Deutsche Forschungsgemeinschaft

Priority Program 1253

Optimization with Partial Differential Equations

L. Banas, M. Klein and A. Prohl

Control of Interface Evolution in Multi-Phase Fluid Flows

October 2012, revised July 2013

Preprint-Number SPP1253-134

http://www.am.uni-erlangen.de/home/spp1253
CONTROL OF INTERFACE EVOLUTION IN MULTI-PHASE FLUID FLOWS

LUBOMÍR BAŇAS, MARKUS KLEIN, AND ANDREAS PROHL

Abstract. We consider an optimal control problem for the interface in a two-dimensional multi-phase fluid problem. The minimization functional consists of two parts: the $L^2$-distance to a given density profile and the interfacial length. We show existence and derive necessary first order optimality conditions for a corresponding phase field approximation of the perimeter functional. An unconditionally stable fully discrete scheme which is based on low order finite elements is proposed, and convergence of corresponding iterates to solutions of the limiting optimality conditions for vanishing discretization parameters is shown. Computational studies are included to validate the model including the phase-field approximation, interface motion, and topological changes, as well as to study relative effects due to discretization, regularization errors, and the relation of both parts of the functional.

1. Introduction

We plan to control the motion of a multi-phase fluid flow in a bounded domain $\Omega \subset \mathbb{R}^2$. A typical application includes the production of aluminium via electrolysis, where liquid aluminium oxid flows on top of liquid aluminium in some container, and aluminium may not come into contact with the electrolytes in some region at the top of the container [13]; as a consequence, oscillatory effects of the interface between the fluids should be avoided. Other applications for the control of interfacial regions between different fluids can be found in microfluidics for material processing, chemistry, biology, and medicine; for more details see [5]. Below, the motion of the fluid is described by the incompressible multi-phase fluid Navier–Stokes equations, cf. e.g. [24], and the objective functional consists of a $L^2$ tracking-type term, together with the perimeter functional. The optimal control problem then reads as follows.

**Problem 1.1.** Let $\Omega \subset \mathbb{R}^2$ be open and bounded, and $\tilde{\rho} : \Omega_T := (0, T) \times \Omega \to \mathbb{R}$ be given. Let $\alpha, \beta, \lambda > 0$, and $T > 0$ be fixed. Find $y, u : \Omega_T \to \mathbb{R}^2$, and $\rho : \Omega_T \to \mathbb{R}$, such that

$$G(\rho, u) := \int_0^T \left\{ \beta H^1(S_{\rho}) + \frac{\lambda}{2} \int_\Omega |\rho - \tilde{\rho}|^2 \, dx + \frac{\alpha}{2} \int_\Omega |u|^2 \, dx \right\} \, dt \quad (1.1)$$

*Date:* July 29, 2013.

2000 Mathematics Subject Classification. 35K55, 49Q10, 65M60, 76D55, 76M10, 76T05, 93C20, 93C95.

Key words and phrases. optimality condition, incompressible Navier–Stokes Equation, Lagrange multiplier, Finite Element, phase field model.

The work of the two last authors was supported by a DFG grant within the Priority Program SPP 1253 (Optimization with Partial Differential Equations).
is minimized subject to the density dependent Navier–Stokes equations,
\begin{align}
\rho y_t + \rho |y \cdot \nabla| y - \text{div}(\mu(\rho) \nabla y) + \nabla p &= \rho u, \quad \text{(1.2a)} \\
\rho_t + [y \cdot \nabla] \rho &= 0, \quad \text{(1.2b)} \\
\text{div } y &= 0, \quad \text{(1.2c)}
\end{align}

together with \( \rho(0,.) = \rho_0, \ y(0,.) = y_0, \) and \( y = 0 \) on \((0,T] \times \partial \Omega).\)

This model tracks given profiles \( \tilde{\rho} \in L^2(\Omega_T) \) while simultaneously minimizing the interface length; see Figure 12 for the regularizing role of the perimeter functional in the optimal control problem.

In Problem 1.1, \( S_\rho \subseteq \Omega \) denotes the jump set of the function \( \rho \), and \( H^1 \) the 1-dimensional Hausdorff measure, see e.g. [2]. The variable \( y \) denotes the velocity, and \( \rho \) the density of the fluid, with \( \mu > 0 \) its local viscosity. A typical situation for the initial density is \( \rho_0 = \rho_1 \chi_{\Omega_1} + \rho_2 \chi_{\Omega_2} \) with \( \Omega_1 \cap \Omega_2 = \emptyset \) and \( 0 < \rho_1 \leq \rho_2 < \infty \).

Minimizing the functional \( G \) without additional constraints has been studied in former works, see e.g. [2, 12], where the perimeter functional is approximated by the following regularization (cf. [12, Chapter 4]),
\begin{equation}
F_\delta(\rho) := \int_0^T \int_\Omega \left\{ \delta |\nabla \rho|^2 + \frac{1}{4\delta} W(\rho) \right\} \, dx \, dt, \tag{1.3}
\end{equation}

where \( \delta > 0 \) and \( W(\rho) = (\rho - \rho_1)^2(\rho - \rho_2)^2 \). It is well-known (for the time independent case) that \( F_\delta \Gamma\)-converges on \( L^2(\Omega) \) to the perimeter functional \( \rho \mapsto H^1(S_\rho) \), which takes finite values in \( GSBV(\Omega) \), see [2, 12]. An important consequence of \( \Gamma \)-convergence is that every sequence of minima of the approximative problem converges to a minimum of the original problem.

The authors are not aware of any work in the literature where \( \Gamma \)-convergence of (sequences of) problems with analytical constraints like differential equations has been studied. A naive way to deal with additional constraints is to set the corresponding functional equal to infinity at points where the constraint does not hold. It is easy to construct examples (even for functionals \( f_n, f : \mathbb{R} \to \mathbb{R} \) with \( f_n \rightharpoonup f \)), such that a corresponding \( \Gamma \)-convergence result will not be inherited in the presence of such additional constraints. The problem here is that by the well-posedness of the equation, the modified functional would have to be changed in “too many points” to infinity.

In order to show existence and derive optimality conditions for a regularized version of Problem 1.1 we have to overcome several difficulties.

1. Since \( \rho \in L^\infty(\Omega_T) \) in (1.2b), and not in \( SBV(\Omega) \) for almost all times in general, neither the jump set \( S_\rho \) is well-defined nor the mapping \( \rho \mapsto S_\rho \) is weakly lower semicontinuous for almost all times. Hence, it is not clear how to construct solutions for Problem 1.1. Moreover, it is not obvious how to derive optimality conditions due to the lack of differentiability of the perimeter functional.

2. The derivation of optimality conditions via the Lagrange multiplier theorem is non-trivial due to the limited regularity of the density in (1.2b), and it is not clear if the minimum is a regular point; as a consequence, the associated Lagrange multiplier
lacks regularity properties; see e.g. [20, Chapter 1] or [26, Chapter 9]. In particular, the Lagrange multiplier to (1.2b) would not be a function that is defined on $\Omega_T$.

In order to handle the first problematic issue, we use the perimeter approximation (1.3) and replace the objective function $G$ in (1.1) by

$$J_\delta(\rho, u) := \frac{\lambda}{2} \int_0^T \int_\Omega |\rho - \rho_0|^2 \, dx \, dt + \frac{\beta}{2} \int_0^T \int_\Omega \left\{ \delta |\nabla \rho|^2 + \frac{1}{4\delta} W(\rho) \right\} \, dx \, dt + \frac{\alpha}{2} \int_0^T \int_\Omega |u|^2 \, dx \, dt,$$

where $\delta > 0$. This phase-field approximation is done in order to construct a well-defined weakly lower semicontinuous functional $J_\delta : L^2(H^1) \times L^2(L^2) \to \mathbb{R}$.

To handle the second problematic issue, we regularize (1.2b) by adding artificial diffusion at a scale $\varepsilon > 0$. This modification improves regularity properties of both, the density $\rho$ and the related Lagrange multipliers, and allows to construct solutions of the modified Problem 4.1.

The idea of adding artificial diffusion is used to solve (1.2) (cf. [24]), and was also used in [22] to derive optimality conditions for the density dependent Stokes equations. It can then be shown that solutions of the modified equation (3.2) converge to those of (1.2). We also refer to a corresponding discussion based on computational studies in Section 9.

Theorem 4.2 asserts solvability of the regularized optimality problem for finite $\varepsilon, \delta > 0$. An open question remains as how to choose pairings $\varepsilon$ and $\delta$ in such a way that the functional is bounded uniformly with respect to $\varepsilon$ and $\delta$. A heuristic choice which is due to standard parabolic a priori estimates is to set $\delta = O(\varepsilon)$ in order to have at least $\mathcal{H}^1(S_\rho) < \infty$ for a limiting density $\rho \in H^1(S_\rho)$. This scaling is supported by computational evidence reported in Section 9: choosing $\varepsilon \gg \delta$ causes highly diffuse interfaces (see Figure 3), as opposed to “parasitic velocities” in the opposite scenario where $\varepsilon \ll \delta$ (see Figure 2).

There are other multiphase fluid flow models which include surface tension terms, and thus avoid highly oscillatory behavior of the surface on physical grounds; see e.g. [1]. In those cases, accordingly, regular interfaces occur due to combined effects of surface tension and the perimeter functional. Hence, we choose the present setup of the optimization problem to decouple effects of the equation from those of the functional, by addressing fluids with negligible surface tension in this work. Moreover, this identification of effects allows for future works with extended functionals of e.g. Willmore energy type, which penalizes areas with large mean curvature on the surface.

The literature on optimal control problems subject to the density dependent Navier–Stokes equations is rare: a main difficulty in Problem 1.1 is the strong coupling between the mass equation and the momentum equation, which leads to a strong coupling in the adjoint equations; another problem comes from the lack of regularity properties of the solution of the mass equation. We mention the work of Kunisch and Lu [22], where an optimal control problem with an $L^2$ tracking type functional subject to the regularized density dependent Stokes equation in $\mathbb{R}^2$ is studied, and optimality conditions are derived. Moreover, by some assumptions on a non-regularized solution, it is shown that minima ($\varepsilon > 0$) converge to a minimum of the limiting problem for $\varepsilon = 0$. We note that a corresponding result seems not clear in the present setting, where $\delta, \varepsilon > 0$: furthermore, as already discussed above, solvability of the limiting problem, Problem 1.1, has to remain an open problem.
The construction of convergent numerical discretizations of \((1.2)\) is a very recent subject. The first work which accomplished this goal is \([25]\), where a discontinuous Galerkin scheme is studied for \((1.2)\), which uses piecewise constant functions for the pressure, in particular. In view of optimal control, corresponding (discrete) optimality conditions couple primal and dual variables, which requires to bound primal variables in stronger norms to show stability of the overall scheme. For this reason, we consider instead the continuous Galerkin scheme \([8]\), where continuous functions for the pressure space are admitted. The discretization for \((1.2)\) in \([8]\) introduces numerical stabilization terms in order to conclude convergence against weak solutions, which fits into the discussion above of a regularized version of Problem \(1.1\) as well. In the present case, since an artificial diffusion term is introduced in \((1.2a)\), we do not need most of the regularization terms as suggested in \([8]\) in our scheme; see \((5.5)\). An alternative strategy to discretize \((1.2)\) is given in \([14]\), where \((1.2)\) is solved numerically using artificial diffusion as well; however, the convergence analysis uses higher order finite elements for the velocity space, which is more expensive on a computational level than the schemes in \([25]\) and \([8]\).

In order to construct necessary optimality conditions by a fully practical discrete scheme for the regularized optimization problem, we use the “first discretize, then optimize” ansatz: We propose a corresponding implementable discrete optimization problem, which involves the discretized equations as discussed above in \([8]\), derive related discrete optimality conditions, and show convergence of the iterates to a solution of the continuous optimality conditions. This method benefits from the available stable, convergent finite element based fully practical discretization of the state equation, and leads to a construction of solutions of the continuous optimality system \((4.1)\). The solvability of discrete optimality conditions follows directly via the Lagrange multiplier theorem. This approach to set up discrete optimality conditions has the advantage of being a natural and structure preserving discretization of the adjoint equation. The main challenging part is the strong coupling of both primal variables \(\rho\) and \(y\), their strong coupling with the two adjoint variables, as well as the coupling between both adjoint variables themselves. In order to address this issue, we first have to derive strong stability properties for the discrete primal variables. The second step is to derive standard parabolic regularity properties for the discrete adjoint variables. Here, we need the regularity of the primal variables and a combined argument: Since derivatives of both adjoint variables are present in both adjoint equations \((6.2a)\) and \((6.2c)\), we have to multiply the adjoint equations simultaneously with different test functions and to consider a proper weighted sum of the resulting inequalities.

From a practical point of view, other ways of controls include a finite dimensional control space where amplitudes of given forces are unknown, or boundary control. However, the distributed control provided here gives hints in which region an optimal control should act. This is relevant in certain engineering applications, such as magnetohydrodynamics, or the control of ferrofluids, cf. \([27]\). We note that all proves can also be done directly in the same manner (even more easier in some points) for a finite dimensional control space, where only amplitudes of given forces are unknown. It is likely that under some assumptions even a boundary control could be possible.

The paper is organized as follows. In Section \(3\) we study the regularized equation \((3.2)\) and prove regularity results, see Theorem \(3.2\). In Section \(4\) Problem \(1.1\) is restated in a proper form: we assert solvability in Theorem \(4.2\) and derive first order necessary optimality
conditions (4.1). In Section 5, we describe the numerical setup, state the numerical scheme for the primal equation (5.5), and show solvability and standard parabolic bounds of the discrete density and velocity in Lemma 5.1. Moreover, we prove boundedness of the discrete density and velocity in stronger norms in Lemmas 5.2 and 5.3. In Section 6, we study the discrete optimization problem and derive discrete optimality conditions (6.2). In Section 7, we derive bounds for the discrete adjoint variables in standard parabolic norms. The proof has to cope with the subtle coupling between both dual variables, i.e., the Lagrange multipliers related to (1.2b) and (1.2a); it is also motivated why we needed that strong bounds on the primal variables before. Finally, in Section 8 we show the main result of this paper in Theorem 8.3: for numerical parameters $h, k \to 0$, a subsequence of solutions of the discrete optimality conditions (5.5)-(6.2) converges to a solution of the continuous optimality conditions (4.1) for fixed $\delta, \varepsilon > 0$. Moreover, the discrete optimal control function $\{U^n\}$ converges to the continuous optimal control function $u$ strongly in $L^2(L^2)$, which will be proven in Theorem 8.4. Lastly, we present several numerical experiments in Section 9. Here, we propose a variable step-size gradient type algorithm for the solution of the discrete problem and study the relative effect of the phase-field formulation of the functional, and the stabilization of the PDE constraint. We also demonstrate qualitatively different behaviors of the fluids for $\beta = 0$ and $\beta > 0$ (see Figure 12) to evidence the regularizing effect of the perimeter functional onto the initial interface. Those experiments are motivated from corresponding behaviors of solutions of the $L^2$-gradient flow of the perimeter functional (cf. [4]).

2. Preliminaries

2.1. General notation. Let $W^{k,p}$, and $H^k := W^{k,2}$ denote standard Sobolev spaces. By $W^{k,p}(W^{m,q}) := W^{k,p}(0, T; W^{m,q})$ we refer to standard Bochner spaces. The space $C(X)$ denotes the space of continuous functions taking values in $X$. Vector-valued functions and spaces containing such functions are written in bold-face notation. We define

$$L^2_0(\Omega) := \left\{ u \in L^2(\Omega) : \int_\Omega u \, dx = 0 \right\}.$$  

The space $V$ (respectively $H$) denotes the closure of $\{v \in C_0^\infty(\Omega) : \text{div} \, v = 0\}$ in the $H^1$-norm (respectively $L^2$-norm). For the scalar products in $L^2$ and $L^2(L^2)$ respectively of $f$ and $g$, we write $(f, g)$ in cases where no confusion arises; otherwise, we add the corresponding space as index to the scalar product. The notation $\|\cdot\|$ stands for the $L^2$ or the $L^2(L^2)$ norm, which will be clear from the context.

The inner product of $X$ and its dual space $X^*$ is written as $\langle \cdot, \cdot \rangle_{X, X^*}$. The space of linear functionals from $X$ to $Y$ is denoted by $L(X, Y)$.

We use $C$ as a generic nonnegative constant; to indicate dependencies, we write $C(\cdot)$.

2.2. Known results. We recall well-known results for the state equation (1.2). For details, we refer to [24].

**Theorem 2.1.** Let $0 < T < \infty$. Assume $u \in L^2(L^2)$, $\Omega \subset \mathbb{R}^d$ ($d = 2, 3$) bounded, open, as well as $y_0 \in H$ and $\rho_0 \in L^\infty(\Omega)$. Then there exists a global weak solution $(y, \rho) \in L^2(V) \times L^\infty(L^\infty)$ of (1.2). Every global weak solution $(y, \rho)$ has the following property: for
every $0 \leq \alpha \leq \beta < \infty$ the measure of $\{ x \in \Omega : \alpha \leq \rho(x, t) \leq \beta \}$ is independent of $t \geq 0$. In particular, we have for almost all $(t, x) \in \Omega_T$

$$0 < \inf_{\Omega} \rho_0 \leq \rho(t, x) \leq \|\rho_0\|_{L^\infty(\Omega)}. \quad (2.1)$$

In addition, the following estimate holds:

$$\|y\|_{L^\infty(H)} + \|y\|_{L^2(V)} \leq C\left(\Omega, T, \|\rho_0\|_{L^\infty}, \|u\|_{L^2(L^2)} \right). \quad (3.1)$$

3. The regularized state equation

The following hypotheses are valid for the rest of the paper. In particular, smoother initial data are required for the following analysis.

**Hypothesis 3.1.** We assume that

1. $0 < T < \infty$.
2. $\Omega \subset \mathbb{R}^2$ is bounded and open, with $\partial \Omega \in C^{1,1}$, or $\Omega$ is polyhedral and convex.
3. $\rho_0 \in H^2(\Omega)$ with $0 < \rho_1 \leq \rho_0 \leq \rho_2 < \infty$.
4. $y_0 \in V$.
5. $\mu(\rho) = \mu > 0$ is constant.

We now want to regularize (1.2). Before doing so, we write the equation in a different way be using the identity (as long as $\rho_t + [y \cdot \nabla] \rho = 0$ -- cf. [25])

$$\rho \left( y_t + [y \cdot \nabla] y \right) = \frac{1}{2} \left( \rho y_t + (\rho y)_t + \rho [y \cdot \nabla] y \right) + \text{div}(\rho y \otimes y). \quad (3.1)$$

This modification is used in order to prove solvability for the numerical scheme (5.5), cf. Lemma 5.1. We note that this modification has no effect in the continuous setting and is only used here to make the continuous optimality system (4.1) and the discrete one (6.2) comparable.

After applying (3.1), we regularize the state equation (1.2) to improve regularity properties of the density $\rho : \Omega_T \to \mathbb{R}$, and of the corresponding Lagrange multiplier $\eta : \Omega_T \to \mathbb{R}$ of the mass equation (3.2b) below. The system then reads (for $\varepsilon > 0$),

$$\frac{1}{2} \rho y_t + \frac{1}{2} (\rho y)_t + \frac{1}{2} \rho [y \cdot \nabla] y + \frac{1}{2} \text{div}(\rho y \otimes y) - \text{div}(\mu(\rho) \nabla y) + \nabla p = \rho u, \quad (3.2a)$$
$$\rho_t + [y \cdot \nabla] \rho - \varepsilon \Delta \rho = 0, \quad (3.2b)$$
$$\text{div} y = 0, \quad (3.2c)$$

together with boundary conditions $y = 0$ and $\partial_n \rho = 0$ on $(0, T) \times \partial \Omega$, as well as the initial conditions $y(0, .) = y_0$, and $\rho(0, .) = \rho_0$.

**Theorem 3.2.** Let $u \in L^2(L^2)$. Then there exists a global weak solution of (3.2) such that

$$y \in Y := L^2(H^2) \cap H^1(H) \subset C(V), \quad (3.3)$$
$$\rho \in R := H^1(H) \cap L^\infty(H^2) \subset C(\Omega_T). \quad (3.4)$$

Furthermore, we have $\rho_1 \leq \rho \leq \rho_2$ a.e. in $\Omega_T$. 

In particular, there exists a constant $C = C(\varepsilon, T, u, y_0, \rho_0) \geq 0$ such that
\[
\|y\|_Y + \|\rho\|_R \leq C.
\]

Bounds on the solution, which are uniform in $\varepsilon > 0$ may be obtained in the following norms:
\[
y \in L^2(V) \cap L^\infty(H) \cap H^1(V^*), \quad \rho \in L^\infty(L^\infty) \cap H^1(H^{-1}). \tag{3.5}
\]

Proof. Existence and the uniform a priori estimates with respect to $\varepsilon > 0$ follow by Schauder’s fixed theorem and standard parabolic theory, similar to [24].
The maximum principle for parabolic equations together with the bounds of $\rho_0$ yields the lower and upper bound for $\rho$.
The regularity of $y$ follows from [24, pp. 32ff], while the regularity of $\rho$ follows from testing (3.2b) with $-\Delta \rho$ and testing the time derivative of (3.2b) with $-\Delta \rho$, respectively. The estimates on $y$ and $\rho$ are both based on Hypothesis 3.1.

To get uniform bounds with respect to $\varepsilon > 0$ in the norms which are indicated in (3.5), we test (3.3) with $y$ and (3.4) with $\rho$, and neglect (non-negative) $\varepsilon$-terms. □

A consequence of Theorem 3.2 is that $p \in L^2(H^1 \cap L^2_0)$.

4. Optimal Control of the Regularized System

Problem 4.1. Let $0 < T < \infty$, and $\varepsilon, \delta > 0$. Minimize $J_\delta$ subject to (3.2).

4.1. Existence.

Theorem 4.2. There exists at least one solution $(\bar{y}, \bar{\rho}, \bar{u}) \in Y \times R \times L^2(L^2)$ of Problem 4.1.

Proof. For every $u \in L^2(L^2)$, Theorem 3.2 ensures the existence of a solution to (3.2); hence the set of feasible points is not empty, and there exists $\bar{J} := \inf J_\delta(y, \rho, u) \geq 0$, where the infimum is taken over all feasible $(y, \rho, u) \in Y \times R \times L^2(L^2)$. Thus, we may consider a minimizing sequence $\{(y_n, \rho_n, u_n)\} \subset Y \times R \times L^2(L^2)$, such that for $n \to \infty$
\[
J_\delta(y_n, \rho_n, u_n) \downarrow \bar{J}.
\]

By the definition of the cost functional $J_\delta$, the sequence $\{u_n\}$ is bounded in $L^2(L^2)$ and – thanks to Theorem 3.2 – the sequences $\{y_n\}$ and $\{\rho_n\}$ are bounded in $Y$ and $R$ respectively. Then there exist $y^* \in Y, \rho^* \in R$, and $u^* \in L^2(L^2)$, such that for corresponding subsequences (not relabeled) and $n \to \infty$
\[
y_n \rightharpoonup y^* \quad \text{weakly in } Y,
\]
\[
\rho_n \rightharpoonup \rho^* \quad \text{weakly in } R,
\]
\[
u_n \rightharpoonup u^* \quad \text{weakly in } L^2(L^2).
\]

We have to show that $(y^*, \rho^*, u^*)$ is a solution of (3.2), and $J_\delta(y^*, \rho^*, u^*) = \bar{J}$.

(1) Following the argumentation in [24, Section 2.4], we can pass to the limit in the momentum equation and in the mass equation (except for the much easier term $\varepsilon \Delta \rho^*$) because of the bounds we have deduced in Theorem 3.2. By the definition of $R$, the distributional limit of the term $-\varepsilon \Delta (\rho^*)$ is $-\varepsilon \Delta \rho^*$. 
have added [15, Section 2.6]. Initial conditions are treated in the same manner as there. Note that we omit boundary conditions in $e$. Optimality conditions.

4.2. Optimality conditions. Next, we show that the Frechet derivative of the side constraints (3.2) is surjective on appropriate spaces, to then derive optimality conditions. These will be compared in Section 8 with the discrete ones from Section 6.

For the next theorem, we use the mapping $e : \left( Y \times L^2(H^1 \cap L^1_0) \times R \times L^2(L^2) \right) \to \left( L^2(L^2) \times L^2(L^2) \times V \times H^2(\Omega) \right)$, which is defined by

$$e(y, p, \rho, u) := \begin{pmatrix}
    e_1(y, p, \rho, u) \\
    e_2(y, p, \rho, u) \\
    e_3(y, p, \rho, u) \\
    a_1(y, p, \rho, u) \\
    a_2(y, p, \rho, u)
\end{pmatrix} = 
\begin{pmatrix}
    \frac{1}{2} \rho y_t + \frac{1}{2} (\rho u)_{t} + \frac{1}{2} \rho |y| y + \frac{1}{2} \text{div}(\rho y \otimes y) - \mu \Delta y - \rho u + \nabla p \\
    \rho p + |y \cdot \nabla| |y| + \frac{1}{2} \text{div}(\rho y \otimes y) - \varepsilon \Delta \rho \\
    y(0, .) - y_0 \\
    \rho(0, .) - \rho_0
\end{pmatrix}.$$

We omit boundary conditions in $e$, which may be treated by standard methods; see e.g. [15, Section 2.6]. Initial conditions are treated in the same manner as there. Note that we have added $\frac{1}{2} \rho \text{div} y = 0$ to the mass equation $e_2$. This term is a standard way to treat the incompressible condition in the context of a finite element approximation, and stabilizes the discrete operator in Section 5. Adding this term here allows us to compare continuous and discrete effects in later sections.

**Theorem 4.3.** Let $0 < T < \infty$, and $\varepsilon, \delta > 0$. The mapping $e$ is well-defined and Frechet differentiable. Moreover, for each $(y, p, \rho, u) \in Y \times L^2(H^1 \cap L^1_0) \times R \times L^2(L^2)$, the derivative $e'(y, p, \rho, u)$ is surjective.

**Proof.** (1) With the estimates in Theorem 3.2 we see that the mapping $e$ is well-defined. (2) The candidate for the derivative of $e$ is

$$\left\langle e'_1(y, p, \rho, u), (\delta y, \delta p, \delta \rho, \delta u) \right\rangle = \rho(\delta y)_t + \frac{1}{2} p(\delta y) + \frac{1}{2} \rho(\delta y \cdot \nabla) y + \frac{1}{2} \rho |y \cdot \nabla| \delta y \\
+ \frac{1}{2} \text{div}(\rho \delta y \otimes y) + \frac{1}{2} \text{div}(\rho \delta y) - \mu \Delta \delta y \\
+ \delta y_t + \frac{1}{2} (\delta \rho)_{t} y + \frac{1}{2} \rho |y \cdot \nabla| y - \frac{1}{2} \delta \rho u \\
+ \frac{1}{2} \text{div}(\rho \delta y \otimes y) - \rho \delta u + \nabla \delta p,$$
\[
\begin{align*}
\langle e'(y, p, \rho, u), (\delta y, \delta p, \delta \rho, \delta u) \rangle &= (\delta \rho)_t + [y \cdot \nabla] \delta \rho + \frac{1}{2} \delta \rho \text{div} y - \varepsilon \Delta \delta \rho \\
&\quad + [\delta y \cdot \nabla] \rho + \frac{1}{2} \rho \text{div} \delta y,
\end{align*}
\]

which is obtained by direct calculation like in \[15\] Section 2.6.

(3) Let \( f, g, h, \varphi, \psi \) \in \( L^2(L^2(\Omega) \times L^2(L^2(\Omega) \times V \times H^2(\Omega)), \) and \( (\varphi, \psi) \in L^2(\Omega) \times L^2(\Omega) \). The existence of solutions \( (\delta y, \delta p, \delta \rho, \delta u) \in Y \times L^2(H^1) \times R \times L^2(L^2) \) for
\[
\langle e'(y, p, \rho, u), (\delta y, \delta p, \delta \rho, \delta u) \rangle = (f, g, h)
\]
\[
\delta y(0, \cdot) = \varphi,
\]
\[
\delta \rho(0, \cdot) = \psi,
\]

together with suitable boundary conditions follows from standard linear parabolic theory, cf. \[23\], and Theorem 3.2.

\[\square\]

Theorem 4.3 allows to apply the Lagrange multiplier theorem (cf. \[26\] Section 9.3), and thus to deduce necessary optimality conditions for Problem 6.1 below. We define the Lagrange functional \( L : Y \times L^2(H^1 \cap L^2) \times R \times L^2(L^2) \times L^2(H^1) \times L^2(L^2) \times L^2(L^2) \rightarrow \mathbb{R} \), via
\[
L(y, p, \rho, u; z, q, \eta) := J(\rho, u) + \langle \eta, \rho_t + [y \cdot \nabla] \rho + \frac{1}{2} \rho \text{div} y - \varepsilon \Delta \rho \rangle_{L^2(L^2), L^2(L^2)}
\]
\[
+ \langle z, \frac{1}{2} \rho y_t + \frac{1}{2} (\delta \rho)_t + \frac{1}{2} \rho [y \cdot \nabla] y - \mu \Delta y + \nabla p - \rho u \rangle_{L^2(L^2), L^2(L^2)}
\]
\[
- \frac{1}{2} \langle \rho [y \cdot \nabla] z, y \rangle_{L^2(H^{-1}), L^2(H^1)} + \langle q, \text{div} y \rangle_{L^2(L^2), L^2(L^2)}
\]

By using the directional derivatives of \( e \) from the proof of Theorem 4.3 together with integration by parts, and setting the derivatives of \( L \) equal to zero then, a straightforward calculation together with methods from \[15\] Section 2.6] leads to the following optimality conditions:

\[
0 = \frac{1}{2} \eta \nabla \rho - \frac{1}{2} \rho \nabla \eta - \frac{1}{2} \rho z_t - \rho z_t + \frac{1}{2} \rho \nabla y z - \frac{1}{2} \nabla \rho \cdot y z - \rho [y \cdot \nabla] z - \frac{1}{2} \rho \nabla y z
\]
\[
0 = \text{div} z,
\]
\[
0 = \lambda (\rho - \rho) - \beta \delta \Delta \rho + \frac{\beta}{8 \delta} W''(\rho) - \eta_t - [y \cdot \nabla] \eta + \varepsilon \Delta \eta
\]
\[
+ \frac{1}{2} z \cdot y_t - \frac{1}{2} y \cdot z_t + \frac{1}{2} [y \cdot \nabla] y \cdot z - \frac{1}{2} [y \cdot \nabla] z \cdot y,
\]
\[
0 = \alpha u - \rho z,
\]
\[ 0 = \rho \dot{y} - \rho [y \cdot \nabla] y - \mu \Delta y - \rho u + \nabla p, \quad (4.1e) \]
\[ 0 = \rho (0, .) = \rho_0, \quad z(T, .) = 0, \quad \eta(T, .) = 0, \quad (4.1f) \]
\[ 0 = \text{div} \ y, \quad (4.1g) \]

together with the initial conditions
\[ y(0, .) = y_0, \quad \rho(0, .) = \rho_0, \quad z(T, .) = 0, \quad \eta(T, .) = 0, \]
and the following homogeneous boundary conditions
\[ y = 0, \quad \partial_n \rho = 0, \quad z = 0, \quad \partial_n \eta = 0 \quad \text{on} \ (0, T] \times \partial \Omega. \]

The Lagrange multiplier theorem assures that this system has at least one solution.

The derivation of initial and boundary conditions is done by a standard argument (cf. [15, Section 2.6]).

Equations (4.1a) and (4.1c) are the adjoint equations, equation (4.1d) is the optimality condition, while equations (4.1e) and (4.1f) are the state equations (3.2a) and (3.2b).

5. Discretization of the state equation

We now consider a discrete version of (3.2), a modification of which is studied in [8].

5.1. Numerical setup and notation. Let \( T_h \) be a quasi-uniform triangulation of \( \Omega \) with \( h := \max_{T \in T_h} \text{diam} \ T \) and
\[ R_h := \{ X_h \in C(\bar{\Omega}) : X_h|_T \in P_k(T) \ \forall T \in T_h \}. \]
We assume that the triangulation is strongly acute, see e.g. [21]. For the finite element approximation, we define the following spaces.

- \( R_h \) for the approximation of the density \( \rho \),
- \( V_h \) and \( M_h \) as an inf-sup stable conforming pair (e.g. Taylor–Hood or MINI Elements) for velocity \( y \) and pressure \( p \), involving zero Dirichlet boundary conditions for \( y \),
- the space of discrete divergence-free functions \( J_h := \{ v_h \in V_h : (\text{div} \ v_h, \chi_h) = 0, \ \forall \chi_h \in M_h \} \).

Recall the discrete Laplace operator \( \Delta_h : R_h \rightarrow R_h \), where
\[ -(\Delta_h V, \Phi) = (\nabla V, \nabla \Phi) \ \forall V, \Phi \in R_h. \]
Analogously, we define the vector-valued discrete Laplacian for the space \( V_h \) by \( \tilde{\Delta}_h : V_h \rightarrow V_h \). The discrete Stokes operator \( A_h \) is defined by \( A_h := -P_h \Delta_h, \) where \( P_h : L^2 \rightarrow J_h \) denotes the \( L^2 \)-projection. For details, we refer to [17, Section 4]. The subset consisting of finite element functions \( V \in R_h \) such that \( \| \Delta_h V \|_{L^2} \leq C < \infty \) with \( C > 0 \) independent of \( h \), will be denoted by \( H^2_{\text{disc}} \subseteq V_h \). The subset \( H^2_{\text{disc}} \subseteq V_h \) is defined in the same way.

Below, we often use the following discrete imbedding and interpolation inequalities (cf. [17, Lemma 4.4.]):
\[ \| \nabla V \|_{L^4} \leq C(\| \Delta_h V \| + \| \nabla V \|), \quad \text{(5.1)} \]
\[ \| \nabla V \|_{L^4} \leq C(\| \nabla V \|^{\frac{1}{2}} (\| \Delta_h V \| + \| \nabla V \|)^{\frac{1}{2}}). \quad \text{(5.2)} \]
\[
\| \nabla V \|_{L^1} \leq C \| \Delta_h V \|, \quad (5.3)
\]
\[
\| \nabla V \|_{L^1} \leq C \| \nabla V \|^{1/2} \| \Delta_h V \|^{1/2}. \quad (5.4)
\]

Let \( t_n := nk \) (for \( n = 0, \ldots, N \), for \( k = \frac{T}{N} \). Let \( (y^0, \rho^0) \in V_h \times R_h \) be the projection of \((y_0, \rho_0)\), with \( \rho_1 \leq R^0 \leq \rho_2 \).

We will use the following notation for discrete functions: The notation \( \{V^n\} \subseteq \mathcal{X}_h \) describes a family of finite element function (in a finite element space \( \mathcal{X}_h \)) evaluated at subsequent times \( t_n \), while \( V : \Omega_T \rightarrow disc \) stands for the piecewise affine, globally continuous time interpolant of \( \{V^n\} \). Moreover, we define the following piecewise constant in time interpolants of \( \{V^n\} \) for \( t \in [t_j, t_{j+1}) \),

\[
V^+(t) := V^{j+1}, \quad V^*(t) := V^j, \quad V^-(t) := V^{j-1}.
\]

For vector-valued functions \( v : \Omega_T \rightarrow \mathbb{R}^n \), we shall write all quantities in boldface style, i.e., \( V^n \) for the discrete iterates, and \( V : \Omega_T \rightarrow \mathbb{R}^n \) for its time interpolant. For the variables \( \rho \) and \( \eta \), we use the capital letters \( \mathcal{R} \) and \( \mathcal{E} \) for the time interpolant, while there should be no confusion with the space \( R \) from Theorem 3.2. The discrete time derivative of the function \( V \) will be denoted as

\[
d_i V^n := \frac{V^n - V^{n-1}}{k}.
\]

The discrete version of (3.2) reads as follows: For \( 1 \leq n \leq N \) find \( (Y^n, P^n, R^n) \in V_h \times M_h \times R_h \) such that for all \( (Z, \Pi, E) \in V_h \times M_h \times R_h \)

\[
\begin{align*}
\left( d_t R^n, E \right) + \varepsilon (\nabla R^n, \nabla E) + \left( [Y^n \cdot \nabla] R^n, E \right) + \frac{1}{2} (R^n \text{ div } Y^n, E) &= 0, \quad (5.5a) \\
\frac{1}{2} (R^{n-1} d_t Y^n, Z) + \frac{1}{2} \left( d_t (R^n Y^n), Z \right) &+ \frac{1}{2} \left( [R^{n-1} Y^{n-1} \cdot \nabla] Y^n, Z \right) - \frac{1}{2} \left( [R^{n-1} Y^{n-1} \cdot \nabla] Z, Y^n \right) + \mu (\nabla Y^n, \nabla Z) + (\nabla P^n, Z) = (R^{n-1} U^n, Z), \quad (5.5b) \\
(\text{div } Y^n, \Pi) &= 0. \quad (5.5c)
\end{align*}
\]

The assumptions on the strongly acute triangulation imply a lower bound for the discrete density. The additional term \( \frac{1}{2} R^n \text{ div } Y^n \) in (5.5a), together with the reformulation in (5.5b) via (3.1) again make the convective operators in (5.5a) and (5.5b) skew-symmetric. For details, we refer to [8]. Just like in [8], we can establish \( \rho_2 \) as an upper bound of \( \{R^n\} \), which is due to the discrete maximum principle.

In the remainder of this section, we derive bounds for the iterates in (5.5), and verify that solutions of (5.5) converge to those of (3.2) for vanishing numerical parameters \( k, h \rightarrow 0 \). We proceed as follows:

(1) Derive uniform bounds for the fully discrete scheme in standard parabolic norms, see Lemma 5.1. By stability of the interpolation, all interpolants inherit these bounds.

(2) Derive uniform bounds in higher norms for the fully discrete scheme for \( \{R^n\} \), i.e., bound \( R \) in \( L^2(H^2_{\text{disc}}) \cap H^1(L^2) \cap L^\infty(H^1) \) uniformly with respect to \( k \) and \( h \), see Lemma 5.2. To do this, we have to test (5.5a) with \( -\Delta_h R^n \) and with \( d_t R^n \).

(3) In Lemma 5.3, we want to bound \( R \) in \( H^1(H^1) \) and \( Y \) in \( L^2(H^2_{\text{disc}}) \cap H^1(L^2) \), which requires a bit more regularity on \( Y \) and \( R \), respectively, than is available from Lemma
In order to conclude, we have to simultaneously test (5.5b) and (5.5a) with different test functions and combine all inequalities. We highlight more details of this strategy at the beginning of the proof of Lemma 5.3. This approach partly mimics ideas from [24, Section 2.2].

(4) With these bounds, and a discrete version of Aubin–Lions’ compactness theorem (see Lemma 5.4), it is possible to derive strong convergence for the affine interpolants.

(5) By arguments from [28], strong convergence also holds for constant in time interpolants. This will lead to the convergence of the scheme (5.5) to (3.2) up to subsequences.

5.2. Stability of the scheme.

Lemma 5.1. Let $0 < T < \infty$ and $\varepsilon > 0$. For every $1 \leq n \leq N$ there exists a solution $(Y^n, P^n, R^n) \in V_h \times M_h \times R_h$ to (5.5) which satisfies

$$\frac{1}{2}d_t \left( \| \sqrt{R^n} Y^n \|^2 \right) + \mu \left( \| \sqrt{R^n-1} \| Y^n \|^2 + \frac{k}{2} \left( \| \sqrt{R^n-1} d_t Y^n \| \| R^n \| \right) \right) = \int_\Omega R^{n-1} U^n Y^n \, dx,$$

$$\frac{1}{2}d_t \| R^n \|^2 + \frac{k}{2} \left( \| d_t R^n \|^2 \right) + \varepsilon \| R^n \|^2 = 0.$$

For small enough $h, k > 0$, and every $1 \leq n \leq N$ there holds

$$0 < \rho_1 \leq R^n \leq \rho_2 < \infty. \quad (5.6)$$

In particular, we have the following uniform bounds for $Y$ and $R$:

$$\| Y \|_{L^\infty(L^1)} + \| Y \|_{L^2(H^1)} + \| R \|_{L^\infty(L^1)} + \| R \|_{L^2(H^1)} \leq C(\varepsilon, u, T).$$

The bounds also hold for $Y^\ast\ast$ and $R^\ast\ast$ respectively.

Proof. This lemma relies on [8] Lemma 3.1 and can be proven with small modifications. □

Lemma 5.2. There holds uniformly with respect to $k, h > 0$

$$\| \Delta_h R^n \|^2_{L^2(L^2)} + \| \nabla R \|^2_{L^\infty(L^2)} + \| d_t R \|^2_{L^2(L^2)} \leq C(\varepsilon, T).$$

Proof. Step 1: Test (5.5a) with $-\Delta_h R^n \in R_h$,

$$(d_t \nabla R^n, \nabla R^n) + \varepsilon \| \Delta_h R^n \|^2 = \left( [Y^n \cdot \nabla] R^n, \Delta_h R^n \right) + \frac{1}{2} (R^n \operatorname{div} Y^n, \Delta_h R^n) =: I_1 + I_2.$$

For $\sigma > 0$, we estimate both terms by

$$I_1 \leq \sigma \| \Delta_h R^n \|^2 + C(\sigma) \| \nabla Y^n \|^2,$$

$$I_2 \leq \sigma \| \Delta_h R^n \|^2 + C(\sigma) \| Y^n \|^2 \| \nabla R^n \|^2 \| \nabla \Delta_h R^n \|^2 \leq \sigma \| \Delta_h R^n \|^2 + C(\sigma) \left( \| Y^n \|^2 \| \nabla Y^n \|^2 + \| Y^n \|^2 \| \nabla R^n \| \right) \| \Delta_h R^n \| \| \nabla R^n \|^2,$$

where we used (5.6). By Lemma 5.1 and an appropriate choice of $\sigma$, we may conclude by Gronwall’s inequality to bound $\Delta_h R^n \in L^2(L^2)$ and $\nabla R \in L^\infty(L^2)$. 

Step 2: In order to show bounds for \( d_t R \), we test \( 5.5a \) with \( d_t R^n \in R_h \) and get
\[
\|d_t R^n\|^2 + \varepsilon (d_t \nabla R^n, \nabla R^n) = -\langle (Y^n \cdot \nabla) R^n, d_t R^n \rangle - \frac{1}{2} (R^n \div Y^n, d_t R^n) =: I_{11} + I_{12}.
\]
Both terms can be estimated as follows for
\[
\|d_t R^n\|^2 \leq C(\sigma) \left( \|Y^n\|^2 \|\nabla Y^n\|^2 + \|\nabla Y^n\|^2 \|\nabla Y^n\|^2 \right) + C(\sigma) \Delta h R^n, \|\Delta_h R^n\|^2,
\]
\[
I_{12} \leq \sigma \|d_t R^n\|^2 + C(\sigma) \|\nabla Y^n\|^2.
\]
Again, we conclude by Gronwall’s inequality, Lemma 5.5, and the first part of this proof.

Lemma 5.3. There holds uniformly in \( k, h > 0 \)
\[
\|\mathcal{Y}\|_{H^1(L^2)} + \|\mathcal{Y}\|_{L^\infty(H^1_0)} + \|\Delta_h \mathcal{Y}\|_{L^2(L^2)} + \|\nabla d_t R\|_{L^2(L^2)} \leq C(\varepsilon, T),
\]
as long as \( k \leq k_0(\Omega, \rho_{min}, \rho_{max}, T, \mu, \varepsilon) \) is sufficiently small.

Proof. Before starting with the technical part let us mention the main difficulties to overcome in the proof: We will first test \( 5.5b \) with \( d_t Y^n \) and \( A_k Y^n \) in order to get positive terms to obtain the desired norms for \( \mathcal{Y} \). In the following calculation, the terms \( I_1 \) and \( K_2 \) are responsible for additional terms with no corresponding positive term, and which are not accessible to a Gronwall type argument. These new bad terms are \( \|\Delta h Y^n\| \) and \( \|\nabla d_t R^n\| \), respectively. Luckily, some of the bad terms are obtained with an arbitrary small constant, which allows to complete the proof at the end. In the first three steps, we will deduce independently three inequalities; in the last step, we will combine them in a proper manner and deduce the desired bounds. Throughout the proof, with the notation \( \hat{J}, \hat{L}, \hat{N} \), we will denote functions which are summable in time by Lemma 5.1 and Lemma 5.2 i.e., \( k \sum \hat{J}, \ldots < \infty \) uniformly with respect to \( k, h > 0 \).

Step 1: Choose \( Z = d_t Y^n \) in \( 5.5b \),
\[
\|\sqrt{R^n-1} d_t Y^n\|^2 + \frac{k}{2} \|d_t \nabla Y^n\|^2 + \frac{k}{2} \|d_t \nabla Y^n\|^2 \\
\leq \frac{1}{2} \left( |(d_t R^n Y^n, d_t Y^n)| + \frac{1}{2} \left| \left( |R^n-1 Y^n\cdot \nabla Y^n, d_t Y^n\right) \right| \right) \\
+ \frac{1}{2} \left| \left( |R^n-1 Y^n\cdot \nabla Y^n, d_t Y^n\right) \right| + \left| (R^n-1 U^n, d_t Y^n) \right| =: I_1 + I_2 + I_3 + I_4.
\]
We derive estimates for each term \( I_1, \ldots, I_4 \) separately. Let \( \sigma, \tau, \theta > 0 \). We calculate
\[
I_1 \leq \sigma \|d_t Y^n\|^2 + C(\sigma) \|d_t R^n\|_{L^4} \|Y^n\|_{L^4}^2
\]
\[
\leq \sigma \|d_t Y^n\|^2 + C(\sigma) \|d_t R^n\| \left( \|d_t \nabla R^n\| + \|d_t R^n\| \right) \|Y^n\| \|\nabla Y^n\|
\]
\[
\leq \sigma \|d_t Y^n\|^2 + \tau \|d_t \nabla R^n\|^2 + C(\sigma, \tau) \|d_t R^n\|^2 \|Y^n\|^2 \|\nabla Y^n\|^2 + C(\sigma, \tau) \|d_t R^n\|^2
\]
\[
=: \sigma \|d_t Y^n\|^2 + \tau \|\nabla d_t R^n\|^2 + C(\sigma, \tau) \|\nabla Y^n\|^2 + C(\sigma, \tau) \hat{J}_1
\]
\[
I_2 \leq \sigma \|d_t Y^n\|^2 + C(\sigma) \|Y^n-1\| \|\nabla Y^n\| - 1 \|\nabla Y^n\| \|\hat{\Delta}_h Y^n\|
\]
\[
\leq \sigma \|d_t Y^n\|^2 + \theta \|\hat{\Delta}_h Y^n\|^2 + C(\sigma, \theta) \|Y^n-1\|^2 \|\nabla Y^n\|^2
\]
\[
=: \sigma \|d_t Y^n\|^2 + \theta \|\hat{\Delta}_h Y^n\|^2 + C(\sigma, \theta) \hat{J}_2 \|\nabla Y^n\|^2,
\]
where we used Sobolev imbeddings and Gagliardo-Nirenberg inequalities. By Lemma 5.2 we have \( k \sum J_1 + J_2 + \tilde{J}_1 \leq C < \infty \) uniformly with respect to \( h, k > 0 \), but depending on \( \varepsilon, T > 0 \). Integration by parts yields
\[
I_3 \leq \left| \left( (\nabla R^{n-1} \cdot Y^{n-1}) d_1 Y^n, Y^n \right) \right| + \left| \left( \left( R^{n-1} \text{div} Y^{n-1} \right) d_1 Y^n, Y^n \right) \right| + I_2 =: I_{3a} + I_{3b} + I_2.
\]

We estimate with (5.2), (5.4), and Sobolev embeddings,
\[
I_{3a} \leq \sigma \| d_1 Y^n \|^2 + C(\sigma) \| \nabla R^{n-1} \|_{L^4} \| Y^{n-1} \|_{L^4} \| Y^n \|_{L^8}^2
\]
\[
\leq \sigma \| d_1 Y^n \|^2 + C(\sigma) \| \nabla R^{n-1} \| \left( \| \Delta_h R^{n-1} \| + \| \nabla R^{n-1} \| \right) \times
\]
\[
\times \| \nabla Y^{n-1} \|^2 || Y^{n-1} \| \| Y^n \|^2 \| Y^n \|^2
\]
\[
\leq \sigma \| d_1 Y^n \|^2 + C(\sigma) \| \nabla R^{n-1} \| \left( \| \Delta_h R^{n-1} \| + \| \nabla R^{n-1} \| \right) \times
\]
\[
\times \| \nabla Y^{n-1} \|^2 || Y^{n-1} \| \| Y^n \|^2 \left( \| \nabla Y^{n-1} \|^2 + \| \nabla Y^n \|^2 \right)
\]
\[
=: \sigma \| d_1 Y^n \|^2 + C(\sigma) J_{3a} \left( \| \nabla Y^{n-1} \|^2 + \| \nabla Y^n \|^2 \right).
\]

\[
I_{3b} \leq \sigma \| d_1 Y^n \|^2 + C(\sigma) \| Y^n \|_{L^4}^2 \| \nabla Y^{n-1} \|_{L^4}^2
\]
\[
\leq \sigma \| d_1 Y^n \|^2 + C(\sigma) \| Y^n \| \| \nabla Y^n \| \| \nabla Y^{n-1} \| \| \Delta_h Y^{n-1} \|
\]
\[
\leq \sigma \| d_1 Y^n \|^2 + \theta \| \Delta_h Y^{n-1} \|^2 + C(\sigma, \theta) \| Y^n \|^2 \| \nabla Y^n \|^2 \| \nabla Y^{n-1} \|^2
\]
\[
=: \sigma \| d_1 Y^n \|^2 + \theta \| \Delta_h Y^{n-1} \|^2 + C(\sigma, \theta) J_{3b} \| \nabla Y^n \|^2,
\]

\[
I_4 \leq \sigma \| d_1 Y^n \|^2 + C(\sigma) \| U^n \|^2,
\]

where we have used the inequality valid in space dimension \( d = 2 \),
\[
\| v \|_{L^8} \leq C \| \nabla v \| \| v \| \quad \forall v \in H_0^1(\Omega).
\]

Like above, \( J_{3a} \) and \( J_{3b} \) are uniformly summable by Lemmas 5.1 and 5.2.

By choosing an approximate \( \sigma > 0 \), we deduce for a summable function \( \tilde{J} \) that
\[
\frac{1}{2} \| \sqrt{R^{n-1}} d_1 Y^n \|^2 + \frac{\mu}{2} d_1 \| \nabla Y^n \|^2 + \frac{\mu}{2} k d_1 \| \nabla Y^n \|^2
\]
\[
\leq \theta \| \Delta_h Y^n \|^2 + \tau \| \nabla d_1 R^n \|^2 + C(\tau, \theta) \tilde{J} \left( \| \nabla Y^{n-1} \|^2 + \| \nabla Y^n \|^2 \right) + \tilde{J}. \tag{5.7}
\]

**Step 2:** By the definition of the projection \( P_h \), we have
\[
\mu(\Delta_h Y^n, P_h \Delta_h Y^n) = \mu \| P_h \Delta_h Y^n \|^2 = \mu \| A_h Y^n \|^2.
\]

By \cite{17} Corollary 4.4, every \( V \in J_h \) satisfies
\[
\| \Delta_h V \| \leq C \| A_h V \|.
\]

We may now test (5.5b) with \( A_h Y^n \) to conclude
\[
C \mu \| \Delta_h Y^n \|^2 \leq \mu \| A_h Y^n \|^2
\]
\[
\leq \frac{1}{2} \left| \langle R^{n-1}d_t Y^n, P_h \Delta_h Y^n \rangle \right| + \frac{1}{2} \left| \langle d_t(Y^n), P_h \Delta_h Y^n \rangle \right|
\]
\[
+ \frac{1}{2} \left| \langle R^{n-1}Y^{n-1} \cdot \nabla Y^n, P_h \Delta_h Y^n \rangle \right| + \frac{1}{2} \left| \langle R^{n-1}Y^{n-1} \cdot \nabla P_h \Delta_h Y^n, Y^n \rangle \right|
\]
\[
+ \langle R^{n-1}U^n, P_h \Delta_h Y^n \rangle \right| =: K_1 + K_2 + K_3 + K_4 + K_5.
\]

Exactly like in the first step, we estimate all terms \(K_1, \ldots, K_5\), use similar argument like in the first step as well as \(\|A_h Y^n\| \leq \|\Delta_h Y^n\|\); then we get for some \(\sigma, \tau > 0\)

\[K_1 \leq \sigma \|\Delta_h Y^n\|^2 + C_Y(\sigma) \|d_t Y^n\|^2,\]
\[K_2 \leq \sigma \|\Delta_h Y^n\|^2 + C_Y(\sigma) \|d_t Y^n\|^2 + C(\sigma) \|d_t R^n\|^2 \|Y^n\|^2 L^4\]
\[\leq \sigma \|\Delta_h Y^n\|^2 + C_Y(\sigma) \|d_t Y^n\|^2 + C(\sigma) \|d_t R^n\|^2 \|Y^n\|^2 L^4\]
\[\leq \sigma \|\Delta_h Y^n\|^2 + C_Y(\sigma) \|d_t Y^n\|^2 + C(\sigma) \|d_t R^n\|^2 \|Y^n\|^2 \|\nabla Y^n\|^2,\]
\[=: \sigma \|\Delta_h Y^n\|^2 + C_Y(\sigma) \|d_t Y^n\|^2 + C(\sigma) \|d_t Y^n\|^2 \|\nabla Y^n\|^2,\]
\[K_3 \leq \sigma \|\Delta_h Y^n\|^2 + C(\sigma) \|Y^{n-1}\|^2 L^4 \|\nabla Y^n\|^2 L^4\]
\[\leq \sigma \|\Delta_h Y^n\|^2 + C(\sigma) \|Y^{n-1}\| \|\nabla Y^n\|^2 \|\nabla Y^n\|^2 \|\Delta_h Y^n\|^2\]
\[\leq \sigma \|\Delta_h Y^n\|^2 + C(\sigma) \|Y^{n-1}\| \|\nabla Y^n\|^2 \|\nabla Y^n\|^2\]
\[=: \sigma \|\Delta_h Y^n\|^2 + C(\sigma) \|\nabla Y^n\|^2,\]
\[K_4 \leq K_3 + \left( \|R^n \cdot Y^{n-1}\Delta_h Y^n, Y^n\| \right) + \left( \|R^n \cdot \nabla Y^n\Delta_h Y^n, Y^n\| \right) =: K_3 + K_{4a} + K_{4b},\]
\[K_{4a} \leq \sigma \|\Delta_h Y^n\|^2 + C(\sigma) \|R^n\|^2 L^4 \|Y^{n-1}\| L^4 \|Y^n\|^2 L^4\]
\[\leq \sigma \|\Delta_h Y^n\|^2 + C(\sigma) \|R^n\|^2 \|Y^{n-1}\| \left( \left( \|\Delta_h R^{n-1}\| + \|\nabla R^{n-1}\| \right) \right)\]
\[\times \|\nabla Y^{n-1}\| \|Y^n\|^2 \|\nabla Y^n\|^2\]
\[=: \sigma \|\Delta_h Y^n\|^2 + C(\sigma) \|\nabla Y^{n-1}\| \|Y^n\|^2 \|\nabla Y^n\|^2,\]
\[K_{4b} \leq \sigma \|\Delta_h Y^n\|^2 + C(\sigma) \|Y^n\|^2 L^4 \|\nabla Y^{n-1}\|^2 L^4\]
\[\leq \sigma \|\Delta_h Y^n\|^2 + C(\sigma) \|Y^n\|^2 \|\nabla Y^n\|^2 \|\Delta_h Y^{n-1}\|\]
\[\leq \sigma \|\Delta_h Y^n\|^2 + \sigma \|\Delta_h Y^{n-1}\|^2 + C(\sigma) \|Y^n\|^2 \|\nabla Y^n\|^2 \|\nabla Y^{n-1}\|^2\]
\[=: \sigma \|\Delta_h Y^n\|^2 + \sigma \|\Delta_h Y^{n-1}\|^2 + C(\sigma) \|\nabla Y^n\|^2,\]
\[K_5 \leq \sigma \|\Delta_h Y^n\|^2 + C(\sigma) \|U^n\|^2,\]

where we used integration by parts for term \(K_4\). For each \(L_i\) we get uniform bounds \(k \sum L_i \leq C < \infty\) by Lemmas \[5.2\] and \[5.1\]. Choosing \(\sigma > 0\) small enough then leads for a summable function \(\hat{L}\) and a constant \(C_Y \in \mathbb{R}\) to the following estimate

\[\|\Delta_h Y^n\|^2 \leq \tau \|\nabla d_t R^n\| + C_Y \|d_t Y^n\|^2 + \hat{L} + C(\tau) \hat{L} \left( \|\nabla Y^{n-1}\|^2 + \|\nabla Y^n\|^2 \right).\]  

(5.8)

**Step 3:** We take the time derivative of (5.5a), which reads as

\[d_t R^n - \varepsilon \Delta_h d_t R^n = -[d_t Y^n \cdot \nabla] R^n - [Y^{n-1} \cdot \nabla] d_t R^n - \frac{1}{2} d_t R^n \div Y^{n-1} - \frac{1}{2} R^n \div d_t Y^n.\]  

(5.9)
We calculate for \( \sigma > 0 \)
\[
M_1 \leq \lambda \|d_t Y^n\|^2 + C(\lambda) \|\nabla R^n\|_{L^4}^2 \|d_t R^n\|^2
\]
\[
\leq \lambda \|d_t Y^n\|^2 + C(\lambda) \|\nabla R^n\| \|\Delta_h R^n\| \|d_t R^n\| \|\nabla d_t R^n\|
\]
\[
\leq \lambda \|d_t Y^n\|^2 + \sigma \|\nabla d_t R^n\|^2 + C(\lambda, \sigma) \|\Delta_h R^n\|^2 \|d_t R^n\|^2
\]
\[
=: \lambda \|d_t Y^n\|^2 + \sigma \|\nabla d_t R^n\|^2 + C(\lambda, \sigma) N_1 \|d_t R^n\|^2,
\]
\[
M_2 \leq \sigma \|\nabla d_t R^n\|^2 + C(\sigma) \|Y^{n-1}\|_{L^4}^2 \|d_t R^n\|^2
\]
\[
\leq \sigma \|\nabla d_t R^n\|^2 + C(\sigma) \|Y^{n-1}\| \|\nabla Y^{n-1}\| \|d_t R^n\| \left( \|\nabla d_t R^n\| + \|\nabla R^n\| \right)
\]
\[
\leq \sigma \|\nabla d_t R^n\|^2 + C(\sigma) \left( \|Y^{n-1}\|^2 + \|\nabla Y^{n-1}\|^2 \right) \|d_t R^n\|^2
\]
\[
=: \sigma \|\nabla d_t R^n\|^2 + C(\sigma) N_2 \|d_t R^n\|^2,
\]
\[
M_3 \leq \|\nabla Y^{n-1}\| \|d_t R^n\|^2_{L^4} \leq \|\nabla Y^{n-1}\| \|d_t R^n\| \left( \|\nabla d_t R^n\| + \|\nabla R^n\| \right)
\]
\[
\leq \sigma \|\nabla d_t R^n\|^2 + C(\sigma) \left( \|\nabla Y^{n-1}\|^2 + \|\nabla Y^{n-1}\| \right) \|d_t R^n\|^2
\]
\[
=: \sigma \|\nabla d_t R^n\|^2 + C(\sigma) N_3 \|d_t R^n\|^2,
\]
\[
M_4 = \frac{1}{2} \langle d_t Y^n, \nabla (R^n d_t R^n) \rangle = \frac{1}{2} \langle d_t Y^n, \nabla R^n d_t R^n \rangle + \frac{1}{2} \langle d_t Y^n, R^n \nabla d_t R^n \rangle =: M_{4a} + M_{4b},
\]
\[
M_{4a} \leq \lambda \|d_t Y^n\|^2 + C(\lambda) \|\nabla R^n\|^2_{L^4} \|d_t R^n\|^2
\]
\[
\leq \lambda \|d_t Y^n\|^2 + C(\lambda) \|\nabla R^n\| \left( \|\Delta_h R^n\| + \|\nabla R^n\| \right) \|d_t R^n\| \left( \|\nabla d_t R^n\| + \|\nabla R^n\| \right)
\]
\[
\leq \sigma \|\nabla d_t R^n\|^2 + \lambda \|d_t Y^n\|^2 + C(\lambda) \left( \|\nabla R^n\|^2 \left( \|\Delta_h R^n\| + \|\nabla R^n\| \right) \right) \|d_t R^n\|^2
\]
\[
+ \|\nabla R^n\| \left( \|\Delta_h R^n\| + \|\nabla R^n\| \right) \|d_t R^n\| \right) \left( \|d_t R^n\|^2 \right)
\]
\[
=: \sigma \|\nabla d_t R^n\|^2 + \lambda \|d_t Y^n\|^2 + C(\sigma, \lambda) N_{4a} \|d_t R^n\|^2,
\]
\[
M_{4b} \leq \sigma \|\nabla d_t R^n\|^2 + C(\sigma) \|d_t Y^n\|^2.
\]
All functions \( N_i \) are summable in time, i.e., we have \( \sum \|N_i\| < \infty \) uniformly in \( k, h > 0 \). For an appropriate choice of \( \sigma \) and \( N \) being a summable function and \( C_{Y, 2} \in \mathbb{R} \), we arrive at
\[
d_t \|d_t R^n\|^2 + \varepsilon \|\nabla d_t R^n\|^2 \leq C_{Y, 2} \|d_t Y^n\|^2 + C \tilde{N} \|d_t R^n\|^2. \tag{5.10}
\]

**Step 4:** We insert (5.8) into (5.7), choose \( \theta > 0 \) small enough, and arrive at
\[
\frac{1}{4} \|\nabla R^{n-1} d_t Y^n\|^2 + \frac{\mu}{2} \|d_t \nabla Y^n\|^2 + \frac{\mu}{2} \|k d_t \nabla Y^n\|^2
\]
We add consisting of all above functions $k \sum \tilde{F} < \infty$ uniformly in $k,h > 0$ where here and below $\tilde{F}$ is generic summable function $k \sum \tilde{F} < \infty$ uniformly in $k,h > 0$ consisting of all above functions $\tilde{F}, \tilde{L}, \tilde{N}$. Inserting (5.12) into (5.8), we get
\[
\|\tilde{\Delta}_h Y^n\|^2 \leq \tau \|d_t R^n\|^2 + \tilde{F} + C(\tau) \tilde{F} \left( \|\nabla Y^{n-1}\|^2 + \|\nabla Y^n\|^2 \right).
\] (5.12)

We conclude with the discrete version of Gronwall’s lemma to obtain for sufficiently small $k \leq k_0(\Omega, \rho_{\text{min}}, \rho_{\text{max}}, T, \mu, \varepsilon)$ all bounds expect for $\tilde{\Delta}_h Y$. This bounds can be derived by inserting all existing bounds into (5.12).

The lemma above is the reason where we need $\mu > 0$ to be constant.

5.3. Convergence of the scheme. In order to pass to the limit, we need a discrete version of Aubin–Lions’ compactness theorem.

Lemma 5.4. Let $\mathcal{T}_h$ be a quasi-uniform triangulation of $\Omega$, let $\{\Phi^n_h\}_{n=0}^N \subset V_h$ and $\mathcal{P}$ is its time interpolation, such that
\[
\|\mathcal{P}\|_{H^1(L^2)} + \|\mathcal{P}\|_{L^2(H^1)} + \|\tilde{\Delta}_h \mathcal{P}\|_{L^2(L^2)} \leq C.
\]

Then, there exist a subsequence $\{\mathcal{P}\}_{k,h} \subset L^2(H^1)$ (not relabeled), and $\mathcal{P} \in L^2(H^1)$ such that $\mathcal{P} \to \mathcal{P}$ in $L^2(H^1)$ for $h,k \to 0$.


The same result holds for the discrete Laplacian $\Delta_h$, which is defined on $R_h$.

Lemma 5.5. There exists a subsequence (not relabeled) such that for $h,k \to 0$ holds

1. $Y^{n/-}, Y, \Delta_h Y \in L^\infty(H^0_0)$ and $R^{n/-}, R, \Delta_h R \in L^\infty(H^1)$,
2. $Y \to y$ in $L^2(H^1_0)$ and $R \to \rho$ in $L^2(H^1)$,
3. div $Y \to$ div $y = 0$ in $L^2(L^2)$,
4. $\mathcal{R} \to \rho$ in $L^q(L^q)$ for $1 \leq q < \infty$.

Proof. (1) This is a consequence of Lemmas 5.3 and 5.2.
(2) This follows from the estimates in Lemmas 5.3 and 5.2 respectively, as well as Lemma 5.4.
(3) This follows from the second part and (5.5c).
(4) Since div $Y \to 0$ in $L^2(L^2)$, we conclude by [8] Lemma 3.2] that $\mathcal{R} \to \rho$ in $L^2(L^2)$.
By the $L^\infty$-bound of $\mathcal{R}$ and this information, it follows easily by Hölder’s inequality that $\mathcal{R} \to \rho$ in $L^q(L^q)$ for every $1 \leq q < \infty$. 

□
The bounds from Lemmas 5.1, 5.2, and 5.3 yield the convergence of all linear terms in (5.5), and the convergence results from Lemma 5.5 are sufficient for the convergence of all nonlinear terms in (5.5); we refer to [8] for details.

In the next sections, we consider a discretization of Problem 4.1 and derive first order optimality conditions in Section 6. The main goal in Section 7 is to verify stability estimates for a solution of the adjoint equation. A main problem to achieve this is that the adjoint equation couples adjoint variables with primal variables. The stability of the adjoint equation together with the bounds from this section are used in Section 8 to practically construct weak solutions of the continuous optimality conditions (4.1).

6. Discrete Optimization Problem

We define a discretization of Problem 4.1 and show existence of a minimum. The finite-dimensional version of the Lagrange multiplier theorem directly yields existence of a solution to the related discrete optimality system (6.2), which corresponds to (4.1). The “first discretize, then optimize” ansatz allows to benefit from the results in the stability analysis for (6.2).

**Problem 6.1.** Let \( \varepsilon, \delta > 0 \), and \( h, k > 0 \). Minimize \( J_{\delta,h,k} : \{ R_n \}_{n=1}^N \times \{ L^2(\Omega) \}_{n=1}^N \) via

\[
J_{\delta,h,k}(R,U) := \frac{\lambda}{2} k \sum_{n=1}^N \int_{\Omega} |R_n - \tilde{\rho}(t_n)|^2 \, dx + \frac{\alpha}{2} k \sum_{n=1}^N \int_{\Omega} |U_n|^2 \, dx + \frac{\beta}{2} k \sum_{n=1}^N \int_{\Omega} \left\{ \delta |\nabla R_n|^2 + \frac{1}{4\delta} W(R_n) \right\} \, dx
\]

(6.1)

subject to (5.5).

A variational discretization for \( \{U^n\}_n \) is used, i.e., every iterate \( U^n \) is in \( L^2(\Omega) \) a priori; however, \( \{U^n\}_n \) is discretized implicitly through (6.2d) by means of \( R \) and \( Z \), and its time interpolant is bounded uniformly in \( L^2(L^2) \). The advantages are an easier analysis and a natural discretization, i.e., we do not have to consider projections in (6.2d). We refer to [18] for details. At the end of Section 8, we show that \( U \rightarrow u \) in \( L^2(L^2) \).

Similar to Theorem 4.2, we can prove the following theorem for the finite-dimensional Problem 6.1.

**Theorem 6.2.** There exists at least one solution of Problem 6.1.

We state the Lagrange functional and consider derivatives of it with respect to all unknowns. Like in Section 4, we can use the Lagrange multiplier theorem in order to derive optimality conditions. We define the Lagrange functional via

\[
\mathcal{L}_{h,k}(Y, P, R, U; Z, Q, E)
\]
\[ := J_{\delta,h,k}(R, U) + k \sum_{n=1}^{N} \left( d_t R^n + [Y^n \cdot \nabla] R^n + \frac{1}{2} R^n \text{ div } Y^n - \varepsilon \Delta_h R^n, E^n \right) \]
\[ + k \sum_{n=1}^{N} \left( \frac{1}{2} R^{n-1} d_t Y^n + \frac{1}{2} d_t (R^n Y^n) + \frac{1}{2} [R^{n-1} Y^{n-1} \cdot \nabla] Y^n, Z^n \right) \]
\[ + k \sum_{n=1}^{N} \left( -\mu \Delta_h Y^n + \nabla P^n - R^{n-1} U^n, Z^n \right) \]
\[ - k \sum_{n=1}^{N} \frac{1}{2} \left( [R^{n-1} Y^{n-1} \cdot \nabla] Z^n, Y^n \right) + k \sum_{n=1}^{N} (\text{div } Y^n, Q^n). \]

The first line stands for equation (5.5a), and the following lines for (5.5b) and (5.5c), respectively. For notational simplicity, let \( L = L_{h,k} \) below.

The derivatives of \( L \) with respect to the Lagrange multipliers \( \{Z^n\}_n \), \( \{P^n\}_n \), and \( \{E^n\}_n \) lead to (5.5). Setting all derivatives of \( L \) equal to zero, we may infer by the Lagrange multiplier theorem and integration by parts in the same manner as in Section 4 that the following system has at least one weak solution:

\[ 0 = \frac{1}{2} E^n \nabla R^n - \frac{1}{2} R^n \nabla E^n - \frac{1}{2} d_t R^n Z^n - R^n d_t Z^{n+1} + \frac{1}{2} R^n \nabla Y^{n+1} \cdot Z^{n+1} \]
\[ - \frac{1}{2} R^n \nabla Z^{n+1} \cdot Y^{n+1} - \frac{1}{2} (\nabla R^n - Y^{n-1}) Z^n - \frac{1}{2} R^{n-1} \text{ div } Y^n - Z^n \]
\[ - [R^{n-1} Y^{n-1} \cdot \nabla] Z^n - \mu \Delta_h Z^n - \nabla Q^n, \]
\[ 0 = - \text{ div } Z^n, \]
\[ 0 = - d_t E^{n+1} - [Y^n \cdot \nabla] E^n + \frac{1}{2} (\text{div } Y^n) E^n - \varepsilon \Delta_h E^{n+1} + \frac{1}{2} d_t Y^{n+1} \cdot Z^{n+1} \]
\[ - \frac{1}{2} [Y^n \cdot d_t Z^{n+1} + \frac{1}{2} [Y^n \cdot \nabla] Y^{n+1} \cdot Z^{n+1} - U^{n+1} \cdot Z^{n+1} - \frac{1}{2} [Y^n \cdot \nabla] Z^{n+1} \cdot Y^{n+1} + \lambda (R^n - \bar{\rho}(t_n)) - \beta \delta \Delta_h R^n + \frac{\beta}{8\delta} W'(R^n), \]
\[ 0 = \alpha U^n - R^{n-1} Z^n, \]

together with the final conditions \( E^{N+1} = 0 \), \( Z^{N+1} = 0 \) and \( Q^{N+1} = 0 \) and zero boundary conditions for \( \{Z^n\}_n \). The complete optimality system includes (5.5).

7. Stability of the discrete adjoint equation

We derive uniform bounds for existing solutions of (6.2). These results are used in Section 8 in order to identify the limit of (6.2) for \( h, k \to 0 \).

**Lemma 7.1.** There holds uniformly in \( k, h > 0 \)
\[ \|Z_t\|_{L^2(L^2)}^2 + \|Z\|_{L^\infty(L^2)} + \|Z\|_{L^2(H^{1+})} + \varepsilon \|\mathcal{E}\|_{L^2(H^{1+})} + \|\mathcal{E}\|_{L^\infty(L^2)} + \|\mathcal{E}\|_{L^2(L^\infty)} \leq C(\varepsilon, T), \]
as long as \( k \leq k_0(\Omega, \rho_{\min}, \rho_{\max}, T, \mu, \varepsilon) \) is sufficiently small.

**Proof.** The proof consists of four steps. We test (6.2a) with \( Z^n \) and \( d_t Z^n \), and we test (6.2c) with \( E^n \). The crucial terms arising in the analysis below are in particular \( I_2 \), \( II_1 \) and \( III_2 \).
By elementary algebraic calculations, we find

\[ I = \frac{1}{2} \left( d_t R^n Z^n, Z^n \right) - \frac{1}{2} \left( E_t \nabla R^n, Z^n \right) + \frac{1}{2} \left( R^n \nabla E^n, Z^n \right) - \frac{1}{2} \left( R^n \nabla Y^{n+1} \cdot Z^{n+1}, Z^n \right) + \frac{1}{2} \left( \nabla R^{-1} \cdot Y^{n-1} \right) Z^n \]

\[ + \left( \left( R^{-1} Y^{n-1} \cdot \nabla \right) Z^n, Z^n \right). \]

By elementary algebraic calculations, we find

\[- \left( R^n d_t Z^{n+1}, Z^n \right) = \frac{1}{2} d_t \sqrt{R^n} Z^{n+1} \parallel Z^n \parallel^2 - \frac{1}{2} \left( d_t R^n Z^n, Z^n \right) + \frac{1}{2} k \sqrt{R^n} d_t Z^n \parallel Z^n \parallel^2. \]

Hence, we get

\[- \frac{1}{2} d_t \sqrt{R^n} Z^{n+1} \parallel Z^n \parallel^2 + \frac{k}{2} \sqrt{R^n} d_t Z^n \parallel Z^n \parallel^2 + \mu \parallel \nabla Z^n \parallel^2 \]

\[ \leq \parallel d_t R^n \parallel \parallel Z^n \parallel^2 + \frac{1}{2} \parallel \nabla R^n \parallel_{L^4} \parallel E^n \parallel_{L^4} \parallel Z^n \parallel + C \parallel \nabla E^n \parallel \parallel Z^n \parallel \]

\[ + C \parallel \nabla Y^{n+1} \parallel \parallel Z^{n+1} \parallel_{L^4} \parallel Z^n \parallel_{L^4} + C \parallel \nabla Z^n \parallel \parallel Y^{n-1} \parallel_{L^\infty} \parallel Z^n \parallel \]

\[ + \frac{1}{2} \parallel \nabla R^{-1} \parallel_{L^4} \parallel Y^{n-1} \parallel_{L^4} \parallel Z^n \parallel_{L^4}^2 + C \parallel \nabla Y^{n-1} \parallel_{L^2} \parallel Z^n \parallel_{L^2}^2 \]

\[ + C \parallel Y^{n-1} \parallel_{L^\infty} \parallel \nabla Z^n \parallel_{L^2} \parallel Z^n \parallel = I_1 + I_2 + \ldots + I_8. \]

For a small \( \sigma, \theta > 0 \), we calculate with the same tools as in the proofs of Lemmas 5.2 and 5.3,

\[ I_1 \leq \sigma \parallel \nabla Z^n \parallel^2 + C(\sigma) \parallel d_t R^n \parallel^2 \parallel Z^n \parallel^2 =: \sigma \parallel \nabla Z^n \parallel^2 + C(\sigma) J_1 \parallel Z^n \parallel^2, \]

\[ I_2 \leq C \parallel \nabla R^n \parallel_{L^4} \parallel (\parallel E^n \parallel + \parallel \nabla E^n \parallel) \parallel Z^n \parallel \leq \theta \parallel E^n \parallel^2 + \theta \parallel \nabla E^n \parallel^2 + C(\theta) \parallel \nabla R^n \parallel_{L^4}^2 \parallel Z^n \parallel^2 \]

\[ =: \theta \parallel E^n \parallel^2 + \theta \parallel \nabla E^n \parallel^2 + C(\theta) J_2 \parallel Z^n \parallel^2, \]

\[ I_3 \leq \theta \parallel \nabla E^n \parallel^2 + C(\theta) \parallel Z^n \parallel^2, \]

\[ I_4 \leq C \parallel \nabla Y^{n+1} \parallel^2 \parallel Z^{n+1} \parallel + C \parallel \nabla Y^{n+1} \parallel^2 \parallel Z^n \parallel \parallel \nabla Z^n \parallel \]

\[ \leq \sigma \parallel \nabla Z^n \parallel^2 + \sigma \parallel \nabla Z^{n+1} \parallel^2 + C(\sigma) \parallel \nabla Y^{n+1} \parallel \left( \parallel Z^{n+1} \parallel^2 + \parallel Z^n \parallel^2 \right) \]

\[ =: \sigma \parallel \nabla Z^n \parallel^2 + \sigma \parallel \nabla Z^{n+1} \parallel^2 + C(\sigma) J_4 \left( \parallel Z^{n+1} \parallel^2 + \parallel Z^n \parallel^2 \right), \]

\[ I_5 \leq \sigma \parallel \nabla Z^n \parallel^2 + C(\sigma) \parallel Y^{n+1} \parallel_{L^\infty} \parallel Z^n \parallel^2 \]

\[ =: \sigma \parallel \nabla Z^{n+1} \parallel^2 + C(\sigma) J_5 \parallel Z^n \parallel^2, \]

\[ I_6 \leq C \parallel \nabla R^n \parallel_{L^4} \parallel Y^{n-1} \parallel_{L^4} \parallel Z^n \parallel \parallel \nabla Z^n \parallel \]

\[ \leq \sigma \parallel \nabla Z^n \parallel^2 + C(\sigma) \parallel \nabla R^n \parallel_{L^4} \parallel Y^{n-1} \parallel_{L^4} \parallel Z^n \parallel^2 \]

\[ =: \sigma \parallel \nabla Z^n \parallel^2 + C(\sigma) J_6 \parallel Z^n \parallel^2, \]

\[ I_7 \leq \sigma \parallel \nabla Z^n \parallel^2 + C(\sigma) \parallel \nabla Y^{n-1} \parallel^2 \parallel Z^n \parallel^2 =: \sigma \parallel \nabla Z^n \parallel^2 + C(\sigma) J_7 \parallel Z^n \parallel^2, \]

\[ I_8 \leq \sigma \parallel \nabla Z^n \parallel^2 + C(\sigma) \parallel Y^{n-1} \parallel_{L^\infty} \parallel Z^n \parallel^2 =: \sigma \parallel \nabla Z^n \parallel^2 + C(\sigma) J_8 \parallel Z^n \parallel^2, \]
where \(k \sum J_t \leq C < \infty\) uniformly with respect to \(h, k > 0\), by Lemmas 5.2 and 5.3. Summing up and choosing \(\sigma > 0\), we get for a summable function \(\tilde{J}\)

\[
- \frac{1}{2} d_t \| \sqrt{R^n} Z^{n+1} \|^2 + \mu \| \nabla Z^n \|^2 + \frac{k}{2} \| \sqrt{R^n} d_t Z^n \|^2 \\
\leq \theta \| E^n \|^2 + \theta \| \nabla E^n \|^2 + C(\theta) \tilde{J} + C(\theta) \tilde{J} \left( \| Z^{n+1} \|^2 + \| Z^n \|^2 \right). 
\]

**Step 2:** We test (6.2c) with \(E^n\) and get

\[
\varepsilon \| \nabla E^{n+1} \|^2 - \frac{1}{2} d_t \| E^{n+1} \|^2 + \frac{k}{2} \| d_t E^{n+1} \|^2 \\
\leq \left\| Y^n \right\|_{L^\infty} \| \nabla E^n \| \| E^n \| + \frac{1}{2} \| \text{div} Y^n \| \| E^n \|_{L^4}^2 + \frac{1}{2} \| d_t Y^{n+1} \| \| Z^{n+1} \|_{L^4} \| E^n \|_{L^4} \\
+ \frac{1}{2} \| Y^n \|_{L^\infty} \| d_t Z^{n+1} \| \| E^n \| + \frac{1}{2} \| Y^n \|_{L^4} \| \nabla Y^{n+1} \|_{L^4} \| Z^{n+1} \|_{L^4} \| E^n \|_{L^4} \\
+ \| U^{n+1} \| \| Z^{n+1} \|_{L^4} \| E^n \|_{L^4} + \frac{1}{2} \| Y^n \|_{L^4} \| \nabla Z^{n+1} \| \| Y^{n+1} \|_{L^4} \| E^n \|_{L^4} \\
+ \lambda (\tilde{R}^n - \tilde{\rho}(t_n), E^n) \right\| + \beta \| \Delta_h R^n \| \| E^n \| + \frac{\beta}{8 \delta} (W'(R^n), E^n) \\
eq: I_{11} + I_{12} + \ldots + I_{110}.
\]

For \(\sigma, \tau, \lambda > 0\) small enough, we derive the following estimates using the techniques from the first part:

\[
I_{11} \leq \sigma \| \nabla E^n \|^2 + C(\sigma) \| Y^n \|_{L^\infty} \| E^n \|^2 \\
=: \sigma \| \nabla E^n \|^2 + C(\sigma) K_1 \| E^n \|^2,
\]

\[
I_{12} \leq C \| \nabla Y^n \| \| E^n \| \left( \| \nabla E^n \| + \| E^n \| \right) \\
\leq \sigma \| \nabla E^n \|^2 + C \left( \| \nabla Y^n \|^2 + \| \nabla Y^n \|^2 \right) \| E^n \|^2,
\]

\[
I_{13} \leq C \| d_t Y^{n+1} \|_{L^2} \| Z^{n+1} \|^2 \left( \left\| \nabla Z^{n+1} \right\|^2 + \| \nabla E^n \|^2 \right) \\
\leq \sigma \| \nabla Z^{n+1} \|^2 + \sigma \| \nabla E^n \|^2 + C(\sigma) \| d_t Y^{n+1} \|^2 \| Z^{n+1} \|^2 \\
+ \frac{C(\sigma)}{2} \left( \| d_t Y^{n+1} \|^2 + \| d_t Y^{n+1} \|^2 \right) \| E^n \|^2 \\
=: \sigma \| \nabla Z^{n+1} \|^2 + \sigma \| \nabla E^n \|^2 + C(\sigma) K_3 \| Z^{n+1} \|^2 + C(\sigma) K_3 \| E^n \|^2,
\]

\[
I_{14} \leq \tau \| d_t Z^{n+1} \|^2 + C(\tau) \| Y^n \|_{L^\infty} \| E^n \|^2 \\
=: \tau \| d_t Z^{n+1} \|^2 + C(\tau) K_4 \| E^n \|^2.
\]

\[
I_{15} \leq C \| Z^{n+1} \| \| \nabla Z^{n+1} \| + C \| Y^n \| \| \nabla Y^n \| \| \nabla Y^{n+1} \| \| \tilde{\Delta}_h Y^{n+1} \| \| E^n \| \left( \| \nabla E^n \| + \| E^n \| \right) \\
\leq \lambda \| \nabla Z^{n+1} \|^2 + \sigma \| \nabla E^n \|^2 + C(\lambda) \left( \| Y^n \|^2 \| \nabla Y^n \|^2 \| \nabla Y^{n+1} \|^2 \| \tilde{\Delta}_h Y^{n+1} \|^2 \\
+ \| Y^n \| \| \nabla Y^n \| \| \nabla Y^{n+1} \| \| \tilde{\Delta}_h Y^{n+1} \| \right) \| E^n \|^2 + C(\lambda) \| Z^{n+1} \|^2 \\
=: \lambda \| \nabla Z^{n+1} \|^2 + \sigma \| \nabla E^n \|^2 + C(\lambda) \| Z^{n+1} \|^2 + C(\lambda) K_5 \| E^n \|^2,
\]

\[
I_{16} \leq C \| U^{n+1} \| \| Z^{n+1} \| \| \nabla Z^{n+1} \| + \| U^{n+1} \| \| E^n \| \left( \| \nabla E^n \| + \| E^n \| \right)
\]
Again, we have derived the following estimate
\[\varepsilon \|\nabla E^{n+1}\|^2 - d_t \|E^{n+1}\|^2 + k \|d_t E^{n+1}\|^2 \leq \lambda \|\nabla Z^{n+1}\|^2 + \tau \|d_t Z^{n+1}\|^2 + C(\lambda, \tau) \|Z^{n+1}\|^2 + C(\lambda, \tau) \|E^n\|^2. \quad (7.2)\]
Choosing \(\lambda, \theta > 0\) small enough and adding (7.1) and (7.2) together, we get
\[\begin{align*}
-d_t \sqrt{R^k} Z^{n+1} &+ \frac{\mu}{2} \|\nabla Z^{n+1}\|^2 - d_t \|E^{n+1}\|^2 + \frac{\varepsilon}{2} \|\nabla E^n\|^2 \\
&\leq \tau \|d_t Z^{n+1}\|^2 + C(\tau) \|Z^{n+1}\|^2 + C(\tau) \|E^n\|^2, \quad (7.3)
\end{align*}\]
where \(\tilde{F}\) is a generic summable function consisting of \(\tilde{J}\) and \(\tilde{K}\) from above.

**Step 3:** We test (6.2) with \(-d_t Z^{n+1}\) and get
\[\|\sqrt{R^k} d_t Z^{n+1}\|^2 - \frac{\mu}{2} \|d_t Z^{n+1}\|^2 + \frac{\mu}{2} k \|d_t \nabla Z^{n+1}\|^2 \leq \frac{1}{2} \|\sqrt{R^k} d_t Z^{n+1}\|^2 + C \|E^n\|^2 \|\nabla R^n\|^2 \|Z^{n+1}\|^2 \|L^4\| \\
+ C \|\nabla Y^{n+1}\|^2 \|L^4\| \|Z^{n+1}\|^2 \|L^4\| + C \|\nabla Z^{n+1}\|^2 \|Y^{n+1}\|^2 \|L^4\| \\
+ C \|\nabla R^n\|^2 \|L^4\| \|Y^{n+1}\|^2 \|L^4\| + C \|\nabla Y^{n+1}\|^2 \|L^4\| \|Z^{n+1}\|^2 \|L^4\| \\
+ C \|Y^{n+1}\|^2 \|L^4\| \|\nabla Z^n\|^2 \leq: \frac{1}{2} \|\sqrt{R^k} d_t Z^{n+1}\|^2 + III_1 + III_2 + \ldots + III_8.
\]
Again, every \(III_i\) can be estimated as follows, using some positive constants \(\theta > 0\):
\[\begin{align*}
III_1 &\leq C \|\nabla R^n\| \left(\|\Delta_h R^n\| + \|\nabla R^n\|\right) \|E^n\| \left(\|\nabla E^n\| + \|E^n\|\right) \\
&\leq \theta \|\nabla E^n\|^2 + C(\theta) \left(\|\nabla R^n\|^2 \left(\|\Delta_h R^n\| + \|\nabla R^n\|\right) + \|\nabla R^n\| \left(\|\Delta_h R^n\| + \|\nabla R^n\|\right)\right) \|E^n\|^2 \\
&=: \theta \|\nabla E^n\|^2 + C(\theta) L_1 \|E^n\|^2, \\
III_2 &\leq C \|\nabla E^n\|^2, \\
III_3 &\leq C \|d_t R^n\| \left(\|\nabla d_t R^n\| + \|d_t R^n\|\right) \|Z^n\| \|\nabla Z^n\| \\
&\leq \|\nabla Z^n\|^2 + C \|d_t R^n\|^2 \left(\|\nabla d_t R^n\| + \|d_t R^n\|\right) \|Z^n\|^2 \|L^4\|^2.
\]

Lemma 8.1. \( \text{optimality system (4.1).} \)

In this section, we pass to the limit for the adjoint variables thanks to the bounds from the previous sections, and identify the limit with a weak solution of the corresponding continuous version of Gronwall's lemma leads then to the assertion of the lemma for a sufficiently small \( \tau \). We consider now the sum \( \sum_{i=1}^k L_i \) are such that \( k \sum L_i \leq C \). We now have shown the following estimate

\[
\| \sqrt{R^n} d_t Z^{n+1} \|^2 - d_t \| \nabla Z^{n+1} \|^2 \\
\leq C E \| \nabla E^n \|^2 + \bar{L} \| E^n \|^2 + \bar{L} \left( \| Z^n \|^2 + \| \nabla Z^n \|^2 + \| Z^{n+1} \|^2 \right) \tag{7.4}
\]

**Step 4:** We consider now the sum \( (2\tau^{-1} C + 1)(7.3) + (7.4) \) and get for a appropriate choice of \( \tau \) that

\[
- d_t \| \sqrt{R^n} Z^{n+1} \|^2 - d_t \| E^{n+1} \|^2 - d_t \| \nabla Z^{n+1} \|^2 + C \| \nabla Z^n \|^2 + \| \sqrt{R^n} d_t Z^{n+1} \|^2 \\
\leq \bar{F} \| Z^{n+1} \|^2 + \bar{F} \| E^n \|^2 + \bar{F} \| \nabla Z^n \|^2
\]

with some generic summable functions \( \bar{F} \) consisting of \( \bar{J}, \bar{K} \) and \( \bar{L} \) from above. The discrete version of Gronwall's lemma leads then to the assertion of the lemma for a sufficiently small \( k \leq k_0(\Omega, \rho_{\min}, \rho_{\max}, T, \mu, \varepsilon) \).

8. Convergence of the Scheme

In this section, we pass to the limit for the adjoint variables thanks to the bounds from the previous sections, and identify the limit with a weak solution of the corresponding continuous optimality system \([4.1]\).

**Lemma 8.1.** There exist functions \( \rho^* \in L^1(H) \cap H^2(V) \cap L^\infty(H^1), \ y^* \in L^\infty(V) \cap L^2(\mathcal{H}^2) \cap H^1(H), \ \eta^* \in H^1(H^{-1}) \cap L^2(H^1) \cap L^\infty(L^2), \ z^* \in H^1(H) \cap L^\infty(H) \cap L^2(V), \ u^* \in L^2(L^2), \) such that a subsequence \( (\{R^n\}, \{U^n\}, \{E^n\}, \{Z^n\}, \{U^n\}) \) converges to their counterparts in the following sense \((h, k \to 0): \)

\[
\begin{align*}
R^* &\rightarrow^* R^* &\text{weakly-star in } L^\infty(L^\infty), \\
\mathbf{y}^* &\rightarrow^* \mathbf{y}^* &\text{strongly in } L^2(H^1), \\
\mathbf{y}^* &\rightarrow \mathbf{y}^* &\text{weakly in } H^1(L^2), \\
\text{div } \mathbf{y}^* &\rightarrow^* \text{div } \mathbf{y}^* &\text{weakly in } L^2(L^2), \\
\mathbf{y}^* &\rightarrow \mathbf{y}^* &\text{weakly-star in } L^\infty(V), \\
\mathbf{y}^* &\rightarrow \mathbf{y}^* &\text{strongly in } L^2(H^1), \\
E^+^* &\rightarrow \eta^* &\text{weakly in } L^2(H^1), \\
E^+^* &\rightarrow \eta^* &\text{strongly in } L^2(L^2),
\end{align*}
\]
\(Z^{+/\ast}, Z \rightarrow z^*\) \(weakly \ in \ H^1(L^2),\)
\(Z^{+/\ast}, Z \rightarrow z^*\) \(weakly \ in \ L^2(H^1),\)
\(Z^{+/\ast}, Z \rightarrow z^*\) \(strongly \ in \ L^2(L^2),\)
\(\text{div } Z^{+/\ast}, \text{div } Z \rightarrow 0\) \(weakly \ in \ L^2(L^2),\)
\(U^{+/\ast}, U \rightarrow u^*\) \(weakly \ in \ L^2(L^2).\)

**Proof.** The weak convergence and weak-star convergence, respectively, follows from Lemmas 5.1, 5.2, 5.3, 5.5, and 7.1. The strong convergence for the affine time interpolants are obtained by the bounds and Lemma 5.4. Since increments are bounded (see Lemmas 5.2, 5.3, and 7.1), all constant in time interpolants inherit the strong convergence as already mentioned at the end of Subsection 5.1.

These convergence properties are sufficient, such that all linear terms in (6.2) will directly converge to their continuous counterparts in a weak sense. It remains to identify weak convergence of all nonlinear parts, where we use in particular the strong convergence results from above. For a simpler notation, we drop the stars from the limits derived above. By a standard density argument, it is enough to identify the limit for used smooth test functions.

**Lemma 8.2.** Let \(\varphi \in C_0^\infty([0,T], C_0^\infty)\) or \(\varphi \in C_0^\infty([0,T], C^\infty)\) respectively. We have

1. \((\nabla R^* \mathcal{E}^* - \nabla \rho \eta, \varphi) \rightarrow 0.\)
2. \((\nabla \mathcal{E}^* R^* - \nabla \eta \rho \varphi) \rightarrow 0.\)
3. \((R^* Z_t - \rho z_t, \varphi) \rightarrow 0.\)
4. \((R_t Z^+ - \rho_t z, \varphi) \rightarrow 0.\)
5. \((\nabla R^+ \cdot \mathcal{Y}^+ \mathcal{Z}^* - \nabla \rho \cdot yz, \varphi) \rightarrow 0.\)
6. \((R^- \text{div } \mathcal{Y}^- \mathcal{Z}^* - \rho \text{div } yz, \varphi) = (R^- \text{div } \mathcal{Y}^- \mathcal{Z}^*, \varphi) \rightarrow 0.\)
7. \((|R^- \mathcal{Y}^- \cdot \nabla| \mathcal{Z}^* - |\rho y \cdot \nabla| z, \varphi) \rightarrow 0.\)
8. \((R^* \nabla \mathcal{Y}^+ \mathcal{Z}^+ - \rho \nabla yz, \varphi) \rightarrow 0.\)
9. \((R^* \nabla \mathcal{Y}^+ \mathcal{Z}^+ - \rho \nabla yz, \varphi) \rightarrow 0.\)
10. \((|\mathcal{Y}^* \cdot \nabla| \mathcal{E}^* - |y \cdot \nabla| \eta, \varphi) \rightarrow 0.\)
11. \((\text{div } \mathcal{Y}^* \mathcal{E}^* - \text{div } y \eta, \varphi) = (\text{div } \mathcal{Y}^* \mathcal{E}^*, \varphi) \rightarrow 0.\)
12. \((\mathcal{Y}_t \cdot \mathcal{Z}^+ - y_t \cdot z, \varphi) \rightarrow 0.\)
13. \((\mathcal{Y}_t \cdot \mathcal{Z}^+ - y_t \cdot z, \varphi) \rightarrow 0.\)
14. \((|\mathcal{Y}^* \cdot \nabla| \mathcal{Y}^+ \cdot \mathcal{Z}^+ - |y \cdot \nabla| y \cdot z, \varphi) \rightarrow 0.\)
15. \((|\mathcal{Y}^* \cdot \nabla| \mathcal{Y}^+ \cdot \mathcal{Y}^+ - |y \cdot \nabla| z \cdot y, \varphi) \rightarrow 0.\)
16. \((U^* \cdot \mathcal{Z}^+ - u \cdot z, \varphi) \rightarrow 0.\)

**Proof.** (1) We write

\((\nabla R^* \mathcal{E}^* - \nabla \rho \eta, \varphi) = (\nabla R^* (\mathcal{E}^* - \eta), \varphi) + (\nabla (R^* - \rho) \eta, \varphi) =: I + II\)

and calculate

\[I \leq \|\nabla R^*\|_{L^2(L^2)} \|\mathcal{E}^* - \eta\|_{L^2(L^2)} \|\varphi\|_{L^\infty(L^\infty)} \rightarrow 0,\]

\[II = (\nabla (R^* - \rho), \eta \varphi) \rightarrow 0.\]
(2) We write
\[(\nabla \mathcal{E}^* R^* - \nabla \eta \rho, \varphi) = (\nabla \mathcal{E}^* (R^* - \rho), \varphi) + (\nabla (\mathcal{E}^* - \eta) \rho, \varphi) =: I + II\]
and calculate
\[I \leq \|\nabla \mathcal{E}^*\|_{L^2(L^2)} \|R^* - \rho\|_{L^2(L^2)} \|\varphi\|_{L^\infty(L^\infty)} \to 0,\]
\[II = (\nabla (\mathcal{E}^* - \eta), \rho \varphi) \to 0.\]

(3) We write
\[(R^* Z_t - \rho z_t, \varphi) = ((R^* - \rho) Z_t, \varphi) + (\rho (Z - z)_t, \varphi) =: I + II\]
and calculate
\[I \leq \|\varphi (R^* - \rho)\|_{L^2(H^1)} \|Z_t\|_{L^2(H^{-1})} \leq \|R^* - \rho\|_{L^2(H^1)} \|Z_t\|_{L^2(H^{-1})} \|\nabla \varphi\|_{L^\infty(L^\infty)} \to 0,\]
\[II = ((Z - z)_t, \rho \varphi) \to 0.\]

(4) We write
\[(R_t Z^* - \rho_t z, \varphi) = (R_t (Z^* - z), \varphi) + ((R - \rho)_t z, \varphi) =: I + II\]
and calculate
\[I \leq \|R_t\|_{L^2(L^2)} \|Z^* - z\|_{L^2(L^2)} \|\varphi\|_{L^\infty(L^\infty)} \to 0,\]
\[II = ((R - \rho)_t, z \cdot \varphi) \to 0.\]

(5) We write
\[(\nabla R^- \cdot \mathcal{Y}^- Z^* - \nabla \rho \cdot y z, \varphi)\]
\[= (\nabla (R^- - \rho) \cdot y z, \varphi) + (\nabla R^- (\mathcal{Y}^- - y z), \varphi) + (\nabla R^- \cdot \mathcal{Y}^- (Z^* - z), \varphi) =: I + II + III\]
and calculate
\[I = (\nabla (R^- - \rho), y \cdot z \varphi) \to 0,\]
\[II \leq \|\nabla R^-\|_{L^\infty(L^2)} \|\mathcal{Y}^- - y\|_{L^2(L^4)} \|z\|_{L^2(L^4)} \|\varphi\|_{L^\infty(L^\infty)} \to 0,\]
\[III \leq \|\nabla R^-\|_{L^2(L^4)} \|\mathcal{Y}^-\|_{L^\infty(L^4)} \|Z^* - z\|_{L^2(L^2)} \|\varphi\|_{L^\infty(L^\infty)}\]
\[\times \|\mathcal{Y}^-\|_{L^\infty(L^4)} \|Z^* - z\|_{L^2(L^2)} \|\varphi\|_{L^\infty(L^\infty)} \to 0.\]

(6) We calculate
\[(R^- \text{ div } \mathcal{Y}^- Z^*, \varphi) \leq \|R^-\|_{L^\infty(L^\infty)} \|\text{ div } \mathcal{Y}^-\|_{L^2(L^2)} \|Z^*\|_{L^2(L^2)} \|\varphi\|_{L^\infty(L^\infty)} \to 0.\]

(7) We write
\[((R^- \mathcal{Y}^- \cdot \nabla) Z^* - |\rho y \cdot \nabla| z, \varphi)\]
\[= - (\nabla R^- \cdot \mathcal{Y}^- Z^* - \nabla \rho \cdot y z, \varphi) - (R^- \text{ div } \mathcal{Y}^- Z^* - \rho \text{ div } y z, \varphi)\]
\[= - (\left< [\nabla \mathcal{Y}^- \cdot \nabla] \varphi, Z^* \right> - \left< |\rho y \cdot \nabla| \varphi, z \right>) =: -I - II + III.\]

The terms $I$ and $II$ are estimated in the last two parts. We rewrite $III$ in the following way:
\[III = (\left< [R^- (\mathcal{Y}^- - y) \cdot \nabla] \varphi, Z^* \right> + (\left< (R^- - \rho) y \cdot \nabla \varphi, Z^* \right> + (\left< |\rho y \cdot \nabla| \varphi, Z^* - z \right>) =: III_a + III_b + III_c.\]
The three terms can be estimated as follows:
\[ III_a \leq ||R^−||L^∞(L^∞)||Z^− - y||L^2(L^2)||\nabla\varphi||L^∞(L^∞)||Z^*||L^2(L^2) \rightarrow 0, \]
\[ III_b \leq ||R^− - \rho||L^2(L^4)||y||L^∞(L^4)||\nabla\varphi||L^∞(L^∞)||Z^*||L^∞(L^∞) \leq C||\nabla(R^− - \rho)||L^2(L^2)||\nabla y||L^∞(L^2)||\nabla\varphi||L^∞(L^∞)||Z^*||L^∞(L^2) \rightarrow 0, \]
\[ III_c \leq ||R^−||L^∞(L^∞)||y||L^2(L^2)||\nabla\varphi||L^∞(L^∞)||Z^* - z||L^2(L^2) \rightarrow 0. \]

(8) We write
\[(R^*\nabla Z^+ Y^+ - \rho\nabla z y, \varphi)\]
\[= ((R^* - \rho)\nabla Z^+ Y^+ + (\rho\nabla(Z^+ - z)y, \varphi) + (\rho\nabla Z^+(Y^+ - y), \varphi) =: I + II + III,\]
and by integration by parts,
\[II = -(Z^+ - z, (\nabla \cdot \varphi y) - (\rho\nabla y(Z^+ - z), \varphi) - (Z^+ - z, \rho \text{ div } \varphi y) =: II_a + II_b + II_c.\]

We calculate
\[ I \leq ||R^* - \rho||L^2(L^4)||\nabla Z^+||L^2(L^2)||Y^+||L^∞(L^∞)||\nabla\varphi||L^∞(L^∞) \rightarrow 0, \]
\[ II_a \rightarrow 0, \]
\[ II_b \leq ||\rho||L^∞(L^∞)||\nabla y||L^2(L^2)||Z^+ - z||L^2(L^2)||\nabla\varphi||L^∞(L^∞) \rightarrow 0, \]
\[ II_c \rightarrow 0, \]
\[ III \leq ||\rho||L^∞(L^∞)||\nabla Z^+||L^2(L^2)||Y^+ - y||L^2(L^2)||\nabla\varphi||L^∞(L^∞) \rightarrow 0. \]

(9) This estimate is just the same like the last one except that \( Z^+ \) and \( Y^+ \) are interchanged and the derivative affects \( Y^+ \), which has an improved regularity in contrast to \( Z^+ \).

(10) We write
\[(|Y^* \cdot \nabla|^2 - [y \cdot \nabla]|\eta, \varphi) = ([|Y^* - y| \cdot \nabla|^2, \varphi] + (y \cdot \nabla)(\varepsilon^* - \eta), \varphi) =: I + II,\]
and calculate
\[ I \leq ||Y^* - y||L^2(L^2)||\nabla\varepsilon^*||L^2(L^2)||\nabla\varphi||L^∞(L^∞) \rightarrow 0, \]
\[ II = -\langle [y \cdot \nabla] |\varphi, \varepsilon^* - \eta \rangle \rightarrow 0, \]
where we used \( \text{div } y = 0 \).

(11) We calculate
\[(\text{div } Y^* \varepsilon^*, \varphi) \leq ||\text{div } Y^*||L^2(L^2)||\varepsilon^*||L^2(L^2)||\nabla\varphi||L^∞(L^∞) \rightarrow 0. \]

(12) We write
\[(Y_t \cdot Z^+ - y_t \cdot z, \varphi) = (Y_t \cdot (Z^+ - z), \varphi) + ((Y - y)_t \cdot z, \varphi) =: I + II,\]
and calculate
\[ I \leq ||Y_t||L^2(L^2)||Z^+ - z||L^2(L^2)||\nabla\varphi||L^∞(L^∞) \rightarrow 0, \]
\[ II = ((Y - y)_t, \varphi z) \rightarrow 0. \]

(13) We write
\[(Y^* \cdot Z_t - y \cdot z_t, \varphi) = ((Y^* - y)Z_t, \varphi) + (y \cdot (Z - z)_t, \varphi) =: I + II,\]
and calculate
\[ I \leq \|\mathbf{Y}^* - \mathbf{y}\|_{L^2(L^2)} \|\mathbf{Z}\|_{L^2(L^2)} \|\varphi\|_{L^\infty(L^\infty)} \to 0, \]
\[ II = ((\mathbf{Z} - \mathbf{z})_t, \varphi \mathbf{y}) \to 0. \]

(14) We write
\[ ((\mathbf{Y}^* \cdot \nabla)(\mathbf{Y}^+ - \mathbf{y} \cdot \nabla)\mathbf{y} \cdot \mathbf{z}, \varphi) = \left( ((\mathbf{Y}^* - \mathbf{y}) \cdot \nabla)(\mathbf{Y}^+ + \mathbf{Z}^+, \varphi) + (\|\mathbf{y} \cdot \nabla)(\mathbf{Y}^+ - \mathbf{y}) \cdot \mathbf{Z}^+, \varphi) \right. \]
\[ \left. + (\mathbf{y} \cdot \nabla\mathbf{y} \cdot (\mathbf{Z}^+ - \mathbf{z}), \varphi) =: I + II + III \right) \]
and calculate
\[ I \leq \|\mathbf{Y}^* - \mathbf{y}\|_{L^2(L^2)} \|\nabla\mathbf{Y}^+\|_{L^\infty(L^2)} \|\mathbf{Z}^+\|_{L^2(L^2)} \|\varphi\|_{L^\infty(L^\infty)} \leq C\|\mathbf{Y}^* - \mathbf{y}\|_{L^2(L^2)} \|\nabla\mathbf{Y}^+\|_{L^\infty(L^2)} \|\mathbf{Z}^+\|_{L^2(L^2)} \|\varphi\|_{L^\infty(L^\infty)} \to 0, \]
\[ II \leq \|\mathbf{y}\|_{L^\infty(L^2)} \|\nabla (\mathbf{Y}^+ - \mathbf{y})\|_{L^2(L^2)} \|\mathbf{Z}^+\|_{L^2(L^2)} \|\varphi\|_{L^\infty(L^\infty)} \leq C\|\mathbf{Y}^*\|_{L^2(L^2)} \|\nabla (\mathbf{Y}^+ - \mathbf{y})\|_{L^2(L^2)} \|\mathbf{Z}^+\|_{L^2(L^2)} \|\varphi\|_{L^\infty(L^\infty)} \to 0, \]
\[ III = (\mathbf{Z}^+ - \mathbf{z}, \varphi \mathbf{y} \cdot \nabla \mathbf{y}) \to 0. \]

(15) Since we have
\[ ((\mathbf{Y}^* \cdot \nabla)(\mathbf{Z}^+ \cdot \mathbf{Y}^+), \varphi) = (\text{div} \mathbf{Y}^* \mathbf{Z}^+ \cdot \mathbf{Y}^+, \varphi) + (\mathbf{Y}^* \cdot \nabla \mathbf{Y}^+ \cdot \mathbf{Z}^+, \varphi) \]
and the same for the continuous counterpart, we have only to show \( (\text{div} \mathbf{Y}^* \mathbf{Z}^+ \cdot \mathbf{Y}^+, \varphi) \to 0 \). The last term has been dealt with in the previous part of the proof. We calculate
\[ (\text{div} \mathbf{Y}^* \mathbf{Z}^+ \cdot \mathbf{Y}^+, \varphi) \leq \|\text{div} \mathbf{Y}^*\|_{L^2(L^2)} \|\mathbf{Z}^+\|_{L^2(L^2)} \|\mathbf{Y}^+\|_{L^\infty(L^4)} \|\varphi\|_{L^\infty(L^\infty)} \leq C\|\text{div} \mathbf{Y}^*\|_{L^2(L^2)} \|\nabla \mathbf{Z}^+\|_{L^2(L^2)} \|\nabla \mathbf{Y}^+\|_{L^\infty(L^2)} \|\varphi\|_{L^\infty(L^\infty)} \to 0. \]

(16) With \( [6.2d] \) and \( [1.1d] \) respectively, we rewrite the terms to
\[ \alpha(\mathbf{U}^+ \cdot \mathbf{Z}^+ - \mathbf{u} \cdot \mathbf{z}, \varphi) = (\mathcal{R}^* \mathbf{Z}^+ \cdot \mathbf{Z}^+ - \rho \mathbf{z} \cdot \mathbf{z}, \varphi) \]
\[ = ((\mathcal{R}^* - \rho) \mathbf{Z}^+ \cdot \mathbf{Z}^+, \varphi) + (\rho (\mathbf{Z}^+ - \mathbf{z}) \cdot \mathbf{Z}^+, \varphi) + (\rho \mathbf{z} \cdot (\mathbf{Z}^+ - \mathbf{z}), \varphi) \]
\[ =: I + II + III. \]

We calculate finally
\[ I \leq \|\mathcal{R}^* - \rho\|_{L^2(L^4)} \|\mathbf{Z}^+\|_{L^\infty(L^2)} \|\mathbf{Z}^+\|_{L^2(L^2)} \|\varphi\|_{L^\infty(L^\infty)} \leq C\|\mathcal{R}^* - \rho\|_{L^2(L^2)} \|\mathbf{Z}^+\|_{L^\infty(L^2)} \|\nabla \mathbf{Z}^+\|_{L^2(L^2)} \|\varphi\|_{L^\infty(L^\infty)} \to 0, \]
\[ II \leq \|\rho\|_{L^\infty(L^\infty)} \|\mathbf{Z}^+ - \mathbf{z}\|_{L^2(L^2)} \|\mathbf{Z}^+\|_{L^2(L^2)} \|\varphi\|_{L^\infty(L^\infty)} \to 0, \]
\[ III \leq \|\rho\|_{L^\infty(L^\infty)} \|\mathbf{z}\|_{L^2(L^2)} \|\mathbf{Z}^+ - \mathbf{z}\|_{L^2(L^2)} \|\varphi\|_{L^\infty(L^\infty)} \to 0. \]

\[ \square \]

**Theorem 8.3.** The functions \( \rho^* \in H^1(H^1) \cap L^\infty(H^2) \), \( \mathbf{y}^* \in L^\infty(V) \cap L^2(H^2) \cap H^1(H) \), \( \eta^* \in H^1(H^{-1}) \cap L^\infty(H^1) \), \( \mathbf{z}^* \in H^1(H) \cap L^\infty(H) \cap L^2(V) \), \( \mathbf{u}^* \in L^2(L^2) \) obtained in Lemma 8.1 solve (4.1).
Proof. With the first nine parts of Lemma 8.2 we deduce that the solutions of (6.2a) converge to solutions of (4.1a) (up to subsequences). With the remaining parts of Lemma 8.2 we deduce that solutions of (6.2c) converge to solutions of (4.1c). All other parts in those equations are linear and we can pass to their limits by the weak convergence results obtained in Lemma 8.1.

Theorem 8.4. For the function \( u^* \in L^2(\mathbf{L}^2) \) obtained in Lemma 8.1 we have \( U^n, U^+ \to u^* \) strongly in \( L^2(\mathbf{L}^2) \) (up to a subsequence) for \( h, k \to 0 \).

Proof. We use (4.1d) and (6.2d), respectively, to rewrite \( U^n \) and \( u \), respectively. We have then to show that

\[ 0 \leftarrow \| u - U^+ \|_{L^2(\mathbf{L}^2)} = \| R^- \mathbf{Z}^* - \rho z \|_{L^2(\mathbf{L}^2)} = \| (\mathbf{Z}^* - z) \rho \|_{L^2(\mathbf{L}^2)} + \| (R^- \rho) \mathbf{Z}^* \|_{L^2(\mathbf{L}^2)}. \]

For the first term, we use \( L^\infty(L^\infty) \) bounds on \( \rho \) and the strong convergence of \( \mathbf{Z}^* \to z \) in \( L^2(\mathbf{L}^2) \).

For the second term, we use Aubin–Lions’ lemma for the spaces \( H^2(\Omega) \hookrightarrow W^{1,4}(\Omega) \hookrightarrow H^1(\Omega) \) and the bounds on \( R^- \rho \) in order to have strong convergence \( R^- \rho \to \rho \) in \( L^2(W^{1,4}) \subseteq L^2(\mathbf{L}^\infty) \). This leads to convergence in the second term together with the bound of \( \mathbf{Z}^* \) in \( L^\infty(\mathbf{L}^2) \).

9. Computational experiments

In this section we present a numerical algorithm and perform computational studies for the system (5.5)-(6.2).

9.1. Implementation details and minimization algorithm. We use the Taylor-Hood finite element pair with vector-valued quadratic finite elements for \( V_h \), standard continuous piecewise linear elements for \( M_h \), and standard linear finite elements for \( R_h \) as discussed in Section 3. It then follows from (6.2d) that for \( n = 1, \ldots, N \) the control variable \( U^n \) belongs to the space of vector-valued conforming piecewise continuous finite elements of degree three.

Due to the advective character of the mass equation, without proper stabilization the numerical solution may exhibit oscillations near the interface. Our experience suggests that one should choose \( \varepsilon_h \) in the stabilization term \( -\varepsilon_h \Delta R^n \) to be dependent on the mesh size and time step values. However, an optimal choice of \( \varepsilon_h \) for this type of stabilization is not clear.

We choose \( \varepsilon_h \| K = \xi h_K \) where \( h_K \) is the diameter of the mesh element \( K \). The constant \( \xi > 0 \) is chosen a priori to minimize the oscillations. For fixed values of \( \xi > 0 \) we still observe oscillations in the values of the computed density in some experiments. In the majority of cases the oscillations did not exceed 2% - 5% of the densities \( \rho_1, \rho_2 \).

Remark 9.1. The oscillations in the density values can always be eliminated by choosing a sufficiently large parameter \( \xi = \| Y \|_{L^\infty(\mathbf{L}^\infty)} \) in \( -\varepsilon_h \Delta R^n \). A more sophisticated alternative is to use an upwind scheme for the mass equation. This corresponds to a space-time adaptive choice of the artificial diffusion parameter \( \varepsilon_h \| K = \gamma h_{K} \| Y \|_{L^\infty(\mathbf{L}^\infty)} \) (along with an appropriate modification of the dual problem); cf. [8] Section 5.1.3. The upwind discretization preserves monotonicity of the computed density \( R \) and introduces less artificial diffusion in the numerical solution than a constant diffusion parameter. However, we experienced difficulties with the convergence of the minimization algorithm for the upwind discretization. In addition, an
adaptive choice of the diffusion parameter may introduce an artificial control mechanism into the system.

The numerical experiments have been performed for different initial conditions for the density, the initial condition for the velocity was always set to zero. We prescribe homogeneous Dirichlet boundary conditions for the equations (5.5b), (6.2a); for equations (5.5a), (6.2c) we prescribe homogeneous Neumann boundary conditions. We use a fixed time-step size and employ a simple mesh refinement strategy. In the regions of the computational domain where
\[ \rho_1 + 10^{-3} < R^n \leq \rho_2 - 10^{-3}, \]
for some \( n = 1, \ldots, N \) (i.e., along the interfacial region) we use a fine mesh size \( h = h_{\text{min}} \) (in most experiments \( h_{\text{min}} = 1/64 \)), elsewhere we use a coarse mesh size \( h_{\text{max}} = 1/16 \). We do not perform any mesh coarsening. Note, that in general the minimum mesh size \( h_{\text{min}} \) should be chosen according to the thickness of the diffuse interface which is determined by the parameter \( \delta \), cf. [6], [7], [9].

A standard approach, see e.g. [19], is to express \( R \equiv R(U) \). Given \( h > 0 \), we note that solutions are unique by a contraction argument for sufficiently small time step sizes \( k > 0 \).

The dependence of the functional (6.1) on the density \( R \) can therefore be eliminated, and the functional can be rewritten as
\[ \hat{J}_{\delta,h,k}(U) = J_{\delta,h,k}(R(U),U). \]

The gradient of \( \hat{J}_{\delta,h,k} \) is equivalent to (6.2d), i.e.,
\[ \nabla U \hat{J}_{\delta,h,k} = \alpha U - \mathcal{R}^{-1}. \]

We then look for the minimum of the reduced functional \( \hat{J}_{\delta,h,k} \), which is equivalent to finding a solution of the system (5.5)-(6.2).

To minimize the functional \( \hat{J}_{\delta,h,k} \) we employ a modified gradient descent algorithm with adaptive mesh size according to the Barzilai and Borwein criterion [10].

1. Choose the constants \( \sigma_{\text{init}}, \sigma_{\min}, \sigma_{\max}, \sigma_{\ast}, \delta_{\text{TOL}} \), set \( U_0 = 0 \).
2. Iterate for \( k = 0, \ldots \)
   (a) If \( k < 2 \) set \( \sigma_k = \sigma_{\text{init}} \); for \( k \geq 2 \) choose step size \( \sigma_k = \sigma_{\ast} \frac{\int_0^T \int (S_k - S_{k-1}) \mathcal{W}_k \, dx \, dt}{\| S_k - S_{k-1} \|^2_{L^2(L^2)}} \)
   where \( S_k = U_k - U_{k-1} \) and \( \mathcal{W}_k = G_k - G_{k-1} \), \( G_k = \nabla U \hat{J}(U_k) \).
   (b) If \( \sigma_k > \sigma_{\max} \) or \( \sigma_k < 0 \) set \( \sigma_k = \sigma_{\min} \).
   (c) Compute \( U_{k+1} = U_k - \sigma_k \nabla U \hat{J}(U_k) \)
3. Stop when \( \| \nabla U \hat{J} \|_{L^2(L^2)}^2 < \delta_{\text{TOL}} \).

In the computational experiments below we choose \( \sigma_{\text{init}} = 10^{-4}, \sigma_{\min} = 100, \sigma_{\max} = 20\sigma_{\min}, \sigma_{\ast} = 0.2, \delta_{\text{TOL}} = 10^{-9} \). For the given tolerance \( \delta_{\text{TOL}} \) the gradient algorithm converges after 200 – 800 steps in all numerical experiments.

We found that the above algorithm with suitably chosen parameters \( \sigma_{\text{init}}, \sigma_{\min}, \sigma_{\max}, \sigma_{\ast} \) was more efficient than a corresponding exact line search algorithm, where the optimal step size was determined by bisection. Note, that we use the steepest descent method for its simplicity. More advanced algorithms, such as SQP, could provide better performance for our problem; cf. [16] for an overview of different algorithms.
Remark 9.2. To evaluate the gradient \( \nabla_{\mathcal{U}} J(\mathcal{U}_k) = \alpha \mathcal{U}_k - \mathcal{R}_k^{-1} \mathcal{Z}_k \) requires the values of \( \mathcal{U}_k \), \( \mathcal{R}_k^{-1} \), and \( \mathcal{Z}_k \). The control \( \mathcal{U}_k \) is known from the previous iteration of the gradient algorithm. The value of \( \mathcal{R}_k^{-1} \) (along with the value of \( \mathcal{Y}_k \)) is obtained by solving the forward equations (5.5) in space and time with the control \( \mathcal{U} = \mathcal{U}_k \) in (5.5b). Then, the dual variable \( \mathcal{Z}_k \) is obtained by solving the backward problem (6.2) with \( \mathcal{Y} \equiv \mathcal{Y}_k \), \( \mathcal{R} \equiv \mathcal{R}_k \).

9.2. Effect of the phase-field term in the energy functional. The numerical experiments in this section demonstrate the effects of different values of the phase-field approximation of the perimeter functional (1.3) in the energy functional \( J_{\delta,h,k} \). If not mentioned otherwise we set \( \lambda = 0 \) in the experiments in this subsection.

The optimal solution for the phase-field approximation for a fixed parameter \( \delta \) can be characterized as follows: The domain \( \Omega \) can be split into two disjoint regions \( \Omega_1, \Omega_2 \), where the density attains distinct values \( \Omega_1 = \{ x : \rho(x) = \rho_1 \} \) and \( \Omega_2 = \{ x : \rho(x) = \rho_2 \} \). The two regions are separated by a thin region \( \Omega_I = \{ x : \rho_1 < \rho(x) < \rho_2 \} \), with a fixed width that depends on the parameter \( \delta \); cf. [3]. Within this diffuse interface region the density varies continuously between the two density values \( \rho_1, \rho_2 \), and the interface between the two fluids is represented by a hypersurface defined by the mean level-set \( \Gamma_I = \{ x : \rho(x) = \frac{1}{2}(\rho_1 + \rho_2) \} \).

The first set experiment examines the effects of the parameters \( \delta \) and \( \varepsilon \). We set \( T = 0.5 \), \( k = 0.05 \), \( \rho_1 = 1 \), \( \rho_2 = 2 \), \( \xi = 10^{-3} \), \( \alpha = 10^{-5} \), \( \beta = 5 \), \( \delta = (3.68 \pi)^{-1} \).

To eliminate the effect of diffusion on the shape of the interfacial region we first take \( \xi = 0.001 \). We start with an initial condition with discontinuous density. To minimize the phase-field energy the discontinuous solution will evolve towards the optimal solution with a diffuse interface area: a diffuse interface layer is produced which leads to an overall decrease of the energy, but the influence on the shape of the interface is undesirable.

We continue with two experiments that demonstrate the effects of different values of the diffusion parameter \( \xi \). We start with a discontinuous density and use a large value for the diffusion coefficient \( \xi = 0.1 \). The energy of the initial iterate of the gradient algorithm is 0.2519 and the energy of the computed optimal solution is 0.2516. As in the previous case, there is very little control involved in the computation, since the interface assumes a diffuse shape due to the large value of the diffusion constant \( \xi > 0 \). Note however, that due to the excessive diffusion, the interface is smeared out, and hence the thickness of the diffuse interface region does not correspond to the optimal thickness for the considered value of the parameter \( \delta > 0 \); see Figure 3. There is no control mechanism included in the model that would allow to counteract the diffusion effects. Consequently, the energy of the optimal solution is larger than in the previous experiment. For \( \xi = 0.5 \) the initial energy is 0.409753, the final energy is 0.409684, and the interface \( \Gamma_{I_{\delta}} \) (represented by a white line) has disappeared; see Figure 3.

Remark 9.3. The behavior that can be observed in Figure 4 is not desirable in practice and can be eliminated by choosing \( \xi > 0 \) sufficiently large. Unless \( \xi > 0 \) is prescribed on physical
grounds, our experience indicates that \( \xi \approx C \delta \) produces satisfactory results. For numerical reasons the constant \( C \) can be chosen according to Remark 9.1.

**Figure 1.** From left to right: Optimal solution with \( \xi = 0.001 \) for the density with discontinuous initial condition at time \( t = 0, 0.25, 0.5 \).

**Figure 2.** Optimal solution with \( \xi = 0.001 \) for velocity at time \( t = 0.05, 0.15, 0.35 \); the vectors at different times are scaled by factors 0.02, 0.06, 0.12, respectively.

**Figure 3.** Optimal solution for the density with large diffusion constant, initial condition and the solution for \( \xi = 0.1 \) and \( \xi = 0.5 \) at time \( t = 0.5 \).

The initial condition for the next experiment has the shape of a square with a diffuse transition region across the interface. We set \( T = 1, \alpha = 10^{-4}, \xi = 0.03 \), the remaining parameters are the same as in previous experiments. The phase-field functional minimizes the perimeter of the interface, and the square initial condition evolves into a circle. In Figure 4, we display the optimal solution at different times. The interface \( \Gamma_I \) is represented by a white line. The optimal velocity field is displayed in Figure 5.

For comparison, we include the tracking-type part in the computation. We set \( \lambda = 10 \) and set the desired state equal to the initial condition for the density \( \hat{\rho}(x) = \rho(0, x) \). The snapshot
Figure 4. Optimal solution at time \( t = 0, 0.25, 0.5, 1 \).

Figure 5. Optimal solution at time \( t = 0.05, 0.25, 0.5 \) (the vectors are scaled by a factor \( 0.3, 1, 30 \)).

of the solution at final time in Figure 3 reveals that for \( \lambda > 0 \) the tracking-type part of the functional forces the interface towards the square shape.

Figure 6. Comparison of the optimal solution for the density at time \( t = 1 \) for \( \lambda = 0 \) (left) and \( \lambda = 10 \) (right).

In the next experiment, the initial condition consists of two circles, the density is discontinuous. We set \( \alpha = 10^{-5}, \lambda = 0, T = 0.5 \), and the remaining parameters have the same values as in the previous experiment. To minimize the phase-field energy, the two circles are forced to join together, and the interface then evolves towards a circle, see Figure 7. The velocity field is displayed in Figure 8, the asymmetry of the results can be accounted to the round-off errors in the numerical approximation. Note that for \( \alpha = 10^{-4} \) the two circles would remain to be separated for \( T = 0.5 \). To further verify the effects of round-off errors, we repeated the experiment with a different linear solver. The results were almost identical with the exception that they were antisymmetric along the \( y \)-axis. This phenomenon can be explained by a possible non-uniqueness of the solution: due to the round-off errors the gradient algorithm may converge towards a different solution with similar energy.
In the next experiment we start with an initial condition in the shape of a dumbbell, set \( \alpha = 10^{-4}, T = 0.5, \lambda = 0 \), and take the remaining parameters as in the previous experiments. The evolution of the computed density is depicted in Figure 9, the dumbbell shape remains connected and the interface evolves towards an ellipsoidal shape. To illustrate the adaptive algorithm we also display the finite element mesh obtained by the adaptive algorithm described in Section 9.1.

In the final experiment from this section we start with a dumbbell shaped initial condition, with rectangular ends that are further apart from each other than in the previous example, and a thinner connecting neck. We choose \( \alpha = 10^{-5}, \lambda = 0, T = 0.5 \), and the other parameters remain unchanged. The evolution of the optimal solution for this setting leads to a pinch-off, the squares become circular and the initially connected dumbbell shape eventually splits up

**Figure 7.** Optimal solution at time \( t = 0, 0.15, 0.5, 1 \) with \( \lambda = 0 \).

**Figure 8.** Optimal solution at time \( t = 0.05, 0.3, 0.5 \) (the velocity vectors are scaled by a factor \( 0.1, 0.3, 0.6 \), respectively).

**Figure 9.** Optimal solution at time \( t = 0, 0.15, 1 \) and the finite element mesh used in the computation.
into two separate bubbles, see Figure 10. In Figure 10 we also display the solution without control, i.e., with zero velocity, which shows that the connecting region becomes thinner due to the effects of diffusion, but the dumbbell remains connected. The optimal velocity and the interface are displayed in Figure 11.

**Figure 10.** From left to right: optimal solution at time $t = 0, 0.15, 0.5$, and the solution at $t = 0.5$ without control; $\lambda = 0$.

**Figure 11.** From left to right: optimal velocity at time $t = 0.1, 0.15, 0.5$ (the velocity vectors are scaled by a factor $0.1, 0.15, 0.4$, respectively); $\lambda = 0$.

9.3. **Effects of the distance and phase-field functionals.** In the next experiment we demonstrate the differences between the optimal solutions for the functional (1.3) and the $L^2$ tracking-type part, i.e., (1.3) with $\beta = 0$. We study a problem with a non-convex initial condition and set $T = 0.5$, $k = 0.05$, $\xi = 0.03$, $\alpha = 10^{-5}$, $\beta = 5$, $\lambda = 5$, $\delta = (3.68\pi)^{-1}$. The remaining parameters were as in the previous experiments. In Figure 12 we display the discontinuous initial condition and the solution at the final time, along with the solution computed by minimizing the distance functional ($\beta = 0$). The solution at the final time for $\beta = 5$ becomes convex, while for the pure $L^2$ distance energy the solution remains non-convex.

**References**

Figure 12. From left to right: Desired state $\tilde{\rho}$, initial condition, density at $t = T$ for $\beta = 5$, density at $t = T$ for $\beta = 0$ ($\lambda = 5$ everywhere).


Department of Mathematics and the Maxwell Institute for Mathematical Sciences, Heriot-Watt University, EH14 4AS Edinburgh, United Kingdom.

E-mail address: L.Banas@hw.ac.uk

Mathematisches Institut, University of Tuebingen, Auf der Morgenstelle 10, 72076 Tuebingen, Germany.

E-mail address: klein@na.uni-tuebingen.de

Mathematisches Institut, University of Tuebingen, Auf der Morgenstelle 10, 72076 Tuebingen, Germany.

E-mail address: prohl@na.uni-tuebingen.de