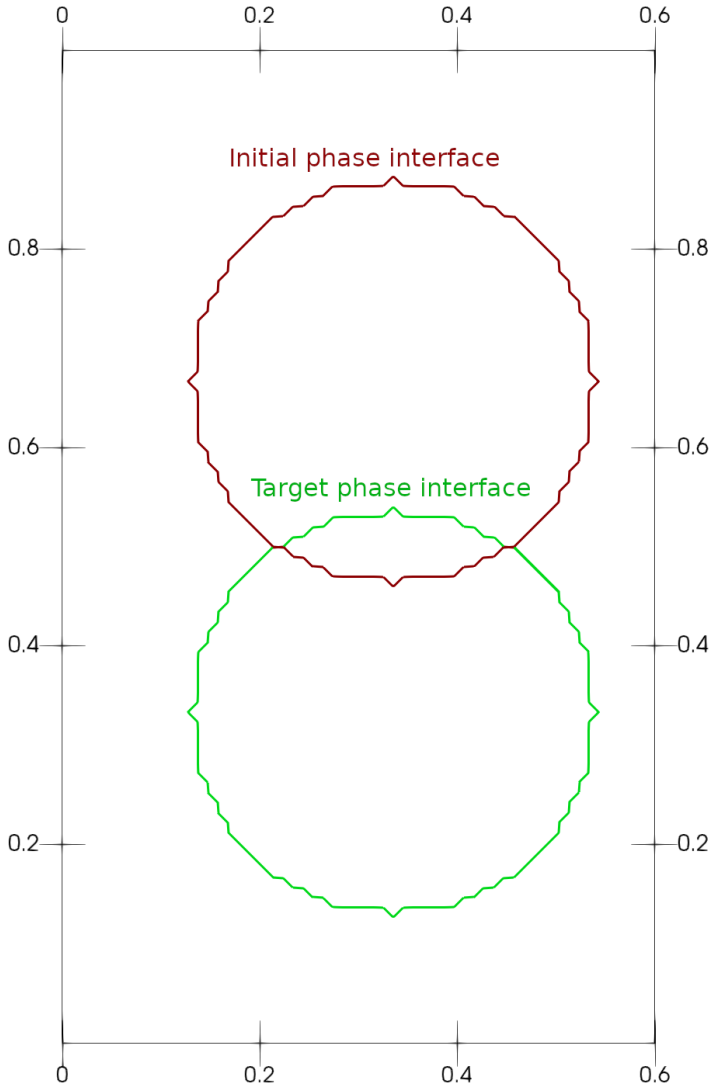

Figure 7.11.: Initial and target phase field interface Γ of experiments 10 and 11.

Settings common for both experiments follow. The objective of the minimization is captured by the target interface Γ_{f}, induced by the target profile p_{f} depicted in Figure 7.11. This figure also contains the initial interface $\Gamma(0)$, which is given by the initial condition p_{ini}. The boundary condition for the phase field p is given by

$$p_{\mathrm{bc}}(t, \boldsymbol{x}) = 0, \ \forall \boldsymbol{x} \in \partial\Omega, \ \forall t \in [0, T). \tag{7.5}$$

The initial guess for the Dirichlet control on the boundary reads

$$\theta_0(t, \boldsymbol{x}) = 1, \ \forall \boldsymbol{x} \in \partial\Omega, \ \forall t \in [0, T). \tag{7.6}$$

Other settings, some of which are not common for both experiments (like the mesh resolution or the final error) can be found in Table 7.5 and Table 7.6.

The optimized control profiles along with the time evolution of the level set $\Gamma(t)$ for experiments 10 and 11 can be reviewed in Figures 7.12 and 7.13. Comparing these two figures shows a qualitative and quantitative difference in the evolution of the phase field p^h between the two. This is to be expected, since (6.6)-(6.11) with (6.12) permits anomalous behaviors. These occur when the values of u^h do not

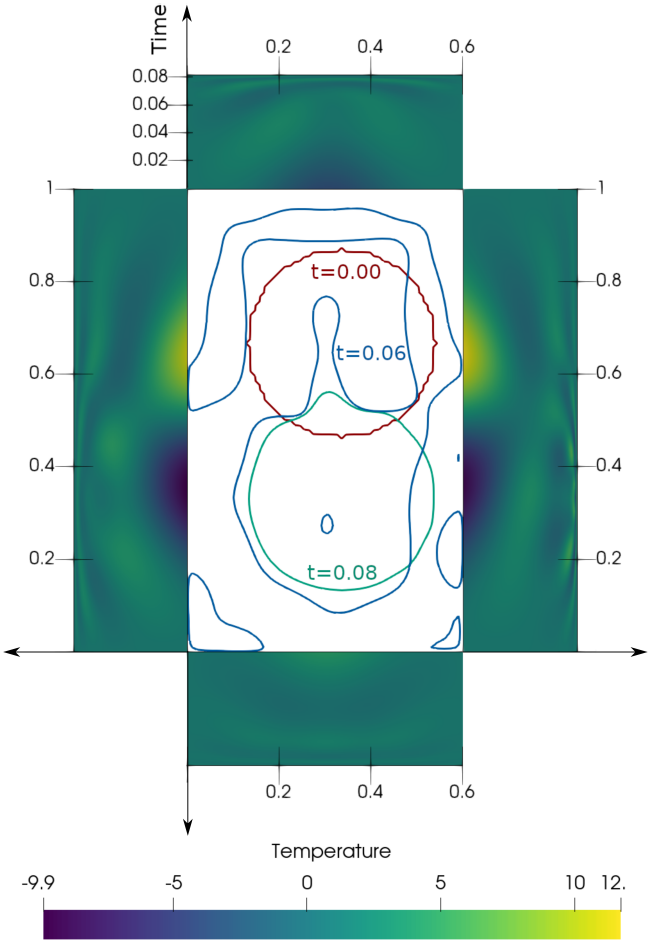

Figure 7.12.: The development of $\Gamma(t)$ for the optimal control θ^h experiment 10. The rectangles attached to the boundary of the domain depict the optimal Dirichlet boundary condition in space-time.

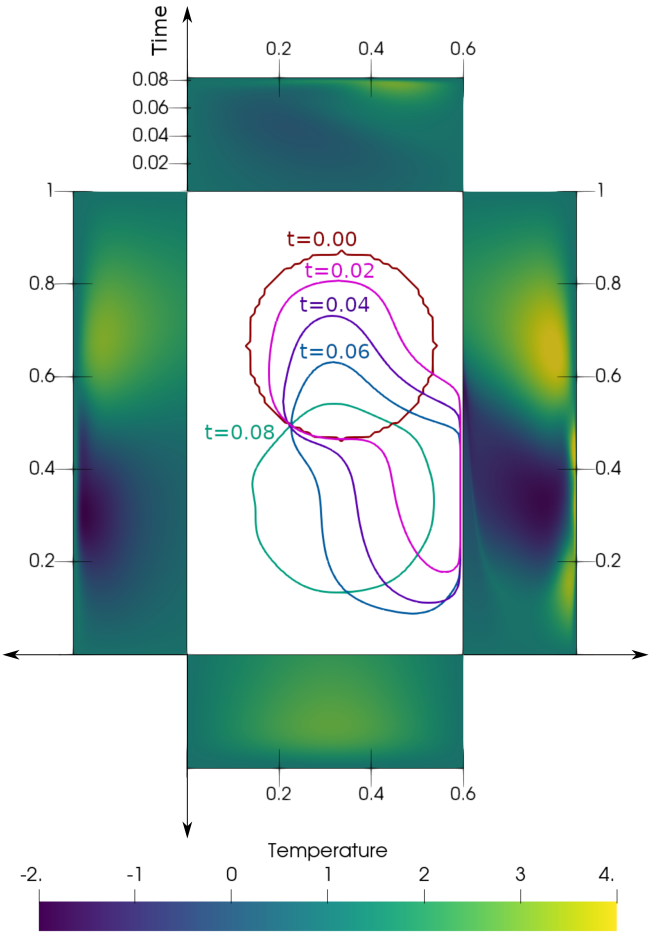

Figure 7.13.: The development of $\Gamma(t)$ for the optimal control θ^h experiment 11. The rectangles attached to the boundary of the domain depict the optimal Dirichlet boundary condition in space-time.

lie in the range given by (6.13), which causes the interface to lose its characteristic profile (Section 3.3). It has been shown that the ΣP1-P reaction term (6.30) does not suffer from such a deficiency (Section 3.3). This is reflected by the evolution of p^h (and consequently by the dynamics of $\Gamma(t)$) in experiment 11. Since the characteristic profile of the interface $\Gamma(t)$ is maintained in Experiment 11 (Figure 7.13) the result can be interpreted as solidification.

In Experiment 10 (Figure 7.12), a highly complex interface shape Γ can be observed at time $t = 0.06$. Soon after (at $t = 0.08$), the crystal nears its final shape. Additional snapshots of the phase field evolution of experiment 10 are shown in Figure 7.14. Completely different behavior is observed in experiment 11 (Figure 7.13). In this case, a process of gradual deformation guides the initial crystal towards the desired crystal shape. A preference to keep the crystal close to the right boundary, where the Dirichlet's boundary condition has the strongest influence, can be observed. This affinity is linked to the non-symmetric initial condition. Close to final time ($t = 0.08$), the crystal separates from the domain boundary and fine adjustments in the control mold it to match the target profile p_f. The resulting interfaces $\Gamma(T)$ along with the prescribed interface Γ_f, that follows from p_f, are shown in Figures 7.15 and 7.16. One may notice that in experiment 10, the crystal has a minor deformity at the top. In experiment 11, the shape is slightly deformed inwards towards the right. Both of these occurrences may be explained by the different processes leading up to the final state.

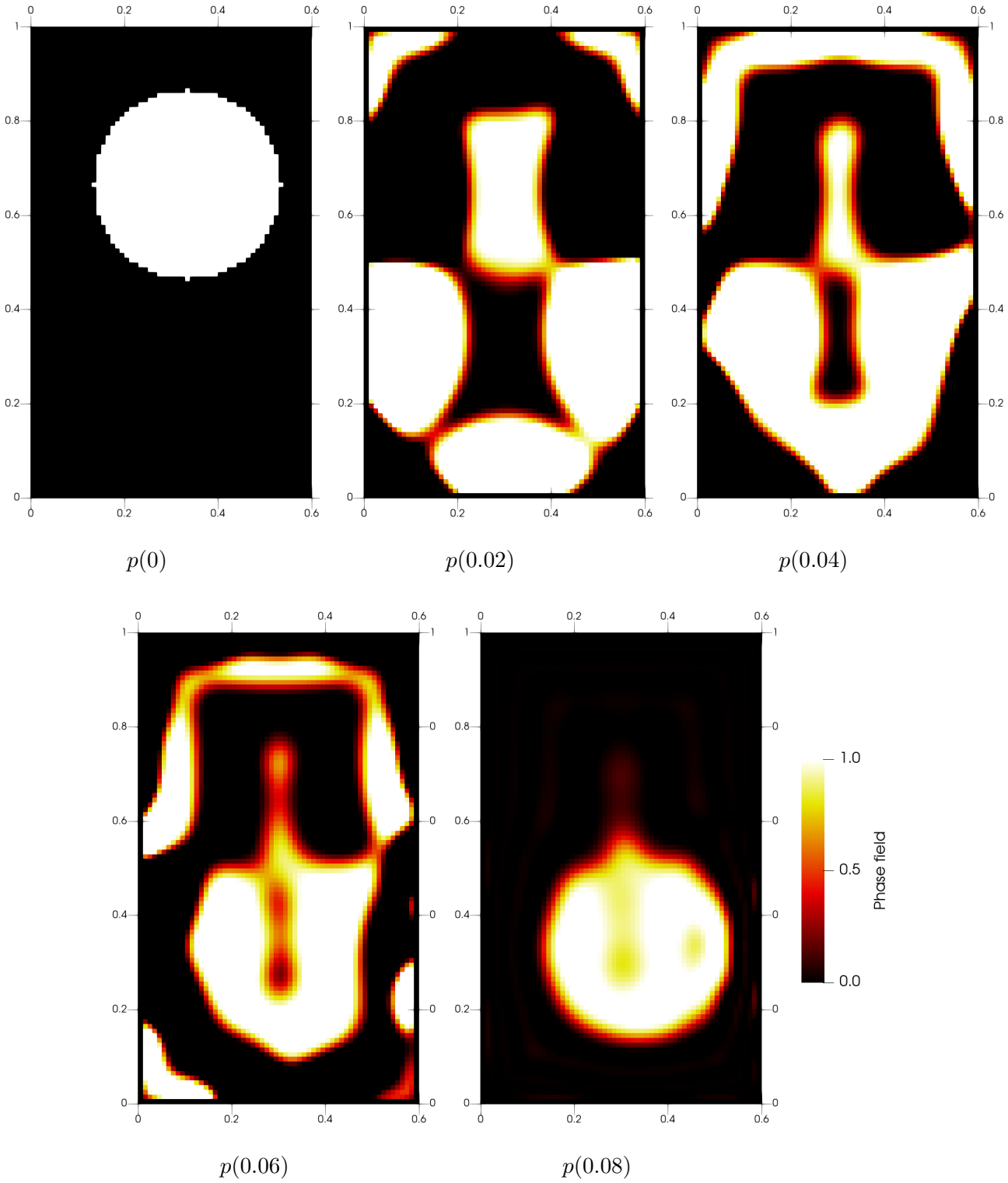

Figure 7.14.: Evolution of the phase field p^h in experiment 10 (linear reaction term)

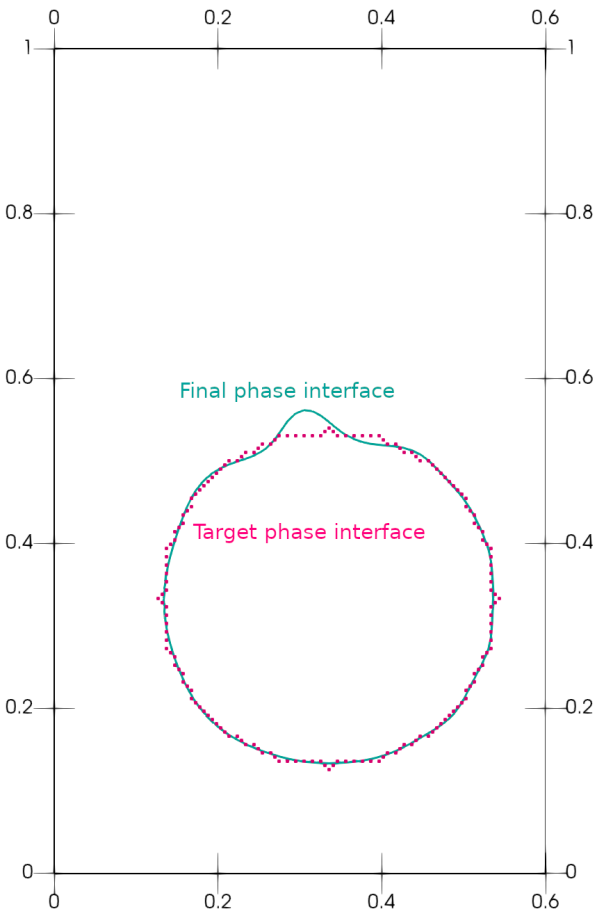

Figure 7.15.: The target (induced by p_{f}) and final (induced by p^h at $t = T$) level sets Γ for experiment 10.

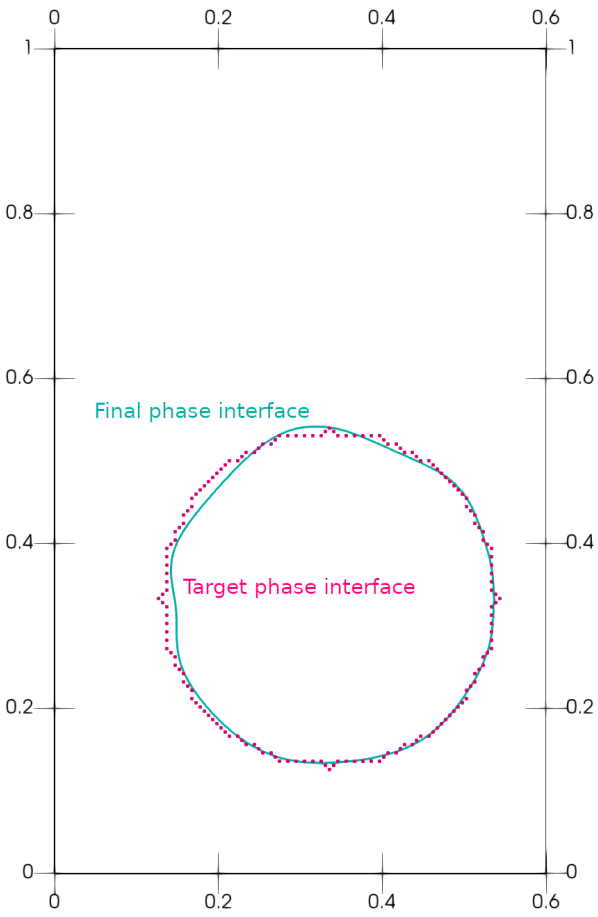

Figure 7.16.: The target (induced by p_{f}) and final (induced by p^h at $t = T$) level sets Γ for experiment 11.

Table 7.6.: Settings for the experiments 10, 11, 12 and the respective values of the difference (error) from the prescribed profile. The first experiment was performed with two reaction terms - linear and ΣP1-P.

Parameter	Simulation	
	Experiments 10 and 11 (moving crystal)	Experiment 12 (crystal split)
number of time steps N_t	$8 \cdot 10^3$	$8 \cdot 10^3$
number of grid points N_{x_1}	60	60
number of grid points N_{x_2}	100	100
initial control given by	(7.6)	(7.8)
final time T	0.081	0.081
regularization parameter γ	0	0
gradient descent step size ε	5.0	7.5
number of iterations	400 (linear), 5000 (ΣP1-P)	1000
$\left\|p^h - P^h p_\mathrm{f}\right\|_2$ at $t = T$	7.66 (linear), 7.39 (ΣP1-P)	7.21

7.2.2. Separating a Crystal with the Improved Reaction Term

In this simulation (labeled experiment 12), the ΣP1-P reaction term is applied to address the problem of crystal separation. Consider a crystal of rectangular shape in the center of domain Ω at initial time $t = T$, described by an appropriate initial condition p_ini. We aim to separate the rectangular crystal and create two circular ones. The level sets $\Gamma(0)$ and Γ_f induced by the initial condition p_ini and target profile p_f, respectively are depicted in figure 7.17. The boundary condition for the phase field p is given by

$$p_\mathrm{bc}(t, \boldsymbol{x}) = 0, \ \forall \boldsymbol{x} \in \partial\Omega, \ \forall t \in [0, T). \tag{7.7}$$

The initial guess for the Dirichlet control is symmetric and reads

$$\theta_0(t, \boldsymbol{x}) = 1, \ \forall \boldsymbol{x} \in \partial\Omega, \ \forall t \in [0, T). \tag{7.8}$$

Other settings for the experiment are listed in Table 7.6. The time evolution of the crystal shape $\Gamma(t)$ and the temporal control profile θ^h are depicted together in Figure 7.18.

Initially, the optimal control θ^h increases the temperature in the vicinity of the points $(x_1, x_2) = (0, 0.5)$ and $(x_1, x_2) = (0.6, 0.5)$. This leads to the crystal separating into two ($t = 0.06$). Following this, subtle adjustments in temperature are applied to finalize the crystal shapes to match the prescribed interface profile Γ_f. As can be seen in Figure 7.19, the two crystals end up being slightly deformed compared to the ones prescribed by Γ_f.

7.3. Performance and Implementation Details

The 1D and 2D solvers used in Section 7, were implemented in MATLAB and C++, respectively. To give the reader an idea about the relative performance demands, some figures for the 2D solver are mentioned. The C++ solver used in Sections 7.2.1-7.2.2 was executed on a single CPU core. A simple benchmark can be reviewed in Table 7.7.

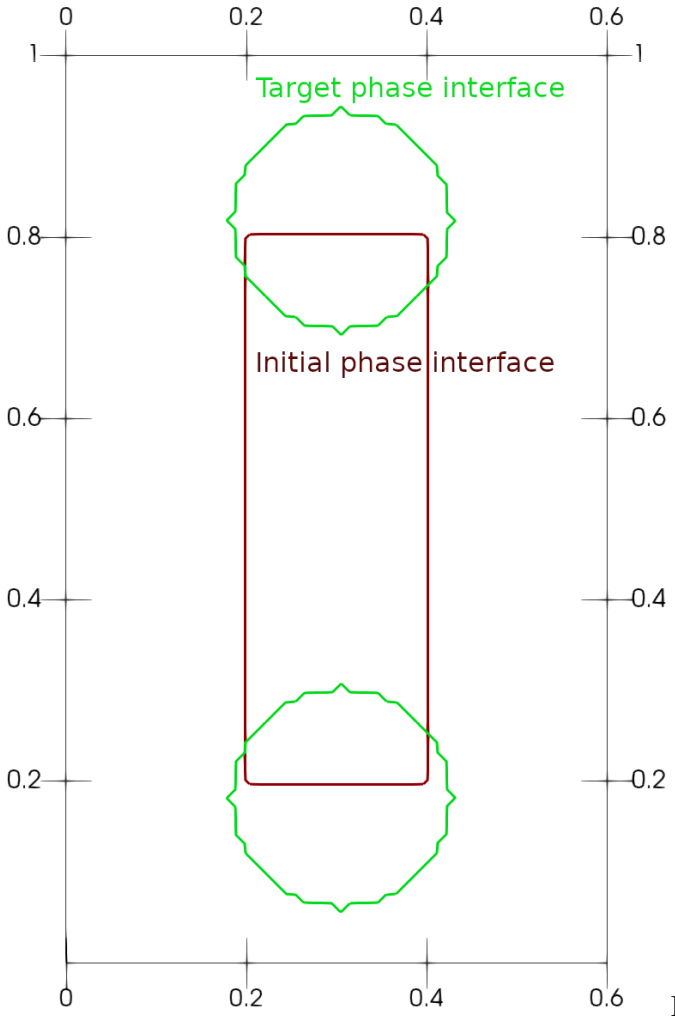

Figure 7.17.: The initial and target interface profiles $\Gamma(0)$ and Γ_f.

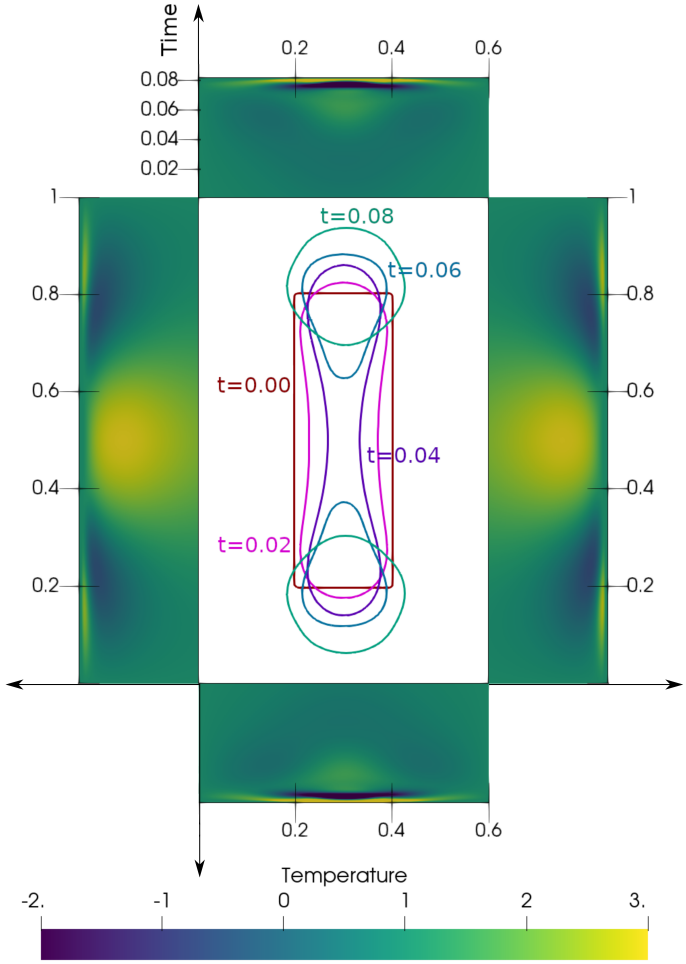

Figure 7.18.: The optimal control θ^h and snapshots of the phase interface $\Gamma(t)$ evolution, when the optimal control θ^h is applied. The rectangles attached to the boundary of the domain depict the optimal Dirichlet boundary condition in space-time.

Table 7.7.: Time to compute 100 iterations of gradient descent for different discretization resolutions. All simulations were performed on desktop with an i7-8700 CPU and 16GB RAM (Fedora 31 Linux). The values of N_t are given in multiples of 10^3.

N_x	N_y	N_t	Computational time [s]
51	51	1	37
51	101	1	78
101	101	4	585
101	101	8	1188
101	201	4	1281

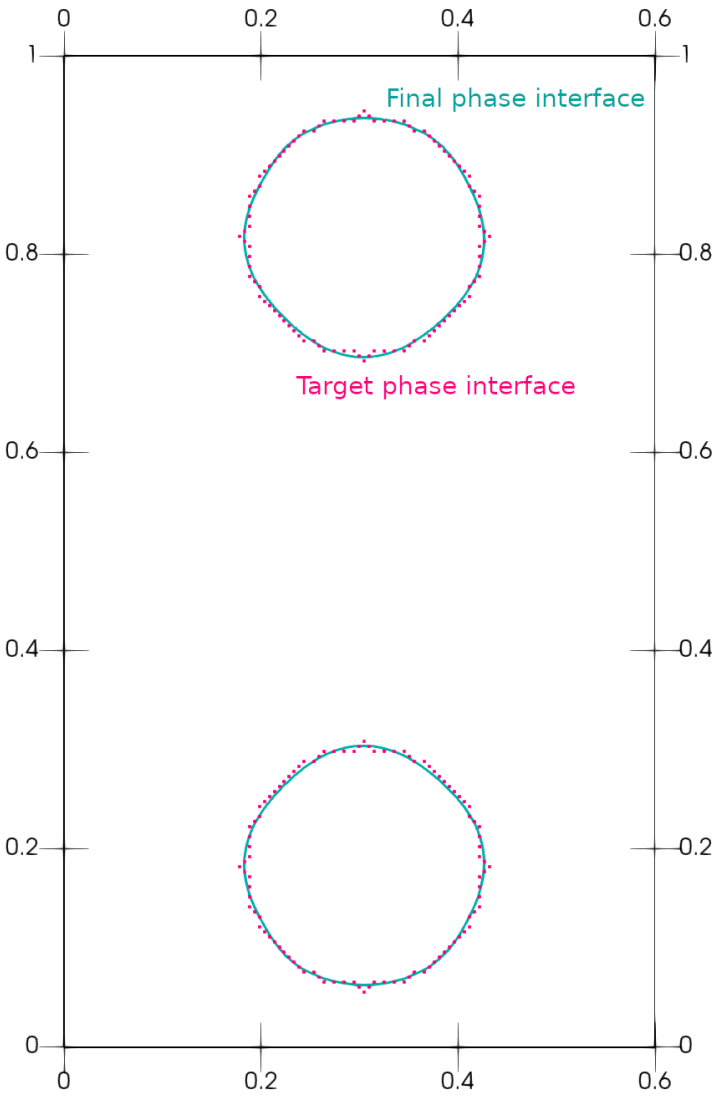

Figure 7.19.: Target and final phase field obtained in the simulation.

Conclusion

Several contributions to phase field modeling were detailed. All of these concern the phase-field model that governs the solidification of a pure melt. Based on a first principles derivation, matched asymptotic analysis and numerical studies, new reaction terms were proposed (Chapter 3) and validated. A PDE-constrained optimization problem, that can be used to find an optimal Dirichlet boundary condition for the heat equation in the phase field model leading to a particular crystal shape at final time, was proposed (Section 5.2). The existence of solution, Fréchet differentiability of the solution operator and optimality conditions for this problem were provided (Section 5.3-5.7). Lastly, the strong form of the aforementioned problem was used to provide framing for numerical simulations (Chapter 6). Using the adjoint problem to give an efficient means of gradient computation, this method was shown to be viable in a wide range of numerical simulations in one and two spatial dimensions. Among other things, it was shown that in the context of optimization the ΣP1-P reaction term (proposed in Chapter 3) provides many advantages. In particular, using the ΣP1-P term ensures that the simulation remains physical for any value of undercooling, while still having good analytical properties, which allow for the use of adjoint techniques for numerical optimization.

A. Appendix

This section provides a summary of the relevant definitions and some of the notation used throughout the thesis. More in depth explanations can be found in [30, 34]. We start by reviewing differentiability for maps between normed linear spaces.

Definition. Let X, Y be normed linear spaces, $f : X \to Y$, $x \in X$ and $v \in X$. We call

$$\delta f(x; v) = \lim_{\theta \to 0} \frac{f(x + \theta v) - f(x)}{h} \tag{A.1}$$

the directional derivative of f at point x in direction v if the limit exists.

Definition. Let X, Y be normed linear spaces, $f : X \to Y$ and $x \in X$. We call f Fréchet differentiable at x if there exists a bounded linear operator $A : X \to Y$ such that

$$\lim_{h \to 0} \frac{\|f(x + h) - f(x) - Ah\|_Y}{\|h\|_X} = 0.$$

We call A the Fréchet derivative of f at x and denote it by $f'(x)$.

Remark. Sometimes $\delta F(\boldsymbol{x})$ will be used to denote the integral representation of the formal Fréchet derivative of functional $F : X \to \mathbb{R}$ at point f, where X is a normed linear space. This results in a integral form representation of $\delta F(f; g)$ defined as

$$\delta F(f; g) = \int_\Omega \delta F(\boldsymbol{x})\, g(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x},$$

where Ω is a bounded domain.

Definition. Let X, Y, Z be normed linear spaces, $f : X \times Y \to Z$, $(a, b) \in X \times Y$ and let $g(x) \equiv f(x, b)$. We say that f has a partial Fréchet derivative with respect to the first coordinate at point (a, b) if $g'(a)$ exists and denote it by $f_x(a, b)$.

Definition. Let H_1, H_2 be a Hilbert space and $A : H_1 \to H_2$ be a bounded linear operator. Then the linear operator $B : H_2 \to H_1$ such that

$$\langle Ax, y \rangle_{H_2} = \langle x, By \rangle_{H_1} \text{ for all } x \in H_1, y \in H_2$$

is called the adjoint operator of A and denoted A^*.

Definition 17. Let $f : E \to \mathbb{R}$, $k \in \mathbb{N}, \gamma \in (0, 1]$ and define the norm

$$\|f\|_{C^{k,\gamma}(\overline{E})} \equiv \|f\|_{C^k(E)} + \sup_{|\beta| = k} \left| D^\beta f \right|_{0,\gamma},$$

where $|g|_{0,\gamma} \equiv \sup_{x,y \in E, x \neq y} \frac{|g(x) - g(y)|}{|x - y|^\gamma}$ for an arbitrary function $g : E \to \mathbb{R}$. Define the set

$$C^{k,\gamma}\left(\overline{E}\right) \equiv \left\{ f \in C^k\left(\overline{E}\right) : \|f\|_{C^{k,\gamma}(\overline{E})} < \infty \right\}.$$

It can be shown that $\left(C^{k,\gamma}\left(\overline{E}\right), \|f\|_{C^{k,\gamma}(\overline{E})} \right)$ is a Banach space.

Lemma 18. *(Grönwall's lemma) Let* $\varrho \in C\left([0,T],\mathbb{R}\right)$*,* $\phi,\psi \in L_1\left([0,T],\mathbb{R}\right)$ *be non-negative and let*

$$\varrho\left(t\right)' \leq \phi\left(t\right)\varrho\left(t\right)+\psi\left(t\right) \quad \text{a.e. in } [0,T].$$

Then

$$\varrho\left(t\right) \leq e^{\int_0^t \phi(r)\mathrm{d}r}\left[\varrho\left(0\right)+\int_0^t \psi\left(r\right)\mathrm{d}r\right] \quad \text{for all } t\in[0,T].$$

Proof. See [34]. □

A.1. Sobolev Spaces $W_p^k\left(E\right), \mathring{W}_p^k\left(E\right)$

Let $E\subset\mathbb{R}^n$ be open and let $k\in\mathbb{N}, 1\leq p\leq\infty$. Let $f:E\to\mathbb{R}$ be arbitrary. Define the Sobolev norm

$$\|f\|_{W_p^k(E)} \equiv \left|\sum_{|\beta|\leq k}\left\|D^\beta f\right\|_{L_p(E)}^p\right|^{\frac{1}{p}} \quad \text{for } 1\leq p<\infty, \tag{A.2}$$

$$\|f\|_{W_\infty^k(E)} \equiv \max_{|\beta|\leq k}\left\|D^\beta f\right\|_{L_\infty(E)}, \tag{A.3}$$

where $D^\beta f$ denotes the weak derivative. The definitions (A.2)-(A.3) can be used to define a number of variants of Sobolev spaces. Define the set

$$W_p^k\left(E\right) \equiv \left\{f\in L_p\left(E\right) : \text{for any } |\beta|\leq k\ D^\beta f\in L_p\left(E\right)\right\}. \tag{A.4}$$

For any $1\leq p\leq\infty$ $\left(W_p^k\left(E\right),\|\cdot\|_{W_p^k(E)}\right)$ is a Banach Space. However, $\left(W_p^k\left(E\right),\|f\|_{W_p^k(E)}\right)$ is reflexive only if $1<p<\infty$. In particular, $\left(W_2^k\left(E\right),\|f\|_{W_2^k(E)}\right)$ is a Hilbert space and the inner product that induces the norm reads

$$(f,g)_{W_2^k(E)} \equiv \sum_{|\beta|\leq k}\left(D^\beta f, D^\beta g\right)_{L_2(M)}, \tag{A.5}$$

where $f,g:E\to\mathbb{R}$.

Another possible Sobolev space is

$$\mathring{W}_p^k\left(E\right) \equiv \left\{f\in W_p^k\left(E\right) : \text{ such that } \mathrm{Tr}\left(f\right)=0\right\}. \tag{A.6}$$

The space $\left(\mathring{W}_p^k\left(E\right),\|\cdot\|_{W_p^k(E)}\right)$ is a Banach space, it is reflexive only if $1<p<\infty$ and $\left(\mathring{W}_2^k\left(E\right),\|f\|_{W_2^k(E)}\right)$ is a Hilbert space with the inner product (A.5).

Both $W_p^k\left(E\right)$ and $\mathring{W}_p^k\left(E\right)$ have commonly used dense subsets when $1<p<\infty$, in particular

$$\overline{C^\infty\left(E\right)}^{W_p^k(E)} = W_p^k\left(E\right), \tag{A.7}$$

$$\overline{C_0^\infty\left(E\right)}^{W_p^k(E)} = \mathring{W}_p^k\left(E\right). \tag{A.8}$$

Let X be a separable Banach space. For $p=2$, the common notation

$$W_2^k\left(E\right) \equiv H^k\left(E\right),$$
$$\mathring{W}_2^k\left(E\right) \equiv H_0^k\left(E\right)$$

is used.

A.1.1. Bochner Spaces $L_p(0,T;Y)$, $W(0,T;p,\widetilde{p},X_0,X,X_1)$ and $W(0,T;p,\widetilde{p},X_0,X_1)$

Let Y be a Banach space. Suppose that $f:[0,T]\to Y$ is a vector valued function and define the norms

$$\|f\|_{L_p(0,T;Y)} \equiv \left(\int_0^T \|f(t)\|_Y^p\, dt\right)^p,$$
$$\|f\|_{L_\infty(0,T;Y)} \equiv \operatorname*{ess\,sup}_{t\in[0,T]} \|f(t)\|_Y.$$

For $1\le p\le\infty$ the Bochner space is defined as

$$L_p(0,T;Y) \equiv \left\{g:[0,T]\to Y \text{ measurable} : \|g\|_{L_p(0,T;Y)}<\infty\right\}.$$

Let X_0, X, X_1 be Banach Spaces, X_0, X_1 be reflexive and

$$X_0 \hookrightarrow\hookrightarrow X \hookrightarrow X_1. \tag{A.9}$$

Additionally, consider $p,\widetilde{p}\in(1,\infty)$ and $T>0$. Then define the set

$$W \equiv W(0,T;p,\widetilde{p};X_0,X_1) \equiv \left\{v\in L_p(0,T;X_0), \dot{v}\in L_{\widetilde{p}}(0,T;X_1)\right\},$$

where the distributional derivative $\dot{v}$ is taken in the space $L_1(0,T;X_1)$. This is possible since for any $f\in L_p(0,T;X_0)$, we get $f\in L_1(0,T;X_0)$ and $f\in L_1(0,T;X_1)$ by (A.9). If $\dot{f}\in L_1(0,T;X_1)$ exists and the additional condition $\dot{f}\in L_{\widetilde{p}}(0,T;X_1)$ is fulfilled, then $f\in W$. In this setting, (A.9) is called a Gelfand (or Hilbert) triple. For any $f:[0,T]\to X_0$ define the norm

$$\|f\|_W \equiv \|f\|_{L_p(0,T;X_0)} + \left\|\dot{f}\right\|_{L_{\widetilde{p}}(0,T;X_1)}. \tag{A.10}$$

The set W along with (A.10) is a Banach space denoted $(W,\|\cdot\|_W)$.

Commonly, the spaces $(W,\|f\|_W)$ are used as the solution space for time dependent partial differential equations (PDEs). Moreover, $(W,\|f\|_W)\hookrightarrow\hookrightarrow L_p(0,T;X)$ (Theorem 22), which comes into play when dealing with non-linear PDEs (Section 5.3). Lastly, an example configuration for the Gelfand triple and space W is presented.

Lastly, other spaces that are associated with vector valued functions are discussed. Let X be a Banach space and $T>0$. Define the set

$$C([0,T];X) \equiv \{g:[0,T]\to X : g \text{ is continuous}\}.$$

The space $C([0,T];X)$ together with the norm

$$\|f\|_{C([0,T];X)} \equiv \sup_{t\in[0,T]} \|f(t)\|_X$$

is a Banach space denoted $\left(C([0,T];X),\|\cdot\|_{C(0,T;X)}\right)$ (since X is complete).

A.1.2. Embedding Theorems

Selected statements about embeddings of functional spaces are listed in this section. The reader should note that even though these embeddings are only discussed for bounded subdomains of $\mathbb{R}^n$, some of these results hold for unbounded domains as well (possibly under further technical assumptions) [1, 34].

Definition 19. Let $\Omega\subset\mathbb{R}^n$ be a bounded domain. We say that Ω has a Lipschitz boundary if every point $x\in\partial\Omega$ has an open neighborhood U_x such that $U_x\cap\partial\Omega$ is the graph of a Lipschitz continuous function.

A. Appendix

Theorem 20. *(Hölder-Sobolev embedding) Let $\Omega \subset \mathbb{R}^n$ be a bounded domain with a Lipschitz boundary. Let $m \in \mathbb{N}$ and $p \in [1, \infty)$. Then for any $k \in \mathbb{N}_0$, $\beta \in (0,1)$ such that*

$$m - \frac{n}{p} \geq k + \beta$$

the embedding

$$W^{m,p}(\Omega) \hookrightarrow C^{k,\beta}(\Omega)$$

holds.

Theorem 21. *(Sobolev compact embedding) Let $\Omega \subset \mathbb{R}^n$ be a bounded domain. Let $m \in \mathbb{N}$, $p \in [1, \infty)$, $q \geq 1$, $k \in \mathbb{N}_0$, $m > k$ and*

$$m - \frac{n}{p} > k - \frac{n}{q}$$

hold. Then

$$W_0^{m,p}(\Omega) \hookrightarrow\hookrightarrow W_0^{k,q}(\Omega).$$

Theorem 22. *(Embedding of Bochner spaces I.) Let X_0, X, X_1 form a Gelfand triple as defined by (A.9). Then for any $p, \widetilde{p} \in (1, \infty)$ and $T > 0$*

$$W(0, T; p, \widetilde{p}; X_0, X_1) \hookrightarrow\hookrightarrow L_p(0, T; X).$$

Theorem 23. *(Embedding of Bochner spaces II.) Let $T > 0$. Then*

$$W\left(0, T; 2, 2; H_0^1(\Omega), L_2(\Omega)\right) \hookrightarrow C\left([0, T]; L_2(\Omega)\right).$$

Bibliography

[1] R.A. Adams and J.J.F. Fournier. *Sobolev Spaces*, volume 140. Elsevier, Amsterdam, 2nd edition, 2003.

[2] S. M. Allen and J. W. Cahn. Coherent and incoherent equilibria in iron-rich iron-aluminum alloys. *Acta Metall.*, 23(9):1017–1026, 1975.

[3] S. M. Allen and J. W. Cahn. On tricritical points resulting from the intersection of lines of higher-order transitions with spinodals. *Scr. Metall.*, 10(5):451–454, 1976.

[4] S. M. Allen and J. W. Cahn. A microscopic theory for antiphase boundary motion and its application to antiphase domain coarsening. *Acta Metall.*, 27:1084–1095, 1979.

[5] N. Arada and J.-P. Raymond. Dirichlet boundary control of semilinear parabolic equations, part 1: Problems with no state constraints. *Appl. Math. Optim.*, 45:125–143, 2002.

[6] A. P. Ayala. Atomistic simulations of the pressure-induced phase transitions in BaF2crystals. *J. Phys.: Condens. Matter*, 13(50):11741–11749, 2001.

[7] B. Aziz. Robust and optimal control problems to a phase-field model for the solidification of a binary alloy with a constant temperature. *J. Dyn. Control Syst.*, 10:453–499, 2004.

[8] J. Barrat and J. Hansen. *Basic Concepts for Simple and Complex Liquids*. Cambridge University Press, 2003.

[9] J. W. Barrett, H. Garcke, and R. Nürnberg. On stable parametric finite element methods for the Stefan problem and the Mullins-Sekerka problem with applications to dendritic growth. *J. Comput. Phys.*, 229:6270–6299, 2010.

[10] F. B. Belgacem, C. Bernardi, and H. E. Fekih. Dirichlet boundary control for a parabolic equation with a final observation I: A space-time mixed formulation and penalization. *Asymptot. Anal.*, 71:101–121, 2011.

[11] F. B. Belgacem, H. E. Fekih, and H. Metoui. Singular perturbation for the Dirichlet boundary control of elliptic problems. *M2AN Math. Model. Numer. Anal.*, 37:850–883, 2003.

[12] G. Bellettini and M. Paolini. Anisotropic motion by mean curvature in the context of Finsler geometry. *Hokkaido Math. J.*, 25(3):537–566, 1996.

[13] M. Beneš. *Phase Field Model or Microstructure Growth in Solidification of Pure Substances*. PhD thesis, Czech Technical University in Prague, 1997.

[14] M. Beneš. Anisotropic phase-field model with focused latent-heat release. *Free Boundary Problems: Theory and Applications II, GAKUTO International Series in Mathematical Sciences and Applications*, 14:18–30, 2000.

[15] M. Beneš. Mathematical analysis of phase-field equations with numerically efficient coupling terms. *Interface. Free. Bound.*, 3:201–221, 2001.

[16] M. Beneš. Mathematical and computational aspects of solidification of pure substances. *Acta Math. Univ. Comenianae*, 70(1):123–151, 2001.

[17] M. Beneš. Diffuse-interface treatment of the anisotropic mean-curvature flow. *Appl. Math-Czech.*, 48(6):437–453, 2003.

[18] M. Beneš. Computational studies of anisotropic diffuse interface model of microstructure formation in solidification. *Acta Math. Univ. Comenianae*, 76:39–59, 2007.

[19] N. Bergeon, R. Trivedi, B. Billia, B. Echebarria, A. Karma, S. Liu, C. Weiss, and N. Mangelinck. Necessity of investigating microstructure formation during directional solidification of transparent alloys in 3D. *Adv. Space Res.*, 36:80–85, 2005.

[20] W. J. Boettinger, S. Coriell, A. L. Greer, A. Karma, W. Kurz, M. Rappaz, and R. Trivedi. Solidification microstructures: Recent developments, future directions. *Acta Mater.*, 48:43–70, 2000.

[21] J. Bragard, A. Karma, and Y. H. Lee. Linking phase-field and atomistic simulations to model dendritic solidification in highly undercooled melts. *Interface Sci.*, 10:121–136, 2002.

[22] G. Caginalp. An analysis of a phase field model of a free boundary. *Arch.RationalMech. Anal. 92, 205-245*, 1986.

[23] G. Caginalp. Stefan and Hele-Shaw type models as asymptotic limits of the phase-field equation. *Phys. Rev. A*, 39:5887–5896, 1989.

[24] A.V. Catalina, A.A. Burbelko, W. Kapturkiewicz, and M. Zhu. Computational models for prediction of solidification microstructure. In *Cast Iron Science and Technology*, volume 1A, 94–105, 2017.

[25] S. Chowdhury, T. Gudi, and A. Nandakumaran. Error bounds for a dirichlet boundary control problem based on energy spaces. *Math. Comput.*, 86:1103–1126, 2015.

[26] J. Christiansen. Numerical solution of ordinary simultaneous differential equations of the 1st order using a method for automatic step change. *Numer. Math.*, 14:317–324, 1970.

[27] P. Colli, G. Gilardi, G. Marinoschi, and E. Rocca. Optimal control for a conserved phase field system with a possibly singular potential. *Evol. Equations Control Theory*, 7(1):95–116, 2018.

[28] P. Colli, G. Gilardi, and Jü Sprekels. Analysis and optimal boundary control of a nonstandard system of phase field equations. *Milan J. Math.*, 80:119–149, 2012.

[29] P. Colli, A. Signori, and Jü. Sprekels. Optimal control of a phase field system modelling tumor growth with chemotaxis and singular potentials. *Appl. Math. Optim.*, 83:2017–2049, 2019.

[30] P. Drábek and J. Milota. *Methods of Nonlinear Analysis*. Birkhäuser, 2000.

[31] C. M. Elliott and A. R. Gardiner. Double obstacle phase field computations of dendritic growth. Technical Report 96/19, University of Sussex at Brighton, 1996.

[32] A. Shaikh et al. Effect of temperature and cooling rates on the alpha+beta morphology of ti-6al-4v alloy. *Procedia Struct. Integrity*, 14:782–790, 2019.

[33] N. Moussa et al. Macroscopic model for solidification in porous media. *Int. J. Heat Mass Transfer*, 113:704–715, 2017.

[34] L. C. Evans. *Partial Differential Equations*. Graduate Studies in Mathematics. American Mathematical Society, 1998.

[35] A. Filipponi, A. Di Cicco, S. De Panfilis, P. Giammatteo, and F. Iesari. Crystalline nucleation in undercooled liquid nickel. *Acta Mater.*, 124:261–267, 2017.

[36] S. Friedli and Y. Velenik. *Statistical Mechanics of Lattice Systems: A Concrete Mathematical Introduction*. Cambridge University Press, 2017.

[37] C. Garetto and M. Ruzhansky. Hyperbolic second order equations with non-regular time dependent coefficients. *Arch. Rational. Mech. Anal.*, 113–154, 2015.

[38] W. Gong, M. Hinze, and Z. Zhou. Finite element method and a priori error estimates for dirichlet boundary control problems governed by parabolic pdes. *J. Sci. Comput.*, 941-967, 2016.

[39] I. Gordeev and S. V. Starikov. Comparison of different methods of atomistic simulation to calculate the temperature of phase transition using the example of zirconium. *J. Exp. Theor. Phys.*, 128(5):747–753, 2019.

[40] M. E. Gurtin. On the two-phase Stefan problem with interfacial energy and entropy. *Arch. Ration. Mech. An.*, 96(3):199–241, 1986.

[41] D. M. Herlach. Non-equilibrium solidification of undercooled metallic melts. *Metals*, 4:196–234, 2014.

[42] C. Hilliard. Free energy of a nonuniform system. I. interfacial free energy. *J. Chem. Phys.*, 1958.

[43] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich. *Optimization with PDE Constraints*. Springer, 2009.

[44] K-H. Hoffman and L. Jiang. Otimal control of a phase field model for solidification. *Numer. Funct. Anal. Optim.*, 13:11–27, 1992.

[45] P.C. Hohenberg and B.I. Halperin. Theory of dynamic critical phenomena. *Rev.Mod.Phys. 49, 435-479*, 1977.

[46] J.J. Hoyt, M. Asta, and A. Karma. Atomistic and continuum modeling of dendritic solidification. *Mater. Sci. Eng. R*, 41:121–163, 2003.

[47] K. Huang. *Statistical Mechanics*. Wiley, 2007.

[48] N. M. Johnson, D. K. Biegelsen, H. C. Tuan, M. D. Moyer, and L. E. Fennell. Single-crystal silicon transistors in laser-crystallized thin films on bulk glass. *IEEE Electron Device Lett.*, 369–372, 1982.

[49] A. Karma and W-J. Rappel. Quantitative phase-field modeling of dendritic growth in two and three dimensions. *Phys. Rev. E*, 4323–4349, 1998.

[50] K. Kawasaki. Diffusion constants near the critical point for time-dependent ising models. i. *Phys-Rev.*, 224–230, 1966.

[51] R. Kobayashi. Modeling and numerical simulations of dendritic crystal growth. *Physica D*, 63:410–423, 1993.

[52] K. Kunisch and B. Vexler. Constrained dirichlet boundary control in l^2 for a class of evolution equations. *SIAM J. Control Optim.*, 46:1726–1753, 01 2007.

[53] A. Kunwar, Y. Amorim Coutinho, J. Hektor, H. Ma, and N. Moelans. Integration of machine learning with phase field method to model the electromigration induced cu6sn5 imc growth at anode side cu/sn interface. *J. Mater. Sci. Technol.*, 59:203–219, 2020.

[54] L.D. Landau. On the anomalous absorption of a sound near to points of phase transition of the second kind. In *Collected Papers of L.D. Landau*. Pergamon, 1965.

[55] G. Leugering, S. Engell, A. Griewank, M. Hinze, R. Rannacher, V. Schulz, M. Ulbrich, and S. Ulbrich. *Constrained Optimization and Optimal Control for Partial Differential Equations.* Birkhäuser, 2012.

[56] J. W. Lum, D. M. Matson, and M. C. Flemings. High-speed imaging and analysis of the solidification of undercooled nickel melts. *Metall. Mater. Trans. B*, 865–870, 1996.

[57] A. M. Meirmanov. *The Stefan Problem.* De Gruyter Expositions in Mathematics. Walter de Gruyter, 1992.

[58] N. Provatas and K. Elder. *Phase-Field Methods in Materials Science and Engineering.* Wiley, 2010.

[59] M. PunKay. Modeling of anisotropic surface energies for quantum dot formation and morphological evolution. In *NNIN REU Research Accomplishments.* University of Michigan, 116–117, 2005.

[60] A.H. Rajkotwala, A. Panda, E.A.J.F. Peters, M.W. Baltussen, C.W.M. van der Geld, J.G.M. Kuerten, and J.A.M. Kuipers. A critical comparison of smooth and sharp interface methods for phase transition. *Int. J. Multiphase Flow*, 120:103093, 2019.

[61] A. Schmidt. Computation of three dimensional dendrites with finite elements. *J. Comput. Phys.*, 125:293–3112, 1996.

[62] J. Sprekels and F. Tröltzsch. Sparse optimal control of a phase field system with singular potentials arising in the modeling of tumor growth. *J. Dyn. Control Syst.*, 5, 10 2020.

[63] P. Strachota. Antidissipative numerical schemes for the anisotropic diffusion operator in problems for the Allen-Cahn equation. In *ALGORITMY 2009 - Proceedings of contributed lectures and posters*, 134–142, 2009.

[64] P. Strachota and M. Beneš. A hybrid parallel numerical algorithm for three-dimensional phase field modeling of crystal growth. In *ALGORITMY 2016, 20th Conference on Scientific Computing, Vysoké Tatry - Podbanské, Slovakia, March 14 - 18, 2016. Proceedings of contributed papers and posters*, 23–32, 2016.

[65] P. Strachota and M. Beneš. Design and verification of the MPFA scheme for three-dimensional phase field model of dendritic crystal growth. In *Numerical Mathematics and Advanced Applications 2011: Proceedings of ENUMATH 2011, the 9th European Conference on Numerical Mathematics and Advanced Applications, Leicester, September 2011*, 459–467, 2013.

[66] P. Strachota, A. Wodecki, and M. Beneš. Focusing the latent heat release in 3D phase field simulations of dendritic crystal growth. *Modelling Simul. Mater. Sci. Eng.*, 29:065009, 2021.

[67] G. Strang and G.J. Fix. *An Analysis of the Finite Element Method, 2ed. Prentice Hall series in automatic computation.* Wellesley-Cambridge Press, 2008.

[68] F. Tröltzsch. *Optimal control of partial differential equations. Theory, methods and applications.* American Mathematical Society, 2010.

[69] R. Wait and A.R. Mitchell. *Finite element analysis and applications.* A Wiley-Interscience publication, 1985.

[70] A. A. Wheeler, W. J. Boettinger, and G. B McFadden. Phase-field model for isothermal phase transitions in binary alloys. *Phys. Rev. A*, 45(10):7424–7440, 1992.

[71] A. A. Wheeler, B. T. Murray, and R. J. Schaefer. Computation of dendrites using a phase field model. *Physica D*, 66:243–262, 1993.

[72] R. Willnecker, D. M. Herlach, and B. Feuerbacher. Evidence of nonequilibrium processes in rapid solidification of undercooled metals. *Phys. Rev. Lett.*, 62(23):2707–2710, 1989.

[73] J-Y. Wu, V. P. Nguyen, C. T. Nguyen, D. Sutula, S. Sinaie, and S. P. A. Bordas. Chapter one - phase-field modeling of fracture. Advances in Applied Mechanics. Elsevier, 1-183, 2020.

[74] W. Yang, J. Sun, and S. Zhang. Analysis of optimal boundary control for a three-dimensional reaction-diffusion system. *SIAM J. Control Optim.*, 325-344, 2017.

[75] X. Zonghong, W. Wei, Z. Ying, W. Yue, and L. Yumei. Optimal control for a phase field model of melting arising from inductive heating. *AIMS Math.*, 121-142, 2022.