Analysis and finite element approximations for distributed optimal control problems for implicit parabolic equations

Konstantinos Chrysafinos *

Abstract

This work concerns analysis and error estimates for optimal control problems related to implicit parabolic equations. The minimization of the tracking functional subject to implicit parabolic equations is examined. Existence of an optimal solution is proved and an optimality system of equations is derived. Semi-discrete (in space) error estimates for the finite element approximations of the optimality system are presented. These estimates are symmetric and applicable for higher order discretizations. Finally, fully-discrete error estimates of arbitrarily high-order are presented based on a discontinuous Galerkin (in time) and conforming (in space) scheme. Two examples related to the Lagrangian-moving mesh Galerkin formulation for the convection diffusion equation are described.

Keywords: Error estimates, Finite Element Methods, Distributed Optimal Control, Implicit Parabolic Equations, Convection-Diffusion Equations, Lagrangian Coordinates, Moving Meshes.

Mathematics Subject Classification: 65M60, 65M15, 49J20, 35B37.

1 Introduction

We consider an optimal control problem related to implicit parabolic equations of the form

$$(M(t)u)_t + A(t)u = F(t), \qquad u|_{\partial\Omega} = 0, \qquad u(0) = u_0. \tag{1.1}$$

Here, Ω is an open, bounded domain of $\mathbb{R}^d$, $d = 2, 3$, with Lipschitz boundary $\partial\Omega$. In the above context, H, U are Hilbert spaces related to the standard pivot construction $U \subset H \approx H^* \subset U^*$, with continuous and dense embedding (say e.g. $H = L^2(\Omega)$, $U = H_0^1(\Omega)$), U^* denotes the dual of U, $M(.) : H \to H$ is a self-adjoint positive definite operator, $A(.) : U \to U^*$ is a linear and continuous map, $F(.) \in U^*$ and $u_0 \in H$.

The main distinction between equation (1.1) and standard parabolic equations is that the time-derivative of the solution is not given explicitly. However, under suitable assumptions on the operators $M(.), A(.)$ the above equations are typically equivalent to "regular" equations of the form

$$M(t)u_t + \tilde{A}(t)u = F(t), \qquad u|_{\partial\Omega} = 0, \qquad u(0) = u_0. \tag{1.2}$$

The equivalence of problems (1.1)-(1.2) is studied in detail into the books of [26, 27] (see also references within).

There are many physical examples of implicit parabolic equations (see e.g. [16, 26, 27]), including several examples related to degenerate parabolic equations. Classical parabolic equations also take the form of (1.1) when a time-dependent change of variables is applied. Typical examples are the diffusion on surfaces which are in motion, and the Lagrange or characteristic Galerkin formulation

*Department of Mathematics, National Technical University of Athens, Zografou Campus, Athens 15780, Greece. Email Address:chrysafinos@math.ntua.gr

of the convection diffusion equation (see e.g. [5, 7, 18]), which is the main motivation for studying approximations of the above type of equations.

The natural setting for implicit parabolic equations, involves time-dependent norms, often called Hilbert scales (see e.g. [26, 27]). We will be using spaces of the form $H(t) = (H, \|.\|_{H(t)})$, $U(t) = (U, \|.\|_{U(t)})$. Here, $\|.\|_{H(t)}$, $\|.\|_{U(t)}$ denote time-dependent norms, and in most cases can be viewed as "weighted" norms of $L^2(\Omega)$ and $H^1_0(\Omega)$ respectively. The temporal regularity of functions with values in $H(.), U(.), U^*(.)$ are denoted in a standard fashion, i.e., $L^2[0,T;U(.)]$, $H^1[0.T;U^*(.)]$, $L^2[0,T;H(.)]$ etc.

The optimal control problem considered here is to minimize a tracking type functional,

$$K(u,g) = (1/2)\int_0^T \|u-z\|^2_{H(.)}dt + (\alpha/2)\int_0^T \|g\|^2_{H(.)}dt \tag{1.3}$$

subject to equation (1.1), with $F(.) = f(.) + g(.)$, i.e., minimize the functional (1.3) subject to,

$$(M(t)u)_t + A(t)u = f + g, \qquad u|_{\partial\Omega} = 0, \qquad u(0) = u_0. \tag{1.4}$$

Here g denotes the control variable, z is the target, and f, u_0 are given forcing and initial data terms. The physical meaning of the tracking optimal control problem under consideration, is to influence the behavior of the system in order to match the solution u of equation (1.4) to a given target z, by using a control function g which acts as a distributed body force. The cost (objective functional) is a quadratic functional which measures the distance between the solution u and the target z, while $\alpha > 0$ can be viewed as a penalty parameter. The second term of (1.3) is used to obtain a bounded control function.

There is an abundant literature concerning the analysis of various optimal control problems having states constrained to evolutionary PDE's. We refer the reader to the books of [9, 11, 16, 17, 24] and the references within, for various theoretical and numerical aspects of distributed optimal control problems. However, there are only few results concerning error estimates for finite element approximations of related optimal control problems. Fully-discrete estimates for a distributed optimal control problem related to the heat equation were given in [23, 25]. In [29], a fully discretized optimal control problem is defined and rigorously analyzed in the context of a general state constrained convex control problem, related to linear parabolic PDE's with possible non-selfadjoint elliptic part. In an earlier work [22], a distributed optimal control problem related to a quasilinear parabolic PDE was studied. Some results related to a-posteriori analysis of optimal control problems constrained to linear parabolic PDE's are developed in [20, 19]. Results related to other type of controls, in particular Neumann and Robin type of controls, are also applicable in case of distributed controls. In [31] a Neumann boundary control is used to minimize the terminal-state tracking functional constrained to linear homogeneous parabolic PDE's, while in [21], a variety of estimates for Neumann boundary control problem having states constrained to linear homogeneous parabolic PDE's are shown. In [15] (see also [16]), a semigroup approach is developed to study various optimal control problems, having states constrained to linear homogeneous parabolic PDE's, and error estimates are presented for finite element approximations. Finally, error estimates of arbitrary order for the semi-discrete approximation of Robin boundary control problems having states constrained to semilinear parabolic PDE's are presented in [3].

The scope of this work is the analysis and finite element approximation of distributed optimal control problems having states constrained to implicit parabolic equations. In particular, we prove the existence of an optimality system of equations, under the assumption that equation (1.4) possess parabolic structure. Then a semi-discrete (in space) approximation scheme and a fully discrete scheme which is discontinuous in time, and conforming in space are formulated and analyzed. The main goal is to show that under certain structural hypotheses on the operators, the error estimates of the corresponding optimality system have the same structure to the estimates of the uncontrolled implicit parabolic equations (see e.g. [4]). The main features of these estimates can be summarized as follows:

- The error estimates are derived under minimal regularity assumptions on the given data, on the energy norm for both state and adjoint variables. For the state variable, we also obtain estimates at arbitrary time-points. These estimates are applicable when higher order elements are being used, provided that the natural parabolic regularity is valid for the solutions of the optimality system.

- The operators $A(.)$ are not assumed to be self-adjoint, contrary to many previous works. For the semi-discrete (in space) approximation the operators $A(.)$ do not need to be strictly coercive. The dependence of various constants appearing in the estimates, on the coercivity and continuity constants are being tracked.

- The fully-discrete scheme is based on the discontinuous (in time) Galerkin approach which allows the use of different subspaces at each (or at every few other) time-steps. In the examples presented here, re-meshing is also necessary in order avoid the degeneracy of the parabolic PDE.

- The parameter α is carefully tracked and does not appear at any exponential.

The rest of the paper is organized as follows. After, introducing the necessary notation and structural assumptions on the operators due to the presence of time-dependent norms, inner-products, etc., we present two examples of implicit parabolic equations in section 2. In section 3, the continuous optimal control problem is analyzed and an optimality system of equations is obtained. In section 4, the semi-discrete (in space) scheme is presented. The estimate of section 4, generalizes the error estimates of [1] to implicit parabolic equations of the form (1.1). In section 5, fully-discrete error estimates are obtained for the optimality system by using techniques of [4] (which were developed for the uncontrolled problem) and a bootstrap argument. The result of section 5, extends to the implicit parabolic case, the previously developed estimate for optimal control problems related to abstract linear parabolic equations (see e.g. [2] and references within). Finally, we present convergence rates for two examples which fall into this category. To our best knowledge these estimates are new.

2 An overview of implicit parabolic equations

For an excellent overview of implicit parabolic equations, we refer the reader to [26, 27] and references within. To formulate the weak problem associated to the implicit parabolic equation (1.1), we introduce time-dependent norms. In particular, the time-dependent (non-selfajdoint) nature of operator $A(t)$, is characterized by introducing equivalent (time-dependent) norms on U, of the form $\|u\|^2_{U(t)} = |u|^2_{U(t)} + |u|^2_{H(t)}$. Here, $|u|_{U(t)}$ denotes a seminorm on U (the principal part), while $|u|_{H(t)} = (M(t)u, u)_H$ is a norm on H, endowed by the symmetric positive operator M(t). Occasionally we adopt the notation, $|u|_{H(t)} = \|u\|_{H(t)}$. The bilinear forms induced by $A(t)$ and $A^*(t)$ are denoted by $a(t; u, v)$ and $a^*(t; u, v)$ respectively, where $A^*(t)$ is the adjoint operator of $A(t)$. The notation of the above operators, bilinear forms, norms will be abbreviated to $A(.)$, $A^*(.)$, $a(.; ., .)$, $\|.\|_{H(.)}$, $\|.\|_{U(.)}$ etc. We will also assume that the following dense and continuous embeddings hold: $U(t) \subset H(t) \subset U^*(t)$. The embedding constants are also assumed to be independent of time. The following structural hypotheses on operators $M(.)$, $A(.)$, are needed.

Assumption 2.1. *The operators $M(t)$ are assumed to satisfy:*

1. *The operators $M(t)$ are nonnegative, self-adjoint and there exist constants $c(t) > 0$ such that $(M(t)u, u)_H \geq c(t)|u|^2_H$.*

2. *(Smoothness) For every $t > 0$ there exists a symmetric bilinear form $\mu(t; ., .)$ satisfying*

$$\frac{d}{dt}(u, v)_{H(t)} = (u_t, v)_{H(t)} + (u, v_t)_{H(t)} + \mu(t; u, v), \qquad \forall\, u, v \in H^1[0, T; H].$$

In addition, there exists C_μ (independent of time) such that

$$\mu(t;u,v) \le C_\mu |u|_{H(t)} |v|_{H(t)}.$$

Remark 2.2. *The assumption 2.1.1 guarantees that for $t \ge 0$, $(M(t)u,v)_H$ is actually an inner-product, which is denoted by $(.,.)_{H(t)}$. Hence, $H(t)$ is a Hilbert space with underlying set H and denoted by $(u,v)_{H(t)} = (M(t)u,v)_H$. The assumption 2.1.2 implies that equation (1.1) is parabolic in nature. Note that if $u,v \in \mathcal{U}$ the weighted inner-product should be replaced by the weighted duality $\langle .,. \rangle_{U*(.),U(.)}$ in assumption 2.1.2.*

A consequence of the structural hypothesis on $M(.)$, is that the norms on $H(t)$ vary continuously with respect to t. In particular, we quote the following lemma from [4, Lemma 2.2].

Lemma 2.3. *Let $v_1, v_2 \in H$ and $s \le t$ then, $e^{C_\mu(s-t)} \le |v_1|^2_{H(t)}/|v_1|^2_{H(s)} \le e^{C_\mu(t-s)}$ and*

$$|(v_1,v_2)_{H(t)} - (v_1,v_2)_{H(s)}| \le (t-s)C_\mu e^{C_\mu(t-s)} |v_1|_{H(t_1)} |v_2|_{H(t_2)}, \qquad t_1, t_2 \in [s,t].$$

An assumption on the equivalence of norms in $U(t)$ follows.

Assumption 2.4. *For every $0 \le \tau \le T$ there exists $C_u > 0$ such that for all $s,t \ge 0$ with $|t-s| < \tau$,*

$$1/C_u \le \frac{\|u\|_{U(t)}}{\|u\|_{U(s)}} \le C_u \qquad \forall\, u \in U.$$

Finally, we quote the basic continuity and coercivity assumptions on the bilinear form and data.

Assumption 2.5. *1. Continuity of bilinear form: There exist $0 \le c_a \le C_a$ such that*

$$\Big| a(t;u,v) \Big| \le \Big(c_a |u|^2_{U(t)} + C_a |u|^2_{H(t)} \Big)^{1/2} \Big(c_a |v|^2_{U(t)} + C_a |v|^2_{H(t)} \Big)^{1/2}, \qquad \forall\, u,v \in U$$

2. There exists a weighted dual norm of $U(.)$ (denoted by $U^*(.)$) such that

$$\Big| \langle F(t), v \rangle \Big| \le \|F(t)\|_* \Big(c_a |u|^2_{U(t)} + C_a |u|^2_{H(t)} \Big)^{1/2}.$$

3. Coercivity: There exist constants $c_\gamma > 0$ and $C_\gamma \in \mathbb{R}$ such that

$$a(t;u,u) \ge c_\gamma |u|^2_{U(.)} - C_\gamma |u|^2_{H(.)}, \qquad \forall\, u \in U.$$

The bilinear form associated to the adjoint operator, will be also assumed to satisfy similar continuity and coercivity properties and in particular that corresponding constants $c_a^*, C_a^*, c_\gamma^*, C_\gamma^*$ are comparable to $c_a, C_a, c_\gamma, C_\gamma$ respectively. In particular, for the examples examined here, the important quantity is the ratio c_a/c_γ, and the following relation holds $c_a/c_\gamma \approx c_a^*/c_\gamma^*$.

Utilizing the above notation and assumptions, the natural weak formulation of (1.4) can be stated as follows: Given, $f \in L^2[0,T;U^*(.)]$, $g \in L^2[0,T;H(.)]$, $u_0 \in H$, we seek a function

$$u \in \mathcal{U} \equiv L^2[0,T;U(.)] \cap H^1[0,T;U^*(.)]$$

such that

$$(u(T),v(T))_{H(T)} + \int_0^T \Big(-(u,v_t)_{H(.)} + a(.;u,v) \Big) = (u_0, v(0))_{H(0)} + \int_0^T \langle F(.), v \rangle \quad \forall\, v \in \mathcal{U}. \quad (2.1)$$

A few remarks with respect to the unique solvability of 2.1 in $\mathcal{U}$ follow. Under the above hypotheses on the operators, it is not evident that $u_t \in L^2[0,T;U^*(.)]$ (it is however true that $(M(.)u(.))_t \in L^2[0,T;U^*(.)]$). Therefore, to formulate the optimal control problem but especially to derive error estimates, we will assume that the PDE has the expected parabolic structure in terms of regularity.

Assumption 2.6. *(Parabolic regularity)*

- *Let* $f \in L^2[0,T;U^*(.)]$, $g \in L^2[0,T;H(.)]$, $u_0 \in H$ *there exists unique* $u \in \mathcal{U}$ *satisfying 2.1.*
- *If* $u_0 \in U$, $f, g \in L^2[0,T;H(.)]$ *then* $u \in H^1[0,T;H(.)]$.

Remark 2.7. *The above regularity assumption is the minimal one to guarantee the presence of time-derivative at the natural energy space and typically corresponds to additional assumptions on the time-differentiation of the operators* $M(.), A(.)$. *Note also that if in addition to assumption 2.5, the operators* $A(.)$ *are regular and self-adjoint (see e.g. [27, Chapter 3] or [26, Chapter 5] for related results), then there exists a unique solution* $u \in L^\infty[0,T;H(.)] \cap L^2[0,T;U(.)]$, *with* $(M(.)u(.))_t \in L^2[0,T;U^*(.)]$, *even dropping the assumption of strict positivity for operators* $M(.)$ *(the case of pseudoparabolic equations). However, in this work, we restrict ourselves to PDE's that possess at least the minimal parabolic regularity.*

2.1 Examples of implicit parabolic equations

Finally, we close this preliminary section by stating examples of parabolic equations, which take the form of implicit parabolic equations after a time-dependent change of variables, and are related with dynamic moving mesh finite element methods.

2.1.1 The convection-diffusion equation in Lagrangian coordinates

Recall that the classical convection diffusion equation takes the form

$$u_t + \mathbf{V}.\nabla u - \epsilon \Delta u = 0, \qquad u|_\Gamma = 0, \qquad u(0,x) = u_0.$$

In most interesting cases, the value of ϵ is small compared to the given velocity field $\mathbf{V}$ resulting to many significant computational and analytical difficulties. A popular strategy to address these problems is to consider the equation in a Lagrangian variable. In particular, let $\tilde{\mathbf{V}}$ denote a numerical approximation of $\mathbf{V}$, and assume that $x = \chi(t,X)$ describes the change of variables defined by the flow map associated with $\tilde{\mathbf{V}}$, i.e., $\dot{x}(t,X) = \tilde{\mathbf{V}}(t, x(t,X))$ with initial data $x(0,X) = X$. Then, if $\bar{u}(t,X) = u(t, x(t,X))$ the convection diffusion equation takes the form

$$\bar{u}_t + (\mathbf{V} - \tilde{\mathbf{V}}).(F^{-T}\nabla_X \bar{u}) - \epsilon(1/J) div_X (J F^{-1} F^{-T} \nabla_X \bar{u}) = 0$$

where $F_{ij} = \partial x_i / \partial X_i$ is the Jacobian of the mapping $x = \chi(t,X)$ and $J = det(F)$. Recall, that $\dot{J} = J div(\tilde{\mathbf{V}})$ and hence

$$\bar{u}_t J = (J\bar{u})_t - J div(\tilde{\mathbf{V}})\bar{u}.$$

Therefore, if $div(\tilde{\mathbf{V}}) \neq 0$ we obtain that the Lagrangian description of the convection diffusion equation takes the form of an implicit parabolic equation, with

$$M(.)\bar{u} = J\bar{u},$$

and

$$A(.)\bar{u} = -div(\tilde{\mathbf{V}})\bar{u}J + (\mathbf{V} - \tilde{\mathbf{V}}).(F^{-T}\nabla_X \bar{u}) - \epsilon div_X (J F^{-1} F^{-T} \nabla_X \bar{u}).$$

The above description generalizes various characteristic Galerkin schemes (see e.g. [7, 18] and references within). The variable X is typically referred as the referential or Lagrangian variable while x denotes the Eulerian variable. Using the the ODE $\dot{J} = J div(\tilde{\mathbf{V}})$, it easy to show that if $0 < c_0 \leq J(0,.) \leq C_0$ then $c_0 e^{-2\|div_x \tilde{\mathbf{V}}\|_{L^\infty} t} \leq J(t,.) \leq C_0 e^{2\|div_x \tilde{\mathbf{V}}\|_{L^\infty} t}$.

The above formulation is related to Lagrangian and dynamic moving-mesh finite element schemes as follows (see e.g. [5, Section 2] and references within): Recall the classical finite element construction uses a reference simplex, denoted by $\hat{K}$ and a mapping $\chi : \hat{K} \to K$, determined by

$x = \chi(X) = \sum_i \hat{l}_i(X)x_i$, $X \in \hat{K}$, where K is the arbitrary cell. Here $\{x_i\}_i$ denote the nodes of K, $\{X_i\}_i$ the nodes of $\hat{K}$ and $\hat{l}_i$ the Lagrange (referential) basis functions. Then, the finite element approximation $u_h(t,x)$ is given by

$$u_h(t, x(X)) = \sum_i \hat{l}_i(X)u_i(t) \equiv \sum_i l_i(x)u_i(t), \qquad X \in \hat{K}.$$

In the above relations, $\{u_i(t)\}_i$ denote the values of u_h at the nodes and $l_i = \hat{l}_i \circ \chi^{-1}$ are the basis functions on K.

When the grid points (nodes) are allowed to move, say $x_i = x_i(t)$ then

$$x = \chi(t, X) = \sum_i \hat{l}_i(X)x_i(t), \qquad \text{and} \qquad \dot{x}(t, X) = \sum_i \hat{l}_i(X)v_i(t)$$

where $v_i(t) = \dot{x}_i(t)$ denotes the velocity of the i^{th} node. Assume that $K(t) = \chi(t, \hat{K})$ is the time-evolved mesh cell, and let $l_i(t, .) : K(t) \to \mathbb{R}$ be defined by $l_i(t, .) = \hat{l}_i \circ \chi^{-1}(t, .)$. Then, we may define the approximate velocity $\tilde{\mathbf{V}}$ on $K(t)$ by $\tilde{\mathbf{V}} = \dot{x} \circ \chi^{-1}$ which implies that $\tilde{\mathbf{V}}(t, x(t, X)) = \dot{x}(t, X)$. This construction, states that when $\dot{x}_i(t) = \mathbf{V}(t, x_i(t))$, then χ is the flow map associated to $\tilde{\mathbf{V}}$ and in addition, $\tilde{\mathbf{V}}$ is the isoparametric interpolant of $\mathbf{V}$ on K. In particular, we have $\tilde{\mathbf{V}}(t, x) = \sum_i l_i(t, x)\mathbf{V}(t, x_i(t))$.

Remark 2.8. *It is evident from the structure of the underlined PDE, that if $\tilde{\mathbf{V}}$ is a good approximation of $\mathbf{V}$ then various constants arising during energy arguments will remain under control. However, we emphasize that while the Jacobian of the transformation satisfies $F(0, X) = I$, its condition number depends exponentially on various quantities of $\tilde{\mathbf{V}}$ and the numerical scheme needs to be re-initialized every few time steps. This is also important in order to maintain the positivity of $J(t, .)$ and hence the parabolic structure of the PDE. Therefore, the re-initialization process implies that different subspaces need to be used every few other steps which give rise to discontinuous Galerkin approximations. For a detailed discussion and error estimates for the uncontrolled problem related to above formulation we refer the reader to [5] (see also the references within).*

2.1.2 Diffusion on manifolds

A more general example of diffusion on manifolds also falls into the category of implicit parabolic equations. We consider the diffusion on a cell membrane, $S(t) \subset \mathbb{R}^3$, which is transported by velocity $\mathbf{V} \equiv \mathbf{V}(\mathbf{t}, \mathbf{x})$ in an ambient fluid (see e.g. [4]). Standard finite element schemes need to construct triangulations (meshes) in each time step, which is computationally expensive. An alternative approach which avoids triangulating on $S(t)$, is to construct a scheme which computes on a reference configuration. Following the notation of [4], let S_r denote the reference configuration, and let $x(t, .) : S_r \to S(t) \subset \mathbb{R}^3$ denote a mapping which relates the reference configuration and $S(t)$. For example, we may take S_r to be $S(0)$ or even the unit sphere S^2. Assume that S_r is locally parameterized by coordinates $X \in U \subset \mathbb{R}^2$, and let σ denote the diffusion constant. Then, the diffusion equation takes the form,

$$u_t - (1/J)div_X(\sigma J(F^T F)^{-1}\nabla_X u) = 0,$$

where F denotes the 3×2 matrix with components $F_{ia} = \partial x_i/\partial X_a$, and $J = \sqrt{det(F^T F)}$ is the determinant of the first fundamental form, which satisfies

$$J_t = J(I - \mathbf{n} \times \mathbf{n}) \cdot (\nabla_x \mathbf{V})$$

where $\mathbf{n}(t, X)$ denotes the normal to $S(t)$. The above equation is an implicit parabolic equation, which can also take the form (1.1), with

$$M(.)u = Ju$$

and

$$A(.)u = -(I - \mathbf{n} \times \mathbf{n}) \cdot (\nabla_x \mathbf{V})uJ - div_X(\sigma J(F^T F)^{-1} \nabla_X u).$$

In this case, it is easy to see that if $0 < c_1 \leq J(0,.) \leq C_1$ then $c_1 e^{-2\|\nabla_x \mathbf{V}\|_{L^\infty} t} \leq J(t,.) \leq C_1 e^{2\|\nabla_x \mathbf{V}\|_{L^\infty} t}$, which establishes the positivity of $M(.)$ of assumption (2.1) and hence the parabolic structure of our problem (the rest of assumptions will be checked in section 4).

3 The continuous optimal control problem

3.1 Existence of an optimal solution

First, we define the set of admissible solutions and the notion of the optimal pair (denoted by $\mathcal{A}_{ad}$ and (u, g) respectively).

Definition 3.1. *Given $\Omega, T > 0$, $u_0 \in H$, $f \in L^2[0, T; U^*(.)]$ and target $z \in L^2[0, T; H(.)]$, $z|_\Gamma = 0$, the pair $(u, g) \in \mathcal{A}_{ad}$ is called an admissible pair if $u \in \mathcal{U}$, $g \in L^2[0, T; H(.)]$ and (u, g) satisfy (2.1).*

Definition 3.2. *Let $T > 0$, $u_0 \in H$, $f \in L^2[0, T; U^*(.)]$, $z \in L^2[0, T; H(.)]$. Then $(u, g) \in \mathcal{A}_{ad}$ is called an optimal pair if $K(u, g) \leq K(v, h) \qquad \forall (v, h) \in \mathcal{A}_{ad}$.*

The target z is typically smoother in many applications. For example z can be the solution of another implicit parabolic equation, and hence to possess higher regularity. For example we may assume that $z \in \mathcal{U}$. Using standard techniques we may prove the existence of an optimal solution, in the sense of Definition (3.2).

Theorem 3.3. *Let $T > 0$, $u_0 \in H$, $z \in L^2[0, T; H(.)]$, $f \in L^2[0, T; U^*(.)]$, and let $U(.) \subset H \subset U^*(.)$ be dense embedding of Hilbert spaces with embedding constants independent of time. Assume that $|.|_{U(.)}$, $|.|_{H(.)}$ are equivalent to $|.|_U$ and $|.|_H$ respectively and let $U \subset H$ with compact embedding. Suppose that Assumptions 2.1,2.4,2.5,2.6 are satisfied, and let $\tilde{u} \in \mathcal{U}$, be the solution of (1.4), when $g(.) \equiv 0$. Then, there exists an optimal pair $(u, g) \in \mathcal{A}_{ad}$.*

Proof. (*Sketch:*) Note that $\mathcal{A}_{ad} \neq 0$, since $(\tilde{u}, 0) \in \mathcal{A}_{ad}$ due to the solvability assumption, and that $K(u, g)$ is bounded below by 0. We denote by $(u_n, g_n) \in \mathcal{A}_{ad}$ a minimizing sequence for the optimal control problem, and note that (u_n, g_n) satisfy (2.1) and by definition, the tracking functional implies that,

$$\|u_n\|^2_{L^2[0,T;H(.)]}, \quad \alpha\|g_n\|^2_{L^2[0,T;H(.)]} < C < \infty,$$

while the parabolic regularity assumption 2.6 (a) guarantees that

$$\|u_n\|_{L^2[0,T;U(.)]}, \quad \|u_{nt}\|_{L^2[0,T;U^*(.)]} < C < \infty.$$

Hence, there exists a subsequence (still denoted by $(u_n, g_n) \in \mathcal{A}_{ad}$) which converges to an element (u, g), in the following sense:

$$u_n \to u \quad \text{weakly in } L^2[0, T; U(.)], \qquad g_n \to g \quad \text{weakly in } L^2[0, T; H(.)]$$

$$u_{nt} \to u_t \quad \text{weakly in } L^2[0, T; U^*(.)], \qquad u_n \to u \quad \text{weakly * in } L^\infty[0, T; H(.)].$$

Recall, that $U(t) \subset H(t) \subset U^*(t)$ with dense and continuous embeddings with constants independent of time, and that the norms $\|.\|_{H(t)} = |.|_{H(.)}$ and semi-norms $|.|_{U(t)}$ are equivalent to the H norm and U semi-norm respectively. Therefore, the compact embedding $U \subset H$ and a well known compactness result for $L^2[0, T; B]$ spaces (see e.g. [28]), implies that

$$u_n \to u \qquad \text{strongly in } L^2[0, T; H].$$

Therefore, we may pass the limit into (2.1), which proves that $(u, g) \in \mathcal{A}_{ad}$. The weak lower semi-continuity of the functional finishes the proof. □

3.2 An optimality system of equations

Adjusting the technique of [14] to the time-dependent norm framework of implicit parabolic equations, we prove that the optimal solution pair satisfies the first-order necessary conditions. First, we show the existence of a Gâteaux derivative at any direction.

Theorem 3.4. *Let $u_0 \in H \equiv H(0)$, $z \in L^2[0,T;H(.)]$, $f \in L^2[0,T;U^*(.)]$ be given and suppose that the Assumptions 2.1-2.4-2.5-2.6 hold. Define a mapping $g \to u(g)$ from $L^2[0,T;H(.)]$ to $L^2[0,T;U(.)]$, where $u(g)$ denotes the solution of (1.4), with $g \in L^2[0,T;H(.)]$ given. Then, there exists a Gâteaux derivative $\left(\frac{Du}{Dg}\right) \cdot h$ in every direction $h \in L^2[0,T;H(.)]$, denoted by $w \equiv w(h) = \left(\frac{Du}{Dg}\right) \cdot h$, satisfying*

$$(M(t)w)_t + A(t)w = h(t), \qquad w|_\Gamma = 0, \qquad w(0,x) = 0. \tag{3.1}$$

In addition, $w \in H^1[0,T;H(.)]$.

Proof. The proof is standard due to the linearity of the operators. □

Using standard techniques, we can show that the optimal solution $(u,g) \in \mathcal{A}_{ad}$ can be located by requiring the Gâteux derivative of Theorem 3.4, to be equal to zero. Next, we derive an explicit formula of the first order necessary condition.

Theorem 3.5. *Suppose the assumptions of Theorem 3.4 hold, and let $w \equiv w(h)$ denote the Gâteaux derivative of Theorem 3.4 at direction $h \in L^2[0,T;H(.)]$. Then, for every $h_2 \in L^2[0,T;H(.)]$,*

$$\int_0^T (h_2, w)_{H(.)} = \int_0^T (\psi, h)_{H(.)}$$

where $\psi \in \mathcal{U}$ satisfies the weak formulation

$$\int_0^T \Big((\psi, v_t)_{H(.)} + \mu(.;\psi,v) + a^*(.;\psi,v) \Big) = -(\psi(0), v(0))_H + \int_0^T (h_2, v)_{H(.)}$$

for every $v \in L^2[0,T;U(.)] \cap H^1[0,T;U^*(.)]$. In addition, $\psi \in H^1[0,T;H(.)]$.

Proof. We begin by noting that the integral $\int_0^T (\psi,h)_{H(.)}$ can be formally computed from (3.1), by using assumption 2.1 and in particular that $M(.)$ is self-adjoint and $w(0,x) = 0$,

$$\begin{aligned}
\int_0^T (\psi, h)_{H(.)} &= \int_0^T (h, \psi)_{H(.)} \\
&= (w(T), \psi(T))_{H(T)} + \int_0^T \Big(-(w, \psi_t)_{H(.)} + a(.;w,\psi) \Big) \\
&= (w(T), \psi(T))_{H(T)} + \int_0^T \Big(-\frac{d}{dt}(w,\psi)_{H(.)} + (w_t, \psi)_{H(.)} + \mu(.;w,\psi) + a(.;w,\psi) \Big) \\
&= \int_0^T \Big((w_t, \psi)_{H(.)} + \mu(.;w,\psi) + a(.;w,\psi) \Big) \equiv \int_0^T (h_2, w)_{H(.)}.
\end{aligned}$$

Here we have used that once more assumption 2.1. The last equality establishes the desired result, after noting that $M(.)$ is self-adjoint, $\mu(.;w,\psi) = \mu(.;\psi,w)$, and using integration by parts. Note that all integration by parts performed are justified due to regularity properties of ψ,w. □

Now we are ready to justify the existence of an optimality system of equations (first order necessary conditions).

Theorem 3.6. *Given $T > 0$, $u_0 \in H$, $z \in L^2[0,T;H(.)]$, $f \in L^2[0,T;U^*(.)]$, and let assumptions of Theorem 3.4 hold. Let (u,g) denote an optimal pair (in the sense of Definition 3.2). Then, $\int_0^T (\psi + \alpha g, h)_{H(.)} = 0$, for all $h \in L^2[0,T;H(.)]$, and where ψ is the solution of:*

$$\int_0^T \Big((\psi, v_t)_{H(.)} + \mu(.;\psi,v) + a^*(.;\psi,v) \Big) + (\psi(0), v(0))_H = \int_0^T (u - z, v)_{H(.)},$$

for every $v \in L^2[0,T;U(.)] \cap H^1[0,T;U^*(.)]$.

Proof. Suppose that (u,g) is an optimal pair, denote by $\left(\frac{DK(u,g)}{Dg}\right) \cdot h$ the Gâteaux derivative of the functional $K(u,g)$ on the direction h. Then, we easily compute

$$\begin{aligned}
\Big(\frac{DK(u(g),g)}{Dg}\Big) \cdot h &= \int_0^T \Big(u - z, \Big(\frac{Du}{Dg}\Big) \cdot h \Big)_{H(.)} + \alpha \int_0^T (g,h)_{H(.)} \\
&= \int_0^T (u - z, w)_{H(.)} + \alpha \int_0^T (g,h)_{H(.)},
\end{aligned}$$

where $w \equiv w(h)$ is defined in Theorem 3.4. Using Theorem 3.5, for $h_2 = u - z \in L^2[0,T;H(.)]$

$$\int_0^T (u - z, w)_{H(.)} = \int_0^T (\psi, h)_{H(.)}$$

Combining the last two equalities, we obtain

$$\Big(\frac{DK(u(g),g)}{Dg}\Big) \cdot h = \int_0^T (\psi + \alpha g, h)_{H(.)},$$

which establishes the conclusion after noting that $\left(\frac{DK(u(g),g)}{Dg}\right) \cdot h = 0$, $\forall\, h \in L^2[0,T;H(.)]$. ☐

Therefore, using the optimality condition to replace, the control from the state equation, the optimality system takes the form, for all $v \in L^2[0,T;U(.)] \cap H^1[0,T;U^*(.)]$,

$$\left\{ \begin{array}{l}
(u(T), v(t))_{H(T)} + \displaystyle\int_0^T \Big(-(u, v_t)_{H(.)} + a(.;u,v) \Big) = (u_0, v(0))_H + \int_0^T \langle f, v \rangle - (1/\alpha)(\psi, v)_{H(.)} \\
\displaystyle\int_0^T \Big((\psi, v_t)_{H(.)} + \mu(.;\psi,v) + a^*(.;\psi,v) \Big) + (\psi(0), v(0))_{H(0)} = \int_0^T (u - z, v)_{H(.)}
\end{array} \right. \tag{3.2}$$

where $u_0, \phi(T) \equiv 0$ are given initial and data terminal data, and f, z denote the forcing term and target function respectively. The above system corresponds to the weak form of the following coupled system of implicit parabolic equations:

$$\left\{ \begin{array}{rll}
(M(t)u)_t + A(t)u & = f - (1/\alpha)\psi & \qquad u(0,x) = u_0 \\
-(M(t)\psi)_t + B(t)\psi & = u - z & \qquad \psi(T) = 0
\end{array} \right. \tag{3.3}$$

where the operator $B(.)$ is induced by, $\langle Bv_1, v_2 \rangle = a^*(,;v_1,v_2) + \mu(.;v_1,v_2)$, for all $v_1, v_2 \in \mathcal{U}$. It is evident that the choice of the L^2 norms into the functional (1.3) leads essentially to an algebraic optimality condition which results to a simpler and more computationally attractive system.

4 The semi-discrete (in space) optimality system

We prove the existence of semi-discrete (in space) approximations of arbitrary order and derive semi-discrete (in space) error estimates for the optimality system based on the operator theoretic approach of Brezzi-Rappaz-Raviart (see e.g. [10]). Within the context of optimal control problems, the Brezzi-Rappaz-Raviart theory was first used in [12] for a boundary control problem related to the stationary Navier-Stokes equations. In order to apply this theory, we need to obtain estimates for a model (uncoupled) implicit parabolic PDE. Contrary to the work of [3], an auxiliary term is included to the model implicit parabolic PDE, in order to overcome the lack of "strict coercivity" of the bilinear forms.

4.1 The discrete optimality system and projection estimates

To simplify the analysis, we set $H(0) \equiv H = L^2(\Omega)$ and $U(0) \equiv U = H^1_0(\Omega)$. First, we introduce the finite element subspaces U_h of $U = H^1_0(\Omega)$ which satisfy the standard approximation properties, constructed over a triangulation $\mathcal{R}_h$ using piecewise polynomials of degree $l \geq 0$. We emphasize that U_h are constructed independent of time, and are subspaces of $U = H^1_0(\Omega)$, while the time-dependent norms capture the structure of the time-dependent operators. Then, the semi-discrete (in space) approximations of the optimality system (3.3) can be written as follows. Given, $f \in L^2[0,T;U^*(.)]$, $u_0 \in H \equiv H(0)$, we seek $u_h, \psi_h \in H^1[0,T;U_h]$ such that for every $v_h \in H^1[0,T;U_h]$,

$$\begin{cases} (u_h(T), v_h(T))_{H(T)} + \displaystyle\int_0^T \Big(-(u_h, v_{ht})_{H(.)} + a(.;u_h, v_h) \Big) \\ \qquad = (u_{h0}, v_h(0))_H + \displaystyle\int_0^T \langle f, v_h \rangle - (1/\alpha)(\psi_h, v_h)_{H(.)} \\ \displaystyle\int_0^T \Big((\psi_h, v_{ht})_{H(.)} + \mu(.;\psi_h, v_h) + a^*(.;\psi_h, v_h) \Big) + (\psi_h(0), v_h(0))_H = \int_0^T (u_h - z, v_h)_{H(.)}, \end{cases} \tag{4.1}$$

where u_{h0} denotes a suitable approximation of the given initial data u_0 and $\psi_h(T) = 0$.

Remark 4.1. *Note that finite element subspaces are constructed in a standard fashion, using standard finite element basis, while the underlying time-dependent subspaces are only implicitly defined through the time-dependent norms and inner products.*

Before proving the existence, stability properties and error estimates of the proposed discretization (4.1), we define the L^2 projections (with time-dependent) norms, which will play a crucial role in the subsequent analysis. We denote the "weighted" projections $P_h(t) : H \to U_h$ by:

$$P_h(t)v \in U_h, \qquad \big(M(t)(v - P_h(t)v), v_h\big)_H = 0 \qquad \forall\, v_h \in U_h. \tag{4.2}$$

The generalized "weighted" L^2 projections will be also needed in order to derive error estimates for the time-derivative. We denote by $Q_h(t) : U^* \to U_h$ the projection which is defined by,

$$Q_h(t)v \in U_h, \qquad \big(M(t)(Q_h(t)v), v_h\big)_H = \langle v, v_h \rangle_{U^*,U} \qquad \forall\, v_h \in U_h. \tag{4.3}$$

Since $U \subset H(t) \subset U^*$ and $H(t)$ is the weighted space with inner product $(u,v)_{H(t)} = (M(t)u, v)_H$, the projection $Q_h(t)$ can be viewed as an extension of $P_h(t)$. Therefore, if $u \in H$ then $P_h(t)u = Q_h(t)u$. In the presence of the structural hypotheses 2.1-2.6 and in particular of the norm and semi-norm equation equivalences (see section 2), we obtain standard approximation properties for the projections P_h, Q_h.

Proposition 4.2. *Let $\{\mathcal{R}_h\}_{h>0}$ be a quasi-regular family of triangulations of the domain Ω, and for each $h > 0$ let $U_h \subset U \equiv H^1_0(\Omega)$ be a classical finite element space constructed over $\mathcal{R}_h$, containing piecewise polynomials of degree $l \geq 0$, on each triangle $K \in \mathcal{R}_h$. Suppose also that there exist constants $C_0(t) > 0, C_1(t) > 0$ such that the following (semi)-norm and norm equivalences hold*

$$\|u\|_{H(0)} e^{-C_0(t)} \leq \|u\|_{H(t)} \leq \|u\|_{H(0)} e^{C_0(t)}, \quad \text{and} \quad |u|_{U(0)} e^{-C_1(t)} \leq |u|_{U(t)} \leq |u|_{U(0)} e^{C_1(t)}.$$

Then,

1. *If $u \in U \cap H^{l+1}(\Omega)$, there exists $C = C(l)$, such that*

$$\|u - P_h(t)u\|_{H(t)} = \|u - Q_h(t)u\|_{H(t)} \leq Ce^{C_0(t)}|u|_{H^{l+1}(\Omega)}h^{l+1}.$$

2. *If the triangulations $\{\mathcal{R}_h\}_{h>0}$ are quasi-uniform then there exists $C = C(l)$ such that the following inverse inequality holds:*

$$|u_h|_{U(t)} \leq (C/h)e^{C_0(t)+C_1(t)}\|u_h\|_{H(t)} \qquad \forall\, u_h \in U_h.$$

3. *If the triangulations $\{\mathcal{R}_h\}_{h>0}$ are quasi-uniform then there exists $C = C(l)$ such that*

$$|u - P_h(t)u|_{U(t)} = |u - Q_h(t)u|_{U(t)} \leq Ce^{2C_0(t)+C_1(t)}|u|_{H^{l+1}(\Omega)}h^l, \qquad \forall\; u \in U \cap H^{l+1}(\Omega).$$

Proof. The proof follows in a verbatim way [5, Lemma 3.4]. □

Remark 4.3. *The Proposition 4.2 states that the classical approximation properties for the weighted L^2 projection hold in both weighted L^2 and H^1 norms, and various constants are quantified. The constants $C_0(t), C_1(t)$ can be explicitly computed in most interesting applications. Note also that an estimate on the dual norm can be derived based on Proposition 4.2. In particular, recall that $U = U(0) = H_0^1(\Omega)$, $H = H(0) = L^2(\Omega)$. Then for $u \in H^{l+1} \cap U$ and for $v \in U^*$, $v_h = P_h(t)v \in U_h$, using Proposition 4.2,*

$$\begin{aligned}
\sup_{0\neq v\in U} \frac{(u - Q_h(t)u, v)_{H(t)}}{\|v\|_{U(t)}} &= \sup_{0\neq v\in U} \frac{(u - Q_h(t)u, v - v_h)_{H(t)}}{\|v\|_{U(t)}} \\
&\leq Ce^{C_0(t)}|u|_{H^{l+1}}h^{l+1}\frac{e^{C_0(t)}|v|_{H^1}h}{|v|_{U(t)}} \\
&\leq Ce^{2C_0(t)+C_1(t)}|u|_{H^{l+1}}h^{l+2},
\end{aligned}$$

by using the semi-norm equivalence.

4.2 Semi-discrete error estimates for a model problem

In order to obtain estimates for the coupled optimality system of equations, we first establish estimates on a model problem, which satisfies the strict coercivity assumption. In particular, we consider the uncontrolled implicit parabolic equation $g(.) \equiv 0$, and we prove semi-discrete error estimates, under minimal regularity assumptions. The auxiliary weak problem is stated as follows: Given $f \in L^2[0,T;U^*(.)]$, $u_0 \in H = H(0)$, we seek $u \in L^2[0,T;U(.)] \cap H^1[0,T;U^*(.)]$ such that for all $v \in L^2[0,T;U(.)] \cap H^1[0,T;U^*(.)]$,

$$(u(T), v(T))_{H(T)} + \int_0^T \Big(-(u, v_t)_{H(.)} + a(.;u,v) + \eta(u,v)_{H(.)} \Big) = (u(0), v(0))_H + \int_0^T \langle f, v\rangle_{U^*,U}. \quad (4.4)$$

Remark 4.4. *The constant $\eta > 0$ will be specified later, and depends on the data $c_a, C_a, c_\gamma, C_\gamma, C_\mu$.*

Similarly, the semi-discrete (in space) problem of the uncontrolled auxiliary problem takes the form: we seek $u_h \in H^1[0,T;U_h]$ such that for all $v_h \in H^1[0,T;U_h]$,

$$\begin{aligned}
&(u_h(T), v_h(T))_{H(T)} + \int_0^T \Big(-(u_h, v_{ht})_{H(.)} + a(.;u_h,v_h) + \eta(u_h,v_h)_{H(.)} \Big) \qquad (4.5)\\
&\quad = (u_h(0), v_h(0))_H + \int_0^T \langle f, v_h\rangle_{U^*,U}.
\end{aligned}$$

The subsequent result will be used to uncouple the optimality system. The key idea is to derive an estimate on the energy norm $\|.\|_{L^2[0,T;U(.)]}$ which is independent of estimates on time-derivative u_t. The proof is based on L^2 projection techniques, and follows [5, Theorem 3.1]. For completeness, we state the relevant result and quantify various constants in terms of the ratio c_a/c_γ. The quantity c_a/c_γ plays important role in applications such as the Lagrangian moving mesh formulation of convection-diffusion equations (see e.g. [5]).

Theorem 4.5. *Let $U_h \subset H_0^1(\Omega)$ be a finite dimensional subspace satisfying the assumptions of Proposition 4.2 and suppose that assumptions 2.1-2.4-2.5 hold. Suppose that $f \in L^2[0,T;U^*(.)]$, $u_0 \in H$, and let u, u_h denote the solutions of (4.4)-(4.5) respectively. Then, for $\eta \geq C_\gamma + (C_\mu/2) + (c_\gamma C_a/c_a)$, the following estimate holds for the error $e = u - u_h$. There exists an algebraic constant $C > 0$ such that:*

$$
\begin{aligned}
c_\gamma |e|^2_{L^2[0,T;U(.)]} + (c_\gamma C_a/c_a)|e|^2_{L^2[0,T;H(.)]} \quad \leq \quad & C\Big(|e(0)|^2_H + (C_a c_a/c_\gamma + c_\gamma C_a/c_a)|e_p|^2_{L^2[0,T;H(.)]} \\
& +(c_a^2/c_\gamma + c_\gamma)|e_p|^2_{L^2[0,T;U(.)]}\Big)
\end{aligned}
$$

with $e_p(.) = u(.) - P_h(.)u(.)$. Here $P_h(t)$ denotes the weighted L^2 orthogonal projection.

Proof. Subtracting (4.5) from (4.4), we obtain the orthogonality condition

$$
(e(T), v_h(T))_{H(T)} + \int_0^T \Big(-(e, v_{ht})_{H(.)} + a(.; e, v_h) + \eta(e, v_h)_{H(.)}\Big) = (e(0), v_h(0))_H. \qquad (4.6)
$$

Decomposing the error as $e(.) = u(.) - u_h(.) = u(.) - P_h(.)u(.) + P_h(.)u(.) - u_h(.)$, denoting $e_p(.) = u(.) - P_h(.)u(.)$, $e_h(.) = P_h(.)u(.) - u_h(.)$ and setting $v_h = e_h$ we obtain

$$
\begin{aligned}
&(e_h(T), e_h(T))_{H(T)} + \int_0^T \Big(-(e_h, e_{ht})_{H(.)} + a(.; e_h, e_h) + \eta(e_h, e_h)_{H(.)}\Big) \\
&\quad = (e_h(0), e_h(0))_{H(0)} - \int_0^T a(.; e_p, e_h).
\end{aligned}
$$

Here we have used the properties of $P_h(.)$. In particular, we emphasize that by construction, $e_h(t) \in U_h$ a.e. $t \in (0,T]$, and since U_h is independent of time, $e_{ht}(t) \in U_h$ too. Hence $\int_0^T (e_p, e_{ht})_{H(.)} = 0$. Using assumption 2.1, we obtain

$$
\begin{aligned}
&|e_h(T)|^2_{H(T)} + \int_0^T \Big(-\frac{1}{2}\frac{d}{dt}(e_h, e_h)_{H(.)} + \frac{1}{2}\mu(.; e_h, e_h) + a(.; e_h, e_h) + \eta|e_h|^2_{H(.)}\Big) \\
&\quad = |e_h(0)|^2_{H(0)} - \int_0^T a(.; e_p, e_h),
\end{aligned}
$$

which leads to (by assumption 2.4)

$$
\begin{aligned}
&(1/2)\Big(|e_h(T)|^2_{H(T)} - |e_h(0)|^2_{H(0)}\Big) + c_\gamma \int_0^T |e_h|^2_{U(.)} + \big(\eta - C_\gamma - (C_\mu/2)\big) \int_0^T |e_h|^2_{H(.)} \\
&\leq \int_0^T \Big(c_a|e_p|^2_{U(.)} + C_a|e_p|^2_{H(.)}\Big)^{1/2} \Big(c_a|e_h|^2_{U(.)} + C_a|e_h|^2_{H(.)}\Big)^{1/2}.
\end{aligned}
$$

Using the identity $ab \leq (1/4\delta)a^2 + \delta b^2$ and selecting $\delta > 0$ to hide the $c_a \int_0^T |e_h|^2_{U(.)}$ to the left, we obtain

$$
\begin{aligned}
&(1/2)\Big(\|e_h(T)\|^2_{H(T)} - \|e_h(0)\|^2_H\Big) + (c_\gamma/2) \int_0^T |e_h|^2_{U(.)} + \big(\eta - C_\gamma - (C_\mu/2)\big) \int_0^T |e_h|^2_{H(.)} \\
&\leq \int_0^T (c_a^2/2c_\gamma)|e_p|^2_{U(.)} + (C_a c_a/2c_\gamma)|e_p|^2_{H(.)} + (c_\gamma C_a/2c_a) \int_0^T |e_h|^2_{H(.)}.
\end{aligned}
$$

Multiplying the last inequality by two, selecting η such that $2\eta - 2C_\gamma - C_\mu - (c_\gamma C_a/c_a) \geq (c_\gamma C_a/c_a)$, and using triangle inequality we obtain the desired estimate. $\square$

Corollary 4.6. *Let the assumptions of Theorem 4.5 hold. Then,*

$$\|e\|^2_{L^2[0,T;U(.)]} \leq C\Big(\|e(0)\|^2_{L^2(\Omega)} + \|e_p\|^2_{L^2[0,T;U(.)]}\Big)$$

where C is a constant depending only on the ratio c_a/c_γ, C_a, C_γ, C_μ.

Proof. Using standard algebra, the estimate of Theorem 4.5 and triangle inequality. $\square$

We close this subsection by proving an error estimate for the time derivative, based on the generalized L^2 projection techniques.

Theorem 4.7. *Suppose that the assumptions of Theorem 4.5 hold, and denote by C_q the stability constant of the projection $Q_h(.)$ with respect to $U(.)$ norm, i.e.,*

$$\|Q_h(.)u\|_{U(.)} \leq C_q\|u\|_{U(.)}, \qquad \textit{and} \qquad \|u - Q_h(.)u\|_{U(.)} \leq C_q\|u\|_{U(.)}. \tag{4.7}$$

Then,

$$\|e_t\|^2_{L^2[0,T;U^*(.)]} \leq C\Big(\|u_t - Q_h(.)u_t\|^2_{L^2[0,T;U^*(.)]} + \|e\|^2_{L^2[0,T;U(.)]}\Big)$$

where the constant C depends on $C_a, C_\mu, C_\gamma, C_q, C_u$ and the ratio c_a/c_γ.

Proof. Working similar to the proof of Theorem 4.5 and integrating by parts the resulting orthogonality condition, we obtain

$$\int_0^T \langle e_t, v_h\rangle_{U^*(.),U(.)} + \mu(.;e,v_h) + a(.;e,v_h) + \eta(e,v_h)_{H(.)} = 0, \tag{4.8}$$

for all $v_h \in H^1[0,T;U_h]$. Note that $e_h \in H^1[0,T;U_h]$. Then adding and subtracting $Q_h(.)v \in U_h$ and using (4.8),

$$\begin{aligned}
\int_0^T \langle e_t, v\rangle_{U*(.),U*(.)} &= \int_0^T \langle e_t, v - Q_h(.)v\rangle_{U*(.),U(.)} + \langle e_t, Q_h(.)v\rangle_{U*(.),U(.)} \\
&= \int_0^T \langle e_t, v - Q_h(.)v\rangle_{U*(.),U(.)} - \int_0^T \Big(\mu(.;e,Q_h(.)v) + a(.;e,Q_h(.)v) + \eta(e,Q_h(.)v)_{H(.)}\Big),
\end{aligned}$$

for all $v \in L^2[0,T;U(.)]$. Here, at the last equality we have also used the fact that $Q_h(.)v(.) \in H^1[0,T;U_h]$. Indeed, the stability inequality (4.7), the definition of $Q_h(.)$, implies that $Q_h(.)v(.) \in U_h$, and hence $(Q_h(.)v(.))_t \in U_h$, since U_h is independent of time (by its construction). For the first term on the right hand side, note that $u_{ht}(.)$ and $Q_h(.)u_t$ belong in U_h, and hence

$$\int_0^T \langle e_t, v - Q_h(.)v\rangle_{U*(.),U(.)} = \int_0^T \langle u_t, v - Q_h(.)v\rangle_{U*(.),U(.)} = \int_0^T \langle u_t - Q_h(.)u_t, v - Q_h(.)v\rangle_{U*(.),U(.)}.$$

Combining the last two equalities, we obtain,

$$\begin{aligned}
\int_0^T \langle e_t, v\rangle_{U*(.),U(.)} \quad &= \quad \int_0^T \langle u_t - Q_h(.)u_t, v - Q_h(.)v\rangle_{U*(.),U(.)} \\
&\quad - \int_0^T \Big(\mu(.;e,Q_h(.)v) + a(.;e,Q_h(.)v) + \eta(e,Q_h(.)v)_{H(.)}\Big).
\end{aligned}$$

Hölder's inequality, and various norm equivalences imply that

$$\int_0^T \langle e_t, v\rangle_{U^*(.),U(.)} \leq C\Big(\int_0^T \|u_t - Q_h(.)u_t\|_{U^*(.)}\|v - Q_h(.)v\|_{U(.)}$$
$$+ \int_0^T C_\mu \|e\|_{H(.)}\|Q_h(.)v\|_{H(.)} + \|e\|_{U(.)}\|Q_h(.)v\|_{U(.)} + \eta\|e\|_{H(.)}\|Q_h(.)v\|_{H(.)}\Big)$$
$$\leq C\Big(\|u_t - Q_h(.)u_t\|_{L^2[0,T;U^*(.)]}\|v - Q_h(.)v\|_{L^2[0,T;U(.)]}$$
$$+\Big((C_\mu + \eta)\|e\|_{L^2[0,T;H(.)]} + \|e\|_{L^2[0,T;U(.)]}\Big)\|Q_h(.)v\|_{L^2[0,T;U(.)]}\Big).$$

Using the stability inequality (4.7), and taking the supremum over $v \in L^2[0,T;U(.)]$ we obtain the desired estimate. □

Various estimates of symmetric form can be deriving by combining the last two results. Recall, that the quantity of interest with respect to the size of the various constants, is c_α/c_γ.

Theorem 4.8. *Suppose that the assumptions of Theorems 4.5-4.7 hold. Then, the following estimate is true:*

$$\|e\|^2_{L^2[0,T;U(.)]} + \|e_t\|^2_{L^2[0,T;U^*(.)]} \leq C\Big(|e(0)|^2_{H(0)} + \|e_p\|^2_{L^2[0,T;U(.)]} + \|u_t - Q_h(.)u_t\|^2_{L^2[0,T;U^*(.)]}\Big),$$

where the constant C depends on $C_a, C_\gamma, C_\mu, C_q, C_u$ and the ratio c_a/c_γ, c_a^2/c_γ.

Once, we have obtained error estimates on the natural energy norm under minimal regularity assumptions for the auxiliary problem, we are ready to rewrite the optimality system into the operator framework of Brezzi-Rappaz-Raviart. First, we quote the main result regarding the Brezzi-Rappaz-Raviart theory, specialized to our needs. For more details the reader can consult [10].

4.3 The Brezzi-Rappaz-Raviart theory

The problems considered are of the following type. Suppose that $\mathcal{X}, \mathcal{Y}$ are Banach spaces and we seek $\chi \in \mathcal{X}$ such that

$$\chi + \mathcal{T}\mathcal{G}\chi = 0, \tag{4.9}$$

where $\mathcal{T} \in \mathcal{L}(\mathcal{Y}, \mathcal{X})$, $\mathcal{G}$ is a C^2 mapping from the solution space $\mathcal{X}$ to the data space $\mathcal{Y}$. We call a solution a regular solution if $\chi + \mathcal{T}\mathcal{G}_\chi(\chi)$ is an isomorphism from $\mathcal{X}$ to $\mathcal{Y}$. Here $\mathcal{G}_\chi$ (or $D\mathcal{G}$) denotes the Fréchet derivative of $\mathcal{G}(.)$. We also assume that there exists another Banach space $\mathcal{Z}$, contained in $\mathcal{Y}$, with continuous embedding, such that

$$\mathcal{G}_\chi(\chi) \in \mathcal{L}(\mathcal{X}, \mathcal{Z}) \qquad \forall\, \chi \in \mathcal{X}. \tag{4.10}$$

Approximations are defined on a subspace $\mathcal{X}_h \subset \mathcal{X}$ based on an approximating operator $\mathcal{T}_h \in \mathcal{L}(\mathcal{Y}, \mathcal{X}_h)$. The discrete problem is to find $\chi_h \in \mathcal{X}_h$ such that

$$\chi_h + \mathcal{T}_h\mathcal{G}\chi_h = 0. \tag{4.11}$$

The approximation operator $\mathcal{T}_h$ needs to satisfy the following properties.

$$\lim_{h\to 0} \|(\mathcal{T}_h - \mathcal{T})y\|_{\mathcal{X}} = 0, \qquad \forall\, y \in \mathcal{Y}, \tag{4.12}$$

and

$$\lim_{h\to 0} \|\mathcal{T}_h - \mathcal{T}\|_{\mathcal{L}(\mathcal{Z},\mathcal{X})} = 0. \tag{4.13}$$

Recall that if $\mathcal{Z} \subset \mathcal{Y}$ with compact embedding then (4.13) follows directly from (4.12). Next, we state the main theorem. In the following statement $D^2\mathcal{G}$ denotes the second Fréchet derivatives.

Theorem 4.9. *Let $\mathcal{X}, \mathcal{Y}$ be Banach spaces. Assume that $\mathcal{G}$ is a C^2 mapping from $\mathcal{X}$ to $\mathcal{Y}$ and that $D^2\mathcal{G}$ is bounded on all bounded sets of $\mathcal{X}$. Suppose also that (4.10)-(4.12)-(4.13) hold and that $\chi \in \mathcal{X}$ is a regular solution of (4.9). Then there exists a neighborhood $\mathcal{O}$ of the origin in $\mathcal{X}$ and for $h \le h_0$ small enough, a unique function $\chi_h \in \mathcal{X}_h$ such that χ_h is a regular solution of (4.11), $\chi_h - \chi \in \mathcal{O}$. Moreover, there exists a constant C independent of h such that*

$$\|\chi_h - \chi\|_{\mathcal{X}} \le C\|(\mathcal{T}_h - \mathcal{T})\mathcal{G}\chi\|_{\mathcal{X}}. \tag{4.14}$$

Proof. [10, Theorem 3.3, pp.307]. □

Remark 4.10. *The main advantage of the using the above framework to the linear optimality system, is that it facilitates the decoupling the forward and backward in time PDE's in presence of the non-selfadjoint operators. Note also that the estimate has a symmetric structure, and relates the error of the coupled optimality system to the error the model (uncoupled) PDE.*

4.4 Brezzi-Rappaz-Raviart framework of the optimality system

In order to apply Theorem 4.9, we need to recast the optimality system, into the Brezzi-Rappaz-Raviart framework. For this purpose, we set $H = L^2(\Omega)$, $U = H_0^1(\Omega)$ and let $H(.)$, $U(.)$ denote the underlying time-dependent spaces, induced by the operators $M(.), A(.)$.

$$X = L^2[0,T;U] \cap H^1[0,T;U^*], \qquad Y = L^2[0,T;U^*] \cap H$$

$$\mathcal{X} = X \times X, \qquad \mathcal{Y} = Y \times Y.$$

All above spaces are endowed with the natural time-dependent norms, e.g.,

$$\begin{aligned} \|(u,\psi)\|^2_{\mathcal{X}} &= c_a|u|^2_{L^2[0,T;U(.)]} + C_a|u|^2_{L^2[0,T;H(.)]} + \|u_t\|^2_{L^2[0,T;U^*(.)]} \\ &\quad + c_a|\psi|^2_{L^2[0,T;U(.)]} + C_a|\psi|^2_{L^2[0,T;H(.)]} + \|\psi_t\|^2_{L^2[0,T;U^*(.)]} \\ \|(f_1,u_1,f_2,u_2)\|^2_{\mathcal{Y}} &= \|f_1\|^2_{L^2[0,T;U^*(.)]} + |u_1|^2_{H(0)} + \|f_2\|^2_{L^2[0,T;U^*(.)]} + |\psi_2|^2_{H(T)}. \end{aligned}$$

We define the operator $T \in \mathcal{L}(\mathcal{Y}, \mathcal{X})$ such that for given data $(\hat{f}_1, \hat{u}_1, \hat{f}_2, \hat{\psi}_2) \in \mathcal{Y}$, $(\hat{u}, \hat{\psi}) = T(\hat{f}_1, \hat{u}_1, \hat{f}_2, \hat{\psi}_2)$ is the unique solution of the problem,

$$\begin{aligned} &(\hat{u}(T), v(T))_{H(T)} + \int_0^T \Big((\hat{u}, v_t)_{H(.)} + a(.;\hat{u}, v) + \eta_1(\hat{u}, v)_{H(.)} \Big) \\ &\quad = (\hat{u}_1, v(0))_H + \int_0^T \langle \hat{f}_1, v \rangle_{U^*(.),U(.)} \qquad \forall\, v \in L^2[0,T;U(.)] \cap H^1[0,T;U^*(.)], \end{aligned} \tag{4.15}$$

$$\begin{aligned} &-(\hat{\psi}_2, v(T))_{H(T)} + \int_0^T \Big((\hat{\psi}, v_t)_{H(.)} + \mu(.;\hat{\psi}, v) + a^*(.;\hat{\psi}, v) + \eta_2(\hat{\psi}, v)_{H(.)} \Big) \\ &\quad = -(\hat{\psi}(0), v(0))_H + \int_0^T \langle \hat{f}_2, v \rangle_{U^*(.),U(.)} \qquad \forall\, v \in L^2[0,T;U(.)] \cap H^1[0,T;U^*(.)]. \end{aligned} \tag{4.16}$$

The parameters η_1, η_2 will be chosen as indicated in the proof of Theorem 4.5. The mapping $\mathcal{G}$ is defined by: Given $(\hat{u}, \hat{\psi}) \in \mathcal{X}$ then $\mathcal{G}(\hat{u}, \hat{\psi}) = (\hat{f}_1, \hat{u}_1, \hat{f}_2, \hat{\psi}_2) \in \mathcal{Y}$ if and only if for all $v \in L^2[0,T;U(.)] \cap H^1[0,T;U^*(.)]$

$$\begin{aligned}
\int_0^T \langle \hat{f}_1, v\rangle_{U^*(.),U(.)} &= -\int_0^T \Big(\langle f, v\rangle_{U^*(.),U(.)} - (1/\alpha)(\hat{\psi}, v)_{H(.)} + \eta_1(\hat{u}, v)_{H(.)} \Big) \\
(\hat{\psi}_1, v(0))_{H(0)} &= (u_0, v(0)) \\
\int_0^T \langle \hat{f}_2, v\rangle_{U^*(.),U(.)} &= -\int_0^T \Big((\hat{u} - z, v)_{H(.)} + \eta_2(\hat{\psi}, v)_{H(.)} \Big) \\
(\hat{\psi}_2, v(T))_{H(T)} &= 0.
\end{aligned}$$

Clearly the pair $(u, \psi) \in \mathcal{X}$ is a solution of the optimality system (3.3) if and only if

$$(u, \psi) + \mathcal{T}\mathcal{G}(u, \psi) = 0.$$

It remains to define the approximating operator $\mathcal{T}_h$. We denote by U_h (independent of time) a finite element subspace of $H_0^1(\Omega)$ satisfying the assumptions of Theorem 4.2, and let $X_h = H^1[0, T; U_h]$. Then,

$$\mathcal{X}_h = X_h \times X_h,$$

and for $(\hat{f}_1, \hat{u}_1, \hat{f}_2, \hat{u}_2) \in \mathcal{Y}$, we define $\mathcal{T}_h(\hat{f}_1, \hat{u}_1, \hat{f}_2, \hat{\psi}_2) = (\hat{u}_h, \hat{\psi}_h) \in \mathcal{X}_h$ if and only if

$$\begin{aligned}
&(\hat{u}_h(T), v_h(T))_{H(T)} + \int_0^T \Big(-\big(\hat{u}_h, v_{ht}\big)_{H(.)} + a(.; \hat{u}_h, v_h) + \eta_1(\hat{u}_h, v_h)_{H(.)} \Big) \\
&\quad = (\hat{u}_1, v_h(0))_H + \int_0^T \langle \hat{f}_1, v_h\rangle_{U^*(.),U(.)} \qquad \forall\, v_h \in H^1[0, T; U_h],
\end{aligned} \tag{4.17}$$

$$\begin{aligned}
&-(\hat{\psi}_2(T), v_h(T))_{H(T)} + \int_0^T \Big((\hat{\psi}_h, v_{ht})_{H(.)} + \mu(.; \hat{\psi}_h, v_h) + a^*(.; \hat{\psi}_h, v_h) + \eta_2(\hat{\psi}_h, v_h)_{H(.)} \Big) \\
&\quad = -(\hat{\psi}_h(0), v_h(0))_H + \int_0^T \langle \hat{f}_2, v_h\rangle_{U^*(.),U(.)} \qquad \forall\, v_h \in H^1[0, T; U_h].
\end{aligned} \tag{4.18}$$

Recall that the approximations are constructed on $U_h \subset H_0^1(\Omega)$ independent of time. Similar to the continuous case, the discrete optimality system now takes the operator form

$$(u_h, \psi_h) + \mathcal{T}_h\mathcal{G}(u_h, \psi_h) = 0.$$

A few remarks about the structure of the operators $\mathcal{T}, \mathcal{T}_h, \mathcal{G}$ follow:

Remark 4.11. *The operator $\mathcal{T}$ contains a forward and a backward in time implicit parabolic equations and it is essentially uncoupled. All coupling terms are contained in operator $\mathcal{G}$. The estimate on $\mathcal{T} - \mathcal{T}_h$ corresponds to two (uncoupled) estimates for the two auxiliary problems (a forward and a backward in time) for suitable choices of η_1, η_2. The addition and subtraction of the term $\eta_i(.,.)_{H(.)}$ into 4.15-4.16 is to guarantee the strict coercivity of the bilinear forms.*

Theorem 4.12. *Let $U_h \subset H_0^1(\Omega)$ be classical finite element subspaces satisfying the standard approximation properties of Proposition 4.2, and $f \in L^2[0, T; U^*(.)]$, $u_0 \in L^2(\Omega)$, $z \in L^2[0, T; H(.)]$ are given data. Let $U(.) \subset H \subset U^*(.)$ be dense embedding of Hilbert spaces with embedding constants independent of t, and let the norm and semi-norm equivalences of Proposition 4.2 hold. Suppose also that the operators $M(.)$,$A(.)$,$A^*(.)$ satisfy Assumptions 2.1,2.4,2.5,2.6. Furthermore, let $(u, \psi) \in \mathcal{X}$ is a regular solution of the optimality system 3.2. Then, there exists a neighborhood of the origin*

$\mathcal{O}$ such that for $h \le h_0$ small enough, $(u_h, \psi_h) \in \mathcal{X}_h$ is a unique solution of the discrete optimality system (4.1), and

$$\|(u,\psi) - (u_h,\psi_h)\|_{\mathcal{X}} \to 0 \qquad \text{as } h \to 0.$$

In addition, if $u, \psi \in L^2[0,T; H^{l+1}(\Omega) \cap U(.)] \cap H^1[0,T; H^{l-1}(\Omega) \cap U(.)]$, then there exists constant $C > 0$, depending to $(1/\alpha), C_a, C_q, C_\mu, C_u$ and the ratio c_a/c_γ such that,

$$\|(u,\psi) - (u_h,\psi_h)\|_{\mathcal{X}} \le Ch^{2l} D^l(u,\psi)$$

where $D^l(u,\psi)$ denote norms of the expected higher parabolic regularity $L^2[0,T; H^{l+1}(\Omega) \cap U(.)] \cap H^1[0,T; H^{l-1}(\Omega) \cap U(.)]$.

Proof. It is clear that $\mathcal{G}$ is a smooth polynomial map from $\mathcal{X}$ to $\mathcal{Y}$. Note also that $D^2\mathcal{G}$ is bounded on all bounded sets of $\mathcal{X}$, and recall that Theorem 4.5, 4.7 imply that

$$\|(\mathcal{T} - \mathcal{T}_h)(\hat{f}_1, \hat{u}_1, \hat{f}_2, \hat{\psi}_1)\|_{\mathcal{X}} \to 0 \qquad \text{as } h \to 0,$$

for appropriate choice of parameters η_1, η_2. Indeed, $\mathcal{T} - \mathcal{T}_h$ essentially compares two uncoupled problems, a forward and an backward in time. Hence, we may apply the estimate of the model problem, with an appropriate choice of $\eta_1 = \eta$ where η is specified in Theorem 4.5, for the forward in time model problem to obtain an estimate of the form Theorem 4.8, while the backward in time problem can be treated similarly. Let $(u,\psi), (\tilde{u}, \tilde{\psi}) \in \mathcal{X}$ and note that the derivative $D\mathcal{G}$ is defined as $D\mathcal{G}(u,\psi) \cdot (\tilde{u}, \tilde{\psi}) = (\tilde{f}_1, \tilde{u}_1, \tilde{f}_2, \tilde{\psi}_1)$ if and only if

$$\begin{aligned}
\int_0^T \langle \tilde{f}_1, v \rangle_{U^*(.),U(.)} &= -\int_0^T \Big(-(1/\alpha)(\tilde{\psi}, v)_{H(.)} + \eta_1 (\tilde{u}, v)_{H(.)} \Big) \\
(\tilde{u}_1, v) &= 0 \\
\int_0^T \langle \tilde{f}_2, v \rangle_{U^*(.),U(.)} &= -\int_0^T \Big((\tilde{u}, v)_{H(.)} + \eta_2 (\tilde{\psi}, v)_{H(.)} \Big) \\
(\tilde{\psi}_1, v) &= 0
\end{aligned}$$

For sufficiently small $\epsilon > 0$, we set $Z = L^2[0,T; L^2(\Omega)] \cap H^1[0,T; H^{-1}(\Omega)] \times H^\epsilon(\Omega)$, and $\mathcal{Z} = \mathcal{Z} \times \mathcal{Z}$. Here, Z is endowed with the time-dependent norm $\|v\|_Z^2 = \|v\|^2_{L^2[0,T;H(.)]} + \|v_t\|^2_{L^2[0,T;U^*(.)]}$. Since $L^2(\Omega) \subset H^{-1}(\Omega)$ with compact embedding, using a standard compactness result in $L^p[0,T;B]$ spaces (see e.g. [28, Theorem 2.1, pp 271]), we obtain that $L^2[0,T; L^2(\Omega)] \cap H^1[0,T; H^{-1}(\Omega)] \subset L^2[0,T; H^{-1}(\Omega)]$ with compact embedding. Therefore, $\mathcal{Z} \subset \mathcal{Y}$ with compact embedding, due to the time-dependent norm, and semi-norm equivalences. Moreover, notice that $D\mathcal{G}(u,\psi) \in \mathcal{Z}$ due to regularity properties of $\tilde{u}, \tilde{\psi}$. Hence, we have verified the assumptions of Theorem 4.9, which clearly imply the desired estimates. □

Remark 4.13. *The above result indicates that the optimality system exhibits the same approximation properties to an uncontrolled model problem, provided that the constants involved to norm equivalences stay under control. However, in many interesting cases such as the Lagrangian moving mesh formulation of convection dominated problems or the dynamic mesh approaches of problems related to the diffusion on manifolds (see e.g. [5] and [4] and relevant discussion within) these constants grow exponentially in terms of various physical variables (see also relevant discussion in section 2) unless re-triangulation of the mesh is performed in every few time-steps. Hence, similar to the uncontrolled case, fully-discrete schemes based on the discontinuous Galerkin (in time) approach are needed to properly model the change of subspaces at every other (or every few) time steps.*

5 Fully-discrete error estimates

In this section we consider approximating optimal control problems which are related to implicit parabolic equations of the form

$$(M(t)u)_t + B(t)u = F, \qquad u|_\Gamma = 0, \qquad u(0,x) = u(0) \tag{5.1}$$

Here, we will assume that the operator $B(.)u : U(.) \to U^*(.)$ induces a strictly coercive bilinear form $b(.;.,.)$ in the sense of Theorem 4.5. In particular, the associated bilinear form $b(.;.,)$ has the following structure:

$$b(.;u,v) = a(.;u,v) + \eta(u,v)_{H(.)}$$

The optimal control problem considered in this section, is to minimize the tracking functional (1.3) subject to equation (5.1) with $F = f + g$. It obvious that the analysis of Sections 3 and 4 is also applicable in this case, while the parameter η can be quantified in terms of constants C_a, C_γ, C_μ and the ratio c_a/c_γ similar to Theorem 4.5. Then, the optimality system of equations take the form, for all $v \in L^2[0,T;U(.)] \cap H^1[0,T;U^*(.)]$,

$$\begin{cases} (u(T),v(T))_{H(T)} + \displaystyle\int_0^T \Big(-(u,v_t)_{H(.)} + a(.;u,v) + \eta(u,v)_{H(.)} \Big) - (u_0, v(0))_H \\ \qquad = \displaystyle\int_0^T \Big(-(1/\alpha)(\psi,v)_{H(.)} + \langle f, v\rangle_{U^*(.),U(.)} \Big) \\ \displaystyle\int_0^T \Big((\psi,v_t)_{H(.)} + a^*(.;\psi,v) + \eta(\psi,v)_{H(.)} + \mu(.;\psi,v) \Big) + (\psi(0),v(0))_H = \displaystyle\int_0^T (u-z,v)_{H(.)} \end{cases} \tag{5.2}$$

Below, we discretize the corresponding optimality system in both space and time, using a discontinuous Galerkin approach. The proposed scheme is discontinuous in time, but conforming in space and the time-discretization is defined in the neighborhood $\mathcal{O}$ of Theorem 4.12 where the corresponding semi-discrete approximation (of arbitrary order) in space was defined. Given a quasi-uniform partition $0 = t^0 < t^1 < ... < t^N = T$ of $[0,T]$ with $\tau = \max_i \tau^i$, $\tau^i \equiv (t^i - t^{i-1})$ and subspaces U_h^n of U, satisfying the standard approximation properties (see e.g. [6]) the DG method constructs approximate solutions $u_h, \psi_h \in \mathcal{U}_h^\tau$ where

$$\mathcal{U}_h^\tau \equiv \ \{u_h \in L^2[0,T;U(.)] \text{ such that } u_h|_{(t^{n-1},t^n]} \in \mathcal{P}_k[t^{n-1},t^n;U_h^n]\}.$$

Here $\mathcal{P}_k[t^{n-1},t^n;U_h^n]$ denotes polynomials of degree k with respect to time and values in U_h^n. Then the fully discrete formulation, is to seek $u_h, \psi_h \in \mathcal{U}_h^\tau$ such that for every $n = 1,...,N$ and $v_h \in \mathcal{P}_k[t^{n-1},t^n;U_h^n]$,

$$\begin{cases} (u^n,v^n)_{H(t^n)} + \displaystyle\int_{t^{n-1}}^{t^n} \Big(-(u_h,v_{ht})_{H(.)} + a(.;u_h,v_h) + \eta(u_h,v_h)_{H(.)} \Big) \\ \qquad -(u^{n-1},v_+^{n-1})_{H(t^{n-1})} = \displaystyle\int_{t^{n-1}}^{t^n} \Big(-(1/\alpha)(\psi_h,v_h)_{H(.)} + \langle f, v_h\rangle_{U^*(.),U(.)} \Big) \\ -(\psi_+^n,v^n)_{H(t^n)} + \displaystyle\int_{t^{n-1}}^{t^n} \Big((\psi_h,v_{ht})_{H(.)} + a^*(.;\psi_h,v_h) + \eta(\psi_h,v_h)_{H(.)} + \mu(.;\psi_h,v_h) \Big) \\ \qquad +(\psi_+^{n-1},v_+^{n-1})_{H(t^{n-1})} = \displaystyle\int_{t^{n-1}}^{t^n} (u_h - z, v_h)_{H(.)}. \end{cases} \tag{5.3}$$

Note that by convention the functions are assumed to be left continuous with right limits, and we denote by u^n, the value of $u_h(t^n) = u_h(t_-^n)$ and by u_+^n the value of $u_h(t_+^n)$. The exact solution are assumed to be $C[0,T;H(.)]$ (see also the parabolic regularity assumption 2.6), so the jump in the error at t^n is denoted by $[e^n] \equiv [u^n] = u_+^n - u^n$. We refer the reader to [30, Chapter 12] for an excellent exposition of the discontinuous (in time) Galerkin methods (see also references within).

Remark 5.1. *The existence of discontinuous Galerkin approximations can be proved easily for low order schemes. For example, recall that when $k = 0$ (piecewise constants in time), the discontinuous Galerkin scheme reduces to the implicit Euler scheme. Alternatively, for the arbitrary k, we first obtain a-priori estimates on u_h, ψ_h at the energy norm and at arbitrary time-points, working similar to the subsequent Theorems 5.4,5.7, and then we proceed by following the approach of [14, Section 3] based on the definition of a fully-discrete optimal control problem. Below, we focus on the derivation of error estimates.*

Next, we derive error estimates of arbitrary order for the above optimality systems of equations. The proof solely relies on suitable projection techniques (see e.g. [4] for the uncontrolled problem) since the lack of regularity for the time-derivative of the discrete problem due to the discontinuities, prohibits the use of the operator theoretic framework of Brezzi-Rappaz-Raviart which was used in the semi-discrete (in space) approximation. These projections, will also allow us to handle the time-dependent norms and inner products associated to our problem. Below, we state the main definitions (see also e.g. [4]).

Definition 5.2.

1. *We denote by $P_n(t) : H(t) \to U_h^n$ the standard weighted L^2 projection from $H(t)$ onto U_h^n, i.e., $(P_n(t)v, v_h)_{H(t)} = (v, v_h)_{H(t)} \quad \forall\, v_h \in U_h^n$.*
2. *We denote by P_n^{loc} the local weighted L^2 projection, by $P_n^{loc} : C[t^{n-1}, t^n; H(.)] \to \mathcal{P}_k[t^{n-1}, t^n; U_h^n]$ that satisfies $(P_n^{loc} v)^n = P_n(t^n)v(t^n)$ and*
$$\int_{t^{n-1}}^{t^n} (v - P_n^{loc} v, v_h)_{H(.)} = 0 \qquad \forall\, v_h \in \mathcal{P}_{k-1}[t^{n-1}, t^n; U_h^n]$$
where we have used the convention that $(P_n^{loc} v)^n \equiv (P_n^{loc} v)(t^n)$.
3. *We denote by P_h^{loc} the projection which consists of all local P_n^{loc} projections, i.e., $P_h^{loc} : C[0, T; H(.)] \to \mathcal{U}_h^\tau$ satisfies $P_h^{loc} v \in \mathcal{U}_h^\tau$ and $(P_n^{loc} v)|_{(t^{n-1}, t^n]} = P_n^{loc}(v|_{[t^{n-1}, t^n]})$.*

Remark 5.3. *For the adjoint equation, the projections of Definition 5.2 (denoted by $P_n^{loc,b}$,$P_h^{loc,b}$ etc) should be modified to handle the backwards in time evolution. In particular, in addition to relation $\int_{t^{n-1}}^{t^n} (v - P_n^{loc} v, v_h)_{H(.)} = 0 \forall\, v_h \in \mathcal{P}_{k-1}[t^{n-1}, t^n; U_h^n]$ we should impose the matching condition at the left end-point. However, note that the approximation properties of both projections are exactly the same, and hence we occasionally use the same notation for both situations. For the basic properties of the above projections we refer the reader to [30, 5] and references within.*

Now, we are ready to prove the main estimate at the energy norm. Throughout the rest of this section we will be tracking the dependence of various constants on the ratio c_a/c_γ and on the parameter α. The proof uses techniques from [4, Theorem 4.3] suitably adjusted for the optimality system.

Theorem 5.4. *Suppose that the Assumptions 2.1, 2.4-2.5-2.6 hold, and let $u, \psi \in L^2[0, T; U(.)] \cap H^1[0, T; U^*(.)]$, $u_h, \psi_h \in \mathcal{U}_h$ denote the solutions of optimality systems (5.2) and (5.3) respectively. Denote by $e = u - u_h$ and $r = \psi - \psi_h$. Then for η satisfying $\eta \geq 4C_\gamma + 4(C_a c_a/c_\gamma) + 6C_\mu + 2C_a$ the* *following estimate holds:*

$$
\begin{aligned}
&|e^N|^2_{H(t^N)} + (1/\alpha)|r^0_+|^2_{H(0)} + \int_0^T \Big(\|e\|^2_{U(.)} + (1/\alpha)\|r(.)\|^2_{U(.)}\Big) + \sum_{i=0}^{N-1} \|[e^i]\|^2_{H(t^i)} + \sum_{i=1}^{N} \|[r^i]\|^2_{H(t^i)} \\
&\le C \max\{1, (1/\alpha^2)\}\Big[\int_0^T \Big(\|(I-P_h^{loc})u\|^2_{U(.)} + \|(I-P_h^{loc})\psi\|^2_{U(.)}\Big) + \int_0^T \|r_p\|^2_{H(.)}dt \\
&\quad + \sum_{i=0}^{N-1} \min\Big\{(C_k/\tau^{i+1}c_\gamma)\|P_{i+1}(I-P_i)u(t^i)\|^2_{U^*(t^i)}, |(I-P_i)u(t^i)|^2_{H(t^i)}\Big\} \\
&\quad + \sum_{i=1}^{N} \min\Big\{(C_k/\tau^{i}c_\gamma)\|P_{i}(I-P_{i+1})\psi(t^i)\|^2_{U^*(t^i)}, |(I-P_i)y(t^i)|^2_{H(t^i)}\Big\} + |e^0|^2_{H(0)}\Big]
\end{aligned}
$$

where C depends on the ratio c_a/c_γ, and on the constants C_a, C_γ, C_μ, C_k, C_u.

Proof. Subtracting (5.3) from (5.2) we obtain the orthogonality condition, for every $n = 1, ..., N$ and $v_h \in \mathcal{P}_k[t^{n-1}, t^n; U_h^n]$,

$$
\begin{cases}
(e^n, v^n)_{H(t^n)} + \int_{t^{n-1}}^{t^n} \Big(-(e, v_{ht})_{H(.)} + a(.; e, v_h) + \eta(e, v_h)_{H(.)}\Big) \\
\qquad -(e^{n-1}, v_+^{n-1})_{H(t^{n-1})} = -(1/\alpha)\int_{t^{n-1}}^{t^n} (r, v_h)_{H(.)} \\
-(r_+^n, v^n)_{H(t^n)} + \int_{t^{n-1}}^{t^n} \Big((r, v_{ht})_{H(.)} + a^*(.; r, v_h) + \eta(r, v_h)_{H(.)} + \mu(.; r, v_h)\Big) \\
\qquad +(r_+^{n-1}, v_+^{n-1})_{H(t^{n-1})} = \int_{t^{n-1}}^{t^n} (e, v_h)_{H(.)}.
\end{cases}
\tag{5.4}
$$

We decompose the error $e = (y - P_n^{loc}y) + (P_n^{loc}y - y_h) \equiv e_p + e_h$, and using the properties of P_n^{loc}, and in particular that $\int_{t^{n-1}}^{t^n}(e_p, v_{ht})_{H(.)} = 0$, since $v_{ht} \in \mathcal{P}_{k-1}[t^{n-1}, t^n; U_h^n]$, we obtain,

$$
\begin{aligned}
&(e_h^n, v^n)_{H(t^n)} + \int_{t^{n-1}}^{t^n} \Big(-(e_h, v_{ht})_{H(.)} + a(.; e_h, v_h) + \eta(e_h, v_h)_{H(.)}\Big) - (e_h^{n-1}, v_+^{n-1})_{H(t^{n-1})} \\
&= ((I - P_{n-1})u(t^{n-1}), v_+^{n-1})_{H(t^{n-1})} - \int_{t^{n-1}}^{t^n} \Big(a(.; e_p, v_h) + \eta(e_p, v_h)_{H(.)}\Big) - (1/\alpha)\int_{t^{n-1}}^{t^n} (r, v_h)_{H(.)}.
\end{aligned}
$$

Setting $v_h = e_h$ in the above equation, and using the assumptions 2.1 and 2.4,

$$
\begin{aligned}
&(1/2)|e_h^n|^2_{H(t^n)} + \int_{t^{n-1}}^{t^n} \Big(c_\gamma|e_h|^2_{U(.)} + (\eta - C_\gamma)|e_h|^2_{H(.)}\Big) + (1/2)|e_{h+}^{n-1} - e_h^{n-1}|^2_{H(t^{n-1})} \\
&\le (1/2)|e_h^{n-1}|^2_{H(t^{n-1})} + \int_{t^{n-1}}^{t^n} \Big(a(.; e_p, e_h) + \eta(e_p, e_h)_{H(.)} + (1/2)\mu(.; e_h, e_h)\Big) \\
&\quad +((I - P_{n-1})u(t^{n-1}), e_+^{n-1})_{H(t^{n-1})} - (1/\alpha)\int_{t^{n-1}}^{t^n} (r, e_h)_{H(.)}.
\end{aligned}
\tag{5.5}
$$

The inequality $ab \le (1/4\delta)a^2 + \delta b^2$, with appropriate $\delta > 0$, and assumptions 2.1,2.4, imply that

$$
\begin{aligned}
&\int_{t^{n-1}}^{t^n} \Big(a(.; e_p, e_h) + \eta(e_p, e_h)_{H(.)} + (1/2)\mu(.; e_h, e_h)\Big) \le \int_{t^{n-1}}^{t^n} \Big((c_\gamma/2)|e_h|^2_{U(.)} + (c_a^2/2c_\gamma)|e_p|^2_{U(.)}\Big) \\
&\quad + \int_{t^{n-1}}^{t^n} \Big(\big((C_a c_a/2c_\gamma) + (\eta/2)\big)|e_p|^2_{H(.)} + \big((c_\gamma C_a/2c_a) + (C_\mu/2) + (\eta/2)\big)|e_h|^2_{H(.)}\Big).
\end{aligned}
$$

Finally for the coupling term, decomposing $r = r_p + r_h$ where $r_p = (I - P_n^{loc,b})\psi$,$r_h = P_n^{loc,b}\psi - \psi_h$, we write $\int_{t^{n-1}}^{t^n} -(1/\alpha)(r, e_h)_{H(.)} = \int_{t^{n-1}}^{t^n} -(1/\alpha)\big((r_p, e_h)_{H(.)} + (r_h, e_h)_{H(.)}\big)$ and note that

$$\Big|\int_{t^{n-1}}^{t^n} -(1/\alpha)(r_p, e_h)_{H(.)}\Big| \leq \int_{t^{n-1}}^{t^n} \Big(1/\alpha^2\eta)|r_p|^2_{H(.)} + (\eta/4)|e_h|^2_{H(.)}\Big).$$

Hence, collecting that last inequalities into (5.5), we arrive to

$$\begin{aligned}
&(1/2)|e_h^n|^2_{H(t^n)} + \int_{t^{n-1}}^{t^n} \Big((c_\gamma/2)|e_h|^2_{U(.)} + ((\eta/4) - C_\gamma - (c_\gamma C_a/2c_a) - (C_\mu/2))|e_h|^2_{H(.)}\Big) \\
&+(1/2)|[e_h^{n-1}]|^2_{H(t^{n-1})} \leq (1/2)|e_h^{n-1}|^2_{H(t^{n-1})} + \int_{t^{n-1}}^{t^n} \Big((c_a^2/2c_\gamma)|e_p|^2_{U(.)} + \big((C_a c_a/2c_\gamma) + (\eta/2)\big)|e_p|^2_{H(.)}\Big) \\
&\quad +((I - P_{n-1})u(t^{n-1}), e_+^{n-1})_{H(t^{n-1})} + \int_{t^{n-1}}^{t^n} (1/\alpha^2\eta)\|r_p\|^2_{H(.)} - (1/\alpha)\int_{t^{n-1}}^{t^n} (r_h, e_h)_{H(.)}. \qquad (5.6)
\end{aligned}$$

It remains to treat the inner product term of the right hand side of equation 5.6. Note that $e_h^{n-1} \in U_h^{n-1}$ and hence

$$\begin{aligned}
((I - P_{n-1})u(t^{n-1}), e_{h+}^{n-1})_{H(t^{n-1})} &= ((I - P_{n-1})u(t^{n-1}), e_{h+}^{n-1} - e_h^{n-1})_{H(t^{n-1})} \\
&\leq |(I - P_{n-1})u(t^{n-1})|^2_{H(t^{n-1})} + (1/4)|[e_h^{n-1}]|^2_{H(t^{n-1})},
\end{aligned}$$

while an alternative bound is obtained by

$$\begin{aligned}
((I - P_{n-1})u(t^{n-1}), e_{h+}^{n-1})_{H(t^{n-1})} &= (P_n(I - P_{n-1})u(t^{n-1}), e_{h+}^{n-1})_{H(t^{n-1})} \\
&\leq \|(P_n(I - P_{n-1})u(t^{n-1})\|_{U^*(t^{t-1})}\|e_{h+}^{n-1}\|_{U(t^{n-1})} \\
&\leq (C(C_k, C_u)/\tau^n c_\gamma)\|(P_n(I - P_{n-1})u(t^{n-1})\|^2_{U^*(t^{t-1})} \\
&\qquad +(c_\gamma/4)\int_{t^{n-1}}^{t^n} \|e_h\|^2_{U(.)}.
\end{aligned}$$

Here, we have used the inverse estimate for functions on $\mathcal{P}_k[t^{n-1}, t^n; U_h^n]$ and the norm equivalence assumption, which states that

$$\|e_{h+}^{n-1}\|^2_{U(t^{n-1})} \leq C_k/\tau^n \int_{t^{n-1}}^{t^n} \|e_h\|^2_{U(t^{n-1})} \leq C(C_k, C_u)/\tau^n \int_{t^{n-1}}^{t^n} \|e_h\|^2_{U(.)}.$$

Now we turn our attention to the backwards in time equation. Decomposing the error of the adjoint variable as $r = r_p + r_h$ with $r_p = (I - P_n^{loc,b})\psi$,$r_h = P_n^{loc,b}\psi - \psi_h$ and working similarly to state variable, we obtain

$$\begin{aligned}
&-(r_{h+}^n, v^n)_{H(t^n)} + \int_{t^{n-1}}^{t^n} \Big((r_h, v_{ht})_{H(.)} + a^*(.; r_h, v_h) + \mu(., r_h, v_h) + \eta(r_h, v_h)_{H(.)}\Big) \\
&= -(r_{h+}^{n-1}, v_+^{n-1})_{H(t^{n-1})} + (r_{p+}^n, v^n)_{H(t^n)} - \int_{t^{n-1}}^{t^n} \Big(a(.; r_p, v_h) + \mu(.; r_p, v_h) + \eta(r_p, v_h)_{H(.)}\Big) \\
&\qquad + \int_{t^{n-1}}^{t^n} (e, v_h)_{H(.)}.
\end{aligned}$$

Setting $v_h = r_h$ and using the assumptions 2.1,2.4 and the inequality $ab \leq (1/4\delta)a^2 + \delta b^2$, for appropriate $\delta > 0$ analogously to the primal variable, we obtain

$$\begin{aligned} &-(1/2)|r_{h+}^{n}|^2_{H(t^n)} + (1/2)|r_{h+}^{n-1}|^2_{H(t^{n-1})} - (1/2)|[r_h^n]|^2_{H(t^n)} \\ &+ \int_{t^{n-1}}^{t^n} \Big((c_\gamma/2)|r_h|^2_{U(.)} + \big((\eta/4) - C_\gamma - (c_\gamma C_a/2c_a) - (3/C_\mu)\big)|r_h|^2_{H(.)}\Big) \\ &\le \int_{t^{n-1}}^{t^n} \Big((c_a^2/2c_\gamma)|r_p|^2_{U(.)} + \big((C_a c_a/2c_\gamma) + (\eta/2) + (C_\mu/2)\big)|r_p|^2_{H(.)}\Big) \\ &\qquad +((I-P_n)\psi(t_+^n), r^n) + \int_{t^{n-1}}^{t^n} (1/\eta)\|e_p\|^2_{H(.)} + \int_{t^{n-1}}^{t^n} (e_h, r_h)_{H(.)}. \end{aligned} \tag{5.7}$$

The inner product term of the right hand side of (5.7) can be treated similarly to (5.6). Multiplying (5.7) by $(1/\alpha)$, adding the resulting inequality to (5.6) and summing from 0 to N, we obtain the desired estimate, after noting the the coupling terms $\int_{t^{n-1}}^{t^n} \big(-(1/\alpha)(r_h, e_h)_{H(.)} + (1/\alpha)(e_h, r_h)_{H(.)}\big)$ are cancelled due to the symmetric property of operators $M(.)$. □

Remark 5.5. *The key feature of the proof is to exploit the presence of the weighted L^2 norm in the functional and the symmetric property of the operator $M(.)$ in order to cancel the terms with the alternative sign.*

Since, we have obtained an estimate on the energy norm $\|.\|_{L^2[0,T;U(.)]}$ for both the state and adjoint variable, the optimality system is now essentially uncoupled. Hence, using a classical "bootstrap" argument, we may obtain estimates at arbitrary time-points, working with each equation separately by applying the techniques of [4, Theorem 4.3]. Recall, that a convenient choice of test functions in order to obtain stability and error estimates at arbitrary time points, is to multiply the equation by $\chi_{[0,t)}u_h$, $\chi_{[0,t)}e_h$ respectively. However, this choice is not available unless t is a partition point. To overcome this difficulty, approximations of such functions need to be constructed. For implicit parabolic equations, this is done in [4, Section 3.1]. The main advantage of this approach within the context of optimal control problems, is that we do not require any additional time-regularity. Below, we state the main results. To simplify the presentation, we consider the interval $[t^{n-1}, t^n)$ and let $t \in [t^{n-1}, t^n)$.

Let $u \in \mathcal{P}_k[t^{n-1}, t^n; U_h^n]$. Then, we define the discrete approximation $\tilde{u}$ of $\chi_{[t^{n-1},t)}u$ such as: $\tilde{u} \in \mathcal{P}_k[t^{n-1}, t^n; U_h^n]$ satisfying

$$\tilde{u}(t^{n-1}) = u(t^{n-1}) \quad \text{and} \quad \int_{t^{n-1}}^{t^n} (\tilde{u}, v)_{H(.)} = \int_{t^{n-1}}^{t} (u, v)_{H(.)}, \qquad \forall\, v \in \mathcal{P}_{k-1}[t^{n-1}, t^n; U_h^n].$$

The following lemma establishes bounds in $L^2[t^{n-1}, t^n; H(.)]$ and $L^2[t^{n-1}, t^n; U(.)]$ norms respectively.

Lemma 5.6. *The mapping $u \to \tilde{u}$ in $\mathcal{P}_k[t^{n-1}, t^n; H(.)]$ is linear, continuous and there exists a constant C_k depending only on k such that*

$$\|\tilde{u}\|_{L^2[t^{n-1},t^n;H(.)]} \le (1 + C_k e^{C_\mu \tau})\|u\|_{L^2[t^{n-1},t^n;H(.)]}.$$

In addition, let the inverse hypothesis constant $C_{inv}(h)$ be defined by

$$C_{inv}(h) = \max_{0\le n\le N} \sup_{u_h \in U_h^n} \sup_{t\in(t^{n-1},t^n]} \frac{|u_h|_{U(t)}}{|u_h|_{H(t)}}.$$

Then, there exists a constant C depending only on C_k, C_μ such that

$$|u|_{L^2[t^{n-1},t^n;U(.)]} \le C(C_k, C_\mu)\Big(C_u^2 |u|_{L^2[t^{n-1},t^n;U(.)]} + \tau C_{inv}(h)|u|_{L^2[t^{n-1},t^n;H(.)]}\Big).$$

Proof. See [4, Lemma 3.2 and Corollary 3.4]. □

Theorem 5.7. *Suppose that the assumptions of Theorem 5.4, and Lemma 5.6 hold. Then, there exists a constant $\hat{C}$ depending on $C(C_k, C_\mu, C_u, \sqrt{c_a}\tau C_{inv}(h))$ and on constant C of Theorem 5.4 such that for $e_h = P_h^{loc}u - u_h$ the following estimate holds:*

$$|e_h(t)|^2_{H(t)} + \int_0^{t^n} c_\gamma |e_h|^2_{U(.)} \leq \hat{C}\max\{1, (1/\alpha^2)\}\Big(|e^0|^2_{H(0)} + \int_0^T \|(I - P_h^{loc})u\|^2_{U(.)}$$
$$+ \sum_{i=0}^{N-1} \min\Big\{(C_k/\tau^{i+1}c_\gamma)\|P_{i+1}(I - P_i)y(t^i)\|^2_{U^*(t^i)}, |(I - P_i)u(t^i)|^2_{H(t^i)}\Big\} + \int_0^T |\psi - \psi_h|^2_{H(.)}\Big)$$

Proof. (*Sketch:*) The proof follows closely the proof of [4, Theorem 4.3], suitably modified to handle the adjoint variable term $\psi - \psi_h$ of the right hand side. Since, we have already obtained an estimate on $\|.\|_{L^2[0,T;U(.)]}$, the estimate at partition points follows easily from (5.5). It suffices to bound $|e(t)|_{H(t)}$ for $t \in (t^{n-1}, t^n]$. Similar to the proof of Theorem 5.4, we decompose the error into $e = (u - P_h^{loc}u) + (P_h^{loc}u - u_h) \equiv e_p + e_h$ to obtain the orthogonality condition,

$$(e_h^n, v^n)_{H(t^n)} + \int_{t^{n-1}}^{t^n} \Big(-(e_h, v_{ht})_{H(.)} + a(.; e_h, v_h) + \eta(e_h, v_h)_{H(.)}\Big) - (e_h^{n-1}, v_+^{n-1})_{H(t^{n-1})}$$
$$= ((I - P_{n-1})u(t^{n-1}), v_+^{n-1})_{H(t^{n-1})} - \int_{t^{n-1}}^{t^n} \Big(a(.; e_p, v_h) + \eta(e_p, v_h)_{H(.)}\Big)$$
$$-(1/\alpha)\int_{t^{n-1}}^{t^n} (\psi - \psi_h, v_h)_{H(.)}.$$

Integrating by parts in time, and using assumption 2.1, we obtain,

$$\int_{t^{n-1}}^{t^n} \Big((e_{ht}, v_h)_{H(.)} + a(.; e_h, v_h) + \eta(e_h, v_h)_{H(.)} + \mu(.; e_h.v_h)\Big) + (e_{h+}^{n-1} - e_h^{n-1}, v_+^{n-1})_{H(t^{n-1})}$$
$$= ((I - P_{n-1})u(t^{n-1}), v_+^{n-1})_{H(t^{n-1})} - \int_{t^{n-1}}^{t^n} \Big(a(.; e_p, v_h) + \eta(e_p, v_h)_{H(.)}\Big)$$
$$-(1/\alpha)\int_{t^{n-1}}^{t^n} (\psi - \psi_h, v_h)_{H(.)}.$$

Setting $v_h \equiv \tilde{e}_h$, where $\tilde{e}_h$ denotes the discrete approximation of $\chi_{[t^{n-1},t)}e_h$, we obtain

$$\int_{t^{n-1}}^{t} (e_{ht}, e_h)_{H(.)} + \int_{t^{n-1}}^{t^n} \mu(.; e_h, \tilde{e}_h) + (e_{h+}^{n-1} - e_h^{n-1}, e_+^{n-1})_{H(t^{n-1})}$$
$$= ((I - P_{n-1})u(t^{n-1}), e_+^{n-1})_{H(t^{n-1})} - \int_{t^{n-1}}^{t^n} \Big(a(.; e_p, \tilde{e}_h) + \eta(e_p, \tilde{e}_h)_{H(.)}\Big)$$
$$-(1/\alpha)\int_{t^{n-1}}^{t^n} (\psi - \psi_h, \tilde{e}_h)_{H(.)} - \int_{t^{n-1}}^{t^n} \Big(a(.; e_p, \tilde{e}_h) + \eta(e_h, \tilde{e}_h)_{H(.)}\Big)$$

Using once more the smoothness assumption 2.1, we finally arrive at

$$(1/2)|e_h(t)|^2_{H(t)} - (1/2)\int_{t^{n-1}}^{t^n} \mu(.; e_h, e_h) + \int_{t^{n-1}}^{t^n} \mu(.; e_h, \tilde{e}_h) - (1/2)|e_{h+}^{n-1}|^2_{H(t^{n-1})} + (1/2)|[e_h^{n-1}]|^2_{H(t^{n-1})}$$
$$= ((I - P_{n-1})u(t^{n-1}), e_+^{n-1})_{H(t^{n-1})} - \int_{t^{n-1}}^{t^n} \Big(a(.; e_p, \tilde{e}_h) + \eta(e_p, \tilde{e}_h)_{H(.)}\Big)$$
$$-(1/\alpha)\int_{t^{n-1}}^{t^n} (\psi - \psi_h, \tilde{e}_h)_{H(.)} - \int_{t^{n-1}}^{t^n} \Big(a(.; e_p, \tilde{e}_h) + \eta(e_h, \tilde{e}_h)_{H(.)}\Big).$$

It remains to bound the terms involving the bilinear form $a(.;.,.)$, the linear form $\mu(.;.,.)$ and the inner products $(.,.)_{H(.)}$. This is done in [4, Theorem 4.3], by using the assumption 2.4, 2.5 combined with the estimates of Lemma 5.6 to bound terms of $\tilde{e}_h$ in terms e_h. The jump terms can be handled similar to Theorem 5.4. It remains to bound the inner product term containing the adjoint variable. For that purpose, note that Cauchy-Schwarz (with $\delta > 0$) and Lemma 5.6 imply

$$\begin{aligned}(1/\alpha)\int_{t^{n-1}}^{t^n}\|\psi-\psi_h\|_{H(.)}\|\tilde{e}_h\|_{H(.)} &\leq (\delta/\alpha)\int_{t^{n-1}}^{t^n}\|\psi-\psi_h\|^2_{H(.)} + (1/4\delta\alpha)\int_{t^{n-1}}^{t^n}\|\tilde{e}_h\|^2_{H(.)} \\ &\leq (\delta/\alpha)\int_{t^{n-1}}^{t^n}\|\psi-\psi_h\|^2_{H(.)} + ((1+C_k e^{C_\mu\tau})/4\delta\alpha)\int_{t^{n-1}}^{t^n}\|e_h\|^2_{H(.)}\end{aligned}$$

Choose $\delta > 0$ such that $((1+C_k e^{C_\mu\tau})/4\delta\alpha) = C_\gamma/4$. The remaining of the proof follows identical to [4, Theorem 4.3]. $\square$

Corollary 5.8. *Under the assumptions of Theorem 5.4, and Lemma 5.6 the following estimate holds:*

$$\begin{aligned}&|e(t)|^2_{H(t)} + \int_0^{t^n} c_a|e|^2_{U(.)} \leq \hat{C}\max\{1,(1/\alpha^2)\}\Big(|e^0|^2_{H(0)} + |e_p(t)|^2_{H(t)} + \int_0^T \|(I-P_h^{loc})u\|^2_{U(.)} \\ &+ \sum_{i=0}^{N-1}\min\Big\{(C_k/\tau^{i+1}c_\gamma)\|P_{i+1}(I-P_i)y(t^i)\|^2_{U^*(t^i)}, |(I-P_i)u(t^i)|^2_{H(t^i)}\Big\} + (1/\alpha^2)\int_0^T|\psi-\psi_h|^2_{H(.)}\Big)\end{aligned}$$

Here, $\hat{C}$ depends on $C(C_k, C_\mu, C_u, \sqrt{c_a}\tau C_{inv}(h))$, the ratio c_a/c_γ and on constant C of Theorem 5.4.

Proof. The proof follows using standard algebra and triangle inequality. $\square$

So far, we have obtained estimates under minimal regularity assumptions for the fully-discrete optimality system of equations, in terms of the projections defined in definition 5.2. We complete this section, by recalling a result from [5, Corollary 4.8] which expresses the error of the projection P_h^{loc} in terms of the local projections P_n^{loc} and hence to the standard weighted L^2 projection.

Lemma 5.9. *Let the spaces $\{H(t)\}_{t=0}^T$ satisfy the assumptions 2.1,2.4, and the inverse hypothesis assumption of definition 5.2. Then, there exist constants C_0, C_1 depending upon $C_0 = C_0(k, e^{C_\mu\tau})$ and $C_1 = C_1(k, C_u, e^{C_\mu\tau}, e^{CT})$ (where C is an algebraic constant) such that the projection $P_h^{loc} : C[0,T;H(.)] \to \mathcal{U}_h^\tau$ of definition 5.2 satisfies,*

$$\begin{aligned}|||u-P_h^{loc}u|||_\infty &\leq C_0\max_{1\leq n\leq N}\Big(\|u-P_n(t^n)u\|_{L^\infty[t^{n-1},t^n;H(t^n)]} + \tau^{k+1}\|u^{(k+1)}\|_{L^\infty[t^{n-1},t^n;H(t^n)]}\Big) \\ &\quad + C_1\sqrt{c_a}(1+C_\mu\tau C_{inv}(h))\Big\{\Big(\sum_{n=1}^N\|u-P_n(t^n)u\|^2_{L^2[t^{n-1},t^n;U(t^n)]}\Big)^{1/2} \\ &\quad + \tau^{k+1}\Big(\sum_{n=1}^N\|u^{(k+1)}\|^2_{L^2[t^{n-1},t^n;U(t^n)]}\Big)^{1/2}\Big\}\end{aligned}$$

when $u^{(k+1)} \in L^\infty[0,T;H] \cap L^2[0,T;U]$, and similarly

$$\begin{aligned}|||u-P_h^{loc}u|||_2 &\leq C_0\Big(\Big(\sum_{n=1}^N\|u-P_n(t^n)u\|^2_{L^2[t^{n-1},t^n;H(t^n)]}\Big)^{1/2} + \tau^{k+1}\Big(\sum_{n=1}^N\|u^{(k+1)}\|^2_{L^2[t^{n-1},t^n;H(t^n)]}\Big)^{1/2}\Big) \\ &\quad + C_1\sqrt{c_a}(1+C_\mu\tau C_{inv}(h))\Big\{\Big(\sum_{n=1}^N\|u-P_n(t^n)u\|^2_{L^2[t^{n-1},t^n;U(t^n)]}\Big)^{1/2} \\ &\quad + \tau^{k+1}\Big(\sum_{n=1}^N\|u^{(k+1)}\|^2_{L^2[t^{n-1},t^n;U(t^n)]}\Big)^{1/2}\Big\}\end{aligned}$$

Here, $u^{(k+1)}$ denotes the $(k+1)^{th}$ time-derivative, and

$$|||u|||_\infty^2 = \sup_{s\in(0,T]} \|u(s)\|_{H(s)}^2 + c_a \int_0^T |e(s)|_{U(s)}^2 ds, \quad |||u|||_2^2 = \int_0^T \|u(s)\|_{H(.)}^2 + c_a \int_0^T \|u(s)\|_{U(.)}^2 ds.$$

Proof. See [5, Corollary 4.8]. □

6 Convergence Rates

Below, we demonstrate the applicability of the results of sections 4 and 5, in the situation of the examples of section 2. Specifically, we state semi-discrete (in space) error estimates for the example 2.1.2 (diffusion on manifolds), and fully-discrete error estimates for the example 2.1.2 (convection diffusion equation in Lagrangian coordinates, with appropriately chosen regularization parameter η). Here we follow the exposition of [4, Section 6].

First, we consider semi-discrete (in space) error estimates for the optimality system related to the example 2.1.2. For that purpose, note that the corresponding time-dependent bilinear forms, inner products, etc are defined by:

$$(u,v)_{H(t)} = (M(t)u,v)_{L^2(S_r)} \equiv \int_{S_r} uvJ$$

and

$$a(.;u,v) = \int_{S_r} \Big(\sigma(\nabla v)^T(F^TF)^{-1}\nabla u - (I - \mathbf{n}\times\mathbf{n})\cdot(\nabla_x \mathbf{V})uv\Big)J.$$

Then, recall that

$$J_t = J(I-\mathbf{n}\times\mathbf{n})\cdot(\nabla_x\mathbf{V}),$$

so if $0 < c_0 \le J(0,.) \le C_0$, we obtain

$$c_0 e^{-Ct} \le J(t,.) \le C_0 e^{Ct}, \qquad \text{with} \qquad C = 2\|\nabla_x \mathbf{V}\|_{L^\infty}.$$

In order to verify Assumptions 2.1-2.4, note first that the semi-norm is defined by

$$|u|_{U(.)}^2 \equiv \int_{S_r} \sigma(\nabla u)^T(F^TF)^{-1}\nabla u J.$$

and that the bilinear form

$$\mu(.;u,v) = \int_{S_r} uvJ_t = \int_{S_r} uv(I-\mathbf{n}\circ\mathbf{n})\cdot\nabla_x\mathbf{V},$$

satisfies Assumption 2.1 with $C_\mu = 2\|\nabla_x\mathbf{V}\|_{L^\infty}$. The norm equivalence on $U(.)$ can be found in [4, Relation 6.2]. Finally, the continuity and coercivity constants, are given by (see also [4, Section 6]),

$$c_a = \sigma, \quad C_a = 2\|\nabla_x\mathbf{V}\|_{L^\infty}, \quad c_\gamma = \sigma, \quad C_\gamma = 2\|\nabla_x\mathbf{V}\|_{L^\infty}.$$

The constants appearing in the adjoint bilinear form $a^*(.;u,v)$ also maintain the same structure, in particular, we point out that the corresponding ratio $c_a/c_\gamma \approx 1$. Combining Theorem 4.12 and the approximation properties of Proposition 4.2, we obtain the following estimate.

Theorem 6.1. *Suppose that $\{\mathcal{R}_h\}_{h>0}$ be a quasi-uniform family of triangulations of S_r, and let $U_h \subset H_0^1(\Omega)$ be a classical finite element space constructed over $\mathcal{T}_h$ by using polynomials of degree $l \ge 0$ on each triangle. Then, under the assumptions of Theorem 4.12, there exists a neighborhood* *of the origin $\mathcal{O}$ such that for $h \leq h_0$ small enough, $(u_h, \psi_h) \in \mathcal{X}_h$ is a unique solution of the discrete optimality system 4.1, and*

$$\|(u,\psi) - (u_h, \psi_h)\|_{\mathcal{X}} \to 0 \qquad \text{as } h \to 0.$$

In addition, if $u, \psi \in L^2[0,T; H^{m+1} \cap U(.)] \cap H^1[0,T; H^{m-1} \cap U(.)]$, then there exists constant $C > 0$, depending to $\|\nabla_x \mathbf{V}\|_{L^\infty}, C_q, C_u, (1/\alpha^2)$ similar to Theorem 4.5 such that,

$$\|(u,\psi) - (u_h, \psi_h)\|_{\mathcal{X}} \leq Ch^{2l} D^l(u,\psi).$$

Here $D^l(u,\psi)$ depends only on norms of u, ψ.

Remark 6.2. *The above estimate states the error for the semi-discrete solution of the optimality system, is as good as the approximation theory enables it to be. Recall, that the Brezzi-Rappaz-Raviart theory compares the estimate of the optimality system to an estimate of a strictly coercive model problem.*

For the example 2.1.1 (convection-diffusion equation in Lagrangian coordinates) we begin by identifying various spaces and constants appearing in our model. First note that we are interested in deriving fully-discrete error estimates in Lagrangian coordinates (t, X) and hence in order to apply our theoretical results we will need to add a regularization parameter $\eta(.,.)_{H(.)}$ to satisfy the strict coercivity of the bilinear form. This is similar to the moving mesh characteristic Galerkin approach of [7, 18]. The spaces $H(t)$ will be taken to be the weighted inner product (in Lagrangian coordinates), i.e.,

$$(\bar{u}, \bar{v})_{H(t)} = \int_{\Omega_r} \bar{u}\bar{v} J(t,.).$$

The continuity of the bilinear form $M(.)$ is understood in the sense

$$\int_{\Omega_r} \bar{u}^{n-1}(t^{n-1}, X)^2 J^{n-1}(t^{n-1}, X) dX = \int_{\Omega} u^2 dx = \int_{\Omega_r} \bar{u}^n(t_+^{n-1}, X)^2 J^n(t_+^{n-1}, X) dX,$$

where $u : \Omega \to \mathbb{R}$ is fixed and $\bar{u}^n(t,X) = u(\chi(t,X))$ denotes the representation of u in the Lagrangian configuration on interval $(t^{n-1}, t^n]$. The weighted semi-norm (principal part) of H^1 is defined by

$$|\bar{u}|^2_{U(.)} = \int_{\Omega_r} |F^{-T} \nabla_X \bar{u}|^2 J dX.$$

Note that performing a changing of variables to the Eulerian coordinates, the last integral equals to $\int_\Omega |\nabla_x u|^2 dx$. The bilinear form $a(.;.,.)$ is identified by

$$a(.;\bar{u},\bar{v}) = \int_{\Omega_r} \Big(- div_x(\mathbf{V})\bar{u}\bar{v} + (\mathbf{V} - \tilde{\mathbf{V}}).(F^{-T}\nabla_X \bar{u})\bar{v} + \epsilon(F^{-T}\nabla_X \bar{u}).(F^{-T}\nabla_X \bar{v})\Big) J.$$

For assumption 2.1, recall that J satisfy $\dot{J} = J div_x(\tilde{\mathbf{V}})$ so in the interval $(t^{n-1}, t^n]$, we easily obtain that

$$e^{-\|div_x(\tilde{\mathbf{V}})\|_{L^\infty}(t-t^{n-1})} \leq J(.,t)/J(.,t_+^{n-1}) \leq e^{\|div_x(\tilde{\mathbf{V}})\|_{L^\infty}(t-t^{n-1})}.$$

Hence, if the reference configuration on each time interval (i.e. the initial condition for the ODE in $(t^{n-1}, t^n]$) is selected to satisfy $\chi(t_+^{n-1}, X) = X$, which implies $J(t_+^{n-1}, .) = 1$ then we easily obtain (see [5, Theorem 4.9]) that

$$c(t)\|\bar{u}\|^2_{L^2(\Omega_r)} \leq (J(.)\bar{u}, \bar{u}) = (M(.)\bar{u}, \bar{u}),$$

with $c(t) = e^{-\|div_x(\tilde{\mathbf{V}})\|_\infty (t-t^{n-1})}$. The norm equivalence on $U(.)$ (assumption 2.4) and the constant C_u can be found in [5, Theorem 4.9]. Finally for constants of assumptions 2.1-2.5, take the form (see [5, Theorem 4.9])

$$c_a = 2\epsilon, \qquad c_\gamma = \epsilon/2, \qquad C_\gamma = \|div_x(\tilde{\mathbf{V}})\|_\infty + \|\mathbf{V} - \tilde{\mathbf{V}}\|_\infty^2/\epsilon$$

and

$$C_a = \max\{1, C_\gamma\}, \qquad C_\mu = \|div_x(\tilde{\mathbf{V}})\|_\infty, \qquad C_{inv}(h) = C(\Omega, k) e^{C_\mu \tau} C_u/h.$$

Theorem 6.3. *Let $\mathcal{U}_h^n$ be a finite element subspaces constructed by piecewise polynomials of degree l in a way to satisfy the classical approximation theory properties as stated in Proposition 4.2. In addition, let $\tau = \max_{i=1,\dots,N} \tau^i$ with $\tau^i = (t^i - t^{i-1})$ be a quasi-uniform partition of the time-interval $[0,T]$. Suppose that the hypotheses of Theorem 5.7 and Lemma 5.6 hold and let the initial value $\bar{u}_h^0$ satisfy $\|\bar{u}_h^0 - P_0(0)\bar{u}(0)\|_{L^2(\Omega_r)} \leq C h^{l+1} \|\bar{u}(0)\|_{H^{l+1}(\Omega_r)}$. Then, for $\eta \geq 4\|div_x(\tilde{\mathbf{V}})\|_\infty + \|\mathbf{V} - \tilde{\mathbf{V}}\|_\infty^2/\epsilon$, there exists a constant C depending upon $C_k, \|div_x(\tilde{\mathbf{V}})\|_\infty$, $\|\mathbf{V} - \tilde{\mathbf{V}}\|_\infty^2/\epsilon$, $e^{\|div_x(\tilde{\mathbf{V}})\|_\infty \tau}$, C_u, C_q and $\sqrt{\epsilon}\tau/h$ such that the approximate solutions of the discrete optimality system computed by the discontinuous Galerkin method satisfy*

$$\begin{aligned} |||\bar{u} - \bar{u}_h|||_2 + (1/\alpha)|||\bar{\psi} - \bar{\psi}_h|||_2 \quad &\leq C \max\{1, (1/\alpha^2)\} \Big\{ h^{l+1} C_1 + \tau^{k+1} C_2 + \sqrt{\epsilon}(1 + \tau/h)\Big(h^l C_3 + \tau^{k+1} C_4\Big) \\ &\quad + h^l \Big(\big(h + \min(h/\sqrt{\tau}, h^2/\tau\sqrt{\epsilon})\big) C_5 + \sqrt{\epsilon}(1 + \tau/h) C_6 \Big) \Big\}. \end{aligned}$$

In addition, at arbitrary points, the following estimate holds:

$$\begin{aligned} |||\bar{u} - \bar{u}_h|||_\infty \quad &\leq C \max\{1, (1/\alpha^2)\} \times \Big\{ \tau^{k+1} \Big(C_7 + \sqrt{\epsilon}(1 + \tau/h) C_4 \Big) \\ &\quad + h^l \Big(\big(h + \min(h/\sqrt{\tau}, h^2/\tau\sqrt{\epsilon})\big) C_5 + \sqrt{\epsilon}(1 + \tau/h) C_6 \Big) \Big\} \end{aligned}$$

where,

$$\begin{aligned}
C_1 &= \Big(\sum_{n=1}^N \|\bar{u}\|^2_{L^2[t^{n-1},t^n;H(t^n)]} \Big)^{1/2} + \Big(\sum_{n=1}^N \|\bar{\psi}\|^2_{L^2[t^{n-1},t^n;H(t^n)]} \Big)^{1/2} \\
C_2 &= \Big(\sum_{n=1}^N \|\bar{u}^{(k+1)}\|^2_{L^2[t^{n-1},t^n;H(t^n)]} \Big)^{1/2} + \Big(\sum_{n=1}^N \|\bar{\psi}^{(k+1)}\|^2_{L^2[t^{n-1},t^n;H(t^n)]} \Big)^{1/2}, \\
C_3 &= \Big(\sum_{n=1}^N \|\bar{u}\|^2_{L^2[t^{n-1},t^n;U(t^n)]} \Big)^{1/2} + \Big(\sum_{n=1}^N \|\bar{\psi}\|^2_{L^2[t^{n-1},t^n;U(t^n)]} \Big)^{1/2}, \\
C_4 &= \Big(\sum_{n=1}^N \|\bar{u}^{(k+1)}\|^2_{L^2[t^{n-1},t^n;U(t^n)]} \Big)^{1/2} + \Big(\sum_{n=1}^N \|\bar{\psi}^{(k+1)}\|^2_{L^2[t^{n-1},t^n;U(t^n)]} \Big)^{1/2}, \\
C_5 &= \max_{1 \leq n \leq N} |\bar{u}|_{L^\infty[t^{n-1},t^n;H^{l+1}(\Omega)]} + \max_{1 \leq n \leq N} |\bar{\psi}|_{L^\infty[t^{n-1},t^n;H^{l+1}(\Omega)]}, \\
C_6 &= \Big(\sum_{n=1}^N |\bar{u}|^2_{L^2[t^{n-1},t^n;H^{l+1}(\Omega_r)]} \Big)^{1/2} + \Big(\sum_{n=1}^N |\bar{\psi}|^2_{L^2[t^{n-1},t^n;H^{l+1}(\Omega_r)]} \Big)^{1/2} \\
C_7 &= \max_{1 \leq n \leq N} \|\bar{u}^{(k+1)}\|_{L^\infty[t^{n-1},t^n;H(t^n)]} + \max_{1 \leq n \leq N} \|\bar{\psi}^{(k+1)}\|_{L^\infty[t^{n-1},t^n;H(t^n)]}.
\end{aligned}$$

Here $\bar{u}^{(k+1)}$, $\bar{\psi}^{(k+1)}$ denotes the $(k+1)^{th}$ time-derivative of $\bar{u}, \bar{\psi}$ respectively.

Proof. The proof follows by combining Theorem 5.4, Corollary 5.7 and Lemma 5.9 and standard algebra. For the jump terms we refer the reader to [5, Corollary 4.10]. Finally, note that $C_{inv}(h) \leq C(\Omega, l)(1/h)$. □

Remark 6.4. *The above estimate is applicable for a variety of choices of discretization parameters τ, h. One of the main features of this approach is that the estimate is still valid even if ϵ is relatively small in terms of τ and h. In case, of fine-meshes, i.e., $h \leq \epsilon$ then the estimate can be simplified. Finally, we would like to point out that the constant C does not depend on exponential terms of $(1/\alpha)$, and hence the estimate remains valid even in case that α is small.*

References

[1] K. Chrysafinos, Moving mesh finite element methods for a distributed optimal control problems for the advection-diffusion equation, *J. Scient. Comput.*, **25** (2005), pp 401-421.

[2] K. Chrysafinos, Discontinuous Galerkin finite element approximations for distributed optimal control problems constrained to parabolic PDE's, *Int. J. Numer. Anal. and Model.*, **4** (2007), pp 690-712.

[3] K. Chrysafinos, M.D. Gunzburger and L.S. Hou, Semidiscrete approximations of optimal Robin boundary control problems constrained by semilinear parabolic PDE, *J. Math. Anal. Appl.*, **323** (2006), pp. 891-912.

[4] K. Chrysafinos and N.J. Walkington, Error estimates for the discontinuous Galerkin methods for implicit parabolic equations, *SIAM J. Numer. Anal.*, **43** (2006), pp. 2478-2499.

[5] K. Chrysafinos and N.J. Walkington, Lagrangian and moving mesh methods for the convection diffusion equations, *ESAIM M^2AN*, **42** (2008), pp. 25-55.

[6] P. Ciarlet, The finite element method for elliptic problems, SIAM Classics, Philadelphia, 2002.

[7] T.F. Dupont and Y. Liu, Symmetric error estimates for moving mesh Galerkin methods for advection-diffusion equations, *SIAM J. Numer. Anal.*, **40** (2002), pp. 914-927.

[8] L. Evans, Partial Differential Equations, AMS, Providence RI, 1998.

[9] A. Fursikov, Optimal control of distributed systems. Theory and applications, AMS, Providence, 2000.

[10] V. Girault and P.-A. Raviart, Finite Element Methods for Navier-Stokes, Springer-Verlag, New York, 1986.

[11] M.D. Gunzburger Perspectives in flow control and optimization, SIAM, Advances in Design and Control, Philadelphia, 2003.

[12] 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.*, **25** (1991), pp. 711-748.

[13] 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), pp. 1913-1945.

[14] 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), pp. 1481-1512.

[15] I. Lasiecka, Rietz-Galerkin approximation of the time optimal boundary control problem for parabolic systems with Dirichlet boundary conditions, *SIAM J. Control and Optim.*, **22** (1984), pp. 477-500.

[16] I. Lasiecka and R. Triggiani, Control theory for partial differential equations, Cambridge University press, 2000.

[17] J.-L. Lions, Some aspects of the control of distributed parameter systems, SIAM publications, 1972.

[18] Y. Liu, R.E. Bank, T.F. Dupont, S. Garcia, and R.F. Santos, Symmetric error estimates for moving mesh mixed methods for advection-diffusion equations, *SIAM J. Numer. Anal.*, **40** (2002), pp. 2270-2291.

[19] W.-B. Liu, H.-P. Ma, T. Tang and N. Yan, A posteriori error estimates for DG time-stepping method for optimal control problems governed by parabolic equations, *SIAM J. Numer. Anal.*, **42** (2004), pp. 1032-1061.

[20] W.-B. Liu and N. Yan, A posteriori error estimates for optimal control problems governed by parabolic equations, *Numer. Math.*, **93** (2003), pp.497-521.

[21] K. Malanowski, Convergence of approximations vs. regularity of solutions for convex, control-constrained optimal-control problems, *Appl. Math. Optim.*, **8** (1981), pp. 69-95.

[22] R.S. McNight and W.E. Bosarge Jr., The Ritz-Galerkin procedure for parabolic control problems, *SIAM J. Control and Optim.*, **11** (1973), pp. 510-524.

[23] D. Meidner and B. Vexler, A priori error estimates for space-time finite element discretization of parabolic optimal control problems. Part I: Problems without control constraints, *SIAM J. Control and Optim.*, **47** (2008), pp. 1150-1177.

[24] P. Neittaanmaki and D. Tiba, Optimal control of nonlinear parabolic systems. Theory, algorithms and applications. M. Dekker, New York, 1994.

[25] A. Rösch, Error estimates for parabolic optimal control problems with control constraints, *Z. Anal. und Anwend.*, **23** (2006), pp. 353-376.

[26] R.E. Showalter, Hilbert space methods for partial differential equations, Pitman, 1979.

[27] R.E. Showalter, Monotone operators in Banach space and nonlinear partial differential equations, AMS publications, Providence, RI 1997.

[28] R. Temam, Navier-Stokes equations, North Holland, 1979.

[29] D. Tiba and F. Tröltzsch, Error estimates for the discretization of the state constrained convex control problems, *Numer. Funct. Anal. Optim.*, **17** (1996), pp. 1005-1028.

[30] V. Thomée, Galerkin finite element methods for parabolic problems, Spinger-Verlag, Berlin, 1997

[31] R. Winther, Initial value methods for parabolic control problems, *Math. Comp.*, **34** (1980), 115-125