DISCRETE AND CONTINUOUS
DYNAMICAL SYSTEMS–SERIES B
Volume **6**, Number **5**, September **2006**

Website: http://AIMsciences.org
pp. **1077–1096**

ERROR ESTIMATES FOR TIME-DISCRETIZATIONS FOR THE VELOCITY TRACKING PROBLEM FOR NAVIER-STOKES FLOWS BY PENALTY METHODS

Konstantinos Chrysafinos

Institut für Numerische Simulation
Universität at Bonn
Wegelerstr 6
Bonn 53115, Germany

(Communicated by Jie Shen)

Abstract. Semi-discrete in time approximations of the velocity tracking problem are studied based on a pseudo-compressibility approach. Two different methods are used for the analysis of the corresponding optimality system. The first one, the classical penalty formulation, leads to estimates of order $k + \varepsilon$, under suitable regularity assumptions. The estimate is based on previously derived results for the solution of the unsteady Navier-Stokes problem by penalty methods (see e.g. Jie Shen [26]) and the Brezzi-Rappaz-Raviart theory (see e.g. [12]). The second one, based on the artificially compressible optimality system, leads to an improved estimate of the form $k + \varepsilon k$ for the linearized system.

1. **Introduction.** The purpose of the velocity tracking problem for Navier-Stokes flows, is to drive the velocity vector field $\mathbf{u}$ to a desired target $\mathbf{U}$ using a distributed control function $\mathbf{f}$. In particular, the optimization problem considered here is to find a suitable pair $(\mathbf{u}, \mathbf{f})$ such that

$$J(\mathbf{u}, \mathbf{f}) = \frac{\alpha}{2} \int_0^T \|\mathbf{u}(t) - \mathbf{U}(t)\|^2_{\mathbf{L}^2(\Omega)} dt + \frac{\beta}{2} \int_0^T \|\mathbf{f}(t)\|^2_{\mathbf{L}^2(\Omega)} dt \tag{1.1}$$

is minimized subject to the constraints:

$$\begin{cases} \mathbf{u}_t - \nu\Delta\mathbf{u} + (\mathbf{u}\cdot\nabla)\mathbf{u} + \nabla p = \mathbf{f} + \mathbf{g} & \text{in } \Omega\times(0,T) \\ \nabla\cdot\mathbf{u} = 0 \quad \text{in } \Omega\times(0,T) \\ \mathbf{u} = \mathbf{0} \quad \text{on } \Gamma\times(0,T) \\ \mathbf{u}(0,x) = \mathbf{u}_0 \quad \text{in } \Omega \end{cases} \tag{1.2}$$

where Ω is a two-dimensional bounded domain with smooth boundary Γ, p denotes the pressure, $\mathbf{u}_0, \mathbf{g}$ are given data, α, β are given parameters and ν denotes the kinematic viscosity.

The scope of this work is to analyze semi-discrete in time schemes based on pseudo-compressible methods. Due to the incompressibility constraint, several difficulties arise when solving the Navier-Stokes system numerically even for the uncontrolled system. Pseudo-compressible methods, such as the penalty method, the

---

2000 *Mathematics Subject Classification.* Primary:65M60, 35B37; Secondary: 49J20.
Key words and phrases. Optimal control, velocity tracking, Navier Stokes equations, artificial compressibility, semidiscrete error estimates.

pressure stabilization method, the artificial compressibility method and the projection methods, aim to overcome this difficulty by relaxing the incompressibility constraint. Several results in case of numerical approximations have been derived before in works of [18, 23, 26, 27, 28], in case of the steady and unsteady Navier-Stokes equations. The best available estimate for first order semi-discrete (in time) approximations using a penalized approach (see e.g. [26]) is of order $k+\varepsilon$, where k, ε denote the time discretization and regularization parameters respectively, while in a recent work of [18], an optimal error estimate of the form $k+\varepsilon+h$, (where h denotes the spatial discretization parameter), has been obtained for a backwards Euler - finite element penalized scheme.

Motivated by the above results, which demonstrate the realization of the optimal choice $\varepsilon \approx k$, we study similar schemes for the time discretization of the optimality system of the velocity tracking problem for Navier-Stokes flows. A class of perturbation problems is defined, based on the penalized and the artificially compressible Navier-Stokes equations and discretized in time. We obtain optimal semi-discrete (in time) error estimates which are in accord with the previously developed theory for pseudo-compressible schemes for the uncontrolled unsteady Navier-Stokes equations.

It is worthmentioning that the optimality system consists of two coupled (a forward and a backward in time) systems of Navier-Stokes type. The above coupling creates many difficulties in the analysis and implementation of numerical schemes. The pseudo-compressible condition is an important asset which facilitates the uncoupling of the state and adjoint variables.

The mathematical literature concerning the analysis of distributed optimal control problems related to evolutionary equations is quite extensive (see [1, 11, 13, 24, 25] and references within) where several results regarding the existence of optimal solutions of various optimal control problems are presented. In [16] analysis and finite element approximations of the velocity tracking problem are presented based on a *"discretize and then optimize"* approach. In particular, a discrete (in time) functional is introduced and the corresponding optimality system is rigorously derived. Then, the optimality system is analyzed based on a gradient algorithm. In case of bounded controls a similar approach is illustrated in [15]. An *"optimize then discretize"* approach is used in [10] to prove semi-discrete (in space) error estimates of optimal order in case of the Taylor-Hood element. Second order methods are studied in [20] for local solutions of optimal control problems for Navier-Stokes flows. Perturbation techniques for the analysis and semi-discrete (in time) discretization for MHD flows have been recently used in [17]. We note also that in [22] an optimal Dirichlet control problem related to elliptic Navier-Stokes equations is studied via a penalized approach. Finally, the artificial compressibility approach for a shape optimization problem was used in [29].

2. **Background.**

2.1. **Notation and definition of the optimal control problem.**

We shall use standard notation for Sobolev spaces (see e.g. [2]), $L^2(\Omega), H^1(\Omega), H^m(\Omega)$, $H^1_0(\Omega) = \{v \in H^1(\Omega) : v|_\Gamma = 0\}$, and we denote by H^{-1} the dual space of $H^1_0(\Omega)$. The corresponding vector valued spaces are denoted by $\mathbf{L}^2(\Omega), \mathbf{H}^1(\Omega), \mathbf{H}^m(\Omega), \mathbf{H}^1_0(\Omega), \mathbf{H}^{-1}(\Omega)$ and we employ the standard notation for inner products, norms, and duality pairings. If X is a Hilbert space, we denote by X^* its dual and by $L^p(0,T;X), H^1(0,T;X)$

the time-space function spaces such that

$$\|v\|^p_{L^p(0,T,X)} \equiv \int_0^T \|v(t)\|^p_X dt < \infty \qquad \forall v \in L^p(0,T;X), \quad 1 \geq p < \infty$$

and

$$\|v\|^2_{H^1(0,T;X)} \equiv \int_0^T (\|v(t)\|^2_X + \|v_t(t)\|^2_X) dt < \infty \qquad \forall v \in H^1(0,T;X),$$

together with the standard modification for $L^\infty(0,T;X)$.

Moreover we define the solenoidal vector spaces

$$\begin{aligned} \mathcal{V}(\Omega) &= \{\mathbf{u} \in (C_0^\infty(\Omega))^2 : \nabla \cdot \mathbf{u} = 0\}, \\ V(\Omega) &= \{\mathbf{u} \in \mathbf{H}^1_0(\Omega) : \nabla \cdot \mathbf{u} = 0\}, \\ W(\Omega) &= \{\mathbf{u} \in \mathbf{L}^2(\Omega) : \nabla \cdot \mathbf{u} = 0, \quad \mathbf{u} \cdot \mathbf{n} = 0\}, \end{aligned}$$

where $\mathbf{n}$ is the unit outer normal. Note that $V(\Omega), W(\Omega)$ are the closures of $\mathcal{V}(\Omega)$ in $\mathbf{H}^1(\Omega)$ and $\mathbf{L}^2(\Omega)$, respectively. We also define,

$$L^2_0(\Omega) = \{p \in L^2(\Omega) : \int_\Omega p dx = 0\}.$$

In order to introduce the weak formulation of the evolutionary Navier-Stokes equations we define the following continuous bilinear forms,

$$a(\mathbf{u}, \mathbf{v}) = \sum_{i,j}^2 \int_\Omega D_{ij}(\mathbf{u}) D_{ij}(\mathbf{v}) dx \qquad \forall\, \mathbf{u}, \mathbf{v} \in \mathbf{H}^1(\Omega),$$

$$b(\mathbf{v}, q) = -\int_\Omega q \nabla \cdot \mathbf{v} dx \qquad \forall\, q \in L^2(\Omega), \mathbf{v} \in \mathbf{H}^1(\Omega),$$

where $D_{ij}(\mathbf{v}) = \frac{1}{2}(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i})$. Moreover we define the continuous trilinear form,

$$c(\mathbf{w}, \mathbf{u}, \mathbf{v}) = \sum_{i,j}^2 \int_\Omega w_j (\frac{\partial u_i}{\partial x_j}) v_i dx \qquad \forall\, \mathbf{w}, \mathbf{u}, \mathbf{v} \in \mathbf{H}^1(\Omega).$$

The target field $\mathbf{U} \in \mathcal{B}$, if and only if $\mathbf{U} \in C(0,T;\mathbf{H}^2(\Omega))$, $F_{\mathbf{U}} = \mathbf{U}_t - \nu \Delta \mathbf{U} + (\mathbf{U} \cdot \nabla)\mathbf{U} \quad \in L^\infty(0,T;\mathbf{L}^2(\Omega))$, $\text{div}\mathbf{U} = 0$ and $\mathbf{U}|_\Gamma = 0$, so that the target $\mathbf{U} \in \mathcal{B}$ has a physical meaning. Then, a weak form for the Navier-Stokes equations can be defined as follows: We seek a velocity $\mathbf{u} \in L^2(0,T;V(\Omega)) \cap H^1(0,T;V(\Omega)^*)$ and a pressure p such that

$$\begin{cases} \langle \mathbf{u}_t, \mathbf{v} \rangle + \nu a(\mathbf{u}, \mathbf{v}) + c(\mathbf{u}, \mathbf{u}, \mathbf{v}) = \langle \mathbf{f} + \mathbf{g}, \mathbf{v} \rangle \qquad \forall \mathbf{v} \in V(\Omega) \\ \mathbf{u}(0, x) = \mathbf{u}_0(x) \qquad \in W(\Omega). \end{cases} \tag{2.1}$$

Assuming $\mathbf{f} + \mathbf{g} \in L^2(0,T;V(\Omega)^*)$ the pressure p satisfies (1.2) in distributional sense. However, if we assume more regular data, a precise regularity statement for $\mathbf{u}_t, p$ is valid (see e.g. [27]).

We define the set of admissible solution pairs $(\mathbf{u}, \mathbf{f})$ and the optimal solution of the control problem (P) as follows:

Definition 2.1. *Given,* $T > 0$, $\mathbf{u}_0 \in V(\Omega)$, $\mathbf{g} \in L^2(0,T;\mathbf{L}^2(\Omega))$, *and* $\mathbf{U} \in \mathcal{B}$,

$$U^1_{ad} = \Big\{ (\mathbf{u}, \mathbf{f}) \in L^2(0,T;\mathbf{H}^1_0(\Omega)) \times L^2(0,T;\mathbf{L}^2(\Omega)), \quad \textit{and there exists a pressure } p \in L^2(0,T;L^2_0(\Omega)) \textit{ such that (2.1) is satisfied} \Big\}.$$

Definition 2.2. *Given $T > 0$, $\mathbf{u}_0 \in V(\Omega)$, $\mathbf{g} \in L^2(0,T;\mathbf{L}^2(\Omega))$, and $\mathbf{U} \in \mathcal{B}$, we seek $(\mathbf{u},\mathbf{f}) \in U^1_{ad}$ such that $J(\mathbf{u},\mathbf{f}) \leq J(\mathbf{w},\phi) \quad \forall (\mathbf{w},\phi) \in U^1_{ad}$.*

Next we state the main existence result [16, Theorem 2.1] for the above optimal control problem (P).

Theorem 2.3. *For $\mathbf{u}_0 \in V(\Omega)$,$\mathbf{g} \in L^2(0,T;\mathbf{L}^2(\Omega))$, and $\mathbf{U} \in \mathcal{B}$, there exists a solution of optimal control problem (P).*

Using techniques of Calculus of Variations, the constrained optimization problem is related to the following optimality system. For a proof of this statement we refer the reader to [16, Section 2]. The weak optimality system corresponding to problem (P), can be written as:

$$\left\{\begin{array}{l} \langle \mathbf{u}_t, \mathbf{v}\rangle + \nu a(\mathbf{u},\mathbf{v}) + c(\mathbf{u},\mathbf{u},\mathbf{v}) + b(\mathbf{v},p) = (\mathbf{f}+\mathbf{g},\mathbf{v}) \qquad \forall\, \mathbf{v} \in \mathbf{H}^1_0(\Omega) \\ b(\mathbf{u},q) = 0 \qquad \forall\, q \in L^2_0(\Omega) \\ (\mathbf{u}(0,x) - \mathbf{u}_0, \mathbf{z}) = 0 \qquad \forall\, \mathbf{z} \in \mathbf{L}^2(\Omega) \end{array}\right. \tag{2.2}$$

$$\left\{\begin{array}{l} -\langle \boldsymbol{\mu}_t, \mathbf{v}\rangle + \nu a(\boldsymbol{\mu},\mathbf{v}) + c(\boldsymbol{\mu},\mathbf{u},\mathbf{v}) + c(\mathbf{u},\boldsymbol{\mu},\mathbf{v}) \\ \qquad\qquad + b(\mathbf{v},r) = \alpha(\mathbf{u}-\mathbf{U},\mathbf{v}) \qquad \forall\, \mathbf{v} \in \mathbf{H}^1_0(\Omega) \\ b(\boldsymbol{\mu},q) = 0 \qquad \forall\, q \in L^2_0(\Omega) \\ (\boldsymbol{\mu}(T,x), \mathbf{z}) = 0 \qquad \forall\, \mathbf{z} \in \mathbf{L}^2(\Omega) \end{array}\right. \tag{2.3}$$

$$\boldsymbol{\mu} = -\beta \mathbf{f}. \tag{2.4}$$

2.2. **Penalized optimal control problems.** We introduce a class for perturbation problems, denoted by (P_ε), based on a penalized weak formulation. For $\varepsilon > 0$, minimize the functional,

$$J(\mathbf{u}^\varepsilon,\mathbf{f}^\varepsilon) = \frac{\alpha}{2}\int_0^T \|\mathbf{u}^\varepsilon(t) - \mathbf{U}(t)\|^2_{\mathbf{L}^2(\Omega)} dt + \frac{\beta}{2}\int_0^T \|\mathbf{f}^\varepsilon(t)\|^2_{\mathbf{L}^2(\Omega)} dt$$

subject to the constraints:

$$\left\{\begin{array}{l} \langle \mathbf{u}^\varepsilon_t, \mathbf{v}\rangle + \nu a(\mathbf{u}^\varepsilon,\mathbf{v}) + \hat{c}(\mathbf{u}^\varepsilon,\mathbf{u}^\varepsilon,\mathbf{v}) + b(\mathbf{v},p^\varepsilon) = (\mathbf{f}^\varepsilon+\mathbf{g},\mathbf{v}) \qquad \forall \mathbf{v} \in \mathbf{H}^1_0(\Omega) \\ \varepsilon(p^\varepsilon,q) - b(\mathbf{u}^\varepsilon,q) = 0 \qquad \forall\, q \in L^2(\Omega) \\ (\mathbf{u}^\varepsilon(0,x), \mathbf{z}) = (\mathbf{u}_0(x), \mathbf{z}) \qquad \mathbf{z} \in \mathbf{L}^2(\Omega) \end{array}\right. \tag{2.5}$$

where $\hat{c}(\mathbf{u},\mathbf{v},\mathbf{w})$ is the modified trilinear term defined by,

$$\hat{c}(\mathbf{u},\mathbf{v},\mathbf{w}) = \frac{1}{2}(c(\mathbf{u},\mathbf{v},\mathbf{w}) - c(\mathbf{u},\mathbf{w},\mathbf{v})) \quad \forall\, \mathbf{u},\mathbf{v},\mathbf{w} \in \mathbf{H}^1_0(\Omega).$$

To guarantee the existence of a unique solution pair of (2.5) on the "natural energy spaces" (see e.g. [28]), $(\mathbf{u}^\varepsilon, p^\varepsilon) \in L^2(0,T;\mathbf{H}^1_0(\Omega)) \cap H^1(0,T;\mathbf{H}^{-1}(\Omega)) \times L^2(0,T;L^2(\Omega))$ the data need only to satisfy: $\mathbf{u}_0 \in \mathbf{L}^2(\Omega)$, $\mathbf{g} \in L^2(0,T;\mathbf{H}^{-1}(\Omega))$. The modified trilinear form satisfies the following properties (see e.g. [9, 27]):

$$\left\{\begin{array}{lcl} \hat{c}(\mathbf{u},\mathbf{v},\mathbf{w}) & = & c(\mathbf{u},\mathbf{v},\mathbf{w}) \qquad \forall\, \mathbf{u} \in V(\Omega), \mathbf{v},\mathbf{w} \in \mathbf{H}^1_0(\Omega), \\ \hat{c}(\mathbf{u},\mathbf{v},\mathbf{w}) & = & -\hat{c}(\mathbf{u},\mathbf{w},\mathbf{v}) \qquad \forall\, \mathbf{u},\mathbf{v},\mathbf{w} \in \mathbf{H}^1_0(\Omega), \\ \hat{c}(\mathbf{u},\mathbf{v},\mathbf{w}) & \leq & C\|\mathbf{u}\|^{\frac{1}{2}}_{\mathbf{L}^2(\Omega)}\|\nabla \mathbf{u}\|^{\frac{1}{2}}_{\mathbf{L}^2(\Omega)} \\ & & \times\|\nabla \mathbf{v}\|_{\mathbf{L}^2(\Omega)}\|\mathbf{w}\|^{\frac{1}{2}}_{\mathbf{L}^2(\Omega)}\|\nabla \mathbf{w}\|^{\frac{1}{2}}_{\mathbf{L}^2(\Omega)} \quad \forall\, \mathbf{u},\mathbf{v},\mathbf{w} \in \mathbf{H}^1_0(\Omega), \\ \hat{c}(\mathbf{u},\mathbf{v},\mathbf{w}) & \leq & C\|\mathbf{u}\|_{\mathbf{H}^1(\Omega)}\|\mathbf{v}\|_{\mathbf{H}^2(\Omega)} \\ & & \times\|\mathbf{w}\|_{\mathbf{L}^2(\Omega)} \quad \forall\, \mathbf{u},\mathbf{w} \in \mathbf{H}^1_0(\Omega), \mathbf{v} \in \mathbf{H}^2(\Omega) \cap \mathbf{H}^1_0(\Omega), \\ \hat{c}(\mathbf{u},\mathbf{v},\mathbf{w}) & \leq & C\|\mathbf{u}\|_{\mathbf{L}^2(\Omega)}\|\mathbf{v}\|_{\mathbf{H}^2(\Omega)} \\ & & \times\|\mathbf{w}\|_{\mathbf{H}^1(\Omega)} \quad \forall\, \mathbf{u},\mathbf{w} \in \mathbf{H}^1_0(\Omega), \mathbf{v} \in \mathbf{H}^2(\Omega) \cap \mathbf{H}^1_0(\Omega). \end{array}\right.$$

Similarly, we may define the admissibility set U^2_{ad} and the optimal control problem (P_ε).

Definition 2.4. *Given $T > 0$, $\mathbf{u}_0 \in V(\Omega)$, $\mathbf{g} \in L^2(0,T;\mathbf{L}^2(\Omega))$, and $\mathbf{U} \in \mathcal{B}$,*

$$U^2_{ad} = \Big\{(\mathbf{u}^\varepsilon, \mathbf{f}^\varepsilon) \in L^2(0,T;\mathbf{H}^1_0(\Omega)) \times L^2(0,T;\mathbf{L}^2(\Omega)), \text{ and there exists a pressure } p^\varepsilon \in L^2(0,T;L^2(\Omega)) \text{ such that (2.5) is satisfied}\Big\}.$$

Definition 2.5. *Given $T > 0$, $\mathbf{u}_0 \in V(\Omega)$, $\mathbf{g} \in L^2(0,T;\mathbf{L}^2(\Omega))$, and $\mathbf{U} \in \mathcal{B}$, we seek $(\mathbf{u}^\varepsilon, \mathbf{f}^\varepsilon) \in U^2_{ad}$ such that $J(\mathbf{u}^\varepsilon, \mathbf{f}^\varepsilon) \le J(\mathbf{w}, \phi) \quad \forall (\mathbf{w}, \phi) \in U^2_{ad}$.*

Below, we state the main result regarding the existence of optimal solutions for the perturbed optimal control problems (P_ε) and their convergence properties as $\varepsilon \to 0$.

Theorem 2.6. *Suppose that $\mathbf{u}_0 \in V(\Omega)$, $\mathbf{g} \in L^2(0,T;\mathbf{L}^2(\Omega))$, and $\mathbf{U} \in \mathcal{B}$. Then, for every ε, there exists an optimal solution $(\mathbf{u}^\varepsilon, p^\varepsilon, \mathbf{f}^\varepsilon)$ of the optimal control problem (P_ε). In addition,*

$$(\mathbf{u}^\varepsilon, p^\varepsilon, \mathbf{f}^\varepsilon) \to (\mathbf{u}, p, \mathbf{f}) \qquad \text{as } \varepsilon \to 0$$

where $(\mathbf{u}, p, \mathbf{f})$ is an optimal solution of (P).

Proof. The existence of an optimal solution of problem (P_ε) can be proven similar to [16, Section 2]. The main convergence result as $\varepsilon \to 0$ can be proven by standard techniques (see e.g. [7, Section 3]). □

Remark 2.7. *In the definition of (P_ε), it is possible to further relax the regularity assumptions on data $\mathbf{u}_0, \mathbf{g}$, to $\mathbf{u}_0 \in \mathbf{L}^2(\Omega), \mathbf{g} \in L^2(0,T;\mathbf{H}^{-1}(\Omega))$. Then, as shown in* [7, 8] *the convergence of optimal solutions as $\varepsilon \to 0$ can be established. However, $p := \lim_{\varepsilon\to 0} p^\varepsilon$ only satisfies the Navier-Stokes equations (1.2) in distributional sense.*

Similar to [16, Section 2], one can show that the optimality system corresponding to (P_ε) is given by:

$$\left\{\begin{array}{l} \langle \mathbf{u}^\varepsilon_t, \mathbf{v}\rangle + \nu a(\mathbf{u}^\varepsilon, \mathbf{v}) + \hat{c}(\mathbf{u}^\varepsilon, \mathbf{u}^\varepsilon, \mathbf{v}) + b(\mathbf{v}, p^\varepsilon) = (\mathbf{f}^\varepsilon + \mathbf{g}, \mathbf{v}) \qquad \forall\, \mathbf{v} \in \mathbf{H}^1_0(\Omega) \\ \varepsilon(p^\varepsilon, q) - b(\mathbf{u}^\varepsilon, q) = 0 \qquad \forall\, q \in L^2(\Omega) \\ (\mathbf{u}^\varepsilon(0,x) - \mathbf{u}_0, \mathbf{z}) = 0 \qquad \forall\, \mathbf{z} \in \mathbf{L}^2(\Omega) \end{array}\right. \tag{2.6}$$

$$\left\{\begin{array}{l} -\langle \boldsymbol{\mu}^\varepsilon_t, \mathbf{v}\rangle + \nu a(\boldsymbol{\mu}^\varepsilon, \mathbf{v}) + \hat{c}(\boldsymbol{\mu}^\varepsilon, \mathbf{u}^\varepsilon, \mathbf{v}) + \hat{c}(\mathbf{u}^\varepsilon, \boldsymbol{\mu}^\varepsilon, \mathbf{v}) \\ \qquad\qquad + b(\mathbf{v}, r^\varepsilon) = \alpha(\mathbf{u}^\varepsilon - \mathbf{U}, \mathbf{v}) \qquad \forall\, \mathbf{v} \in \mathbf{H}^1_0(\Omega) \\ -\varepsilon(r^\varepsilon, q) - b(\boldsymbol{\mu}^\varepsilon, q) = 0 \qquad \forall\, q \in L^2(\Omega) \\ (\boldsymbol{\mu}^\varepsilon(T,x), \mathbf{z}) = 0 \qquad \forall\, \mathbf{z} \in \mathbf{L}^2(\Omega). \end{array}\right. \tag{2.7}$$

$$\boldsymbol{\mu}^\varepsilon = -\beta \mathbf{f}^\varepsilon. \tag{2.8}$$

We further quote regularity results regarding the solvability of problem (2.5).

Proposition 2.8. *Let $\mathbf{f}^\varepsilon_1 \equiv \mathbf{f}^\varepsilon + \mathbf{g}$ and suppose that $\mathbf{f}^\varepsilon_1 \in L^2(0,T;\mathbf{L}^2(\Omega)), \mathbf{u}_0 \in V(\Omega)$ with norms bounded independent of ε. Then, there exists a unique solution pair of (2.5), such that*

$$(\mathbf{u}^\varepsilon, p^\varepsilon) \in L^\infty(0,T;\mathbf{H}^1_0(\Omega)) \cap L^2(0,T;\mathbf{H}^2(\Omega) \cap \mathbf{H}^1_0(\Omega)) \times L^2(0,T;H^1(\Omega))$$

with norms bounded independent of ε.

In addition, if we assume that $\mathbf{f}_1^\varepsilon \in L^\infty(0,T;\mathbf{L}^2(\Omega)))$, $\mathbf{f}_{1t}^\varepsilon \in L^2(0,T;\mathbf{H}^{-1}(\Omega))$, $\mathbf{f}_1^\varepsilon(0) \in \mathbf{L}^2(\Omega)$, *and* $\mathbf{u}_0 \in \mathbf{H}^2(\Omega) \cap V(\Omega)$, *with norms bounded independent of* ε, *then there exists a unique solution pair* $(\mathbf{u}^\varepsilon, p^\varepsilon)$ *of (2.5) such that*

$$\begin{cases} \mathbf{u}^\varepsilon \in L^\infty(0,T;\mathbf{H}^2(\Omega) \cap \mathbf{H}_0^1(\Omega)), \mathbf{u}_t^\varepsilon \in L^2(0,T;\mathbf{H}_0^1(\Omega)) \cap L^\infty(0,T;\mathbf{L}^2(\Omega)) \\ p^\varepsilon \in L^\infty(0,T;H^1(\Omega)) \end{cases}$$

with norms bounded independent of ε.

Proof. The first result is stated in [26, Section 2], and it is proven similar to [27]. The second regularity result is proven in [27, Theorems 3.5, 3.6 and Remark 3.8]. For the dependence on ε see [26, Lemma 5.1]. □

Remark 2.9. *Assuming additional regularity on* $\mathbf{g}, \mathbf{u}_0, \mathbf{U}$, *we can apply a "bootstrap" regularity argument to the optimality system (2.6)-(2.7)-(2.8). After noting that* $\|\mathbf{f}^\varepsilon\|_{L^2(0,T;\mathbf{L}^2(\Omega))}$ *is bounded independent of* ε *(see e.g.* [7, Chapter 3]*), then using equation (2.6), we may also obtain that* $(\mathbf{u}^\varepsilon, p^\varepsilon)$ *is bounded independent of* ε *in the* $L^2(0,T;\mathbf{H}^2(\Omega) \cap \mathbf{H}_0^1(\Omega)) \cap L^\infty(0,T;\mathbf{H}_0^1(\Omega)) \cap H^1(0,T;\mathbf{L}^2(\Omega))$ *and* $L^2(0,T;H^1(\Omega))$ *norms respectively. Then, if* $\mathbf{u}_0, \mathbf{U}(0,x)$ $\in \mathbf{H}^2(\Omega) \cap V(\Omega)$, *equation (2.7) and the enhanced regularity at the right hand side, i.e.* $\mathbf{U}, \mathbf{u}^\varepsilon \in L^\infty(0,T;\mathbf{L}^2(\Omega)), \mathbf{U}_t, \mathbf{u}_t^\varepsilon \in L^2(0,T;\mathbf{H}^{-1}(\Omega))$, *implies bounds independent of* ε *for* $\boldsymbol{\mu}^\varepsilon, r^\varepsilon$ *in the* $L^\infty(0,T;\mathbf{H}^2(\Omega) \cap \mathbf{H}_0^1(\Omega)) \cap H^1(0,T;\mathbf{H}_0^1(\Omega)) \times L^\infty(0,T;H^1(\Omega))$ *norms respectively.*

3. Convergence of semi-discrete approximations.

3.1. The discrete optimal control problem.

Next we discretize (in time) the optimality system in the following manner. Let $t_0 = 0 < t_1 < t_2 < \dots < t_N = T$ be a partition of $[0,T]$ into equal subintervals with $k = T/N$. To each function v or vector valued function $\mathbf{v}$ and for every fixed N, we associate the approximate function $\mathbf{v}^N \equiv \mathbf{v}^N(t,x) = \mathbf{v}^n(x), t \in (t_{n-1}, t_n], n = 1, \dots, N$ (sometimes also denoted by the set $\{\mathbf{v}^n(x)\}_{n=1}^N$) and a continuous piecewise (in time) linear function $\mathbf{v}_{pwl}^N = \mathbf{v}_{pwl}^N(t,x)$ by the interpolation conditions $\mathbf{v}_{pwl}^N(t_n,x) = \mathbf{v}^n(x)$ for every $n = 1, \dots, N$. Similarly, on the same partition, we define the discrete (in time) target and data as $\mathbf{U}^n(x) = \mathbf{U}(t_n,x)$ and $\mathbf{g}^n(x) = \mathbf{g}(t_n,x)$ for $n = 1, \dots, N$ respectively.

Then, the discrete (in time) optimality system can be written as: For every $n = 1, \dots, N$,

$$\begin{cases} \frac{1}{k}(\mathbf{u}_\varepsilon^n - \mathbf{u}_\varepsilon^{n-1}, \mathbf{v}) + \nu a(\mathbf{u}_\varepsilon^n, \mathbf{v}) + \hat{c}(\mathbf{u}_\varepsilon^n, \mathbf{u}_\varepsilon^n, \mathbf{v}) + b(\mathbf{v}, p_\varepsilon^n) = (\mathbf{f}_\varepsilon^n + \mathbf{g}^n, \mathbf{v}) \\ \qquad \forall \mathbf{v} \in \mathbf{H}_0^1(\Omega) \\ \varepsilon(p_\varepsilon^n, q) - b(\mathbf{u}_\varepsilon^n, q) = 0 \qquad \forall q \in L^2(\Omega) \\ \mathbf{u}_\varepsilon^n(x) = 0, \qquad \text{on } \Gamma \end{cases} \tag{3.1}$$

$$\begin{cases} -\frac{1}{k}(\boldsymbol{\mu}_\varepsilon^n - \boldsymbol{\mu}_\varepsilon^{n-1}, \mathbf{v}) + \nu a(\mathbf{v}, \boldsymbol{\mu}_\varepsilon^{n-1}) + \hat{c}(\mathbf{u}_\varepsilon^n, \mathbf{v}, \boldsymbol{\mu}_\varepsilon^{n-1}) + \hat{c}(\mathbf{v}, \mathbf{u}_\varepsilon^n, \boldsymbol{\mu}_\varepsilon^{n-1}) \\ \qquad +b(\mathbf{v}, r_\varepsilon^{n-1}) = \alpha(\mathbf{u}_\varepsilon^n - \mathbf{U}^n, \mathbf{v}) \qquad \forall \mathbf{v} \in \mathbf{H}_0^1(\Omega) \\ -\varepsilon(r_\varepsilon^{n-1}, q) - b(\boldsymbol{\mu}_\varepsilon^{n-1}, q) = 0 \qquad \forall q \in L^2(\Omega) \\ \boldsymbol{\mu}_\varepsilon^{n-1}(x) = 0, \qquad \text{on } \Gamma \end{cases} \tag{3.2}$$

$$\mathbf{f}_\varepsilon^n = -\frac{1}{\beta}\boldsymbol{\mu}_\varepsilon^{n-1}. \tag{3.3}$$

The above semi-discrete (in time) optimality system can be viewed as the optimality system corresponding to the following discrete optimal control problem.

Definition 3.1. *Given* $k = T/N$, $\mathbf{u}_0 \in V(\Omega)$, $\mathbf{g} \in L^2(0,T;\mathbf{L}^2(\Omega))$, *and* $\mathbf{U} \in \mathcal{B}$, *we seek* $\{\mathbf{u}_\varepsilon^n, p_\varepsilon^n, \mathbf{f}_\varepsilon^n\}_{n=1}^N \in \mathbf{H}_0^1(\Omega) \times L^2(\Omega) \times \mathbf{L}^2(\Omega)$ *such that the discrete functional*

$$J^N(\{\mathbf{u}_\varepsilon^n\}_{n=1}^N, \{\mathbf{f}_\varepsilon^n\}_{n=1}^N) = \frac{\alpha k}{2}\sum_{n=1}^N \|\mathbf{u}_\varepsilon^n - \mathbf{U}^n\|^2_{\mathbf{L}^2(\Omega)} + \frac{\beta k}{2}\sum_{n=1}^N \|\mathbf{f}_\varepsilon^n\|^2_{\mathbf{L}^2(\Omega)} \tag{3.4}$$

is minimized subject to the constraints: $\forall n = 1, ..., N$,

$$\begin{cases} \frac{1}{k}(\mathbf{u}_\varepsilon^n - \mathbf{u}_\varepsilon^{n-1}, \mathbf{v}) + \nu a(\mathbf{u}_\varepsilon^n, \mathbf{v}) + \hat{c}(\mathbf{u}_\varepsilon^n, \mathbf{u}_\varepsilon^n, \mathbf{v}) + b(\mathbf{v}, p_\varepsilon^n) = (\mathbf{f}_\varepsilon^n + \mathbf{g}^n, \mathbf{v}) \forall \mathbf{v} \in \mathbf{H}_0^1(\Omega) \\ \varepsilon(p_\varepsilon^n, q) - b(\mathbf{u}_\varepsilon^n, q) = 0 \qquad \forall q \in L^2(\Omega) \\ \mathbf{u}_\varepsilon^n(x) = 0, \qquad on\ \Gamma. \end{cases} \tag{3.5}$$

The set of the admissible solutions U_{ad}^d of the discrete optimal control problem can be defined similar to U_{ad}^2, i.e., given $k = T/N$, $\mathbf{u}_0 \in V(\Omega)$, $\mathbf{g} \in L^2(0,T;\mathbf{L}^2(\Omega))$ and $U \in \mathcal{B}$, then

$$U_{ad}^d = \{\{\mathbf{u}_\varepsilon^n, p_\varepsilon^n, \mathbf{f}_\varepsilon^n\}_{n=1}^N \in \mathbf{H}_0^1(\Omega) \times L^2(\Omega) \times \mathbf{L}^2(\Omega) \text{ such that (3.5) is satisfied}\}$$

Note that as $k \to 0$ the functional (3.4) tends to the functional of the corresponding continuous problem and the initial value $\mathbf{f}_\varepsilon^0$ is not involved in the above formulation, so it can be chosen arbitrary.

Next we prove the existence of a solution for the discrete optimal control problem and we clarify the dependance on ε of various norms. Then, the optimality system (3.1)-(3.2)-(3.3) follows using techniques similar to [16, Section 3].

Lemma 3.2. *Given* $k = T/N$, $\mathbf{u}_0 \in V(\Omega)$, $\mathbf{g} \in L^2(0,T;\mathbf{L}^2(\Omega))$, *and* $\mathbf{U} \in \mathcal{B}$ *then for a solution sequence* $\{\mathbf{u}_\varepsilon^n, p_\varepsilon^n, \mathbf{f}_\varepsilon^n\}_{n=1}^N$ *of the discrete optimal control problem we obtain:*$\forall n = 1, .., N$,

$$\begin{aligned} &\|\mathbf{u}_\varepsilon^n\|^2_{\mathbf{L}^2(\Omega)} + k\nu \sum_{i=1}^n \|\mathbf{u}_\varepsilon^i\|^2_{\mathbf{H}_0^1(\Omega)} + \varepsilon k \sum_{i=1}^n \|p_\varepsilon^i\|^2_{\mathbf{L}^2(\Omega)} \\ \le\ & C(\frac{1}{\nu})k \sum_{i=1}^n (\|\mathbf{f}_\varepsilon^i\|^2_{\mathbf{L}^2(\Omega)} + \|\mathbf{g}^i\|^2_{\mathbf{L}^2(\Omega)}) < \infty. \end{aligned}$$

Proof. Multiply the (3.5) by $2\mathbf{u}_\varepsilon^n$ and $2p_\varepsilon^n$ respectively and adding the corresponding equalities we obtain:

$$\begin{aligned} &\|\mathbf{u}_\varepsilon^n\|^2_{\mathbf{L}^2(\Omega)} - \|\mathbf{u}_\varepsilon^{n-1}\|^2_{\mathbf{L}^2(\Omega)} + \|\mathbf{u}_\varepsilon^n - \mathbf{u}_\varepsilon^{n-1}\|^2_{\mathbf{L}^2(\Omega)} \\ &+ C\nu k \|\mathbf{u}_\varepsilon^n\|^2_{\mathbf{H}_0^1(\Omega)} + 2\varepsilon k \|p_\varepsilon^n\|^2_{\mathbf{L}^2(\Omega)} \\ \le\ & \frac{2Ck}{\nu}\big(\|\mathbf{f}_\varepsilon^n\|^2_{\mathbf{L}^2(\Omega)} + \|\mathbf{g}_\varepsilon^n\|^2_{\mathbf{L}^2(\Omega)}\big). \end{aligned} \tag{3.6}$$

Summing the above inequalities from 1 to n we obtain the desired estimate. □

Lemma 3.3. *Let* $\varepsilon > 0$, $k = T/N$ *and suppose that* $\mathbf{u}_0 \in V(\Omega)$, $\mathbf{g} \in L^2(0,T;\mathbf{L}^2(\Omega))$, $\mathbf{U} \in \mathcal{B}$. *Then for a solution sequence* $\{\mathbf{u}_\varepsilon^n, p_\varepsilon^n, \mathbf{f}_\varepsilon^n\}_{n=1}^N$ *of the discrete optimal control problem we obtain:*

$$\beta k \sum_{n=1}^N \|\mathbf{f}_\varepsilon^n\|^2_{\mathbf{L}^2(\Omega)}, \quad \nu k \sum_{n=1}^N \|\mathbf{u}_\varepsilon^n\|^2_{\mathbf{H}_0^1(\Omega)},$$

$$\|\mathbf{u}_\varepsilon^N\|^2_{\mathbf{L}^2(\Omega)} + \varepsilon k \sum_{n=1}^N \|p_\varepsilon^n\|^2_{L^2(\Omega)} \le C(\frac{1}{\nu}, T, \alpha) < \infty. \tag{3.7}$$

where $C(\frac{1}{\nu}, T, \alpha)$ is independent of ε, k.

Proof. For every ε let $\{\tilde{\mathbf{u}}^n_\varepsilon, \tilde{p}^n_\varepsilon\}_{n=1}^N$ be the solution of

$$\begin{cases} \frac{1}{k}(\tilde{\mathbf{u}}^n_\varepsilon - \tilde{\mathbf{u}}^{n-1}_\varepsilon, \mathbf{v}) + \nu a(\tilde{\mathbf{u}}^n_\varepsilon, \mathbf{v}) + \hat{c}(\tilde{\mathbf{u}}^n_\varepsilon, \tilde{\mathbf{u}}^n_\varepsilon, \mathbf{v}) + b(\mathbf{v}, \tilde{p}^n_\varepsilon) = (\mathbf{g}^n, \mathbf{v}) \quad \forall\, \mathbf{v} \in \mathbf{H}^1_0(\Omega) \\ \varepsilon(\tilde{p}^n_\varepsilon, q) - b(\tilde{\mathbf{u}}^n_\varepsilon, q) = 0 \quad \forall\, q \in L^2(\Omega) \\ \tilde{\mathbf{u}}^n_\varepsilon(x) = 0, \qquad \text{on } \Gamma. \end{cases} \tag{3.8}$$

Then $\{\tilde{\mathbf{u}}^n_\varepsilon, \tilde{p}^n_\varepsilon, \mathbf{0}\}_{n=1}^N$ belongs to the discrete admissibility set U^d_{ad} and moreover,

$$\begin{aligned} J^N(\{\mathbf{u}^n_\varepsilon\}_{n=1}^N, \{\mathbf{f}^n_\varepsilon\}_{n=1}^N) &= \frac{\alpha k}{2} \sum_{n=1}^N \|\mathbf{u}^n_\varepsilon - \mathbf{U}^n\|^2_{\mathbf{L}^2(\Omega)} + \frac{\beta k}{2} \sum_{n=1}^N \|\mathbf{f}^n_\varepsilon\|^2_{\mathbf{L}^2(\Omega)} \\ &\leq J^N(\{\tilde{\mathbf{u}}^n_\varepsilon\}_{n=1}^N, 0) = \frac{\alpha k}{2} \sum_{n=1}^N \|\tilde{\mathbf{u}}^n_\varepsilon - \mathbf{U}^n\|^2_{\mathbf{L}^2(\Omega)}. \end{aligned}$$

The later sum is bounded independent of ε, k due to Lemma 3.2 applied to equation (3.8), which clearly implies the first result. Returning back to the estimates of Lemma 3.2, we easily obtain the last two bounds. □

Next we prove the existence of an optimal solution for the discrete problem for every fixed ε.

Theorem 3.4. *Let $\varepsilon > 0$, $k = T/N$ and suppose that $\mathbf{u}_0 \in V(\Omega), \mathbf{g} \in L^2(0,T; \mathbf{L}^2(\Omega))$, $\mathbf{U} \in \mathcal{B}$. Then, there exists*

$$\{\mathbf{u}^n_\varepsilon, p^n_\varepsilon, \mathbf{f}^n_\varepsilon\}_{n=1}^N \in \mathbf{H}^1_0(\Omega) \times L^2(\Omega) \times \mathbf{L}^2(\Omega)$$

such that the discrete functional (3.4) is minimized subject to (3.5).

Proof. For simplicity we drop the ε notation. First note that due to Lemma 3.2 for every ε there exists a solution of the corresponding semi-discrete equations, i.e., $U^d_{ad} \neq \emptyset$. Therefore, we may obtain a minimizing sequence denoted by $\mathbf{u}_m, \mathbf{f}_m$ of the functional (3.4), satisfying: for all $m = 1, 2, ...$ and $n = 1, ..., N$,

$$\begin{cases} \frac{1}{k}(\mathbf{u}^n_m - \mathbf{u}^{n-1}_m, \mathbf{v}) + \nu a(\mathbf{u}^n_m, \mathbf{v}) + c(\mathbf{u}^n_m, \mathbf{u}^n_m, \mathbf{v}) + b(\mathbf{v}, p^n_m) = (\mathbf{f}^n_m + \mathbf{g}^n, \mathbf{v}) \\ \qquad \forall\, \mathbf{v} \in \mathbf{H}^1_0(\Omega) \\ \varepsilon(p^n_m, q) - b(\mathbf{u}^n_m, q) = 0 \quad \forall\, q \in L^2(\Omega) \\ \mathbf{u}^n_m(x) = 0 \quad \text{on } \Gamma. \end{cases} \tag{3.9}$$

It is easy to see that as $m \to \infty$ for every $n = 1, ..., N$

$$\begin{aligned} &\mathbf{u}^n_m \to \mathbf{u}^n \quad \text{in } \mathbf{H}^1_0(\Omega) \text{ weakly}, \quad \mathbf{f}^n_m \to \mathbf{f}^n \quad \text{in } \mathbf{L}^2(\Omega) \text{ weakly} \\ &p^n_m \to p^n \quad \text{in } L^2(\Omega) \text{ weakly}, \quad \mathbf{u}^n_m \to \mathbf{u}^n \quad \text{in } \mathbf{L}^2(\Omega) \text{ strongly}, \end{aligned}$$

where the last result follows from the compact embedding $\mathbf{H}^1_0(\Omega) \subset \mathbf{L}^2(\Omega)$. Due to the strong convergence result, we may pass the limit through the nonlinear term since for any $\mathbf{w} \in C^\infty_0(\Omega)$ we have that

$$\lim_{m \to \infty} \hat{c}(\mathbf{u}^n_m, \mathbf{u}^n_m, \mathbf{w}) = \hat{c}(\mathbf{u}^n, \mathbf{u}^n, \mathbf{w}).$$

A well known density argument, implies that this is still true for $\mathbf{w} \in \mathbf{H}^1_0(\Omega)$. Therefore, we may pass the limit to obtain equations (3.5). The lower semi-continuity of functional (3.4), finishes the proof. □

Finally, we prove that as $k \to 0$ the solution of the semi-discrete optimal control problem converges to the solution of the continuous problem (P_ε).

Theorem 3.5. *Suppose that* $\mathbf{u}_0 \in V(\Omega), \mathbf{g} \in L^2(0,T;\mathbf{L}^2(\Omega)), \mathbf{U} \in \mathcal{B}$ *and let* $k = T/N$*. Then, the solution of the semi-discrete optimal control problem*

$$\{\mathbf{u}_\varepsilon^n, p_\varepsilon^n, \mathbf{f}_\varepsilon^n\}_{n=1}^N \in \mathbf{H}_0^1(\Omega) \times L^2(\Omega) \times \mathbf{L}^2(\Omega)$$

converges to the solution of the optimal control problem (P_ε) *as* $k \to 0$.

Proof. Set $\bar{\mathbf{u}}_\varepsilon^n \equiv (\mathbf{u}_\varepsilon^n - \mathbf{u}_\varepsilon^{n-1})/k$. Lemma 3.2 implies that

$$\|\mathbf{u}_\varepsilon^n\|_{\mathbf{L}^2(\Omega)}, \quad k\sum_{n=1}^N \|\mathbf{u}_\varepsilon^n\|_{\mathbf{H}_0^1(\Omega)}^2, \quad k\sum_{n=1}^N \|\mathbf{f}_\varepsilon^n\|_{\mathbf{L}^2(\Omega)}^2, \quad \varepsilon k \sum_{n=1}^N \|p_\varepsilon^n\|_{L^2(\Omega)}^2 < \infty$$

are bounded independent of ε, k. In addition note that $\varepsilon k \sum_{n=1}^N \|\bar{\mathbf{u}}_\varepsilon^n\|_{\mathbf{H}^{-1}(\Omega)}^2$ is also bounded independent of ε, k since,

$$\begin{aligned}\|\bar{\mathbf{u}}_\varepsilon^n\|_{\mathbf{H}^{-1}(\Omega)} &\leq \sup_{\mathbf{v}\in\mathbf{H}_0^1(\Omega)} \frac{\nu|a(\mathbf{u}_\varepsilon^n,\mathbf{v})| + |\hat{c}(\mathbf{u}_\varepsilon^n,\mathbf{u}_\varepsilon^n,\mathbf{v})| + |b(\mathbf{v},p_\varepsilon^n)| + |(\mathbf{f}_\varepsilon^n + \mathbf{g}^n,\mathbf{v})|}{\|\mathbf{v}\|_{\mathbf{H}_0^1(\Omega)}} \\ &\leq C\big(\|\mathbf{u}_\varepsilon^n\|_{\mathbf{H}_0^1(\Omega)} + \|\mathbf{u}_\varepsilon^n\|_{\mathbf{L}^2(\Omega)}\|\mathbf{u}_\varepsilon^n\|_{\mathbf{H}_0^1(\Omega)} \\ &\quad + \|p_\varepsilon^n\|_{L^2(\Omega)} + \|\mathbf{f}_\varepsilon^n\|_{\mathbf{L}^2(\Omega)} + \|\mathbf{g}^n\|_{\mathbf{L}^2(\Omega)}\big).\end{aligned}$$

The last inequality clearly implies the desired bound, after squaring both sides, multiplying by ε, k, summing from $n = 1$ to $n = N$, and using the previously derived a-priori bounds. Therefore, using standard techniques regarding the semi-discrete approximation (in time) for the Navier-Stokes equations (see e.g. [27, Part III, Chapter 4]), we obtain that $\mathbf{u}_\varepsilon^N$, p_ε^N, $\mathbf{f}_\varepsilon^N$ and $\frac{d}{dt}\mathbf{u}_{\varepsilon,pwl}^N$ are bounded (independent of k), in $L^\infty(0,T;\mathbf{L}^2(\Omega)) \cap L^2(0,T;\mathbf{H}_0^1(\Omega))$, $L^2(0,T;L^2(\Omega))$, $L^2(0,T;\mathbf{L}^2(\Omega))$ and $L^2(0,T;\mathbf{H}^{-1}(\Omega))$ respectively. Therefore, we may extract subsequences such that

$$\begin{aligned}&\mathbf{u}_\varepsilon^N \to \mathbf{u}^\varepsilon \text{ in } L^2(0,T;\mathbf{H}_0^1(\Omega)) \text{ weakly} \qquad \mathbf{u}_\varepsilon^N \to \mathbf{u}^\varepsilon \text{ in } L^\infty(0,T;\mathbf{L}^2(\Omega)) \text{ weakly-*} \\ &p_\varepsilon^N \to p^\varepsilon \text{ in } L^2(0,T;L^2(\Omega) \text{ weakly} \qquad \mathbf{f}_\varepsilon^N \to \mathbf{f}^\varepsilon \text{ in } L^2(0,T;\mathbf{L}^2(\Omega)) \text{ weakly} \\ &\frac{d}{dt}\mathbf{u}_{\varepsilon,pwl}^N \to \mathbf{u}_t^\varepsilon \text{ in } L^2(0,T;\mathbf{H}^{-1}(\Omega)) \text{ weakly.}\end{aligned}$$

Due to the compact embedding of $L^2(0,T;\mathbf{H}_0^1(\Omega)) \cap H^1(0,T;\mathbf{H}^{-1}(\Omega)) \subset L^2(0,T;\mathbf{L}^2(\Omega))$ (see [27, Chapter III, Theorem 2.1]), we obtain

$$\mathbf{u}_\varepsilon^N \to \mathbf{u}^\varepsilon \text{ in } L^2(0,T;\mathbf{L}^2(\Omega)) \text{ strongly}$$

The theorem now follows using standard techniques. □

Remark 3.6. *To rigorously derive the semi-discrete optimality system, we may now directly apply the techniques of* [16, Section 3]*, after noting that* $k\sum_{n=1}^N \|\mathbf{f}_\varepsilon^n\|_{\mathbf{L}^2(\Omega)}^2$ *is bounded independent of* ε.

3.2. **Semi-discrete (in time) estimates for a model problem.** The auxiliary linear problem, considered here, is the semi-discrete (in time) approximations of the following problem: Given data $\mathbf{f}_1$, $\mathbf{u}_0$ we seek a solution pair $\mathbf{u}^\varepsilon, p^\varepsilon$ satisfying

$$\begin{cases}(\mathbf{u}_t^\varepsilon,\mathbf{v}) + \nu a(\mathbf{u}^\varepsilon,\mathbf{v}) + b(\mathbf{v},p^\varepsilon) = (\mathbf{f}_1,\mathbf{v}) & \forall\, \mathbf{v} \in \mathbf{H}_0^1(\Omega) \\ \varepsilon(p^\varepsilon,q) - b(\mathbf{u}^\varepsilon,q) = 0 & \forall\, q \in L^2(\Omega) \\ (\mathbf{u}(0,x),\mathbf{z}) = (\mathbf{u}_0,\mathbf{z}) & \forall\, \mathbf{z} \in \mathbf{L}^2(\Omega).\end{cases} \tag{3.10}$$

Next we define spaces suitable for the semi-discrete in time approximations for the model problem (3.10), associated to optimality system (3.1)-(3.2)-(3.3). As before, let $t_0 = 0 < t_1 < t_2 < ... < t_N = T$ be a partition of $[0,T]$ into equal subintervals with $k = T/N$. Recall that to each function or vector valued function $\mathbf{v}$ and for every fixed N, we associate the approximate function $\mathbf{v}^N$ and its set $\{\mathbf{v}^n(x)\}_{n=0}^N$ defined by $\mathbf{v}^N = \mathbf{v}^n(x), t \in (t_{n-1}, t_n], n = 1, ..., N$ and a continuous piecewise linear function $\mathbf{v}_{pwl}^N = \mathbf{v}_{pwl}^N(t,x)$ defined by interpolation conditions $\mathbf{v}^N(t_n, x) = \mathbf{v}^n(x)$ for every $n = 1, ..., N$. Suppose that the data satisfy $\mathbf{f}_1 \in Y$, $\mathbf{u}_0 \in Y_0$ where,

$$\begin{aligned} Y &:= Y^N \\ &= \{\mathbf{f}_1 \in L^\infty(0,T;\mathbf{L}^2(\Omega)), \mathbf{f}_{1t} \in L^2(0,T;\mathbf{H}^{-1}(\Omega)), \\ &\quad\ \mathbf{f}_1(0) \in \mathbf{L}^2(\Omega), \mathbf{f}_1|_{(t_{n-1},t_n]} \in C^0\}, \\ Y_0 &:= \mathbf{H}^2(\Omega) \cap V(\Omega). \end{aligned}$$

We define the solution space $X = X_1 \times X_2 \times M$ where

$$\begin{aligned} X_1 &:= X_1^N = \{\mathbf{u} \in L^\infty(0,T;\mathbf{H}^2(\Omega) \cap \mathbf{H}_0^1(\Omega)) : \mathbf{u}|_{(t_{n-1},t_n])} \in C^0\}, \\ X_2 &:= X_2^N = \{\mathbf{u}_t \in L^\infty(0,T;\mathbf{L}^2(\Omega)) : \mathbf{u}_t|_{(t_{n-1},t_n])} \in C^0\}, \\ M &:= M^N = \{p \in L^\infty(0,T;H^1(\Omega)) : p|_{(t_{n-1},t_n)} \in C^0\}. \end{aligned}$$

and we seek $(\mathbf{u}_\varepsilon^N, \frac{d}{dt}\mathbf{u}_{\varepsilon,pwl}^N, p_\varepsilon^N) \in X$ such that for every $n = 1, ..., N$.

$$\begin{cases} \frac{1}{k}(\mathbf{u}_\varepsilon^n - \mathbf{u}_\varepsilon^{n-1}, \mathbf{v}) + \nu a(\mathbf{u}_\varepsilon^n, \mathbf{v}) + b(\mathbf{v}, p_\varepsilon^n) = (\mathbf{f}_1^n, \mathbf{v}) \quad \forall \mathbf{v} \in \mathbf{H}_0^1(\Omega) \\ \varepsilon(p_\varepsilon^n, q) - b(\mathbf{u}_\varepsilon^n, q) = 0 \quad \forall q \in L^2(\Omega) \\ \mathbf{u}_\varepsilon^n(x) = 0, \qquad \text{on } \Gamma. \end{cases} \tag{3.11}$$

Finally, we quote the main result of [26] regarding semi-discrete in time approximations of the linearized model problem.

Proposition 3.7. *Suppose that $\mathbf{f}_1 \in Y, \mathbf{u}_0 \in Y_0$, and $(\mathbf{u}^\varepsilon, p^\varepsilon)$, $\{\mathbf{u}_\varepsilon^n, p_\varepsilon^n\}_{n=1}^N$ satisfy (3.10) and (3.11) respectively. Then the following estimate holds:* $\forall\, n = 1, ..., N$,

$$\|\mathbf{e}^n\|_{\mathbf{H}_0^1(\Omega)}^2 + \nu k \sum_{i=1}^n \|\mathbf{u}^i\|_{\mathbf{H}^2(\Omega)}^2 \le C(\frac{1}{\nu}, T)k^2$$

$$\nu k \sum_{i=1}^n \|\frac{\mathbf{e}^i - \mathbf{e}^{i-1}}{k}\|_{\mathbf{L}^2(\Omega)}^2 + \nu k \sum_{i=1}^n \|p^\varepsilon(t_i) - p_\varepsilon^i\|_{H^1(\Omega)}^2 \le C(\frac{1}{\nu}, T)k^2$$

where $C(\frac{1}{\nu}, T)$ is a constant independent of ε, and $\mathbf{e}^n := \mathbf{u}^\varepsilon(t_n) - \mathbf{u}_\varepsilon^n$.

Proof. (Sketch) Note that $p^\varepsilon = \frac{1}{\varepsilon}$ div $\mathbf{u}^\varepsilon$, and hence, for data $\mathbf{f}_1 \in Y, \mathbf{u}_0 \in Y_0$ the linear model problem (2.5) can be written into the strong form:

$$\frac{1}{k}(\mathbf{u}_\varepsilon^n - \mathbf{u}_\varepsilon^{n-1}) + \mathcal{A}^\varepsilon \mathbf{u}^n = \mathbf{f}_1^n$$

where the operator $\mathcal{A}^\varepsilon \equiv -\nu\Delta\mathbf{u}^\varepsilon - \frac{1}{\varepsilon}\nabla$ div $\mathbf{u}^\varepsilon$ is a positive self-adjoint operator from $\mathbf{H}^2(\Omega) \cap \mathbf{H}_0^1(\Omega)$ to $\mathbf{L}^2(\Omega)$ with powers $(\mathcal{A}^\varepsilon)^s, s \in \mathbb{R}$ well defined. Therefore, with $\mathbf{e}^n = \mathbf{u}^\varepsilon(t_n) - \mathbf{u}_\varepsilon^n$, the error equation takes the form of the linear version of [26, Equations 5.6,5.7]

$$\frac{1}{k}(\mathbf{e}^n - \mathbf{e}^{n-1}) + \mathcal{A}^\varepsilon \mathbf{e}^n = \frac{1}{k}\int_{t_{n-1}}^{t_n} (t - t_{n-1})\mathbf{u}_{tt}^\varepsilon dt. \tag{3.12}$$

The first inequality is proven in [26, Lemma 5.2]. The last inequality is essentially contained in the proof of [26, Lemma 5.2, Remark 5.1]. For the first term, we obtain the estimate $\nu k \sum_{i=1}^n \|\frac{\mathbf{e}^i - \mathbf{e}^{i-1}}{k}\|_{\mathbf{L}^2(\Omega)}^2 \le C(\frac{1}{\nu}, T)k^2$.

Finally, for the pressure term note that $p^\varepsilon = \frac{1}{\varepsilon}$ div $\mathbf{u}^\varepsilon$, $p^i_\varepsilon = \frac{1}{\varepsilon}$ div $\mathbf{u}^i_\varepsilon$ so the error equation (3.12) implies,

$$\|p^\varepsilon(t_i) - p^i_\varepsilon\|_{H^1(\Omega)} \leq C\Big(\|\frac{\mathbf{e}^i - \mathbf{e}^{i-1}}{k}\|_{\mathbf{L}^2(\Omega)} + \|\mathbf{e}^i\|_{\mathbf{H}^2(\Omega)} + \|\frac{1}{k}\int_{t_{i-1}}^{t_i}(t - t_{i-1})\mathbf{u}^\varepsilon_{tt} dt\|_{\mathbf{L}^2(\Omega)}\Big).$$

The estimate then follows from the first inequality together with the estimate on the time derivative, after taking the squares, summing from 0 to n muptiplying by νk and using the bound, $k\sum_{n=0}^N \|\frac{1}{k}\int_{t_{n-1}}^{t_n}(t - t_{n-1})\mathbf{u}^\varepsilon_{tt} dt\|^2_{\mathbf{L}^2(\Omega)} \leq k^2\int_0^T \|\mathbf{u}^\varepsilon_{tt}\|^2_{\mathbf{L}^2(\Omega)} dt$. Note that $\mathbf{u}^\varepsilon_{tt} \in L^2(0, T; \mathbf{L}^2(\Omega))$ with norm bounded independent of ε (see also [26, Lemma 5.1]). □

Remark 3.8. *The estimate on* $\frac{1}{k}(\mathbf{e}^n - \mathbf{e}^{n-1})$ *can be used to obtain estimates in various norms. Suppose that* $\mathbf{u}^\varepsilon_t \in C((t_{n-1}, t_n]; \mathbf{L}^2(\Omega)) \forall n = 1, .., N$. *Then, for every* $t \in (t_{n-1}, t_n]$ *note that* $u_t(t) - \frac{d}{dt}\mathbf{u}^N_{\varepsilon,pwl}(t) = \mathbf{u}^\varepsilon_t(t) - \frac{1}{k}(\mathbf{u}^n_\varepsilon - \mathbf{u}^{n-1}_\varepsilon)$, *and using standard algebra,*

$$\frac{1}{k}\int_{t_{n-1}}^{t_n}(t - t_{n-1})\mathbf{u}^\varepsilon_{tt}(t) dt = \mathbf{u}^\varepsilon_t(t_n) - \frac{1}{k}(\mathbf{e}^n - \mathbf{e}^{n-1}) - \frac{1}{k}(\mathbf{u}^n_\varepsilon - \mathbf{u}^{n-1}_\varepsilon)$$

Combining the last two equalities we obtain

$$\mathbf{u}^\varepsilon_t(t) - \frac{d}{dt}\mathbf{u}^N_{\varepsilon,pwl}(t) = \mathbf{u}^\varepsilon_t(t) - \mathbf{u}^\varepsilon_t(t_n) + \frac{1}{k}(\mathbf{e}^n - \mathbf{e}^{n-1}) + \frac{1}{k}\int_{t_{n-1}}^{t_n}(t - t_{n-1})\mathbf{u}^\varepsilon_{tt}(t) dt,$$

which clearly implies that $\|u^\varepsilon_t(t) - \frac{d}{dt}\mathbf{u}^N_{\varepsilon,pwl}\|_{L^\infty((t^{n-1}, t_n]; \mathbf{L}^2(\Omega))} \to 0$.

3.3. **Some results concerning the approximation of a class of nonlinear problems.** Next we describe the main results concerning the Brezzi-Rappaz-Raviart (BRR) theory, introduced in [4]. To our knowledge within the context of optimal control problems, BRR theory was first used in [14] to handle the nonlinear terms and to uncouple the discrete state and adjoint equations in case of a Dirichlet boundary optimal control problem for the stationary Navier-Stokes equations. The essence of this theory is that under certain hypotheses the error of the approximation of the coupled problem is of the same order of a related uncoupled linear one. For a more extensive presentation of the BRR theory one may consult [12]. The problems considered in [12], specialized to our needs, are of the following type: Let $\mathcal{X}, \mathcal{Y}$ be Banach spaces. We seek a $\psi \in \mathcal{X}$ such that

$$\psi + \mathcal{T}\mathcal{G}(\lambda, \psi) = 0, \tag{3.13}$$

where $\mathcal{T} \in \mathcal{L}(\mathcal{Y}, \mathcal{X}), \mathcal{G}$ is a C^2 mapping from $\Lambda \times \mathcal{X}$ into $\mathcal{Y}$, and Λ is a compact interval of $\mathbb{R}$. The set $\{\lambda, \psi(\lambda) | \lambda \in \Lambda\}$ is called a branch of solutions of (3.13) if $\lambda \to \psi(\lambda)$ is a continuous function from Λ into $\mathcal{X}$ such that (3.13) is satisfied. The solution branch ψ (depending on λ) is called regular if we also have that $I + \mathcal{T}\mathcal{G}_\psi(\lambda, \psi)$ is an isomorphism from $\mathcal{X}$ to $\mathcal{X}$, where $\mathcal{G}_\psi$ denotes the Fréchet derivative with respect to the ψ and I the identity mapping. We assume that there exists another Banach space $\mathcal{Z}$, contained in $\mathcal{Y}$, with continuous embedding, such that

$$\psi \to \mathcal{G}_\psi(\lambda, \psi) \in \mathcal{L}(\mathcal{X}, \mathcal{Z}) \qquad \forall \lambda \in \Lambda, \psi \in \mathcal{X}. \tag{3.14}$$

Approximations are defined by introducing a subspace $\mathcal{X}^N \subset \mathcal{X}$ and an approximating (semi-discrete in time, in our case) operator $\mathcal{T}^N \in \mathcal{L}(\mathcal{Y}, \mathcal{X}^N)$. Then, we seek $\psi^N \in \mathcal{X}^N$ such that

$$\psi^N + \mathcal{T}^N\mathcal{G}(\lambda, \psi^N) = 0. \tag{3.15}$$

Concerning the linear operator we assume the approximation properties:

$$\lim_{N\to\infty} \|(\mathcal{T}^N - \mathcal{T})w\|_{\mathcal{X}} = 0 \qquad \forall\, w \in \mathcal{Y} \tag{3.16}$$

and

$$\lim_{N\to\infty} \|\mathcal{T}^N - \mathcal{T}\|_{\mathcal{L}(\mathcal{Z},\mathcal{X})} = 0. \tag{3.17}$$

Note that whenever the imbedding $\mathcal{Z} \subset \mathcal{Y}$ is compact, the last relation follows from (3.16), and moreover the operator $\psi \to \mathcal{T}\mathcal{G}_\psi(\psi) \in \mathcal{L}(\mathcal{X},\mathcal{Y})$ is compact. The main theorem can be stated as follows:

Theorem 3.9. *Let $\mathcal{X}$ and $\mathcal{Y}$ be Banach spaces. Assume that $\mathcal{G}$ is a C^2 mapping from $\Lambda \times \mathcal{X}$ to $\mathcal{Y}$ and that $\mathcal{D}^2\mathcal{G}$is bounded on all bounded sets of $\Lambda \times \mathcal{X}$. Assume that (3.14)-(3.16)-(3.17) hold and that ψ is a regular solution of (3.13). Then there exists a neighborhood $\mathcal{O}$ of the origin in $\mathcal{X}$ and for $N \ge N_0$ big enough a unique function $\psi^N(\lambda) \in \mathcal{X}^N$ such that $\psi^N(\lambda)$ is a regular solution of (3.15), and $\psi^N(\lambda) - \psi(\lambda) \in \mathcal{O}$. Moreover, there exists a constant $C > 0$, independent of N and λ such that*

$$\|\psi^N(\lambda) - \psi(\lambda)\|_{\mathcal{X}} \le C\|(\mathcal{T}^N - \mathcal{T})\mathcal{G}(\lambda,\psi)\|_{\mathcal{X}}. \tag{3.18}$$

Proof. See [12, pp 306-307]. □

3.4. **Error estimates for the optimality System.** In this section, we recast optimality system (2.6)-(2.7)-(2.8) into a form that allow us to use Theorem 3.9. For this purpose set $\lambda \equiv \frac{1}{\nu}$ and adopting the notation of Section 3.2, we define the spaces

$$X := X^N = X_1^N \times X_2^N \times M^N, \qquad \mathcal{X} = X \times X$$

and

$$\mathcal{Y} := Y^N \times Y_0 \times Y^N \times Y_0 \qquad \mathcal{Z} := \mathcal{Y}$$

endowed with the natural norms. Then, we define the linear operator $\mathcal{T} : \mathcal{Y} \to \mathcal{X}$ for every $(\mathbf{f}_1, \mathbf{u}_1, \mathbf{f}_2, \boldsymbol{\mu}_1) \in \mathcal{Y}$ as follows: $T(\mathbf{f}_1, \mathbf{u}_1, \mathbf{f}_2, \boldsymbol{\mu}_1) = (\mathbf{u}^\varepsilon, \mathbf{u}_t^\varepsilon, p^\varepsilon, \boldsymbol{\mu}^\varepsilon, \boldsymbol{\mu}_t^\varepsilon, r^\varepsilon) \in \mathcal{X}$ if and only if

$$\begin{cases} \langle \mathbf{u}_t^\varepsilon, \mathbf{v}\rangle + a(\mathbf{u}^\varepsilon, \mathbf{v}) + b(\mathbf{v}, p^\varepsilon) = (\mathbf{f}_1, \mathbf{v}) & \forall\, \mathbf{v} \in \mathbf{H}_0^1(\Omega) \\ \varepsilon(p^\varepsilon, q) - b(u^\varepsilon, q) = 0 \quad \forall\, q \in L^2(\Omega) \\ (\mathbf{u}^\varepsilon(0,x), \mathbf{z}) = (\mathbf{u}_1, \mathbf{z}) \quad \forall\, \mathbf{z} \in \mathbf{L}^2(\Omega) \end{cases}$$

$$\begin{cases} -\langle \boldsymbol{\mu}_t^\varepsilon, \mathbf{v}\rangle + a(\boldsymbol{\mu}^\varepsilon, \mathbf{v}) + b(\mathbf{v}, r^\varepsilon) = (\mathbf{f}_2, \mathbf{v}) & \forall\, \mathbf{v} \in \mathbf{H}_0^1(\Omega) \\ -\varepsilon(r^\varepsilon, q) - b(\boldsymbol{\mu}^\varepsilon, q) = 0 \quad \forall\, q \in L^2(\Omega) \\ (\boldsymbol{\mu}^\varepsilon(T,x), \mathbf{z}) = (\boldsymbol{\mu}_1, \mathbf{z}) \quad \forall\, \mathbf{z} \in \mathbf{L}^2(\Omega). \end{cases}$$

Analogously, we define the semi-discrete (in time) operator $\mathcal{T}^N \in \mathcal{L}(Y^N, X^N)$, such that $T(\mathbf{f}_1, \mathbf{u}_1, \mathbf{f}_2, \boldsymbol{\mu}_1) = (\mathbf{u}_\varepsilon^N, \frac{d}{dt}\mathbf{u}_{\varepsilon,pwl}^N, p_\varepsilon^N, \boldsymbol{\mu}_\varepsilon^N, \frac{d}{dt}\boldsymbol{\mu}_{\varepsilon,pwl}^N, r_\varepsilon^N) \in X^N \times X^N := \mathcal{X}$ if and only if

$$\begin{cases} (\frac{d}{dt}\mathbf{u}_{\varepsilon,pwl}^N, \mathbf{v}) + a(\mathbf{u}_\varepsilon^N, \mathbf{v}) + b(\mathbf{v}, p_\varepsilon^N) = (\mathbf{f}_1, \mathbf{v}) \quad \forall\, \mathbf{v} \in \mathbf{H}_0^1(\Omega) \\ \varepsilon(p_\varepsilon^N, q) - b(\mathbf{u}_\varepsilon^N, q) = 0 \quad \forall\, q \in L^2(\Omega) \\ \mathbf{u}_\varepsilon^N(x) = 0, \qquad \text{on } \Gamma \qquad \mathbf{u}_\varepsilon^N(0,x) = \mathbf{u}_1 \end{cases} \tag{3.19}$$

$$\begin{cases} -(\frac{d}{dt}\boldsymbol{\mu}_{\varepsilon,pwl}^N, \mathbf{v}) + a(\boldsymbol{\mu}_\varepsilon^N, \mathbf{v}) + b(\mathbf{v}, r_\varepsilon^N) = (\mathbf{f}_2, \mathbf{v}) \quad \forall\, \mathbf{v} \in \mathbf{H}_0^1(\Omega) \\ -\varepsilon(r_\varepsilon^N, q) - b(\boldsymbol{\mu}_\varepsilon^N, q) = 0 \quad \forall\, q \in L^2(\Omega) \\ \boldsymbol{\mu}_\varepsilon^N(x) = 0, \qquad \text{on } \Gamma \qquad \boldsymbol{\mu}_\varepsilon^N(T,x) = \boldsymbol{\mu}_1. \end{cases} \tag{3.20}$$

Moreover, we denote by $\mathcal{G} : \Lambda \times \mathcal{X} \to \mathcal{Y}$ the mapping containing all *coupled* terms, i.e., $\mathcal{G}(\lambda, \mathbf{u}^\varepsilon, \mathbf{u}^\varepsilon_t, p^\varepsilon, \boldsymbol{\mu}^\varepsilon, \boldsymbol{\mu}^\varepsilon_t, r^\varepsilon) = (\mathbf{f}_1, \mathbf{u}_1, \mathbf{f}_2, \boldsymbol{\mu}_1)$ if and only if

$$\begin{aligned}
(\mathbf{f}_1, \mathbf{v}) &= -\lambda\big(-\frac{1}{\beta}(\boldsymbol{\mu}^\varepsilon, \mathbf{v}) + (\mathbf{g}, \mathbf{v}) - \hat{c}(\mathbf{u}^\varepsilon, \mathbf{u}^\varepsilon, \mathbf{v})\big) \quad \forall \mathbf{v} \in \mathbf{H}^1_0(\Omega),\\
(\mathbf{u}_1, \mathbf{z}) &= (\mathbf{u}_0, \mathbf{z}) \qquad \forall \mathbf{z} \in \mathbf{L}^2(\Omega),\\
(\mathbf{f}_2, \mathbf{v}) &= -\lambda\big(\alpha(\mathbf{u}^\varepsilon - \mathbf{U}, \mathbf{v}) - \hat{c}(\mathbf{u}^\varepsilon, \boldsymbol{\mu}^\varepsilon, \mathbf{v}) - \hat{c}(\boldsymbol{\mu}^\varepsilon, \mathbf{u}^\varepsilon, \mathbf{v})\big) \quad \forall \mathbf{v} \in \mathbf{H}^1_0(\Omega),\\
(\boldsymbol{\mu}_1, \mathbf{z}) &= 0 \qquad \forall \mathbf{z} \in \mathbf{L}^2(\Omega).
\end{aligned}$$

Clearly, the continuous optimality system (2.6)-(2.7)-(2.8) is equivalent to

$$(\mathbf{u}^\varepsilon, \lambda\mathbf{u}^\varepsilon_t, \lambda p^\varepsilon, \boldsymbol{\mu}^\varepsilon, \lambda\boldsymbol{\mu}^\varepsilon_t, \lambda r^\varepsilon) + \mathcal{T}\mathcal{G}(\lambda, \mathbf{u}^\varepsilon, \lambda\mathbf{u}^\varepsilon_t, \lambda p^\varepsilon, \boldsymbol{\mu}^\varepsilon, \lambda\boldsymbol{\mu}^\varepsilon_t, \lambda r^\varepsilon) = 0,$$

and the semidiscrete optimality system (3.1)-(3.2)-(3.3) is equivalent to

$$\begin{aligned}
&(\mathbf{u}^N_\varepsilon, \lambda\frac{d}{dt}\mathbf{u}^N_{\varepsilon,pwl}, \lambda p^N_\varepsilon, \boldsymbol{\mu}^N_\varepsilon, \lambda\frac{d}{dt}\boldsymbol{\mu}^N_{\varepsilon,pwl}, \lambda r^N_\varepsilon)\\
&+\mathcal{T}^N\mathcal{G}(\lambda, \mathbf{u}^N_\varepsilon, \lambda\frac{d}{dt}\mathbf{u}^N_{\varepsilon,pwl}, \lambda p^N_\varepsilon, \boldsymbol{\mu}^N_\varepsilon, \lambda\frac{d}{dt}\boldsymbol{\mu}^N_{\varepsilon,pwl}, \lambda r^N_\varepsilon) = 0.
\end{aligned}$$

Therefore, we have recast the continuous and semidiscrete optimality system into a form that enables us to apply BRR theory.

Theorem 3.10. *Suppose that $\lambda = \frac{1}{\nu}$ is chosen in way that $(\mathbf{u}^\varepsilon, \lambda p^\varepsilon) - (\boldsymbol{\mu}^\varepsilon, \lambda r^\varepsilon)$ is a regular branch of solutions of the optimality system (2.6)-(2.7)-(2.8), and assume that the given data $\mathbf{g}, \mathbf{u}_0$ and the control function $\mathbf{f}^\varepsilon$ satisfy the regularity properties of proposition (2.8). Then there exists a neighborhood of the origin $\mathcal{O}$ in $\mathcal{X}$ and for $N \geq N_0$ a unique regular solution $\mathbf{u}^N_\varepsilon(\lambda), \frac{d}{dt}\mathbf{u}^N_{\varepsilon,pwl}(\lambda), p^N_\varepsilon(\lambda), \boldsymbol{\mu}^N_\varepsilon(\lambda), \frac{d}{dt}\boldsymbol{\mu}^N_{\varepsilon,pwl}(\lambda)$, $r^N_\varepsilon(\lambda)$ and a positive constant C independent of ε, k, such that $(\mathbf{u}^N_\varepsilon(\lambda), \frac{d}{dt}\mathbf{u}^N_{\varepsilon,pwl}(\lambda)$, $p^N_\varepsilon(\lambda)$, $\boldsymbol{\mu}^N_\varepsilon(\lambda), \frac{d}{dt}\boldsymbol{\mu}^N_{\varepsilon,pwl}(\lambda), r^N_\varepsilon(\lambda)) \in \mathcal{O}$ and*

$$\begin{aligned}
&\|(\mathbf{u}^\varepsilon(\lambda), \mathbf{u}^\varepsilon_t(\lambda), p^\varepsilon(\lambda), \boldsymbol{\mu}_\varepsilon(\lambda), \boldsymbol{\mu}^\varepsilon_t(\lambda), r_\varepsilon(\lambda))\\
&-(\mathbf{u}^N_\varepsilon(\lambda), \frac{d}{dt}\mathbf{u}^N_{\varepsilon,pwl}(\lambda), p^N_\varepsilon(\lambda), \boldsymbol{\mu}^N_\varepsilon(\lambda), \frac{d}{dt}\boldsymbol{\mu}^N_{\varepsilon,pwl}(\lambda), r^N_\varepsilon(\lambda))\|_{\mathcal{X}} \to 0.
\end{aligned}$$

In addition if $\mathbf{u}^\varepsilon, \boldsymbol{\mu}^\varepsilon \in C^1(0,T; \mathbf{H}^2(\Omega) \cap \mathbf{H}^1_0(\Omega))$, $\mathbf{u}^\varepsilon_t, \boldsymbol{\mu}^\varepsilon_t \in C^1(0,T; \mathbf{L}^2(\Omega))$ then

$$\begin{aligned}
&\|(\mathbf{u}^\varepsilon(\lambda), \mathbf{u}^\varepsilon_t(\lambda), p^\varepsilon(\lambda), \boldsymbol{\mu}_\varepsilon(\lambda), \boldsymbol{\mu}^\varepsilon_t(\lambda), r_\varepsilon(\lambda))\\
&-(\mathbf{u}^N_\varepsilon(\lambda), \frac{d}{dt}\mathbf{u}^N_{\varepsilon,pwl}(\lambda), p^N_\varepsilon(\lambda), \boldsymbol{\mu}^N_\varepsilon(\lambda), \frac{d}{dt}\boldsymbol{\mu}^N_{\varepsilon,pwl}(\lambda), r^N_\varepsilon(\lambda))\|_{\mathcal{X}} \leq C(\frac{1}{\nu}, T)k.
\end{aligned}$$

Proof. We follow arguments similar to [17, Theorem 5.2]. For simplicity, we drop the ε, λ from the notation. It is easy to see that D^2G is a C^∞ mapping, and bounded on all bounded sets of $\mathcal{X}$. A well known regularity result (see e.g. Proposition 2.8) for the continuous and semi-discrete problems imply that $\mathbf{u}, \boldsymbol{\mu} \in C((t_{n-1}, t_n]; \mathbf{H}^2(\Omega) \cap \mathbf{H}^1_0(\Omega))$, $\mathbf{u}_{tt}, \boldsymbol{\mu}_{tt} \in L^2(0,T; \mathbf{L}^2(\Omega))$ and whence,

$$\begin{aligned}
&\|(\mathbf{u}^N - \mathbf{u}, \boldsymbol{\mu}^N - \boldsymbol{\mu})\|_{X^N_1 \times X^N_1}\\
\leq\ &\max_{n=1,\ldots,N} \sup_{t\in(t_{n-1},t_n]} \Big(\|\mathbf{u}(t) - \mathbf{u}(t_n)\|_{\mathbf{H}^2(\Omega)} + \|\boldsymbol{\mu}(t) - \boldsymbol{\mu}(t_n)\|_{\mathbf{H}^2(\Omega)}\Big)\\
&+ \max_{n=1,..,N} \Big(\|\mathbf{u}(t_n) - \mathbf{u}^n\|_{\mathbf{H}^2(\Omega)} + \|\boldsymbol{\mu}(t_n) - \boldsymbol{\mu}^n\|_{\mathbf{H}^2(\Omega)}\Big). \qquad (3.21)
\end{aligned}$$

Employing estimate (3.12), and the fact that $\mathbf{u}, \boldsymbol{\mu}$ are piecewise continuous (in time), we obtain $\|(\mathbf{u}^N - \mathbf{u}, \boldsymbol{\mu}^N - \boldsymbol{\mu})\|_{X^N_1 \times X^N_1} \to 0$. Similarly, we can establish convergence for the time derivative norm after noting that $\mathbf{u}_t, \boldsymbol{\mu}_t \in C((t_{n-1}, t_n]; \mathbf{L}^2(\Omega))$ (see also Remark 3.8). Therefore, Proposition 3.7, implies that

$$\lim_{N\to\infty} \|(\mathcal{T}^N - \mathcal{T})\mathbf{w}\|_{\mathcal{L}(\mathcal{Y},\mathcal{X})} \to 0 \qquad \forall \mathbf{w} \in \mathcal{Y}.$$

The additional regularity assumption on $\mathbf{u}, \boldsymbol{\mu}$, i.e., $\mathbf{u}, \boldsymbol{\mu} \in C^1(0, T; \mathbf{H}^2(\Omega) \cap \mathbf{H}^1_0(\Omega))$ also imply that the first part of (3.21) is also of order $\mathcal{O}(k)$. Working similarly for the time derivative and pressure terms, using the enhanced regularity and once more Proposition 3.7, we obtain

$$\|\mathcal{T}^N - \mathcal{T}\|_{\mathcal{L}(\mathcal{Y},\mathcal{X})} \le Ck.$$

Finally, we need to prove the condition for the derivative $D\mathcal{G}$. $D\mathcal{G}(\mathbf{u}, p, \boldsymbol{\mu}, r) \cdot (\tilde{\mathbf{u}}, \tilde{p}, \tilde{\boldsymbol{\mu}}, \tilde{r}) = (\tilde{\boldsymbol{\zeta}}_1, \tilde{\mathbf{u}}_1, \tilde{\boldsymbol{\eta}}_1, \tilde{\boldsymbol{\mu}}_1)$ if and only if,

$$\begin{aligned}
(\tilde{\boldsymbol{\zeta}}_1, \mathbf{v}) &= -\Big(-\frac{1}{\beta}(\boldsymbol{\mu}, \mathbf{v}) - \hat{c}(\mathbf{u}, \tilde{\mathbf{u}}, \mathbf{v}) - \hat{c}(\tilde{\mathbf{u}}, \mathbf{u}, \mathbf{v})\Big) \quad \forall\, \mathbf{v} \in \mathbf{H}^1_0(\Omega),\\
(\tilde{\mathbf{u}}_1, \mathbf{z}) &= 0 \quad \forall\, \mathbf{z} \in \mathbf{L}^2(\Omega),\\
(\tilde{\boldsymbol{\eta}}_1, \mathbf{v}) &= -\Big(\alpha(\mathbf{u}, \mathbf{v}) + \hat{c}(\tilde{\mathbf{u}}, \mathbf{v}, \boldsymbol{\mu}) + \hat{c}(\mathbf{u}, \mathbf{v}, \tilde{\boldsymbol{\mu}})\\
&\quad + \hat{c}(\mathbf{u}, \tilde{\mathbf{u}}, \mathbf{v}) + \hat{c}(\tilde{\mathbf{u}}, \mathbf{u}, \mathbf{v})\Big) \quad \forall\, \mathbf{v} \in \mathbf{H}^1_0(\Omega),\\
(\tilde{\boldsymbol{\mu}}_1, \mathbf{z}) &= 0 \quad \forall\, \mathbf{z} \in \mathbf{L}^2(\Omega).
\end{aligned}$$

Rewrite $\tilde{\boldsymbol{\zeta}}_1 = \tilde{\boldsymbol{\zeta}}^1_1 + \tilde{\boldsymbol{\zeta}}^2_1 + \tilde{\boldsymbol{\zeta}}^3_1$, where $\langle \tilde{\boldsymbol{\zeta}}^1_1, \mathbf{v}\rangle = \frac{1}{\beta}(\boldsymbol{\mu}, \mathbf{v})$, $\langle \tilde{\boldsymbol{\zeta}}^2_1, \mathbf{v}\rangle = \hat{c}(\mathbf{u}, \tilde{\mathbf{u}}, \mathbf{v})$, $\langle \tilde{\boldsymbol{\zeta}}^3_1, \mathbf{v}\rangle = \hat{c}(\tilde{\mathbf{u}}, \mathbf{u}, \mathbf{v})$, $\forall\, \mathbf{v} \in \mathbf{H}^1_0(\Omega)$. Note that the regularity of $\boldsymbol{\mu}$ implies that $\tilde{\boldsymbol{\zeta}}^1_1 \in \mathcal{Z} := \mathcal{Y}$. For the $\tilde{\boldsymbol{\zeta}}^2_1$ we use standard techniques for the trilinear form (see Section 2.2) to obtain:

$$\begin{aligned}
\|\tilde{\boldsymbol{\zeta}}^2_1\|_{\mathbf{L}^2(\Omega)} &= \sup_{\mathbf{v}\in\mathbf{L}^2(\Omega)} \frac{|\hat{c}(\mathbf{u}, \tilde{\mathbf{u}}, \mathbf{v})|}{\|\mathbf{v}\|_{\mathbf{L}^2(\Omega)}}\\
&\le C \sup_{\mathbf{v}\in\mathbf{L}^2(\Omega)} \frac{\|\mathbf{u}\|_{\mathbf{H}^1_0(\Omega)}\|\tilde{\mathbf{u}}\|_{\mathbf{H}^2(\Omega)}\|\mathbf{v}\|_{\mathbf{L}^2(\Omega)}}{\|\mathbf{v}\|_{\mathbf{L}^2(\Omega)}} \le C\|\mathbf{u}\|_{\mathbf{H}^1_0(\Omega)}\|\tilde{\mathbf{u}}\|_{\mathbf{H}^2(\Omega)}.
\end{aligned}$$

From the last inequality and regularity properties of $\mathbf{u}, \tilde{\mathbf{u}}$ follows that $\tilde{\boldsymbol{\zeta}}^2_1 \in \mathcal{Y}$. For the time derivative term note that

$$\begin{aligned}
\|\tilde{\boldsymbol{\zeta}}^2_{1t}\|_{\mathbf{H}^{-1}(\Omega)} &= \sup_{\mathbf{v}\in\mathbf{H}^1_0(\Omega)} \frac{|\hat{c}(\mathbf{u}_t, \tilde{\mathbf{u}}, \mathbf{v})| + |\hat{c}(\mathbf{u}, \tilde{\mathbf{u}}_t, \mathbf{v})|}{\|\mathbf{v}\|_{\mathbf{H}^1_0(\Omega)}}\\
&\le C\big(\|\mathbf{u}_t\|_{\mathbf{L}^2(\Omega)}\|\tilde{\mathbf{u}}\|_{\mathbf{H}^2(\Omega)} + \|\mathbf{u}\|_{\mathbf{H}^2(\Omega)}\|\tilde{\mathbf{u}}_t\|_{\mathbf{L}^2(\Omega)}\big).
\end{aligned}$$

Taking the squares, integrating with respect to time, and using regularity properties of $\mathbf{u}, \tilde{\mathbf{u}}$ we also obtain that $\tilde{\boldsymbol{\zeta}}^2_{1t} \in L^2(0, T; \mathbf{H}^{-1}(\Omega))$. Working similarly for the other terms we obtain the desired estimate, which concludes our proof. □

Remark 3.11. *The constant C appearing in Theorem (3.9) is independent of $\lambda = \frac{1}{\nu}$ but depends on α, β through various norms of $I + \mathcal{T} D_\psi \mathcal{G}(\lambda, \psi)$. However, note that to estimate $\|(\mathcal{T}^n - \mathcal{T})\mathcal{G}\psi\|_{\mathcal{X}}, \psi \in \mathcal{Y}$ we have applied Proposition 3.7, where a constant depending on $\frac{1}{\nu}, T$ is introduced.*

Remark 3.12. *The additional regularity appears to be necessary in order to guarantee a rate of convergence. Note however, that only additional regularity in time is needed. In the two-dimensional case we may easily prove the existence of a regular solution (see e.g.* [9, 27]*). The BRR theory can be applied three dimensional* *problems, where uniqueness and existence of regular solutions hold only for certain values of the parameter ν, i.e., for small intervals Λ.*

Finally, we conclude this section by stating the main result comparing the semi-discrete penalized system (3.1)-(3.2)-(3.3) to the solution of the original optimality system (2.2)-(2.3)-(2.4).

Proposition 3.13. *Let $(\mathbf{u}, p, \boldsymbol{\mu}, r)$ $(\mathbf{u}_\varepsilon^N, p_\varepsilon^N, \boldsymbol{\mu}_\varepsilon^N, r_\varepsilon^N)$ be the solutions of (2.2)-(2.3)-(2.4) and (3.1)-(3.2)-(3.3) respectively, satisfying the assumptions of Theorem (3.10). Then, there estists a positive constant C independent of ε, k such that following estimate holds:*

$$\nu\|\mathbf{u}(t_n) - \mathbf{u}_\varepsilon^n\|_{\mathbf{H}_0^1(\Omega)} + \nu\|\boldsymbol{\mu}(t_n) - \boldsymbol{\mu}_\varepsilon^n\|_{\mathbf{H}_0^1(\Omega)} \leq C(\frac{1}{\nu}, \alpha, \beta)(k + \varepsilon). \tag{3.22}$$

Proof. Using standard techniques (see e.g. [7, Chapter 3]), $\|\mathbf{f}^\varepsilon\|_{L^2(0,T;\mathbf{L}^2(\Omega))}$ is bounded independent of ε. Then, subtracting (2.2)-(2.3)-(2.4) from (2.6)-(2.7)-(2.8) yields:

$$\begin{cases} \langle \mathbf{u}_t^\varepsilon - \mathbf{u}, \mathbf{v}\rangle + \nu a(\mathbf{u}^\varepsilon - \mathbf{u}, \mathbf{v}) + \hat{c}(\mathbf{u}^\varepsilon, \mathbf{u}^\varepsilon, \mathbf{v}) - \hat{c}(\mathbf{u}, \mathbf{u}, \mathbf{v}) \\ \qquad + b(\mathbf{v}, p^\varepsilon - p) = (\mathbf{f}^\varepsilon - \mathbf{f}, \mathbf{v}) \qquad \forall\, \mathbf{v} \in \mathbf{H}_0^1(\Omega) \\ \varepsilon(p^\varepsilon, q) - b(\mathbf{u}^\varepsilon - \mathbf{u}, q) = 0 \qquad \forall\, q \in L^2(\Omega) \\ (\mathbf{u}^\varepsilon(0,x) - \mathbf{u}_0, \mathbf{z}) = 0 \qquad \forall\, \mathbf{z} \in \mathbf{L}^2(\Omega) \end{cases} \tag{3.23}$$

$$\begin{cases} -\langle \boldsymbol{\mu}_t^\varepsilon - \boldsymbol{\mu}, \mathbf{v}\rangle + \nu a(\boldsymbol{\mu}^\varepsilon - \boldsymbol{\mu}, \mathbf{v}) + \hat{c}(\boldsymbol{\mu}^\varepsilon, \mathbf{u}^\varepsilon, \mathbf{v}) + \hat{c}(\mathbf{u}^\varepsilon, \boldsymbol{\mu}^\varepsilon, \mathbf{v}) \\ \qquad -\hat{c}(\boldsymbol{\mu}, \mathbf{u}, \mathbf{v}) - \hat{c}(\mathbf{u}, \boldsymbol{\mu}, \mathbf{v}) + b(\mathbf{v}, r^\varepsilon - r) = \alpha(\mathbf{u}^\varepsilon - \mathbf{u}, \mathbf{v}) \qquad \forall\, \mathbf{v} \in \mathbf{H}_0^1(\Omega) \\ -\varepsilon(r^\varepsilon, q) - b(\boldsymbol{\mu}^\varepsilon - \boldsymbol{\mu}, q) = 0 \qquad \forall\, q \in L^2(\Omega) \\ (\boldsymbol{\mu}^\varepsilon(T,x), \mathbf{z}) = 0 \qquad \forall\, \mathbf{z} \in \mathbf{L}^2(\Omega). \end{cases} \tag{3.24}$$

$$\boldsymbol{\mu}^\varepsilon - \boldsymbol{\mu} = -\beta(\mathbf{f}^\varepsilon - \mathbf{f}). \tag{3.25}$$

Therefore, we may apply [26, Theorem 3.1] into equation (3.23), to obtain the estimate $\|\mathbf{u} - \mathbf{u}^\varepsilon\|_{L^2(0,T;\mathbf{L}^2(\Omega))} \leq C\varepsilon^{1/2}$. Then, the regularity assumptions on $\mathbf{u} - \mathbf{u}^\varepsilon$ imply that the assumptions of [26, Theorem 4.1] are satisfied (see also Proposition 2.8, and Remark 2.9) and hence using the equation (3.24), we obtain the estimate $\|\boldsymbol{\mu} - \boldsymbol{\mu}^\varepsilon\|_{L^\infty(0,T;\mathbf{H}_0^1(\Omega))} \leq C\varepsilon$. Returning back to equation (3.23), and using the enhanced regularity of $\boldsymbol{\mu} - \boldsymbol{\mu}^\varepsilon$ term, we also obtain that $\|\mathbf{u} - \mathbf{u}^\varepsilon\|_{L^\infty(0,T;\mathbf{H}_0^1(\Omega))} \leq C\varepsilon$. The theorem follows directly from triangle inequality and results of Theorem 3.10. ☐

4. **An improved estimate for the linearized optimality system.** Next we introduce an alternative scheme based on the time discretization of the artificially compressible Navier-Stokes. An alternative class of pertubation problems, based on the artificial compressibility condition is defined in a similar manner: Minimize the tracking functional $J(\mathbf{u}^\varepsilon, \mathbf{f}^\varepsilon)$ subject to the constraints:

$$\begin{cases} \langle \mathbf{u}_t^\varepsilon, \mathbf{v}\rangle + \nu a(\mathbf{u}^\varepsilon, \mathbf{v}) + \hat{c}(\mathbf{u}^\varepsilon, \mathbf{u}^\varepsilon, \mathbf{v}) + b(\mathbf{v}, p^\varepsilon) = (\mathbf{f}^\varepsilon + \mathbf{g}, \mathbf{v}) \qquad \forall \mathbf{v} \in \mathbf{H}_0^1(\Omega) \\ \varepsilon k \frac{d}{dt}(p^\varepsilon, q) - b(\mathbf{u}^\varepsilon, q) = 0 \qquad \forall\, q \in L^2(\Omega) \\ (\mathbf{u}^\varepsilon(0,x), \mathbf{z}) = (\mathbf{u}_0(x), \mathbf{z}) \quad (p^\varepsilon(0,x), q) = (p_0(x), q) \qquad \forall\, \mathbf{z} \in \mathbf{L}^2(\Omega), q \in L^2(\Omega), \end{cases} \tag{4.1}$$

where $\mathbf{u}_0, p_0, g, \mathbf{U}$ are given data. Similar theorems regarding existence of optimal solutions and their convergence as $\varepsilon \to 0$ can be also proved (see [8, Section 3]). For analysis and fully-discrete finite element error estimates for the uncontrolled system, one may consult the classical work of [27, Part III, Section 8].

In this section we only treat the linearized optimality system. Adopting similar notation as in the previous section, we seek sequences $\{\mathbf{u}_\varepsilon\}_{n=1}^N, \{p_\varepsilon\}_{n=1}^N$ and $\{\boldsymbol{\mu}_\varepsilon\}_{n=1}^N, \{r_\varepsilon\}_{n=1}^N$ such that that:

$$\left\{\begin{array}{l} \frac{1}{k}(\mathbf{u}_\varepsilon^n - \mathbf{u}_\varepsilon^{n-1}, \mathbf{v}) + \nu a(\mathbf{u}_\varepsilon^n, \mathbf{v}) + b(\mathbf{v}, p_\varepsilon^n) = (\mathbf{f}_\varepsilon^n + \mathbf{g}^n, \mathbf{v}) \quad \forall\, \mathbf{v} \in \mathbf{H}_0^1(\Omega) \\ \varepsilon(p_\varepsilon^n - p_\varepsilon^{n-1}, q) - b(\mathbf{u}_\varepsilon^n, q) = 0 \qquad \forall\, q \in L^2(\Omega) \\ \mathbf{u}_\varepsilon^n(x) = 0, \qquad \text{on } \Gamma \end{array}\right. \tag{4.2}$$

$$\left\{\begin{array}{l} -\frac{1}{k}(\boldsymbol{\mu}_\varepsilon^n - \mu_\varepsilon^{n-1}, \mathbf{v}) + \nu a(\mathbf{v}, \boldsymbol{\mu}_\varepsilon^{n-1}) + b(\mathbf{v}, r_\varepsilon^{n-1}) = \alpha(\mathbf{u}_\varepsilon^n - \mathbf{U}^n, \mathbf{v}) \quad \forall\, \mathbf{v} \in \mathbf{H}_0^1(\Omega) \\ -\varepsilon(r_\varepsilon^n - r_\varepsilon^{n-1}, q) - b(\boldsymbol{\mu}_\varepsilon^{n-1}, q) = 0 \qquad \forall\, q \in L^2(\Omega) \\ \boldsymbol{\mu}_\varepsilon^{n-1}(x) = 0, \qquad \text{on } \Gamma \end{array}\right. \tag{4.3}$$

$$\mathbf{f}_\varepsilon^n = -\frac{1}{\beta}\boldsymbol{\mu}_\varepsilon^{n-1}. \tag{4.4}$$

The above semi-discrete (in time) discretization corresponds to the following optimality system, and it is based on the artificial compressibility.

$$\left\{\begin{array}{l} \langle \mathbf{u}_t^\varepsilon, \mathbf{v}\rangle + \nu a(\mathbf{u}^\varepsilon, \mathbf{v}) + b(\mathbf{v}, p^\varepsilon) = (-\frac{1}{\beta}\boldsymbol{\mu}^\varepsilon + \mathbf{g}, \mathbf{v}) \qquad \forall\, \mathbf{v} \in \mathbf{H}_0^1(\Omega) \\ \varepsilon k \frac{d}{dt}(p^\varepsilon, q) - b(\mathbf{u}^\varepsilon, q) = 0 \qquad \forall\, q \in L^2(\Omega) \\ (\mathbf{u}^\varepsilon(x), \mathbf{z}) = (\mathbf{u}_0, \mathbf{z}), \quad (p^\varepsilon(0), q) = (p_0, q) \qquad \forall \mathbf{z} \in \mathbf{L}^2(\Omega), q \in L^2(\Omega) \end{array}\right. \tag{4.5}$$

$$\left\{\begin{array}{l} \langle \boldsymbol{\mu}_t^\varepsilon, \mathbf{v}\rangle + \nu a(\boldsymbol{\mu}^\varepsilon, \mathbf{v}) + b(\mathbf{v}, r^\varepsilon) = \alpha(\boldsymbol{\mu}^\varepsilon - \mathbf{U}, \mathbf{v}) \qquad \forall\, \mathbf{v} \in \mathbf{H}_0^1(\Omega) \\ -\varepsilon k \frac{d}{dt}(r^\varepsilon, q) - b(\boldsymbol{\mu}^\varepsilon, q) = 0 \qquad \forall\, q \in L^2(\Omega) \\ \boldsymbol{\mu}^\varepsilon(T) = 0, \qquad r^\varepsilon(T) = 0. \end{array}\right. \tag{4.6}$$

Below, we compare the solution of systems (2.2)-(2.3)-(2.4), (4.2)-(4.3)-(4.4).

Theorem 4.1. *Given data $\mathbf{g} \in L^2(0,T;\mathbf{L}^2(\Omega)) \cap H^1(0,T;\mathbf{L}^2(\Omega))$, $\mathbf{u}_0 \in V(\Omega)$, $\mathbf{U} \in \mathcal{B}$, suppose that $(\mathbf{u}, p)$, $(\boldsymbol{\mu}, r)$ is the solution of problem (2.2)-(2.3)-(2.4) and $\{\mathbf{u}_\varepsilon^n, p_\varepsilon^n\}_{n=1}^N$, $\{\boldsymbol{\mu}_\varepsilon^n, r_\varepsilon^n\}_{n=1}^N$ the solution of its semi-discrete approximation (4.2)-(4.3)-(4.4). In addition, let $\mathbf{u}_{tt}, \boldsymbol{\mu}_{tt} \in L^2(0,T;\mathbf{H}^{-1}(\Omega))$, $p_t, r_t \in C(0,T;L^2(\Omega))$ $p_{tt}, r_{tt} \in C(0,T;L^2(\Omega))$. Then, the following estimate holds:*

$$\|\mathbf{e}^N\|_{\mathbf{L}^2(\Omega)}^2 + \varepsilon\|e_p^N\|_{L^2(\Omega)}^2 + k\nu \sum_{n=1}^N \|\nabla \mathbf{e}^n\|_{\mathbf{L}^2(\Omega)}^2 \le C\Big(k^2 + \varepsilon^2 k^2\Big)$$

$$\|\mathbf{d}^0\|_{\mathbf{L}^2(\Omega)}^2 + \varepsilon\|d_p^0\|_{L^2(\Omega)}^2 + k\nu \sum_{n=0}^{N-1} \|\nabla \mathbf{d}^n\|_{\mathbf{L}^2(\Omega)}^2 \le C\Big(k^2 + \varepsilon^2 k^2\Big)$$

where $\mathbf{e}^n = \mathbf{u}(t_n) - \mathbf{u}_\varepsilon^n, e_p^n = p(t_n) - p_\varepsilon^n$ and $\mathbf{d}^n = \boldsymbol{\mu}(t_n) - \boldsymbol{\mu}_\varepsilon^n$, $d_p^n = r(t_n) - r_\varepsilon^n$ and C is constant depending on $\frac{1}{\nu^2}\max\{\frac{\alpha^2}{\beta\nu}, \frac{\alpha}{\beta^2\nu}\}, \min\{\frac{1}{\beta}, \alpha\}, T$ and on domain Ω.

Proof. For convenience, we omit ε from $\mathbf{u}_\varepsilon^n, p_\varepsilon^n, \boldsymbol{\mu}_\varepsilon^n, r_\varepsilon^n$. First, we introduce an auxiliary optimality system. Suppose that $\{\hat{\mathbf{u}}^n\}_{n=1}^N, \{\hat{p}^n\}_{n=1}^N, \{\hat{\boldsymbol{\mu}}^n\}_{n=1}^N, \{\hat{r}^n\}_{n=1}^N$ are the solutions of the optimality system: $\forall\, n = 1, ..., N$,

$$\left\{\begin{array}{l} \frac{1}{k}(\hat{\mathbf{u}}^n - \hat{\mathbf{u}}^{n-1}) + \nu a(\hat{\mathbf{u}}^n, \mathbf{v}) + b(\mathbf{v}, \hat{p}^n) = (\mathbf{f}(t_n) + \mathbf{g}(t_n), \mathbf{v}) \quad \forall\, \mathbf{v} \in \mathbf{H}_0^1(\Omega) \\ \varepsilon(\hat{p}^n - \hat{p}^{n-1}, q) - b(\hat{\mathbf{u}}^n, q) = 0 \qquad \forall\, q \in L^2(\Omega) \\ \hat{\mathbf{u}}^n(x) = 0, \qquad \text{on } \Gamma \end{array}\right. \tag{4.7}$$

$$\begin{cases} -\frac{1}{k}(\hat{\boldsymbol{\mu}}^n - \hat{\boldsymbol{\mu}}^{n-1}, \mathbf{v}) + \nu a(\mathbf{v}, \hat{\boldsymbol{\mu}}^{n-1}) + b(\mathbf{v}, \hat{r}^{n-1}) = \alpha(\mathbf{u}(t_n) - \mathbf{U}(t_n), \mathbf{v}) \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \forall\, \mathbf{v} \in \mathbf{H}_0^1(\Omega) \\ -\varepsilon(\hat{r}^n - \hat{r}^{n-1}, q) - b(\hat{\boldsymbol{\mu}}^{n-1}, q) = 0 \qquad \forall\, q \in L^2(\Omega) \\ \hat{\boldsymbol{\mu}}^{n-1}(x) = 0, \qquad \text{on } \Gamma. \end{cases} \tag{4.8}$$

We denote by $\hat{\mathbf{e}}^n = \mathbf{u}(t_n) - \hat{\mathbf{u}}^n, \hat{e}_p^n = p(t_n) - \hat{p}^n$, $\hat{\mathbf{d}}^n = \boldsymbol{\mu}(t_n) - \hat{\boldsymbol{\mu}}^n, \hat{d}_p^n = r(t_n) - \hat{r}^n$ and note that we may directly apply results of [26, Proposition 5.1] since the above system is essentially uncoupled. Indeed, note that the regularity assumptions of [26, Proposition 5.1] are satisfied, and subtracting (4.7)-(4.8) from (2.2)-(2.3)-(2.4) we obtain

$$\begin{cases} \frac{1}{k}(\hat{\mathbf{e}}^n - \hat{\mathbf{e}}^{n-1}, \mathbf{v}) + \nu a(\hat{\mathbf{e}}^n, \mathbf{v}) + b(\mathbf{v}, \hat{e}_p^n) = (\mathbf{R}_1^n, \mathbf{v}) \qquad \forall \mathbf{v} \in \mathbf{H}_0^1(\Omega) \\ \varepsilon(\hat{e}_p^n - \hat{e}_p^{n-1}, q) - b(\hat{\mathbf{e}}^n, q) = (R_2^n, q) \quad \forall\, q \in L^2(\Omega) \end{cases} \tag{4.9}$$

$$\begin{cases} -\frac{1}{k}(\hat{\mathbf{d}}^n - \hat{\mathbf{d}}^{n-1}, \mathbf{v}) + \nu a(\hat{\mathbf{d}}^{n-1}, \mathbf{v}) + b(\mathbf{v}, \hat{d}_p^{n-1}) = (\mathbf{R}_3^n, \mathbf{v}) \\ -\varepsilon(\hat{d}_p^n - \hat{d}_p^{n-1}, q) - b(\hat{\mathbf{d}}^{n-1}, q) = (R_4^n, q) \quad \forall\, q \in L^2(\Omega) \end{cases} \tag{4.10}$$

where

$$\mathbf{R}_1^n = \frac{1}{k}\int_{t_{n-1}}^{t^n} (t - t_{n-1})\mathbf{u}_{tt} dt, \quad R_2^n = \varepsilon \int_{t_{n-1}}^{t^n} p_t dt$$

and

$$\mathbf{R}_3^n = -\frac{1}{k}\int_{t_{n-1}}^{t^n} (t - t_{n-1})\boldsymbol{\mu}_{tt} dt, \quad R_4^n = -\varepsilon \int_{t_{n-1}}^{t^n} r_t dt.$$

Then, we apply [26, Proposition 5.1] directly to (4.9) to obtain an estimate

$$\|\hat{\mathbf{e}}^n\|^2_{\mathbf{L}^2(\Omega)} + k\nu \sum_{i=0}^n \|\nabla \hat{\mathbf{e}}^i\|^2_{\mathbf{L}^2(\Omega)} \leq C(\frac{1}{\nu}, T)(k^2 + \varepsilon^2 k^2), \qquad \forall\, n = 1, ..., N \tag{4.11}$$

and once again to (4.10) to obtain the estimate,

$$\|\hat{\mathbf{d}}^n\|^2_{\mathbf{L}^2(\Omega)} + k\nu \sum_{i=1}^n \|\nabla \hat{\mathbf{d}}^i\|^2_{\mathbf{L}^2(\Omega)} \leq C(\frac{1}{\nu}, T)(k^2 + \varepsilon^2 k^2), \qquad \forall\, n = 0, ..., N-1 \tag{4.12}$$

where C is independent of $\varepsilon, k, \alpha, \beta$. It remains to compare systems (4.7)-(4.8) to (4.2)-(4.3)-(4.4). For this purpose, we denote by $\tilde{\mathbf{e}}^n = \hat{\mathbf{u}}^n - \mathbf{u}^n, \tilde{e}_p^n = \hat{p}^n - p^n$, $\tilde{\mathbf{d}}^n = \hat{\boldsymbol{\mu}}^n - \boldsymbol{\mu}^n, \tilde{d}_p^n = \hat{r}^n - r^n$. Subtracting (4.2)-(4.3)-(4.4) from (4.7)-(4.8) we obtain:

$$\begin{cases} \frac{1}{k}(\tilde{\mathbf{e}}^n - \tilde{\mathbf{e}}^{n-1}, \mathbf{v}) + \nu a(\tilde{\mathbf{e}}^n, \mathbf{v}) + b(\mathbf{v}, \tilde{e}_p^n) = (\mathbf{f}(t_n) - \mathbf{f}^n, \mathbf{v}) \qquad \forall \mathbf{v} \in \mathbf{H}_0^1(\Omega) \\ \varepsilon(\tilde{e}_p^n - \tilde{e}_p^{n-1}, q) - b(\tilde{\mathbf{e}}^n, q) = 0 \quad \forall\, q \in L^2(\Omega) \end{cases} \tag{4.13}$$

$$\begin{cases} -\frac{1}{k}(\tilde{\mathbf{d}}^n - \tilde{\mathbf{d}}^{n-1}, \mathbf{v}) + \nu a(\tilde{\mathbf{d}}^{n-1}, \mathbf{v}) + b(\mathbf{v}, \tilde{d}_p^{n-1}) = \alpha(\mathbf{u}(t_n) - \mathbf{u}^n, \mathbf{v}) \\ -\varepsilon(\tilde{d}_p^n - \tilde{d}_p^{n-1}, q) - b(\tilde{\mathbf{d}}^{n-1}, q) = 0 \quad \forall\, q \in L^2(\Omega) \end{cases} \tag{4.14}$$

Note that using (2.4)-(4.4) we may write $\mathbf{f}(t_n) - \mathbf{f}^n = -\frac{1}{\beta}(\boldsymbol{\mu}(t_n) - \boldsymbol{\mu}(t_{n-1})) - \frac{1}{\beta}(\boldsymbol{\mu}(t_{n-1}) - \boldsymbol{\mu}^{n-1})$, with $\boldsymbol{\mu}(t_{n-1}) - \boldsymbol{\mu}^{n-1} = \hat{\mathbf{d}}^{n-1} + \tilde{\mathbf{d}}^{n-1}$, and $\mathbf{u}(t^n) - \mathbf{u}^n = \hat{\mathbf{e}}^n + \tilde{\mathbf{e}}^n$. Setting $2k\tilde{\mathbf{e}}^n, 2k\tilde{e}_p^n$ respectively into (4.13), we obtain:

$$\begin{aligned} & \|\tilde{\mathbf{e}}^n\|^2_{\mathbf{L}^2(\Omega)} - \|\tilde{\mathbf{e}}^{n-1}\|^2_{\mathbf{L}^2(\Omega)} + \|\tilde{\mathbf{e}}^n - \tilde{\mathbf{e}}^{n-1}\|^2_{\mathbf{L}^2(\Omega)} + 2k\nu\|\nabla\tilde{\mathbf{e}}^n\|^2_{\mathbf{L}^2(\Omega)} \\ & \quad +\varepsilon(\|\tilde{e}_p^n\|^2_{L^2(\Omega)} - \|\tilde{e}_p^{n-1}\|^2_{L^2(\Omega)} + \|\tilde{e}_p^n - \tilde{e}_p^{n-1}\|^2_{L^2(\Omega)}) \\ = \quad & -\frac{2k}{\beta}\big((\tilde{\mathbf{d}}^{n-1}, \tilde{\mathbf{e}}^n) + (\hat{\mathbf{d}}^{n-1}, \tilde{\mathbf{e}}^n) + (\boldsymbol{\mu}(t_n) - \boldsymbol{\mu}(t_{n-1}), \tilde{\mathbf{e}}^n)\big) \end{aligned} \tag{4.15}$$

and similarly multilplying (4.14) by $2k\tilde{\mathbf{d}}^{n-1}, 2k\tilde{d}_p^{n-1}$,

$$\begin{aligned}
&\|\tilde{\mathbf{d}}^{n-1}\|^2_{\mathbf{L}^2(\Omega)} - \|\tilde{\mathbf{d}}^n\|^2_{\mathbf{L}^2(\Omega)} + \|\tilde{\mathbf{d}}^n - \tilde{\mathbf{d}}^{n-1}\|^2_{\mathbf{L}^2(\Omega)} + 2k\nu\|\nabla\tilde{\mathbf{d}}^{n-1}\|^2_{\mathbf{L}^2(\Omega)} \\
&\quad +\varepsilon(\|\tilde{d}_p^{n-1}\|^2_{L^2(\Omega)} - \|\tilde{d}_p^n\|^2_{L^2(\Omega)} + \|\tilde{d}^n - \tilde{d}^{n-1}\|^2_{L^2(\Omega)}) \\
= \;& 2\alpha k\big((\hat{\mathbf{e}}^n, \tilde{\mathbf{d}}^{n-1}) + (\tilde{\mathbf{e}}^n, \tilde{\mathbf{d}}^{n-1})\big). \qquad (4.16)
\end{aligned}$$

Multiplying (4.15) by α, and (4.16) by $\frac{1}{\beta}$ and adding the resulting equations, we may cancel the coupling term $(\tilde{\mathbf{e}}^n, \tilde{\mathbf{d}}^{n-1}))$ on the right hand side to obtain:

$$\begin{aligned}
&\alpha\big(\|\tilde{\mathbf{e}}^n\|^2_{\mathbf{L}^2(\Omega)} - \|\tilde{\mathbf{e}}^{n-1}\|^2_{\mathbf{L}^2(\Omega)} + \|\tilde{\mathbf{e}}^n - \tilde{\mathbf{e}}^{n-1}\|^2_{\mathbf{L}^2(\Omega)} + 2k\nu\|\nabla\tilde{\mathbf{e}}^n\|^2_{\mathbf{L}^2(\Omega)}\big) \\
&+\varepsilon\alpha\big(\|\tilde{e}_p^n\|^2_{L^2(\Omega)} - \|\tilde{e}_p^{n-1}\|^2_{L^2(\Omega)} + \|\tilde{e}_p^n - \tilde{e}_p^{n-1}\|^2_{L^2(\Omega)}\big) \\
&+\frac{1}{\beta}\big(\|\tilde{\mathbf{d}}^{n-1}\|^2_{\mathbf{L}^2(\Omega)} - \|\tilde{\mathbf{d}}^n\|^2_{\mathbf{L}^2(\Omega)} + \|\tilde{\mathbf{d}}^n - \tilde{\mathbf{d}}^{n-1}\|^2_{\mathbf{L}^2(\Omega)} + 2k\nu\|\nabla\tilde{\mathbf{d}}^{n-1}\|^2_{\mathbf{L}^2(\Omega)}\big) \\
&+\frac{\varepsilon}{\beta}\big(\|\tilde{d}_p^{n-1}\|^2_{L^2(\Omega)} - \|\tilde{d}_p^n\|^2_{L^2(\Omega)} + \|\tilde{d}^n - \tilde{d}^{n-1}\|^2_{L^2(\Omega)}\big) \\
\le \;& \frac{2\alpha k}{\beta}(\hat{\mathbf{e}}^n, 2\tilde{\mathbf{d}}^{n-1}) - \frac{2\alpha k}{\beta}(\hat{\mathbf{d}}^{n-1}, \tilde{\mathbf{e}}^n) - \frac{2\alpha k}{\beta}(\boldsymbol{\mu}(t_n) - \boldsymbol{\mu}(t_{n-1}), \tilde{\mathbf{e}}^n). \qquad (4.17)
\end{aligned}$$

It remains to bound the last three terms. Using Cauchy's inequality with appropriate constants, where C depends only on the geometry,

$$\frac{2\alpha k}{\beta}(\hat{\mathbf{e}}^n, \tilde{\mathbf{d}}^{n-1}) \le \frac{k\nu}{\beta}\|\nabla\tilde{\mathbf{d}}^{n-1}\|^2_{\mathbf{L}^2(\Omega)} + \frac{\alpha^2 kC}{\beta\nu}\|\hat{\mathbf{e}}^n\|^2_{\mathbf{L}^2(\Omega)}$$

$$\frac{2\alpha k}{\beta}(\hat{\mathbf{d}}^{n-1}, \tilde{\mathbf{e}}^n) \le \frac{2\alpha kC}{\beta^2\nu}\|\hat{\mathbf{d}}^{n-1}\|^2_{\mathbf{L}^2(\Omega)} + \frac{\alpha\nu k}{2}\|\nabla\tilde{\mathbf{e}}^n\|^2_{\mathbf{L}^2(\Omega)}.$$

In order to bound the last term note that

$$\boldsymbol{\mu}(t_n) - \boldsymbol{\mu}(t_{n-1}) = k\boldsymbol{\mu}_t(t_n) - \int_{t_{n-1}}^{t_n}(t - t_{n-1})\boldsymbol{\mu}_{tt}dt$$

and

$$\frac{2\alpha k}{\beta}(k\boldsymbol{\mu}_t(t_n), \tilde{\mathbf{e}}^n) \le \frac{\alpha kC}{\beta^2\nu}\|k\boldsymbol{\mu}_t\|^2_{\mathbf{L}^2(\Omega)} + \frac{\alpha\nu k}{4}\|\nabla\tilde{\mathbf{e}}^n\|^2_{\mathbf{L}^2(\Omega)},$$

$$\begin{aligned}
&\frac{2\alpha k}{\beta}\Big(\int_{t_{n-1}}^{t_n}(t - t_{n-1})\boldsymbol{\mu}_{tt}dt, \tilde{\mathbf{e}}^n\Big) \\
\le \;& \frac{2\alpha kC}{\beta^2\nu}\Big\|\int_{t_{n-1}}^{t_n}(t - t_{n-1})\boldsymbol{\mu}_{tt}dt\Big\|^2_{\mathbf{H}^{-1}(\Omega)} + \frac{\alpha\nu k}{4}\|\nabla\tilde{\mathbf{e}}^n\|^2_{\mathbf{L}^2(\Omega)}.
\end{aligned}$$

Collecting the last three inequalities, setting $\gamma = \min\{\alpha, \frac{1}{\beta}\}$, and using (4.17) recursively, we easily obtain

$$\begin{aligned}
&\gamma\nu k\sum_{n=1}^{N}\|\nabla\tilde{\mathbf{e}}^n\|^2_{\mathbf{L}^2(\Omega)} + \gamma\nu k\sum_{n=0}^{N-1}\|\nabla\tilde{\mathbf{d}}^{n-1}\|^2_{\mathbf{L}^2(\Omega)} \\
\le \;& C(\alpha,\beta,\nu)k\Big(\sum_{n=1}^{N}\|\nabla\hat{\mathbf{e}}^n\|^2_{\mathbf{L}^2(\Omega)} + \sum_{n=0}^{N-1}\|\nabla\hat{\mathbf{d}}^{n-1}\|^2_{\mathbf{L}^2(\Omega)}\Big) \\
&+C(\alpha,\beta,\nu)\Big(k^3\sum_{n=0}^{N}\|\boldsymbol{\mu}_t(t_n)\|^2_{\mathbf{L}^2(\Omega)} + k^2\int_0^T\|\boldsymbol{\mu}_{tt}\|^2_{\mathbf{H}^{-1}(\Omega)}dt\Big).
\end{aligned}$$

where $C(\alpha,\beta,\nu) = \max\{\frac{\alpha^2}{\beta\nu}, \frac{\alpha}{\beta^2\nu}\}$. The the estimate easily follows by triangle inequality, after (4.11)-(4.12), after noting that
$k^3 \sum_{n=0}^{N} \|\boldsymbol{\mu}_t(t_n)\|^2_{\mathbf{L}^2(\Omega)} \leq k^2 T \|\boldsymbol{\mu}_t\|^2_{L^\infty(0,T;\mathbf{L}^2(\Omega))}$. □

Remark 4.2. *It is evident that the penalization based on the artificial compressibility formulation leads to an estimate of the form $k + \varepsilon k$, unlike to the standard penalization algorithm which usually leads to estimates of the form $k + \varepsilon$. For the standard penalization algorithm the obvious choice $k \approx \varepsilon$ leads to systems with large condition numbers. However, discretization based on the artificial compressibility allows greater flexibility to the choice of ε.*

Acknowledgements. The author acknowledges the financial support provided through the European Community's Human Potential Programme under contract HPRN–CT–2002–00286 (Breaking Complexity)

REFERENCES

[1] F. Abergel and R.Temam, *On some control problems in fluid mechanics*, Theor. Comput. Fluid Dyn., **1**(1990), 303-326.
[2] R. Adams, *Sobolev Spaces*, Academic Press, New York, 1975.
[3] F. Brezzi and M. Fortin, *Mixed and Hybrid Finite Element Methods*, Springer-Verlag, 1991.
[4] F. Brezzi, J. Rappaz and P.-A. Raviart, *Finite dimensional approximation of nonlinear problems. Part I: Branches of nonsingular solutions*, Numer. Math., **3**6(1980), 1-25.
[5] A.J. Chorin, *Numerical solution to the Navier-Stokes equations* Math. Comp., **2**2 (1968), 745-762.
[6] M. Fortin and R. Glowinski, *Augmented Lagrangian Methods for Navier-Stokes Equations*, North Holland, Amsterdanm, 1983.
[7] K. Chrysafinos, *Analysis and finite element approximations for the velocity tracking problem for Stokes flows via a penalized formulation*, ESAIM Control Optim. Calc. Var., **1**0 (2004), 574-592.
[8] K. Chrysafinos, *Error estimates for the velocity tracking problem for Navier-Stokes flows based on the artificial compressibility formulation*, Numer. Funct. Anal. Optim., **2**6 (2005), 773-812.
[9] P. Constantin and C. Foias, *Navier-Stokes equations*, Chicago Lectures in Mathematics, 1988.
[10] K. Dechelnick and M. Hinze, *Semidiscretization and error estimates for distributed control of the instationary Navier-Stokes equations*, Numer. Math., **9**7 (2004), 297-320.
[11] A. Fursikov, *Optimal control of distributed systems. Theory and applications*, AMS Providence, 2000.
[12] V. Girault and P.-A. Raviart, *Finite Element Methods for Navier-Stokes*, Springer-Verlag, New York, 1986.
[13] M.D. Gunzburger, *Perspectives in Flow Control and Optimization*, Advances in Design and Control, SIAM, Philadelphia 2003.
[14] M.D. Gunzburger, L.S. Hou and T. Svobodny, *Analysis and finite element approximation of optimal control problems for the stationary Navier-Stokes equations with Dirichlet controls*, RAIRO Model. Math. Anal. Numer., **2**5 (1991), 711-748.
[15] M.D. Gunzburger and S. Manservisi, *The velocity tracking problem for Navier-Stokes flows with bounded distributed control*, SIAM J.Control and Optim., **37** (2000), 1913-1945.
[16] M.D. Gunzburger and S. Manservisi, *Analysis and approximation of the velocity tracking problem for Navier-Stokes flows with distributed control*, SIAM J. Numer. Anal., **3**7 (2000), 1481-1512.
[17] M.D. Gunzburger and C. Trenchea, *Analysis and discretization of an optimal control problem for the time-periodic MHD Equations*, J. Math. Anal. Appl., **3**08 (2005), 440-466.
[18] Y. He, *Optimal error estimate of the penalty finite element method for the time-dependent Navier-Stokes equations*, Math. Comp., **7**4 (2005), 1201-1216.
[19] J.G. Heywood and R. Rannacher, *Finite element approximation of the nonstationary Navier-Stokes problem. I. Regularity of solutions and second-order error estimates for spatial discretization*, SIAM. J. Num. Anal., **1**9 (1982), 275-311.

[20] M. Hinze and K. Kunisch, *Second order methods for optimal control of time-dependent fluid flow*, SIAM J. Control and Optim., **4**0 (2001), 925-946.
[21] L.S. Hou, *Error estimates for semidiscrete finite element approximation of the Stokes equations under minimal regularity assumptions*, J. Sci. Comput., **1**6 (2001), 287-317.
[22] L.S. Hou and S.S. Ravindran, *A penalized Neumann control approach for solving an optimal Dirichlet control problem for the Navier-Stokes equations*, SIAM J.Control and Optim., **3**6 (1998), 1795-1814.
[23] T.J.R. Hughes, W.T. Liu and A.J. Brooks, *Finite element analysis of incompressible flows by the penalty function formulation*, J. Comp. Phys., **3**0 (1979), 1-60.
[24] J.-L. Lions, *Control of distributed singular systems*, Bordas, Paris 1985.
[25] J.-L. Lions, *Optimal Control of Distributed Parameter Systems*, SIAM, Philadelphia, 1972.
[26] J. Shen, *On error estimates of the penalty method for unsteady Navier-Stokes equations*, SIAM J. Numer. Anal., **3**2 (1995), 386-403.
[27] R. Temam, *Navier-Stokes equations*, North Holland, Amsterdam, 1979.
[28] R. Temam, *Une méthode d'approximation de la solution des equations de Navier-Stokes*, Bull. Soc. Math. Fracce, **9**8 (1968) 115-152.
[29] B.A. Ton, *Optimal shape control problems for the Navier-Stokes equations*, SIAM J. Control and Optim., **4**1 (2003), 1733-1747.

Received May 2005; revised March 2006.

E-mail address: chrysafinos@ins.uni-bonn.de