ESAIM: COCV
October 2004, Vol. 10, 574–592
DOI: 10.1051/cocv:2004021

ESAIM: Control, Optimisation and Calculus of Variations

ANALYSIS AND FINITE ELEMENT ERROR ESTIMATES FOR THE VELOCITY TRACKING PROBLEM FOR STOKES FLOWS *VIA* A PENALIZED FORMULATION

Konstantinos Chrysafinos[1]

Abstract. A distributed optimal control problem for evolutionary Stokes flows is studied *via* a pseudocompressibility formulation. Several results concerning the analysis of the velocity tracking problem are presented. Semidiscrete finite element error estimates for the corresponding optimality system are derived based on estimates for the penalized Stokes problem and the BRR (Brezzi-Rappaz-Raviart) theory. Finally, the convergence of the solutions of the penalized optimality systems as $\varepsilon \to 0$ is examined.

Mathematics Subject Classification. 35B37, 65M60, 49J20.

Received April 3, 2003. Revised December 16, 2003.

1. Introduction

We consider the following optimal control problem: minimize

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

subject to the constrains:

$$\begin{cases} u_t - \nu\Delta u + \nabla p = f + g & \text{in } \Omega\times(0,T) \\ \nabla\cdot u = 0 & \text{in } \Omega\times(0,T) \\ u = 0 & \text{on } \Gamma\times(0,T) \\ u(0,x) = u_0 & \text{in } \Omega \end{cases} \tag{1.2}$$

where Ω is a bounded domain, with boundary Γ. From the physical point of view, the main objective is to steer the velocity vector field u of a Stokes flow to a prescribed target U using a distributed control function f. The cost functional consists of two parts. The first norm measures the effectiveness of the control process while the second one the cost of the control function. Adjusting the parameters α, β we can balance the effectiveness with the cost.

Flow control problems have been studied before both analytically and numerically (see, *e.g.* [3, 5–7, 9] and references within). Several results concerning the analysis of flow control problems, including the existence of

Keywords and phrases. Optimal control, velocity tracking, finite elements, semidiscrete error estimates, Stokes equations, penalized formulation.

[1] Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh PA 15213, USA;
e-mail: `kchrysaf@zxc2.math.cmu.edu`

optimal solutions and the derivation of an optimality system, can be found in [3] (see also references within). In [5–7], analysis and finite element approximations of optimal control problems are studied based on an optimality system approach. Specifically, in [7] analysis and approximations of the velocity tracking problem for Navier-Stokes flows using a distributed control are presented. A gradient method for the solution of the fully discrete equations of the optimality system, and its convergence are examined. In [6], a similar approach is illustrated in case of a bounded distributed control. Finally, in [5] several results concerning the velocity tracking problem related to elliptic Navier-Stokes equations are presented using a boundary control.

The scope of this paper is to derive semidiscrete finite element error estimates for the optimality system, including estimates for the pressure p and for the time derivative u_t of the velocity vector field. In order to derive such estimates, it is important to address the issue of regularity of weak solutions of problem (1.1), *i.e.*, to determine the regularity of u, u_t, p, under minimal regularity assumptions on data f, g, u_0. In particular, given $f + g \in L^2(0,T;\mathbf{H}^{-1}(\Omega))$ the "natural" space to seek convergence is

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

Furthermore, the use of the above spaces is also desirable for uncoupling the state and the adjoint variables of the optimality system. Despite the extensive literature concerning the regularity of weak solutions of (1.1) such result is not available (see, *e.g.* [8, 11]).

To overcome this obstacle, we introduce an auxiliary optimal control problem based on the penalized Stokes equations, *i.e.*, we minimize $J(\cdot,\cdot)$ subject to:

$$\begin{cases} u_t - \nu\Delta u + \nabla p = f + g \quad & \text{in } \Omega \times (0,T) \\ \varepsilon p + \nabla \cdot u = 0 \quad & \text{in } \Omega \times (0,T) \\ u = 0 \quad & \text{on } \Gamma \times (0,T) \\ u(0,x) = u_0 \quad & \text{in } \Omega. \end{cases} \tag{1.3}$$

This approach is motivated from [9], where a boundary control problem for the elliptic Navier-Stokes equations is studied *via* a penalized formulation and it can be viewed as an attempt to regularize our system. In [13] an optimal shape control problem related to the evolutionary Navier-Stokes equations is also studied based on an artificial compressibility approach. Pseudocompressible methods have been studied extensively in [10–12] both analytically and numerically. The penalized formulation (1.3) was first analyzed in [12] where it was proven that under certain assumptions the solution of (1.3) converge as $\varepsilon \to 0$ to the solution of (1.2) in an appropriate norm. The main asset of this method within the optimal control setting is that it provides the means to derive error estimates on the "natural" space, which subsequently facilitates the derivation of semidiscrete error estimates for the optimality system. Furthermore, another advantage of this approach is that the finite element subspaces do not need to satisfy the inf-sup condition. Of course, the limit case has to be carefully examined.

The rest of this paper is as follows: after providing the basic notation and definitions, in Section 3 we analyze the existence of an optimal solution and its convergence as $\varepsilon \to 0$. In Section 4, we derive semidiscrete error estimates on the natural norm for the penalized Stokes equations. Our emphasis is on the regularity assumptions for the given data. To our knowledge such estimates are new. Using these estimates together with the BRR (Brezzi-Rappaz-Raviart) theory (see *e.g.* [4]) we obtain semidiscrete error estimates for the corresponding optimality system. Finally, in Section 5, we examine the convergence as $\varepsilon \to 0$ of the solution of the penalized optimality system to the solution of the optimality system corresponding to the original optimal control problem.

2. Preliminaries

We shall use the standard notation for the Sobolev spaces. Throughout this paper u, μ, f, g, U denote vector fields, p, r, q scalar functions and C, D constants depending only on the domain Ω. We denote by $L^2(\Omega)$ the space of all Lebesgue square integrable functions defined on Ω and by $H^1(\Omega) = \{v \in \Omega : \frac{\partial v}{\partial x_i} \in L^2(\Omega), i = 1, ..., d\}$

where $d = 2, 3$. Note that $H_0^1(\Omega) = \{v \in H^1(\Omega) : v|_\Gamma = 0\}$. Similarly we denote by $H^m(\Omega)$ the standard Sobolev spaces which contain weak partial derivatives of order m (positive integer). For the definition of fractional ordered Sobolev spaces see [1]. We also denote by H^{-1} the dual space of $H_0^1(\Omega)$. We employ the standard notation for the $L^2(\Omega)$ inner product *i.e.* $(\cdot,\cdot)$, together with the corresponding norm notation $\|\cdot\|_{L^2(\Omega)}$. Similar notation also holds for the norms of $H^m(\Omega)$. If X is a Banach space, we denote by X^* its dual. In addition, we use the notation $\langle\cdot,\cdot\rangle \equiv \langle\cdot,\cdot\rangle_{(H^{-1}(\Omega),H_0^1(\Omega))}$ for the standard duality pairing. Vector valued Sobolev spaces in Ω are denoted by $\mathbf{H}^m(\Omega)$ with norms and inner products denoted in the same way as above. If X is a Hilbert space, we denote by $L^2(0,T;X), H^1(0,T;X)$ the time-space function spaces such that

$$\|v\|^2_{L^2(0,T,X)} \equiv \int_0^T \|v(t)\|_X^2 \mathrm{d}t < \infty \qquad \forall v \in L^2(0,T;X),$$

and

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

together with the appropriate modification in case of $L^\infty(0,T;X)$. Moreover we define the solenoidal vector spaces

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

where n denotes the unit outer normal in Γ. Note that $V(\Omega)$ is the closure of $\mathcal{V}(\Omega)$ in $\mathbf{H}^1(\Omega)$. Furthermore, we equip the above spaces with norms given by $\|\cdot\|_{V(\Omega)} \equiv \|\cdot\|_{\mathbf{H}^1(\Omega)}$ and $\|\cdot\|_{W(\Omega)} \equiv \|\cdot\|_{\mathbf{L}^2(\Omega)}$ respectively. We also define,

$$L_0^2(\Omega) = \{p \in L^2(\Omega) : \int_\Omega p\mathrm{d}x = 0\}\cdot$$

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

$$\begin{aligned} a(u,v) &= \nu\sum_{i,j}\int_\Omega D_{ij}(u)D_{ij}(v)\mathrm{d}x \qquad \forall\, u,v \in \mathbf{H}_0^1(\Omega),\\ b(v,q) &= -\int_\Omega q\nabla\cdot v\mathrm{d}x \qquad \forall\, q \in L^2(\Omega), v \in \mathbf{H}^1(\Omega), \end{aligned}$$

where $D_{ij}(v) = \frac{1}{2}(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i})$. Note that the bilinear form satisfy the following coercivity property:

$$a(u,u) \geq C\|u\|^2_{H_0^1(\Omega)}, \qquad \forall\, u \in \mathbf{H}_0^1(\Omega).$$

Finally we denote by $\mathcal{B}$ the set of all $f \in L^2(0,T;\mathbf{L}^2(\Omega))$. We consider target velocity fields $U \in \mathcal{B}$, but typically the divergence free condition as well as the homogeneous boundary condition needs also to be satisfied. Furthermore, we assume $F_U \equiv U_t - \nu\Delta U \in L^\infty(0,T;\mathbf{L}^2(\Omega))$ so that the target U has a physical meaning. For the mathematical analysis none of these constraints are necessary. For the rest of this paper we assume that the data satisfy,

$$g \in L^2(0,T;\mathbf{H}^{-1}(\Omega)), \qquad u_0 \in W(\Omega).$$

Then, a weak form for the evolutionary Stokes equations can be defined as follows: we seek a velocity u such that

$$\begin{cases} \langle u_t, v\rangle + a(u,v) = \langle f+g, v\rangle & \forall v \in V(\Omega)\\ u(0,x) = u_0(x) & \in W(\Omega). \end{cases} \tag{2.1}$$

Furthermore, there exists an associated pressure p such that (1.2) is satisfied in the sense of distributions. We recall that if Γ is Lipschitz and $g \in L^2(0,T;\mathbf{H}^{-1}(\Omega)), u_0 \in W(\Omega)$ then the solutions of (2.1) satisfy,

$$u \in L^\infty(0,T;W(\Omega)) \cap L^2(0,T;V(\Omega)), \qquad u_t \in L^2(0,T;(V(\Omega))^*).$$

The admissible set is defined as follows,

Definition 2.1. Given, $T > 0, u_0 \in W(\Omega), g \in L^2(0,T;\mathbf{H}^{-1}(\Omega))$,

$$U_{ad}^1 = \Big\{(u,f) \in L^2(0,T;\mathbf{H}_0^1(\Omega)) \times L^2(0,T;\mathbf{L}^2(\Omega)), J(u,f) < \infty \text{ and such that (2.1) is satisfied}\Big\}\cdot$$

The definition of an optimal solution of problem (P_1) is a local one:

Definition 2.2. We seek an optimal solution $(u,f) \in U_{ad}^1$ such that $J(u,f) \le J(w,h) \quad \forall (w,h) \in U_{ad}^1$ satisfying $\|h - f\|_{L^2(0,T;\mathbf{L}^2(\Omega))} \le \epsilon$ for some $\epsilon > 0$.

We would like to emphasize a fundamental difficulty involved in the above weak formulation. As mentioned earlier, the desired weak formulation setting is that given a forcing term in $L^2(0,T;\mathbf{H}^{-1}(\Omega))$ and initial data in $W(\Omega)$, seek a solution pair

$$(u,p) \in L^2(0,T;\mathbf{H}_0^1(\Omega)) \cap H^1(0,T;\mathbf{H}^{-1}(\Omega)) \times L^2(0,T;L_0^2(\Omega)),$$

such that

$$\begin{cases} \langle u_t, v\rangle + a(u,v) + b(v,p) = \langle f+g, v\rangle \quad \forall v \in \mathbf{H}_0^1(\Omega) \\ b(u,q) = 0 \qquad \forall\, q \in L_0^2(\Omega) \\ u(0,x) = u_0(x) \qquad \in W(\Omega). \end{cases}$$

From the numerical analysis viewpoint this is the natural space to show the convergence of semidiscrete solutions. Unfortunately, such an existence theorem is not available in the literature. Therefore, in order to regularize our system and "uncouple" p from u_t we use a penalized weak formulation. The problem (P_2) we consider is defined as follows:

Definition 2.3. For fixed, $\varepsilon > 0$, minimize the functional,

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

subject to the constraint:

$$\begin{cases} \langle u_t^\varepsilon, v\rangle + a(u^\varepsilon,v) + b(v,p^\varepsilon) = \langle f^\varepsilon + g, v\rangle \qquad \forall v \in \mathbf{H}_0^1(\Omega) \\ \varepsilon(p^\varepsilon, q) - b(u^\varepsilon, q) = 0 \qquad \forall\, q \in L^2(\Omega) \\ (u^\varepsilon(0,x), z) = (u_0(x), z) \qquad \forall\, z \in \mathbf{L}^2(\Omega). \end{cases} \tag{2.2}$$

Similarly, we may define the admissibility set U_{ad}^2 for the optimal control problem (P_2) as:

Definition 2.4. Given $T > 0, u_0 \in W(\Omega), g \in L^2(0,T;\mathbf{H}^{-1}(\Omega))$

$$U_{ad}^2 = \Big\{(u^\varepsilon, f^\varepsilon) \in L^2(0,T;\mathbf{H}_0^1(\Omega)) \times L^2(0,T;\mathbf{L}^2(\Omega)), J(u^\varepsilon, f^\varepsilon) < \infty \text{ and}$$
$$\text{there exists a pressure } p^\varepsilon \in L^2(0,T;L^2(\Omega)) \text{ such that (2.2) is satisfied}\Big\}\cdot$$

3. Analysis and limiting behaviour of optimal control problem (P_2)

3.1. Existence of an optimal solution

In this section, for fixed ε we prove the existence of a solution pair $(u^\varepsilon, f^\varepsilon) \in U^2_{ad}$ and a corresponding pressure $p^\varepsilon \in L^2(0,T;L^2(\Omega))$ such that $J(\cdot,\cdot)$ is minimized subject to (2.2). First we state some properties concerning the solvability of problem (2.2) for given data f^ε, g, u_0. The proof of this statement can be found in [12].

Lemma 3.1. *For every $\varepsilon > 0$ and for $f^\varepsilon \in L^2(0,T;\mathbf{L}^2(\Omega)), g \in L^2(0,T;\mathbf{H}^{-1}(\Omega))$, $u_0 \in W(\Omega)$ there exists a unique solution $(u^\varepsilon, p^\varepsilon)$ such that*

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

Moreover, u^ε are weakly continuous from $[0,T]$ into $\mathbf{L}^2(\Omega)$.

Proof. See [12], Theorem (I.1). □

Remark 3.2. [12], Theorem (I.1), is proven for the penalized Navier-Stokes equations in case that $g \in L^2(0,T;\mathbf{L}^2(\Omega)), u_0 \in W(\Omega)$. However, it is easy to see to that for the corresponding linear problem, an existence and uniqueness theorem can be proven for $g \in L^2(0,T;\mathbf{H}^{-1}(\Omega))$ using exactly the same techniques (see also Lem. 3.3).

Next we prove some useful *a priori* estimates which will be subsequently used to show the existence of an optimal solution and its convergence as $\varepsilon \to 0$. The following estimate is equivalent to [12], estimate (I; 2.7).

Lemma 3.3. *For fixed $\varepsilon > 0$ let $f^\varepsilon \in L^2(0,T;\mathbf{L}^2(\Omega)), g \in L^2(0,T;\mathbf{H}^{-1}(\Omega))$, and $u_0 \in W(\Omega)$. Then the solution pair $(u^\varepsilon, p^\varepsilon)$ of (2.2) satisfies the following a priori estimate.*

$$\begin{aligned} &\|u^\varepsilon\|^2_{L^\infty(0,T;\mathbf{L}^2(\Omega))} + \varepsilon\|p^\varepsilon\|^2_{L^2(0,T;L^2(\Omega))} + \|u^\varepsilon\|^2_{L^2(0,T;\mathbf{H}^1_0(\Omega))} \\ &\quad \le C\left(\|f^\varepsilon\|^2_{L^2(0,T;\mathbf{L}^2(\Omega))} + \|u_0\|^2_{\mathbf{L}^2(\Omega)} + \|g\|^2_{L^2(0,T;\mathbf{H}^{-1}(\Omega))}\right). \end{aligned} \tag{3.1}$$

Proof. Set $v = u^\varepsilon, q = p^\varepsilon$ into (2.2). Then,

$$\begin{cases} \dfrac{1}{2}\dfrac{\mathrm{d}}{\mathrm{d}t}\|u^\varepsilon\|^2_{\mathbf{L}^2(\Omega)} + C\|u^\varepsilon\|^2_{\mathbf{H}^1(\Omega)} + b(u^\varepsilon, p^\varepsilon) = \langle f^\varepsilon + g, u^\varepsilon\rangle \\ \varepsilon\|p^\varepsilon\|^2_{L^2(\Omega)} - b(u^\varepsilon, p^\varepsilon) = 0. \end{cases}$$

Summing the above equations and using standard techniques we obtain the desired estimate. □

Theorem 3.4. *For every $\varepsilon > 0$ there exists a solution of the optimal control problem (P_2).*

Proof. For convenience we drop the ε notation from functions $u^\varepsilon, p^\varepsilon, f^\varepsilon$. From Lemma 3.3 it is obvious that for fixed ε there exists a solution of (2.2). Therefore $U^2_{ad} \neq \emptyset$, so we may choose a minimizing sequence (u^m, f^m) satisfying (2.2), *i.e.*,

$$\begin{cases} \langle u^m_t, v\rangle + a(u^m, v) + b(v, p^m) = \langle f^m + g, v\rangle & \forall\, v \in \mathbf{H}^1_0(\Omega) \\ \varepsilon(p^m, q) - b(u^m, q) = 0 \qquad \forall\, q \in L^2(\Omega) \\ (u^m(0,x), z) = (u_0, z) \qquad \forall\, z \in \mathbf{L}^2(\Omega) \end{cases} \tag{3.2}$$

and

$$\lim_{m\to 0} J(u^m, f^m) = \inf_{(u,f)\in U^2_{ad}} \Big\{ J(u,f) : \text{there exists } p \text{ satisfying (2.2)} \Big\}.$$

The boundedness of $J(u^m, f^m)$ implies the boundedness of $\|f^m\|_{L^2(0,T;\mathbf{L}^2(\Omega))}$ and $\|u^m\|_{L^2(0,T;\mathbf{L}^2(\Omega))}$. Therefore, Lemma 3.3 guarantees the boundedness of $\|u^m\|_{L^2(0,T;\mathbf{H}_0^1(\Omega))}$, $\|u^m\|_{L^\infty(0,T;\mathbf{L}^2(\Omega))}$ and $\|p^m\|_{L^2(0,T;L^2(\Omega))}$. Thus, we may extract subsequences still denoted by (u^m, p^m, f^m) such that:

$$\begin{aligned}
&f^m \to f^\varepsilon \text{ weakly in } L^2(0,T;\mathbf{L}^2(\Omega)), \quad u^m \to u^\varepsilon \text{ weakly in } L^2(0,T;\mathbf{H}_0^1(\Omega)),\\
&p^m \to p^\varepsilon \text{ weakly in } L^2(0,T;L^2(\Omega)), \quad u^m \to u^\varepsilon \text{ weakly-* in } L^\infty(0,T;\mathbf{L}^2(\Omega)),\\
&u^m \to u^\varepsilon \text{ strongly in } L^2(0,T;\mathbf{L}^2(\Omega)),
\end{aligned}$$

where the last convergence result follows from a well known compactness embedding $L^2(0,T;\mathbf{H}_0^1(\Omega)) \cap H^1(0,T;\mathbf{H}^{-1}(\Omega)) \subset L^2(0,T;\mathbf{L}^2(\Omega))$ (see [11], Th. 2.1, p. 271), after noting that u^m remains bounded on $L^2(0,T;\mathbf{H}_0^1(\Omega)) \cap H^1(0,T;\mathbf{H}^{-1}(\Omega))$ (see also [12], Relation (I; 3.9)). The weak lower semicontinuity of the functional guarantees that

$$J(u^\varepsilon, f^\varepsilon) \le \liminf_{m\to 0} J(u^m, f^m).$$

It remains to show that the limit $(u^\varepsilon, p^\varepsilon, f^\varepsilon)$ defined as above satisfies the weak formulation (2.2). For that purpose, let ψ be a continuously differentiable scalar function on $[0,T]$, with $\psi(T) = 0$. Setting into the first equation of system (3.2) $\psi(t)w$, $w \in \mathbf{C}_0^\infty(\Omega)$ and integrating by parts with respect to time, we obtain:

$$-\int_0^T (u^m, \psi_t(t)w)\mathrm{d}t + \int_0^T a(u^m, w\psi(t))\mathrm{d}t + \int_0^T b(v, p^m)\mathrm{d}t = (u^m(0,x), \psi(0)w) + \int_0^T \langle f^m + g, w\psi(t)\rangle \mathrm{d}t. \quad (3.3)$$

The convergence results allows us to pass the limit into (3.3).

$$-\int_0^T (u^\varepsilon, \psi_t(t)w)\mathrm{d}t + \int_0^T a(u^\varepsilon, w\psi(t))\mathrm{d}t + \int_0^T b(v, p^\varepsilon)\mathrm{d}t = (u(0,x), \psi(0)w) + \int_0^T \langle f^\varepsilon + g, w\psi(t)\rangle \mathrm{d}t.$$

Integration by parts in time once more, and a well known density argument imply that the first equation of (2.2) holds for all $v \in \mathbf{H}_0^1(\Omega)$. Similarly, it is easy to see that we may pass the limit into the second equation of (3.2) to obtain,

$$\varepsilon(p^\varepsilon, q) - b(u^\varepsilon, q) = 0 \qquad \forall\, q \in L^2(\Omega).$$

Therefore, for every ε there exists an optimal solution pair $(u^\varepsilon, f^\varepsilon)$. □

3.2. Convergence of optimal solutions as $\varepsilon \to 0$

In this section, we examine the convergence of $(u^\varepsilon, p^\varepsilon, f^\varepsilon)$ as ε approaches to zero. We prove the following convergence results.

Theorem 3.5. *For fixed $\varepsilon > 0$, let $(u^\varepsilon, p^\varepsilon, f^\varepsilon)$ be an optimal solution for (P_2). Then, there exists a subsequence ε_k, $(u^{\varepsilon_k}, p^{\varepsilon_k}, f^{\varepsilon_k})$ such that:*

$$\begin{aligned}
&u^{\varepsilon_k} \to \hat{u} \text{ weakly in } L^2(0,T;\mathbf{H}_0^1(\Omega)), \quad u^{\varepsilon_k} \to \hat{u} \text{ weakly -* in } L^\infty(0,T;\mathbf{L}^2(\Omega)),\\
&f^{\varepsilon_k} \to \hat{f} \text{ weakly in } L^2(0,T;\mathbf{L}^2(\Omega)), \quad \sqrt{\varepsilon} p^{\varepsilon_k} \to \hat{p} \text{ weakly in } L^2(0,T;L_0^2(\Omega)),\\
&u^{\varepsilon_k} \to \hat{u} \text{ strongly in } L^2(0,T;\mathbf{L}^2(\Omega)).
\end{aligned}$$

Proof. Let $(\hat{u}^\varepsilon, \hat{p}^\varepsilon)$ be the solution of,

$$\left\{\begin{array}{ll}
\langle \hat{u}_t^\varepsilon, v\rangle + a(\hat{u}^\varepsilon, v) + b(v, \hat{p}^\varepsilon) = \langle g, v\rangle & \forall\, v \in \mathbf{H}_0^1(\Omega)\\
\varepsilon(\hat{p}^\varepsilon, q) - b(\hat{u}^\varepsilon, q) = 0 & \forall\, q \in L^2(\Omega)\\
(\hat{u}^\varepsilon(0,x), z) = (u_0, z) & \forall\, z \in \mathbf{L}^2(\Omega).
\end{array}\right. \quad (3.4)$$

Then, using estimates of Lemma 3.3 for system (3.4), we obtain,

$$\|\hat{u}^\varepsilon\|^2_{L^\infty(0,T;\mathbf{L}^2(\Omega))} + \varepsilon\|\hat{p}^\varepsilon\|^2_{L^2(0,T;L^2(\Omega))} + \|\hat{u}^\varepsilon\|^2_{L^2(0,T;\mathbf{H}^1_0(\Omega))} \leq \|u_0\|^2_{\mathbf{L}^2(\Omega)} + \|g\|^2_{L^2(0,T;\mathbf{H}^{-1}(\Omega))}. \tag{3.5}$$

Note also that the optimality of $(u^\varepsilon, f^\varepsilon)$ implies that $J(u^\varepsilon, f^\varepsilon) \leq J(\hat{u}^\varepsilon, 0)$, *i.e.*,

$$\frac{\alpha}{2}\int_0^T \|u^\varepsilon - U\|^2_{\mathbf{L}^2(\Omega)}\mathrm{d}t + \frac{\beta}{2}\int_0^T \|f^\varepsilon\|^2_{\mathbf{L}^2(\Omega)}\mathrm{d}t \leq \frac{\alpha}{2}\int_0^T \|\hat{u}^\varepsilon - U\|^2_{\mathbf{L}^2(\Omega)}\mathrm{d}t. \tag{3.6}$$

But the later integral in (3.6) is bounded independent of ε due to (3.5). Thus (3.6) clearly guarantees that $\|f^\varepsilon\|^2_{L^2(0,T;\mathbf{L}^2(\Omega))}$, $\|u^\varepsilon\|^2_{L^2(0,T;\mathbf{L}^2(\Omega))}$ are bounded independent of ε. Returning back to the *a priori* estimate of Lemma 3.3 we also obtain that

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

are bounded independed of ε. Therefore, there exists a subsequence denoted by $(u^{\varepsilon_k}, p^{\varepsilon_k}, f^{\varepsilon_k})$ such that the first three convergence results hold. Since $\|f^\varepsilon\|^2_{L^2(0,T;\mathbf{L}^2(\Omega))}$ is bounded independent of ε we may apply a standard argument (see [12], Relation (I; 3.11)) to obtain a uniform bound of a fractional derivative in time independent of ε. Then, using a compactness theorem [11], Theorem 2.1, p. 271, we obtain the strong convergence result (see also [12], Th. I.2). □

Next we need to show that the limit $(\hat{u}, \hat{f})$, together with the corresponding pressure $\hat{p}$ (defined as in Th. 3.5) are indeed solutions of the original optimal control problem (P_1). For that purpose we prove the following two propositions.

Proposition 3.6. *The limit* $(\hat{u}, \hat{f}) \in U^1_{ad}$ *and* $\hat{u}_t \in L^2(0,T;V(\Omega)^*)$.

Proof. First note that $(u^{\varepsilon_k}, f^{\varepsilon_k}) \in U^2_{ad}$ and p^{ε_k} satisfy:

$$\begin{cases} \langle u^{\varepsilon_k}_t, v\rangle + a(u^{\varepsilon_k}, v) + b(v, p^{\varepsilon_k}) = \langle f^{\varepsilon_k} + g, v\rangle & \forall\, v \in \mathbf{H}^1_0(\Omega) \\ \varepsilon_k(p^{\varepsilon_k}, q) - b(u^{\varepsilon_k}, q) = 0 \qquad \forall\, q \in L^2(\Omega) \\ (u^{\varepsilon_k}(0,x), z) = (u_0, z) \qquad \forall\, z \in \mathbf{L}^2(\Omega). \end{cases} \tag{3.7}$$

Moreover, the weak convergence of $\sqrt{\varepsilon}p^{\varepsilon_k}$ to $\hat{p}$ in $L^2(0,T;L^2(\Omega))$ implies that

$$\varepsilon\int_0^T (p^{\varepsilon_k}, q)\mathrm{d}t \to 0 \qquad \forall\, q \in L^2(0,T;L^2_0(\Omega)).$$

The above result together with (3.7) and convergence results of Theorem 3.5 clearly imply that $\hat{u} \in L^2(0,T;V(\Omega))$. Setting $v = \psi(t)w$ into (3.7) where ψ is a smooth scalar function with $\psi(T) = 0$, $w \in \mathcal{V}(\Omega)$,

$$-\int_0^T (u^{\varepsilon_k}, \psi_t(t)w)\mathrm{d}t + \int_0^T a(u^{\varepsilon_k}, \psi(t)w)\mathrm{d}t = (u_0, \psi(0)w) + \int_0^T \langle f^{\varepsilon_k} + g, \psi(t)w\rangle\mathrm{d}t. \tag{3.8}$$

It it obvious that we may pass the limit through (3.8), to obtain,

$$-\int_0^T (\hat{u}, \psi_t(t)w)\mathrm{d}t + \int_0^T a(\hat{u}, \psi(t)w)\mathrm{d}t = (\hat{u}_0, \psi(0)w) + \int_0^T \langle \hat{f} + g, \psi(t)w\rangle\mathrm{d}t.$$

Integrating by parts in time once more and using a standard density argument, we conclude our proof. In addition, (2.1) and the regularity of $\hat{u} \in L^2(0,T;V(\Omega))$ imply that $\hat{u}_t \in L^2(0,T;V(\Omega)^*)$. □

Proposition 3.7. $J(\hat{u},\hat{f}) \leq J(u,f) \quad \forall\,(u,f) \in U^1_{ad}$.

Proof. Suppose that $(u,f) \in U^1_{ad}$ satisfy

$$\begin{cases} \langle u_t, v\rangle + a(u,v) = \langle f+g, v\rangle & \forall\, v \in V(\Omega) \\ u(0,x) = u_0 \quad \in W(\Omega). \end{cases} \tag{3.9}$$

We define $(\tilde{u}^\varepsilon, \tilde{p}^\varepsilon)$ to be the solution of

$$\begin{cases} \langle \tilde{u}^\varepsilon_t, v\rangle + a(\tilde{u}^\varepsilon, v) + b(v, \tilde{p}^\varepsilon) = \langle f+g, v\rangle & \forall\, v \in \mathbf{H}^1_0(\Omega) \\ \varepsilon(\tilde{p}^\varepsilon, q) - b(\tilde{u}^\varepsilon, q) = 0 \quad \forall\, q \in L^2(\Omega) \\ (\tilde{u}^\varepsilon(0,x), z) = (u_0, z) \quad \forall\, z \in \mathbf{L}^2(\Omega). \end{cases} \tag{3.10}$$

Note that $(\tilde{u}^\varepsilon, f)$ together with the corresponding pressure $\tilde{p}^\varepsilon$ belong to the admissible set U^2_{ad} of problem (P_2) therefore, using the weak lower semicontinuity of $J(\cdot,\cdot)$ after possibly passing to a subsequence,

$$\begin{aligned} J(\hat{u},\hat{f}) &\leq \liminf_{\varepsilon\to 0} J(u^\varepsilon, f^\varepsilon) \leq J(\tilde{u}^\varepsilon, f) = \frac{\alpha}{2}\int_0^T \|\tilde{u}^\varepsilon(s) - U(s)\|^2_{\mathbf{L}^2(\Omega)}\mathrm{d}s + \frac{\beta}{2}\int_0^T \|f(s)\|^2_{\mathbf{L}^2(\Omega)}\mathrm{d}s \\ &\leq \frac{\alpha}{2}\int_0^T \|u(s) - \tilde{u}^\varepsilon(s)\|^2_{\mathbf{L}^2(\Omega)}\mathrm{d}s + J(u,f) + \alpha\int_0^T \|u(s) - \tilde{u}^\varepsilon(s)\|_{\mathbf{L}^2(\Omega)}\|u(s) - U(s)\|_{\mathbf{L}^2(\Omega)}\mathrm{d}s, \end{aligned}$$

where at the last step we used the triangle inequality. It remains to show that

$$\int_0^T \|u(s) - \tilde{u}^\varepsilon(s)\|^2_{\mathbf{L}^2(\Omega)}\mathrm{d}s \to 0 \text{ as } \varepsilon \to 0.$$

Indeed, using considerations similar to ones of Theorem 3.5 and Proposition 3.6 for problems (3.9)–(3.10) and compactness (see also [12], Th. (I.2)), we obtain that

$$\tilde{u}^\varepsilon - u \to 0 \qquad \text{weakly in } L^2(0,T;\mathbf{H}^1_0(\Omega)), \qquad \tilde{u}^\varepsilon - u \to 0 \qquad \text{strongly in } L^2(0,T;\mathbf{L}^2(\Omega))\,. \qquad \square$$

Remark 3.8. In order to obtain an actual rate of convergence, it appears that further regularity assumptions are needed. Suppose that u, u^ε are solutions of (1.2),(1.3) respectively and that $f+g \in L^\infty(0,T;\mathbf{L}^2(\Omega)), u_0 \in V(\Omega)$. Then [10], Lemma 3.1 of guarantees that $\|u - u^\varepsilon\|_{L^2(0,T;\mathbf{H}^1(\Omega))} \leq C\sqrt{\varepsilon}$. An analogous result for the optimality system will be proven in Section 5.

4. Finite element approximations of the penalized optimality system

4.1. **The discrete optimality system**

In previous section we established the existence of an optimal solution for the penalized Stokes system and we proved that as $\varepsilon \to 0$, $(u^\varepsilon, f^\varepsilon)$ becomes an optimal solution for the original problem. Therefore, instead of solving the optimality system corresponding to the optimal control problem (P_1), we may fix ε small enough and use the optimality system of the approximate problem (P_2). This approach leads to semidiscrete error estimates for the finite element approximations and it is based on the derivation of semidiscrete error estimates for an appropriate model problem as well as the BRR (Brezzi-Rappaz-Raviart) theory. Note also that due to penalization, our finite element subspaces do not need to satisfy the classical inf-sup condition. Using standard techniques of Calculus of Variations, see [3, 7], the optimality system has the following weak form

$$\begin{cases} \langle u_t^\varepsilon, v\rangle + a(u^\varepsilon, v) + b(v, p^\varepsilon) = \langle f + g, v\rangle & \forall\, v \in \mathbf{H}_0^1(\Omega) \\ \varepsilon(p^\varepsilon, q) - b(u^\varepsilon, q) = 0 & \forall\, q \in L^2(\Omega) \\ (u^\varepsilon(0,x) - u_0, z) = 0 & \forall\, z \in \mathbf{L}^2(\Omega) \end{cases} \tag{4.1}$$

$$\begin{cases} -\langle \mu_t^\varepsilon, v\rangle + a(\mu^\varepsilon, v) + b(v, r^\varepsilon) = \alpha(u - U, v) & \forall\, v \in \mathbf{H}_0^1(\Omega) \\ \varepsilon(r^\varepsilon, q) - b(\mu^\varepsilon, q) = 0 & \forall\, q \in L^2(\Omega) \\ (\mu^\varepsilon(T), z) = 0 & \forall\, z \in \mathbf{L}^2(\Omega) \end{cases} \tag{4.2}$$

$$\mu^\varepsilon = -\beta f^\varepsilon. \tag{4.3}$$

Of course, using the last equality (4.3), we may replace the control term from the state equation (4.1) and obtain the optimality system (4.2)–(4.4).

$$\begin{cases} \langle u_t^\varepsilon, v\rangle + a(u^\varepsilon, v) + b(v, p^\varepsilon) = \langle -\frac{1}{\beta}\mu^\varepsilon + g, v\rangle & \forall\, v \in \mathbf{H}_0^1(\Omega) \\ \varepsilon(p^\varepsilon, q) - b(u^\varepsilon, q) = 0 & \forall\, q \in L^2(\Omega) \\ (u^\varepsilon(0,x) - u_0, z) = 0 & \forall\, z \in \mathbf{L}^2(\Omega). \end{cases} \tag{4.4}$$

Next we define the finite element approximation of the optimality system (4.2)–(4.4). For simplicity we drop the ε notation from the optimality system. We choose finite element subspaces $V^h \subset H_0^1(\Omega), M^h \subset L^2(\Omega)$. We introduce the spaces $\Phi_0^r \equiv H_0^{\min\{1,r\}}$, for all real r, equipped with the $H^{\min\{1,r\}}(\Omega)$ norm, *i.e.*,

$$\Phi_0^r(\Omega) = \begin{cases} H_0^1(\Omega) & \text{if } r \geq 1, \\ H_0^r(\Omega) & \text{if } \frac{1}{2} < r < 1, \\ H^r(\Omega) & \text{if } r \leq \frac{1}{2}. \end{cases}$$

The standard approximation properties hold for V^h, M^h, *i.e.*, there exists an integer k, and a constant C, independent of h such that,

$$\inf_{v^h \in L^2(0,T;V^h)} \|v - v^h\|_{L^2(0,T;H^s(\Omega))} \leq Ch^{r+1-s}\|v\|_{L^2(0,T;H^{r+1}(\Omega))}$$
$$\forall\, v \in L^2(0,T;H^{r+1}(\Omega) \cap \Phi_0^{r+1}(\Omega)), -2 \leq r \leq k, s = -1, 0, 1 \tag{4.5}$$

$$\inf_{q^h \in L^2(0,T;M^h)} \|q - q^h\|_{L^2(0,T;L^2(\Omega))} \leq Ch^{r+1}\|q\|_{L^2(0,T;H^{r+1}(\Omega))}$$
$$\forall\, q \in L^2(0,T;H^{r+1}(\Omega)), -1 \leq r \leq k. \tag{4.6}$$

In addition, we assume that the following convergence results hold: for $s = -1, 0, 1$,

$$\inf_{v^h \in L^2(0,T;V^h)} \|v - v^h\|_{L^2(0,T;H^s(\Omega))} \to 0, \qquad \forall\, v \in L^2(0,T;H^s(\Omega)), \tag{4.7}$$

and

$$\inf_{q^h \in L^2(0,T;M^h)} \|q - q^h\|_{L^2(0,T;L^2(\Omega))} \to 0, \qquad \forall\, q \in L^2(0,T;L^2(\Omega)). \tag{4.8}$$

We denote the corresponding vector valued spaces by $\mathbf{V}^h$. Moreover, we denote by P^h the $\mathbf{L}^2(\Omega)$ projection from $\mathbf{L}^2(\Omega)$ to $\mathbf{V}^h$ such that

$$(P^h v, w^h) = (v, w^h) \qquad \forall\, w^h \in \mathbf{V}^h,$$

and by Q^h the generalized $\mathbf{L}^2(\Omega)$ projection, *i.e.*, $Q^h : \mathbf{H}^{-1}(\Omega) \to \mathbf{V}^h$ such that,

$$\langle Q^h v, w^h\rangle = \langle v, w^h\rangle \qquad \forall\, w^h \in \mathbf{V}^h.$$

Note that $Q^h v = P^h v \quad \forall v \in \mathbf{L}^2(\Omega)$. Similarly, we denote by P_p^h the $L^2(\Omega)$ projection from $L^2(\Omega)$ to M^h,

$$(P_p^h q, q^h) = (q, q^h) \qquad \forall\, q^h \in M^h.$$

Our next goal is to analyze the approximation properties of the above projections.

Lemma 4.1. *Let V^h, M^h be a family of finite element subspaces of $H_0^1(\Omega), L^2(\Omega)$ respectively satisfying (4.5)–(4.8), and let $\mathbf{V}^h$ be the vector valued version of V^h. Then the following approximation properties hold:*

$$\|v - P^h v\|^2_{L^2(0,T;\mathbf{H}^1(\Omega))} \to 0 \text{ as } h \to 0, \qquad \forall v \in L^2(0,T;\mathbf{H}_0^1(\Omega)), \tag{4.9}$$

$$\|v - Q^h v\|^2_{L^2(0,T;\mathbf{H}^{-1}(\Omega))} \to 0 \text{ as } h \to 0, \qquad \forall v \in L^2(0,T;\mathbf{H}^{-1}(\Omega)), \tag{4.10}$$

$$\|v - P^h v\|_{L^2(0,T;\mathbf{H}^1(\Omega))} \le Ch^r \|v\|_{L^2(0,T;\mathbf{H}^{r+1}(\Omega))}$$
$$\forall\, v \in L^2(0,T;\mathbf{H}^{r+1}(\Omega) \cap \mathbf{H}_0^1(\Omega)), 0 \le r \le k, \tag{4.11}$$

and

$$\|v - Q^h v\|_{L^2(0,T;\mathbf{H}^{-1}(\Omega))} \le Ch^{r+2} \|v\|_{L^2(0,T;\mathbf{H}^{r+1}(\Omega))}$$
$$\forall\, v \in L^2(0,T;\mathbf{H}^{r+1}(\Omega) \cap \Phi_0^r(\Omega)), -1 \le r \le k. \tag{4.12}$$

Furthermore, the following inequalities hold:

$$\|v - P^h v\|_{L^2(0,T;\mathbf{H}^1(\Omega))} \le C\|v - v^h\|_{L^2(0,T;\mathbf{H}^1(\Omega))}$$
$$\forall\, v \in L^2(0,T;\mathbf{H}_0^1(\Omega)), \forall\, v^h \in L^2(0,T;\mathbf{V}^h), \tag{4.13}$$

$$\|v - Q^h v\|_{L^2(0,T;\mathbf{H}^{-1}(\Omega))} \le C\|v - v^h\|_{L^2(0,T;\mathbf{H}^{-1}(\Omega))}$$
$$\forall\, v \in L^2(0,T;\mathbf{H}^{-1}(\Omega)), \forall\, v^h \in L^2(0,T;\mathbf{V}^h). \tag{4.14}$$

In addition similar properties also hold for the projection P_p, i.e.,

$$\|q - P_p^h q\|^2_{L^2(0,T;L^2(\Omega))} \to 0 \text{ as } h \to 0, \qquad \forall q \in L^2(0,T;L^2(\Omega)), \tag{4.15}$$

$$\|q - P_p^h q\|_{L^2(0,T;L^2(\Omega))} \le Ch^{r+1} \|q\|_{L^2(0,T;H^{r+1}(\Omega))}$$
$$\forall\, q \in L^2(0,T;H^{r+1}(\Omega)). -1 \le r \le k. \tag{4.16}$$

Proof. The proof of (4.9)–(4.14) can be found in [2]. For (4.15), (4.16) note that the definition of the P_p^h projection implies that $(q - P_p^h q, r^h) = 0 \quad \forall r^h \in M^h$. Thus, adding and subtracting $q^h \in M^h$ we obtain, $(q^h - P_p^h q, r^h) = (q^h - q, r^h)\ \forall r^h \in M^h$. Setting $r^h = q^h - P_p^h q$ into the above equality and using standard techniques together with approximation properties (4.6)–(4.8) we obtain the desired estimates. □

Now, we are ready to define finite element approximations of the optimality system: given $f \in L^2(0,T;\mathbf{L}^2(\Omega))$, $g \in L^2(0,T;\mathbf{H}^{-1}(\Omega)), u_0^h \in \mathbf{V^h}$ we seek $u^h, \mu^h \in H^1(0,T;\mathbf{V}^h)$, $p^h, r^h \in L^2(0,T;M^h)$ such that

$$\begin{cases} \langle u_t^h, v^h\rangle + a(u^h, v^h) + b(v^h, p^h) = -\frac{1}{\beta}(\mu^h, v^h) + \langle g, v^h\rangle \qquad \forall\, v^h \in \mathbf{V}^h \\ \varepsilon(p^h, q^h) - b(u^h, q^h) = 0 \qquad \forall\, q^h \in M^h \\ (u^h(0,x) - u_0^h, v^h) = 0 \qquad \forall\, v^h \in \mathbf{V}^h \end{cases} \tag{4.17}$$

$$\begin{cases} -\langle \mu_t^h, v^h\rangle + a(\mu^h, v^h) + b(v^h, r^h) = \alpha(u^h - U, v^h) \qquad \forall\, v^h \in \mathbf{V}^h \\ \varepsilon(r^h, q^h) - b(\mu^h, q^h) = 0 \qquad \forall\, q^h \in M^h \\ (\mu^h(T,x), v^h) = 0 \qquad \forall\, v^h \in \mathbf{V}^h. \end{cases} \tag{4.18}$$

4.2. Some results concerning the approximation of a class of nonlinear problems

Next we describe the main results concerning the BRR theory. In [5] BRR theory is used to handle nonlinear terms and to uncouple the discrete state and adjoint equations of an optimality system related to a boundary optimal control problem for the stationary Navier-Stokes equations. In our case, we use BRR theory to uncouple the state and the adjoint variables of the optimality system. The problems considered in BRR theory (see *e.g.* [4]) are of the following type. We seek a $\psi \in \mathcal{X}$ such that

$$\psi + \mathcal{T}\mathcal{G}(\psi) = 0, \tag{4.19}$$

where $\mathcal{T} \in \mathcal{L}(\mathcal{Y}, \mathcal{X}), \mathcal{G}$is a C^2 mapping from $\mathcal{X}$ into $\mathcal{Y}$ and $\mathcal{X}, \mathcal{Y}$ are Banach spaces. ψ is called a regular solution if we also have that $\psi + \mathcal{T}\mathcal{G}_\psi(\psi)$ is an isomorphism from $\mathcal{X}$ to $\mathcal{X}$, where $\mathcal{G}_\psi$ denotes the Frechet derivative. We assume that there exists another Banach space $\mathcal{Z}$, contained in $\mathcal{Y}$, with continuous embedding, such that the mapping

$$\psi \to \mathcal{G}_\psi(\psi) \in \mathcal{L}(\mathcal{X}, \mathcal{Z}) \qquad \forall\, \psi \in \mathcal{X}. \tag{4.20}$$

Approximations are defined by introducing a subspace $\mathcal{X}^h \subset \mathcal{X}$ and an approximating operator $\mathcal{T}^h \in \mathcal{L}(\mathcal{Y}, \mathcal{X}^h)$. We seek $\psi^h \in \mathcal{X}^h$ such that

$$\psi^h + \mathcal{T}^h\mathcal{G}(\psi^h) = 0. \tag{4.21}$$

Concerning the linear operator we assume the approximation properties:

$$\lim_{h\to 0} \|(\mathcal{T}^h - \mathcal{T})w\|_{\mathcal{X}} = 0 \qquad \forall\, w \in \mathcal{Y} \tag{4.22}$$

and

$$\lim_{h\to 0} \|\mathcal{T}^h - \mathcal{T}\|_{\mathcal{L}(\mathcal{Z},\mathcal{X})} = 0. \tag{4.23}$$

Note that whenever the imbedding $\mathcal{Z} \subset \mathcal{Y}$ is compact, the last relation follows from (4.22), and moreover the operator $\mathcal{T}\mathcal{G}_\psi \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ is compact. The main theorem can be stated as follows:

Theorem 4.2. *Let $\mathcal{X}$ and $\mathcal{Y}$ be Banach spaces. Assume that $\mathcal{G}$ is a C^2 mapping from $\mathcal{X}$ to $\mathcal{Y}$ and that $\mathcal{D}^2\mathcal{G}$is bounded on all bounded sets of $\mathcal{X}$. Assume that (4.20)–(4.22) and (4.23) hold and that ψ is a regular solution of (4.19). Then there exists a neighborhood $\mathcal{O}$ of the origin in $\mathcal{X}$ and for $h \le h_0$ small enough a and unique function $\psi^h \in \mathcal{X}^h$, such that ψ^h is a regular solution (4.21) and $\psi^h - \psi \in \mathcal{O}$. Moreover, there exists a constant $C > 0$, independent of h such that*

$$\|\psi^h - \psi\|_{\mathcal{X}} \le C\|(\mathcal{T}^h - \mathcal{T})\mathcal{G}(\psi)\|_{\mathcal{X}}.$$

Proof. See [4], pp. 306-307. □

Remark 4.3. The essence of this theory is that under certain hypotheses the error of the approximation of the nonlinear problem is of the same order of a related linear one. In order to apply the results of the above theorem, one needs to establish semidiscrete error estimates under minimal regularity assumptions for the related linear problem. The penalty method is an important asset in this proof. For more details concerning BRR theory, one may consult [4].

4.3. Semidiscrete finite element error estimates for the penalized Stokes equations

In this section we derive semidiscrete error estimates for the penalized Stokes equations. First note that [12], Theorem I.1 implies that for every $\varepsilon > 0$, $f \in L^2(0,T;\mathbf{L}^2(\Omega)), g \in L^2(0,T;\mathbf{H}^{-1}(\Omega))$, $u_0 \in W(\Omega)$ there exists a solution pair (u,p) such that

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

satisfying

$$\begin{cases} \langle u_t, v\rangle + a(u,v) + b(v,p) = \langle f+g, v\rangle \qquad \forall\, v \in \mathbf{H}^1_0(\Omega) \\ \varepsilon(p,q) - b(u,q) = 0 \qquad \forall\, q \in L^2(\Omega) \\ (u(0,x) - u_0, z) = 0 \qquad \forall\, z \in \mathbf{L}^2(\Omega). \end{cases} \tag{4.24}$$

The discrete weak problem is defined as follows: find $u^h \in H^1(0,T;\mathbf{V}^h), p^h \in L^2(0,T;M^h)$ such that

$$\begin{cases} \langle u^h_t, v^h\rangle + a(u^h,v^h) + b(v^h,p^h) = \langle f+g, v^h\rangle \qquad \forall\, v^h \in \mathbf{V}^h \\ \varepsilon(p^h,q^h) - b(u^h,q^h) = 0 \qquad \forall\, q^h \in M^h \\ (u^h(0,x) - u^h_0, v^h) = 0 \qquad \forall\, v^h \in \mathbf{V}^h. \end{cases} \tag{4.25}$$

First, we prove the following estimate:

Theorem 4.4. *Let $(u,p), (u^h,p^h)$ be the solutions of problems (4.24)-(4.25) respectively. Then the following estimate holds:*

$$\begin{aligned} \|u-u^h\|^2_{L^\infty(0,T;\mathbf{L}^2(\Omega))} &+ \frac{\varepsilon}{4}\|p-p^h\|^2_{L^2(0,T;L(\Omega))} + \|u-u^h\|^2_{L^2(0,T;\mathbf{H}^1(\Omega))} \\ &\le C\Big(\|u-P^hu\|^2_{L^\infty(0,T;\mathbf{L}^2(\Omega))} + \max\{\varepsilon,1\}\|p-P^h_p p\|^2_{L^2(0,T;L^2(\Omega))} \\ &\qquad + \max\{\frac{1}{\varepsilon},1\}\|u-P^hu\|^2_{L^2(0,T;\mathbf{H}^1(\Omega))}\Big). \end{aligned} \tag{4.26}$$

In addition, if $u \in L^2(0,T;\mathbf{H}^{m+1}(\Omega)) \cap H^1(0,T;\mathbf{H}^{m-1}(\Omega))$, $p \in L^2(0,T;H^m(\Omega)), 0 \le m \le k$ and $u^h_0 = P^hu_0$ then,

$$\begin{aligned} \|u-u^h\|^2_{L^\infty(0,T;\mathbf{L}^2(\Omega))} &+ \varepsilon\|p-p^h\|^2_{L^2(0,T;L^2(\Omega))} + \|u-u^h\|^2_{L^2(0,T;\mathbf{H}^1(\Omega))} \\ &\le Ch^{2m}\left(\max\left\{\frac{1}{\varepsilon},1\right\}\|u\|^2_{L^2(0,T;\mathbf{H}^{m+1}(\Omega))} + \max\{\varepsilon,1\}\|p\|^2_{L^2(0,T;H^m(\Omega))} + \|u_t\|^2_{L^2(0,T;\mathbf{H}^{m-1}(\Omega))}\right). \end{aligned} \tag{4.27}$$

Proof. Subtracting (4.25) from (4.24) we obtain the following orthogonality condition. For almost every t, $\forall\, v^h \in \mathbf{V}^h$,

$$\langle u_t(t) - u^h_t(t), v^h\rangle + a(u(t)-u^h(t), v^h) + b(v^h, p(t)-p^h(t)) = 0 \forall\, v^h \in \mathbf{V}^h \tag{4.28}$$

$$\varepsilon(p(t)-p^h(t), q^h) - b(u(t)-u^h(t), q^h) = 0 \qquad \forall\, q^h \in M^h. \tag{4.29}$$

Adding and subtracting appropriate terms and using the orthogonality condition two times, for $v^h = u^h(t)$ and $v^h = P^hu(t)$ respectively,

$$\begin{aligned} &\langle u_t(t) - u^h_t(t), u(t)-u^h(t)\rangle + a(u(t)-u^h(t), u(t)-u^h(t)) + b(u(t)-u^h(t), p(t)-p^h(t)) = \\ &\quad = \langle u_t(t)-u^h_t(t), u(t)\rangle + a(u(t)-u^h(t), u(t)) + b(u(t), p(t)-p^h(t)) \\ &\quad = \langle u_t(t)-u^h_t(t), u(t)-P^hu(t)\rangle + a(u(t)-u^h(t), u(t)-P^hu(t)) \\ &\qquad + b(u(t)-P^hu(t), p(t)-p^h(t)). \end{aligned} \tag{4.30}$$

Note that the definition of the projection implies that $P^hu(t)$ is defined for almost every t, and P^hu has the same regularity as u. Hence,

$$\begin{aligned}&\langle u_t(t)-u_t^h(t), u(t)-P^hu(t)\rangle = \langle u_t(t), u(t)-P^hu(t)\rangle\\ &= \langle u_t(t)-(P^hu)_t(t), u(t)-P^hu(t)\rangle = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u(t)-P^hu(t)\|^2_{\mathbf{L}^2(\Omega)}.\end{aligned} \tag{4.31}$$

Similarly, using (4.29) for $q^h=p^h(t)$ and for $q^h=P^h_pp(t)$

$$\begin{aligned}&\varepsilon(p(t)-p^h(t), p(t)-p^h(t)) - b(u(t)-u^h(t), p(t)-p^h(t))\\ &= \varepsilon(p(t)-p^h(t), p(t)) - b(u(t)-u^h(t), p(t))\\ &= \varepsilon(p(t)-p^h(t), p(t)-P^h_pp(t)) - b(u(t)-u^h(t), p(t)-P^h_pp(t)).\end{aligned} \tag{4.32}$$

Finally, adding (4.30)–(4.32), after using (4.31) and the coercivity condition,

$$\begin{aligned}&\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u(t)-u^h(t)\|^2_{\mathbf{L}^2(\Omega)} + \varepsilon\|p(t)-p^h(t)\|^2_{L^2(\Omega)} + C\|u(t)-u^h(t)\|^2_{\mathbf{H}^1(\Omega)}\\ &\leq \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u(t)-P^hu(t)\|^2_{\mathbf{L}^2(\Omega)} + a(u(t)-u^h(t), u(t)-P^hu(t))\\ &\quad + b(u(t)-P^hu(t), p(t)-p^h(t)) + \varepsilon(p(t)-p^h(t), p(t)-P^h_pp(t))\\ &\qquad - b(u(t)-u^h(t), p(t)-P^h_pp(t)).\end{aligned} \tag{4.33}$$

We may bound the last four terms in a standard way. Then, the usual Gronwall's lemma techniques lead to (4.26). Finally, (4.26) together with the approximation properties of the projections (Lem. 4.1), and the continuous embedding $u\in L^2(0,T;\mathbf{H}^{m+1}(\Omega))\cap H^1(0,T;\mathbf{H}^{m-1}(\Omega))\subset C(0,T;\mathbf{H}^m(\Omega))$ lead to (4.27). □

Remark 4.5. Using the projection techniques, we are able to derive error estimates on the "natural" energy norm that do not require additional approximation and regularity properties on the time derivative. For semi discrete in time and fully discrete error estimates, one may consult [10, 12] respectively.

In order to complete our estimates, we consider the approximation of time derivatives. As mentioned earlier, we are interested in deriving error estimates for the natural norms, under minimal regularity assumptions, *i.e.*, estimates for $\|u_t-u_t^h\|_{L^2(0,T;\mathbf{H}^{-1}(\Omega))}$.

Theorem 4.6. *Let $(u,p),(u^h,p^h)$ be the solutions of (4.24)-(4.25) respectively. Then,*

$$\|u_t-u_t^h\|^2_{L^2(0,T;\mathbf{H}^{-1}(\Omega))} \leq D\Big(\|u_t-Q^hu_t\|^2_{L^2(0,T;\mathbf{H}^{-1}(\Omega))} + \|u-u^h\|^2_{L^2(0,T;\mathbf{H}^1(\Omega))} + \|p-p^h\|^2_{L^2(0,T;L^2(\Omega))}\Big). \tag{4.34}$$

In addition if, $u\in L^2(0,T;\mathbf{H}^{m+1}(\Omega))\cap H^1(0,T;\mathbf{H}^{m-1}(\Omega))$, $p\in L^2(0,T;H^m(\Omega))$ $0\leq m\leq k$, and $u_0^h=P^hu_0$ then,

$$\begin{aligned}\|u_t-u_t^h\|^2_{L^2(0,T;\mathbf{H}^{-1}(\Omega))} \leq Dh^{2m}\Big(&\max\left\{\frac{1}{\varepsilon},\frac{1}{\varepsilon^2}\right\}\|u\|^2_{L^2(0,T;\mathbf{H}^{m+1}(\Omega))} + \max\left\{\frac{1}{\varepsilon},1\right\}\|p\|^2_{L^2(0,T;H^m(\Omega))}\\ &+ \max\left\{\frac{1}{\varepsilon},1\right\}\|u_t\|^2_{L^2(0,T;\mathbf{H}^{m-1}(\Omega))}\Big).\end{aligned} \tag{4.35}$$

Proof. In this proof, we make extensive use of the generalized projection Q^h following ideas and techniques similar [2, 8]. First note, that adding and subtracting appropriate terms and using the orthogonality condition:

$$\begin{aligned}&\langle u_t(t)-u_t^h(t), v\rangle = \langle u_t(t)-u_t^h(t), v-v^h\rangle + \langle u_t(t)-u_t^h(t), v^h\rangle\\ &= \langle u_t(t)-u_t^h(t), v-v^h\rangle - a(u(t)-u^h(t), v^h) - b(v^h, p(t)-p^h(t)).\end{aligned}$$

Setting $v^h = Q^h v$ into the above equality,

$$\langle u_t(t) - u_t^h(t), v\rangle = \langle u_t(t) - u_t^h(t), v - Q^h v\rangle - a(u(t) - u^h(t), Q^h v) - b(Q^h v, p(t) - p^h(t)).$$

Note that the definition of the Q^h projection imply that

$$\langle u_t(t) - u_t^h(t), v - Q^h v\rangle = \langle u_t(t), v - Q^h v\rangle = \langle u_t(t) - Q^h u_t(t), v - Q^h v\rangle.$$

Taking the supremum over $v \in \mathbf{H}_0^1(\Omega)$, $\|v\|_{\mathbf{H}^1(\Omega)} = 1$ and noting that $\|P^h v - v\|_{\mathbf{H}^1(\Omega)} \leq C\|v\|_{\mathbf{H}^1(\Omega)}$, $\|P^h v\|_{\mathbf{H}^1(\Omega)} \leq C\|v\|_{\mathbf{H}^1(\Omega)}$, and $P^h v = Q^h v \quad \forall\, v \in \mathbf{L}^2(\Omega)$ we obtain,

$$\|u_t(t) - u_t^h(t)\|_{\mathbf{H}^{-1}(\Omega)} \leq C\Big(\|u_t(t) - Q^h u_t(t)\|_{\mathbf{H}^{-1}(\Omega)} + \|u(t) - u^h(t)\|_{\mathbf{H}^1(\Omega)} + \|p(t) - p^h(t)\|_{L^2(\Omega)}\Big).$$

The above inequality leads to (4.34). Approximation properties of the projections, (4.27) and (4.34) imply (4.35). □

4.4. Semidiscrete error estimates for the penalized optimality system

Finally, using ideas and theorems established in previous sections, we prove semidiscrete error estimates for the penalized optimality system. The main ingredient will be the BRR theory. First, we rewrite the optimality system in the form

$$(u, p, \mu, r) + \mathcal{T}\mathcal{G}(u, p, \mu, r) = 0.$$

$\mathcal{T}$ is the operator that contains the "model" penalized Stokes problem, with arbitrary given data satisfying minimal regularity assumptions (similar to Sect. 4.3), while $\mathcal{G}$ is the mapping that contains all *coupled* terms. For that purpose, we set

$$X \equiv L^2(0, T; \mathbf{H}_0^1(\Omega)) \cap H^1(0, T; \mathbf{H}^{-1}(\Omega)), M \equiv L^2(0, T; L^2(\Omega))$$

$$\mathcal{X} = X \times M \times X \times M,$$

with the norm,

$$\begin{aligned}\|(u, p, \mu, r)\|_{\mathcal{X}}^2 = \|u\|^2_{L^2(0,T;\mathbf{H}^1(\Omega))} + \|u_t\|^2_{L^2(0,T;\mathbf{H}^{-1}(\Omega))} + \|p\|^2_{L^2(0,T;L^2(\Omega))} \\ + \|\mu\|^2_{L^2(0,T;\mathbf{H}^1(\Omega))} + \|\mu_t\|^2_{L^2(O,T;\mathbf{H}^{-1}(\Omega))} + \|r\|^2_{L^2(0,T;L^2(\Omega))}\end{aligned}$$

and

$$Y \equiv L^2(0, T; \mathbf{H}^{-1}(\Omega)) \times \mathbf{L}^2(\Omega), \qquad \mathcal{Y} = Y \times Y$$

with the norm,

$$\|(f_1, u_1, f_2, \mu_1)\|^2_{\mathcal{Y}} = \|f_1\|^2_{L^2(0,T;\mathbf{H}^{-1}(\Omega))} + \|u_1\|^2_{\mathbf{L}^2(\Omega)} + \|f_2\|^2_{L^2(0,T;\mathbf{H}^{-1}(\Omega))} + \|\mu_1\|^2_{\mathbf{L}^2(\Omega)}.$$

Then, we define the linear operator $\mathcal{T} : \mathcal{Y} \to \mathcal{X}$ for every $(f_1, u_1, f_2, \mu_1) \in \mathcal{Y}$ as follows: $\mathcal{T}(f_1, u_1, f_2, \mu_1) = (u, p, \mu, r) \in \mathcal{X}$ if and only if

$$\begin{cases} \langle u_t, v\rangle + a(u, v) + b(v, p) = \langle f_1, v\rangle & \forall\, v \in \mathbf{H}_0^1(\Omega) \\ \varepsilon(p, q) - b(u, q) = 0 & \forall\, q \in L^2(\Omega) \\ (u(0, x), z) = (u_1, z) & \forall\, z \in \mathbf{L}^2(\Omega) \end{cases}$$

$$\begin{cases} -\langle \mu_t, v\rangle + a(\mu, v) + b(v, r) = \langle f_2, v\rangle & \forall\, v \in \mathbf{H}_0^1(\Omega) \\ \varepsilon(r, q) - b(\mu, q) = 0 & \forall\, q \in L^2(\Omega) \\ (\mu(T, x), z) = (\mu_1, z) & \forall\, z \in \mathbf{L}^2(\Omega). \end{cases}$$

Furthermore, we define by

$$X^h \equiv H^1(0,T;\mathbf{V}^h) \times L^2(0,T;M^h), \quad \mathcal{X}^h \equiv X^h \times X^h$$

and by $\mathcal{T}^h : \mathcal{Y} \to \mathcal{X}^h$ the semidiscrete solution operator, *i.e.*, $\mathcal{T}^h(f_1, u_1, f_2, \mu_1) = (u^h, p^h, \mu^h, r^h)$ if and only if

$$\begin{cases} \langle u^h_t, v^h\rangle + a(u^h, v^h) + b(v^h, p^h) = \langle f_1, v^h\rangle \qquad \forall\, v^h \in \mathbf{V}^h \\ \varepsilon(p^h, q^h) - b(u^h, q^h) = 0 \qquad \forall\, q^h \in M^h \\ (u^h(0,x), v^h) = (u_1, v^h) \qquad \forall\, v^h \in \mathbf{V}^h \end{cases}$$

$$\begin{cases} -\langle \mu^h_t, v^h\rangle + a(\mu^h, v^h) + b(v^h, r^h) = \langle f_2, v^h\rangle \qquad \forall\, v^h \in \mathbf{V}^h \\ \varepsilon(r^h, q^h) - b(\mu^h, q^h) = 0 \qquad \forall\, q^h \in M^h \\ (\mu^h(T,x), v^h) = (\mu_1, v^h) \qquad \forall\, v^h \in \mathbf{V}^h. \end{cases}$$

Moreover, we denote by $\mathcal{G} : \mathcal{X} \to \mathcal{Y}$ the mapping containing all *coupled* terms, *i.e.*, $\mathcal{G}(u, p, \mu, r) = (f_1, u_1, f_2, \mu_1)$ if and only if

$$\begin{aligned} &\langle f_1, v\rangle = -\left(-\frac{1}{\beta}(\mu, v) + \langle g, v\rangle\right) \quad \forall\, v \in \mathbf{H}^1_0(\Omega) \\ &(u_1, z) = (u_0, z) \qquad \forall\, z \in \mathbf{L}^2(\Omega) \\ &(f_2, v) = -\alpha(u - U, v) \quad \forall\, v \in \mathbf{H}^1_0(\Omega) \\ &(\mu_1, z) = 0 \qquad \forall\, z \in \mathbf{L}^2(\Omega). \end{aligned}$$

Clearly, the continuous optimality system is equivalent to

$$(u, p, \mu, r) + \mathcal{T}\mathcal{G}(u, p, \mu, r) = 0,$$

and the semidiscrete optimality system is equivalent to

$$(u^h, p^h, \mu^h, r^h) + \mathcal{T}^h\mathcal{G}(u^h, p^h, \mu^h, r^h) = 0.$$

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

Theorem 4.7. *Assume that $\{(u, p, \mu, r) \in \mathcal{X}\}$ are nonsingular solutions of the optimality system. Furthermore, let V^h, M^h satisfy the approximations properties (4.5)–(4.8). Then, there exists a neighborhood of the origin in $\mathcal{X}$, and for $h \le h_0$ small enough, a unique solution $\{(u^h, p^h, \mu^h, r^h) \in \mathcal{X}\}$ of solutions of the discrete optimality system such that for fixed $\varepsilon > 0$,*

$$\|(u, p, \mu, r) - (u^h, p^h, \mu^h, r^h)\|^2_{\mathcal{X}} \to 0, \quad \text{as } h \to 0. \tag{4.36}$$

In addition, if $u, \mu \in L^2(0,T;\mathbf{H}^{m+1}(\Omega)) \cap H^1(0,T;\mathbf{H}^{m-1}(\Omega))$, $p, r \in L^2(0,T;H^m(\Omega))$ for $0 \le m \le k$, and $u^h_0 = P^h u_0$ then,

$$\begin{aligned} \|(u, p, \mu, r) - (u^h, p^h, \mu^h, r^h)\|^2_{\mathcal{X}} \le Ch^{2m}\Big(&\max\left\{\frac{1}{\varepsilon}, \frac{1}{\varepsilon^2}\right\}\|u\|^2_{L^2(0,T;\mathbf{H}^{m+1}(\Omega))} \\ &+ \max\left\{\frac{1}{\varepsilon}, 1\right\}\|u_t\|^2_{L^2(0,T;\mathbf{H}^{m-1}(\Omega))} + \max\left\{\frac{1}{\varepsilon}, \frac{1}{\varepsilon^2}\right\}\|\mu\|^2_{L^2(0,T;\mathbf{H}^{m+1}(\Omega))} \\ &+ \max\left\{\frac{1}{\varepsilon}, 1\right\}\|\mu_t\|^2_{L^2(0,T;\mathbf{H}^{m-1}(\Omega))} + \max\left\{\frac{1}{\varepsilon}, 1\right\}\|p\|^2_{L^2(0,T;H^m(\Omega))} \\ &+ \max\left\{\frac{1}{\varepsilon}, 1\right\}\|r\|^2_{L^2(0,T;H^m(\Omega))}\Big). \end{aligned} \tag{4.37}$$

Proof. First note that $\mathcal{G}$ is a smooth polynomial map from $\mathcal{X}$ into $\mathcal{Y}$ and $D^2\mathcal{G}$ is bounded on all bounded sets of $\mathcal{X}$. Furthermore, Theorems 4.4–4.6 imply that

$$\|(\mathcal{T}-\mathcal{T}^h)(f_1,u_1,f_2,\mu_1)\|_{\mathcal{X}} \to 0 \text{ as } h\to 0$$

for all $(f_1,u_1,f_2,\mu_1)\in\mathcal{Y}$. Next we need to satisfy the condition for the derivative $D\mathcal{G}$. $D\mathcal{G}(u,p,\mu,r)\cdot(u,p,\mu,r)=(\tilde{\zeta},\tilde{u}_1,\tilde{\eta},\tilde{\mu}_1)$ if and only if,

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

Now for sufficiently small $\delta>0$, set $Z=L^2(0,T;\mathbf{L}^2(\Omega))\cap H^1(0,T;\mathbf{H}^{-1}(\Omega))\times\mathbf{H}^\delta(\Omega)$, and $\mathcal{Z}=Z\times Z$. Note that $\mathbf{L}^2(\Omega)\subset\mathbf{H}^{-1}(\Omega)$ with compact embedding, so using [11], Theorem 2.1, p. 271, we obtain that $L^2(0,T;\mathbf{L}^2(\Omega))\cap H^1(0,T;\mathbf{H}^{-1}(\Omega))\subset L^2(0,T;\mathbf{H}^{-1}(\Omega))$ with compact embedding. Hence $\mathcal{Z}\subset\mathcal{Y}$ with compact embedding. Moreover, it is easy to see that $DG(u,p,\mu,r)\in\mathcal{Z}$ due to regularity properties of u,μ. Therefore, Theorems 4.2–4.4–4.6 imply (4.36)–(4.37). □

5. Convergence of optimality systems as $\varepsilon\to 0$

Finally, in this section we examine the convergence of the optimality systems (4.2)–(4.4) and (4.17)–(4.18) as $\varepsilon\to 0$. First note that the optimality system corresponding to problem (P_1) has the following form (see [7], Sect. 2.4):

$$\begin{cases}
\langle u_t,v\rangle + a(u,v)+b(v,p)=\langle -\frac{1}{\beta}\mu+g,v\rangle \qquad \forall\, v\in\mathbf{H}_0^1(\Omega)\\
b(u,q)=0 \qquad \forall\, q\in L^2(\Omega)\\
(u(0,x)-u_0,z)=0 \qquad \forall\, z\in\mathbf{L}^2(\Omega)
\end{cases} \tag{5.1}$$

$$\begin{cases}
-\langle \mu_t,v\rangle + a(\mu,v)+b(v,r)=\alpha(u-U,v) \qquad \forall\, v\in\mathbf{H}_0^1(\Omega)\\
b(\mu,q)=0 \qquad \forall\, q\in L^2(\Omega)\\
(\mu(T),z)=0 \qquad \forall\, z\in\mathbf{L}^2(\Omega).
\end{cases} \tag{5.2}$$

The scope of the next theorem is to obtain an actual rate of convergence as $\varepsilon\to 0$ and it is analogous to [10], Lemma 3.1.

Theorem 5.1. *Suppose that $(u^\varepsilon,p^\varepsilon,\mu^\varepsilon,r^\varepsilon)$, (u,p,μ,r) are the solutions of optimality systems (4.2)–(4.4) and (5.1)–(5.2) respectively. In addition, suppose that $p\in L^2(0,T;L_0^2(\Omega))$. Then, the following estimate holds:*

$$\begin{aligned}
&\|u^\varepsilon-u\|^2_{L^2(0,T;\mathbf{H}_0^1(\Omega))}+\|\mu^\varepsilon-\mu\|^2_{L^2(0,T;\mathbf{H}_0^1(\Omega))}\\
&\qquad+\varepsilon\|p^\varepsilon-p\|^2_{L^2(0,T;L^2(\Omega))}+\varepsilon\|r^\varepsilon-r\|^2_{L^2(0,T;L^2(\Omega))}\le\mathcal{O}(\varepsilon),\\
&\|u_t^\varepsilon-u_t\|^2_{L^2(0,T;V(\Omega)^*)}+\|\mu_t^\varepsilon-\mu_t\|^2_{L^2(0,T;V(\Omega)^*)}\le\mathcal{O}(\varepsilon).
\end{aligned} \tag{5.3}$$

Proof. Subtracting (5.1) from (4.4) and (5.2) from (4.2) we obtain:

$$\begin{cases}
\langle u_t^\varepsilon-u_t,v\rangle + a(u^\varepsilon-u,v)+b(v,p^\varepsilon-p)=-\dfrac{1}{\beta}(\mu^\varepsilon-\mu,v) \quad \forall\, v\in\mathbf{H}_0^1(\Omega)\\
\varepsilon(p^\varepsilon,q)-b(u^\varepsilon-u,q)=0 \qquad \forall\, q\in L^2(\Omega)\\
(u^\varepsilon(0,x)-u_0,z)=0 \qquad \forall\, z\in\mathbf{L}^2(\Omega)
\end{cases} \tag{5.4}$$

$$\begin{cases} -\langle \mu^\varepsilon_t - \mu_t, v\rangle + a(\mu^\varepsilon - \mu, v) + b(v, r^\varepsilon - r) = \alpha(u^\varepsilon - u, v) & \forall\, v \in \mathbf{H}^1_0(\Omega) \\ \varepsilon(r^\varepsilon, q) - b(\mu^\varepsilon - \mu, q) = 0 & \forall\, q \in L^2(\Omega) \\ (\mu^\varepsilon(T) - \mu(T), z) = 0 & \forall\, z \in \mathbf{L}^2(\Omega). \end{cases} \tag{5.5}$$

Setting $v = \alpha(u^\varepsilon - u), q = \alpha(p^\varepsilon - p)$ into (5.4) and $v = \frac{1}{\beta}(\mu^\varepsilon - \mu), q = \frac{1}{\beta}(r^\varepsilon - r)$ into (5.5) and adding the corresponding equalities we arrive to:

$$\begin{aligned} &\frac{\alpha}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u^\varepsilon - u\|^2_{\mathbf{L}^2(\Omega)} - \frac{1}{2\beta}\frac{\mathrm{d}}{\mathrm{d}t}\|\mu^\varepsilon - \mu\|^2_{\mathbf{L}^2(\Omega)} + \alpha a(u^\varepsilon - u, u^\varepsilon - u) \\ &\quad + \frac{1}{\beta}a(\mu^\varepsilon - \mu, \mu^\varepsilon - \mu) + \alpha\varepsilon(p^\varepsilon, p^\varepsilon - p) + \frac{\varepsilon}{\beta}(r^\varepsilon, r^\varepsilon - r) = 0. \end{aligned} \tag{5.6}$$

Hence, adding and subtracting appropriate terms into (5.6),

$$\begin{aligned} &\frac{\alpha}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u^\varepsilon - u\|^2_{\mathbf{L}^2(\Omega)} - \frac{1}{2\beta}\frac{\mathrm{d}}{\mathrm{d}t}\|\mu^\varepsilon - \mu\|^2_{\mathbf{L}^2(\Omega)} + \alpha\|u^\varepsilon - u\|^2_{\mathbf{H}^1_0(\Omega)} + \frac{1}{\beta}\|\mu^\varepsilon - \mu\|^2_{\mathbf{H}^1_0(\Omega)} \\ &\quad + \frac{\varepsilon\alpha}{2}\|p^\varepsilon - p\|^2_{L^2(\Omega)} + \frac{\varepsilon}{2\beta}\|r^\varepsilon - r\|^2_{L^2(\Omega)} \le \frac{\varepsilon\alpha}{2}\|p\|^2_{L^2(\Omega)} + \frac{\varepsilon}{2\beta}\|r\|^2_{L^2(\Omega)}. \end{aligned}$$

Integrating the above inequality in time we obtain the desired estimate. □

Remark 5.2. Note that a standard regularity argument for system (5.2) (see [11], Prop. 1.2, p. 267) and the regularity property $u - U \in L^2(0,T;\mathbf{L}^2(\Omega))$, we obtain that $\mu \in L^2(0,T;\mathbf{H}^2(\Omega) \cap \mathbf{H}^1_0(\Omega)) \cap H^1(0,T;\mathbf{L}^2(\Omega)), r \in L^2(0,T;H^1(\Omega))$. Assuming $g \in L^2(0,T;\mathbf{L}^2(\Omega)), u_0 \in V(\Omega)$ similar regularity results holds for (u,p).

We are now ready to present some results regarding the convergence of the solution of the discrete optimality system (4.17)–(4.18) of problem (P_2) to the solution of the discrete optimality system of problem (P_1):

$$\begin{cases} \langle u^h_t, v^h\rangle + a(u^h, v^h) + b(v^h, p^h) = \langle -\frac{1}{\beta}\mu^h + g, v^h\rangle & \forall\, v^h \in \mathbf{V}^h \\ b(u^h, q^h) = 0 & \forall\, q^h \in M^h \\ (u^h(0,x) - u^h_0, z^h) = 0 & \forall\, z^h \in \mathbf{V}^h \end{cases} \tag{5.7}$$

$$\begin{cases} -\langle \mu^h_t, v^h\rangle + a(\mu^h, v^h) + b(v^h, r^h) = \alpha(u^h - U, v^h) & \forall\, v^h \in \mathbf{V}^h \\ b(\mu^h, q^h) = 0 & \forall\, q^h \in M^h \\ (\mu^h(T), z^h) = 0 & \forall\, z^h \in \mathbf{V}^h. \end{cases} \tag{5.8}$$

Theorem 5.3. *Suppose that* $(u^{\varepsilon,h}, p^{\varepsilon,h}, \mu^{\varepsilon,h}, r^{\varepsilon,h})$, (u^h, p^h, μ^h, r^h) *are the solutions of optimality systems (4.17)–(4.18) and (5.7)–(5.8) respectively. Then,*

$$\begin{aligned} &\|u^{\varepsilon,h} - u^h\|^2_{L^2(0,T;\mathbf{H}^1_0(\Omega))} + \|\mu^{\varepsilon,h} - \mu^h\|^2_{L^2(0,T;\mathbf{H}^1_0(\Omega))} \\ &\quad + \varepsilon\|p^{\varepsilon,h} - p^h\|^2_{L^2(0,T;L^2(\Omega))} + \varepsilon\|r^{\varepsilon,h} - r^h\|^2_{L^2(0,T;L^2(\Omega))} \le \mathcal{O}(\varepsilon) \\ &\varepsilon\|u^{\varepsilon,h}_t - u^h_t\|^2_{L^2(0,T;\mathbf{H}^{-1}(\Omega))} + \varepsilon\|\mu^{\varepsilon,h}_t - \mu^h_t\|^2_{L^2(0,T;\mathbf{H}^{-1}(\Omega))} \le \mathcal{O}(\varepsilon). \end{aligned} \tag{5.9}$$

Proof. The first inequality follows similar to Theorem 5.1. For the estimate on the time derivative, we subtract (5.7) from (4.17) to obtain:

$$\langle u^{\varepsilon,h}_t - u^h_t, v^h\rangle + a(u^{\varepsilon,h} - u^h, v^h) + b(v^h, p^{\varepsilon,h} - p^h) = -\frac{1}{\beta}(\mu^{\varepsilon,h} - \mu^h, v^h) \qquad \forall\, v^h \in \mathbf{V}^h.$$

Then $\forall\, v \in \mathbf{H}^{-1}(\Omega)$,

$$\begin{aligned} \langle u^{\varepsilon,h}_t - u^h_t, v\rangle &= \langle u^{\varepsilon,h}_t - u^h_t, v - v^h\rangle + \langle u^{\varepsilon,h}_t - u^h_t, v^h\rangle = \langle u^{\varepsilon,h}_t - u^h_t, v - v^h\rangle \\ &\quad - \Big(a(u^{\varepsilon,h} - u^h, v^h) + b(v^h, p^{\varepsilon,h} - p^h) + \frac{1}{\beta}(\mu^{\varepsilon,h} - \mu^h, v^h)\Big). \end{aligned}$$

Choosing $v^h = Q^h v$ (see Sect. 4.1) and using the definition of the projection, we finally arrive to

$$\langle u_t^{\varepsilon,h} - u_t^h, v\rangle = -a(u^{\varepsilon,h} - u^h, Q^h v) - b(Q^h v, p^{\varepsilon,h} - p^h) - \frac{1}{\beta}(\mu^{\varepsilon,h} - \mu^h, Q^h v).$$

Taking the supremum over $v \in \mathbf{H}_0^1(\Omega), \|v\|_{\mathbf{H}^1(\Omega)} = 1$, and using similar techniques as in Theorem 4.6, together with (5.9) and Theorem 5.1 we obtain the desired estimate. For the estimate on the $\mu^{\varepsilon,h} - \mu^h$ we follow exactly the same techniques. □

Finally combining Theorems 5.1–5.3 together with results of Section 4, we may derive error estimates for the discrete optimality system of problem (P_1). To simplify the presentation we define the following norm.

$$|||(u,\mu)||| \equiv \|u\|_{L^2(0,T;\mathbf{H}^1(\Omega))} + \|u_t\|_{L^2(0,T;\mathbf{H}^{-1}(\Omega))} + \|\mu\|_{L^2(0,T;\mathbf{H}^1(\Omega))} + \|\mu_t\|_{L^2(0,T;\mathbf{H}^{-1}(\Omega))}.$$

Theorem 5.4. *Let $(u,p,\mu,r),(u^h,p^h,\mu^h,r^h)$ be the solutions of (5.1)–(5.2) and (5.7)–(5.8). Suppose also that $p \in L^2(0,T;L_0^2(\Omega))$. Then the following estimate holds:*

$$\begin{aligned}\|u - u^h\|^2_{L^2(0,T;\mathbf{H}^1(\Omega))} + \|\mu - \mu^h\|^2_{L^2(0,T;\mathbf{H}^1(\Omega))} \le \mathcal{O}(\varepsilon) \\ + \|u^{\varepsilon,h} - u^\varepsilon\|^2_{L^2(0,T;\mathbf{H}^1(\Omega))} + \|\mu^{\varepsilon,h} - \mu^\varepsilon\|^2_{L^2(0,T;\mathbf{H}^1(\Omega))}.\end{aligned} \tag{5.10}$$

In addition, if $(u^\varepsilon,p^\varepsilon),(\mu^\varepsilon,r^\varepsilon),(u,p),(\mu,r) \in L^2(0,T;\mathbf{H}^{m+1}(\Omega)) \cap H^1(0,T;\mathbf{H}^{m-1}(\Omega)) \times L^2(0,T;H^m(\Omega))$ and $u_0^h = P^h u_0$ then,

$$\begin{aligned}|||(u,\mu) - (u^h,\mu^h)|||^2 \le \mathcal{O}(\varepsilon) + Ch^{2m}\Big(&\max\left\{\frac{1}{\varepsilon},\frac{1}{\varepsilon^2}\right\}\|u^\varepsilon\|^2_{L^2(0,T;\mathbf{H}^{m+1}(\Omega))} \\ &+ \max\left\{\frac{1}{\varepsilon},1\right\}\|u_t^\varepsilon\|^2_{L^2(0,T;\mathbf{H}^{m-1}(\Omega))} + \max\left\{\frac{1}{\varepsilon},\frac{1}{\varepsilon^2}\right\}\|\mu^\varepsilon\|^2_{L^2(0,T;\mathbf{H}^{m+1}(\Omega))} \\ &+ \max\left\{\frac{1}{\varepsilon},1\right\}\|\mu_t^\varepsilon\|^2_{L^2(0,T;\mathbf{H}^{m-1}(\Omega))} + \max\left\{\frac{1}{\varepsilon},1\right\}\|p^\varepsilon\|^2_{L^2(0,T;H^m(\Omega))} \\ &+ \max\left\{\frac{1}{\varepsilon},1\right\}\|r^\varepsilon\|^2_{L^2(0,T;H^m(\Omega))} + \|u_t\|^2_{L^2(0,T;\mathbf{H}^{m-1}(\Omega))} \\ &+ \|\mu_t\|^2_{L^2(0,T;\mathbf{H}^{m-1}(\Omega))} + \|p\|^2_{L^2(0,T;H^m(\Omega))} + \|r\|^2_{L^2(0,T;H^m(\Omega))}\Big).\end{aligned} \tag{5.11}$$

Moreover, if $\mathbf{V}^h, M^h$ satisfy the discrete inf-sup condition (see e.g. [4]) then

$$\|p - p^h\|_{L^2(0,T;L^2(\Omega))} + \|r - r^h\|_{L^2(0,T;L^2(\Omega))} \le C|||(u,\mu) - (u^h,\mu^h)|||. \tag{5.12}$$

Proof. The first inequality follows directly from Theorems 4.7, 5.1, 5.2 and triangle inequality. For the second estimate, we define the "discrete divergence free" $\mathbf{H}^{-1}(\Omega)$ projection (see [8]), by

$$\mathcal{Q}^h v \in Z^h, \qquad (\mathcal{Q}^h v, v^h) = \langle v, v^h\rangle \qquad \forall\, v^h \in Z^h$$

where $Z^h \equiv \{u^h \in \mathbf{V}^h : b(u^h,q^h) = 0, \forall\, q^h \in M^h\}$. Subtracting (5.1) from (5.7) we obtain, $\forall\, v^h \in \mathbf{V^h}$, and for a.e. t,

$$\langle u_t^h(t) - u_t(t), v^h\rangle + a(u^h(t) - u(t), v^h) + b(v^h, p^h(t) - p(t)) = -\frac{1}{\beta}(\mu^h(t) - \mu(t), v^h). \tag{5.13}$$

Therefore,

$$\begin{aligned}\langle u_t^h(t) - u_t(t), v\rangle = \langle u_t^h(t) - u_t(t), v - v^h\rangle + \langle u_t^h(t) - u_t(t), v^h\rangle = \langle u_t^h(t) - u_t(t), v - v^h\rangle \\ -\Big(a(u^h(t) - u(t), v^h) + b(v^h, p^h(t) - p(t)) - \frac{1}{\beta}(\mu^h(t) - \mu(t), v^h)\Big).\end{aligned}$$

Setting $v^h = \mathcal{Q}^h v$, taking into account that $b(\mathcal{Q}^h v, p^h - p) = b(\mathcal{Q}^h v, q^h - p)$, and $\langle u_t^h - u_t, v - \mathcal{Q}^h v\rangle = \langle \mathcal{Q}^h u_t - u_t, v - v^h\rangle$, we obtain similar to Theorem 4.6,

$$\|u_t^h - u_t\|_{L^2(0,T;\mathbf{H}^{-1}(\Omega))} \le C\big(\|\mathcal{Q}^h u_t - u_t\|_{L^2(0,T;\mathbf{H}^{-1}(\Omega))} + \|u^h - u\|_{L^2(0,T;\mathbf{H}^1(\Omega))} + \|\mu^h - \mu\|_{L^2(0,T;\mathbf{L}^2(\Omega))} + \|q^h - p\|_{L^2(0,T;L^2(\Omega))}\big)$$

where $q^h \in M^h$ is an arbitrary element. Therefore the estimate follows directly from the above inequality, (5.10) and Theorem 4.7. Finally, using the discrete inf-sup condition, and (5.13),

$$\|p^h - q^h\|_{L^2(\Omega)} \le \sup_{v^h \in \mathbf{V}^h} \frac{|b(v^h, p^h(t) - p(t)) + b(v^h, p(t) - q^h(t))|}{\|v^h\|_{\mathbf{H}^1(\Omega)}}$$
$$\le C\Big(\|p(t) - q^h(t)\|_{L^2(\Omega)} + \|u(t) - u^h(t)\|_{\mathbf{H}^1(\Omega)} + \|\mu(t) - \mu^h(t)\|_{\mathbf{L}^2(\Omega)} + \|u_t(t) - u_t^h(t)\|_{\mathbf{H}^{-1}(\Omega)}\Big).$$

An analogous proof, establishes estimates for $\mu_t^h - \mu_t, r^h - r$. □

Remark 5.5. Note that if $g \in L^2(0,T;\mathbf{L}^2(\Omega)), u_0 \in V(\Omega)$ then $\|u^\varepsilon\|_{L^2(0,T;\mathbf{H}^2(\Omega))} \le C$, independent of ε (see [10], p. 389), since $\|f^\varepsilon\|_{L^2(0,T;\mathbf{L}^2(\Omega))} \le C$. Then, taking the orthogonal projection $P_{W(\Omega)}$ into (1.3), we easily obtain that $\|u_t^\varepsilon\|_{L^2(0,T;\mathbf{L}^2(\Omega))}$. Returning back to (1.3) we also obtain that $\|p^\varepsilon\|_{L^2(0,T;H^1(\Omega))} \le C$. Similar results also hold for the adjoint system. Therefore, estimate (5.11) reads as: $|||(u,\mu) - (u^h,\mu^h)|||^2 \le \mathcal{O}(\varepsilon) + \frac{h^2}{\varepsilon^2}$.

Remark 5.6. The key elements of the above proof are contained in [8].

Acknowledgements. The author would like to thank the referees for many useful comments.

References

[1] R. Adams, *Sobolev Spaces*. Academic Press, New York (1975).
[2] K. Chrysafinos and L.S. Hou, Error estimates for semidiscrete finite element approximations of linear and semilinear parabolic equations under minimal regularity assumptions. *SIAM J. Numer. Anal.* **40** (2002) 282-306.
[3] A. Fursikov, *Optimal control of distributed systems. Theories and Applications*. AMS Providence (2000).
[4] V. Girault and P.-A. Raviart, *Finite Element Methods for Navier-Stokes*. Springer-Verlag, New York (1986).
[5] 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. *ESAIM: M2AN* **25** (1991) 711-748.
[6] M.D. Gunzburger and S. Manservisi, The velocity tracking problem for Navier-Stokes flows with bounded distributed control. *SIAM J. Control Optim.* **37** (2000) 1913-1945.
[7] 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.* **37** (2000) 1481-1512.
[8] L.S. Hou, Error estimates for semidiscrete finite element approximation of the Stokes equations under minimal regularity assumptions. *J. Sci. Comput.* **16** (2001) 287-317.
[9] 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 Optim.* **36** (1998) 1795-1814.
[10] Jie Shen, On error estimates of the penalty method for unsteady Navier-Stokes equations. *SIAM J. Numer. Anal.* **32** (1995) 386-403.
[11] R. Temam, *Navier-Stokes equations*. North-Holland, Amsterdam (1979).
[12] R. Temam, Une méthode d'approximation de la solution des équations de Navier-Stokes. *Bull. Soc. Math. France* **98** (1968) 115-152.
[13] B.A. Ton, Optimal shape control problems for the Navier-Stokes equations. *SIAM J. Control Optim.* **41** (2003) 1733-1747.