SYMMETRIC ERROR ESTIMATES FOR DISCONTINUOUS GALERKIN APPROXIMATIONS FOR AN OPTIMAL CONTROL PROBLEM ASSOCIATED TO SEMILINEAR PARABOLIC PDE'S

KONSTANTINOS CHRYSAFINOS*AND EFTHIMIOS N. KARATZAS†

Abstract. A discontinuous Galerkin finite element method for an optimal control problem having states constrained to semilinear parabolic PDE's is examined. The schemes under consideration are discontinuous in time but conforming in space. It is shown that under suitable assumptions, the error estimates of the corresponding optimality system are of the same order to the standard linear (uncontrolled) parabolic problem. These estimates have symmetric structure and are also applicable for higher order elements.

1. Introduction. The optimal control problem considered here, is associated to the minimization of the tracking functional subject to semi-linear parabolic PDEs. In particular, we seek states y and controls g (of distributed type) such that

$$J(y,g) = \frac{1}{2}\int_0^T \|y-U\|^2_{L^2(\Omega)}dt + \frac{\alpha}{2}\int_0^T \|g\|^2_{L^2(\Omega)}dt \tag{1.1}$$

is minimized subject to the constraints,

$$\begin{cases} y_t - \ \text{div}[A(x)\nabla y] + \phi(y) & = f+g \qquad \text{in } (0,T)\times\Omega \\ y & = 0 \qquad \text{on } (0,T)\times\Gamma \\ y(0,x) & = y_0 \qquad \text{in } \Omega. \end{cases} \tag{1.2}$$

The physical meaning of the optimization problem under consideration is to seek states y and controls g such that y is as close as possible to a given target U. Here, Ω denotes a bounded domain in $\mathbb{R}^2$, with Lipschitz boundary Γ, y_0, f denote the initial data and the forcing term respectively, and α is a penalty parameter which measures the size of the control. The nonlinear mapping ϕ satisfies certain continuity and monotonicity properties, and $A(x) \in C^1(\bar{\Omega})$ is a symmetric matrix valued function that is uniformly positive definite. The scope of this work is the error analysis of the first order necessary conditions (optimality system) of the above optimal control problem by using a discontinuous (in time), Galerkin (DG) scheme. The corresponding optimality system consists of a primal (forward in time) equation and an adjoint (backwards in time) equation which are coupled through an optimality condition, and nonlinear terms (see, e.g [26, 29, 38, 47, 53]).

The main aim is to show that the DG approximations of the optimality system exhibit similar approximation properties to the standard linear (uncontrolled) parabolic

†AMS subject classification:65M60,49J20
Keywords and phrases: Symmetric Error Estimates, Discontinuous Galerkin, Distributed Control, Semi-linear Parabolic PDE's.
*Address: National Technical University of Athens, School of Applied Mathematical and Physical Sciences, Department of Mathematics, Zografou Campus, Athens 15780, Greece, chrysafinos@math.ntua.gr
†Address: National Technical University of Athens, School of Applied Mathematical and Physical Sciences, Department of Mathematics, Zografou Campus, Athens 15780, Greece, karmakis@central.ntua.gr
‡DCDS-b0832: Revised version: Submitted 19 December 2011.

equation. In particular, it is shown that the error of the DG approximations is as good as the regularity of the solutions and the approximation properties of the subspaces enables it to be, for suitable data f, y_0, U.

This is achieved by proving the following symmetric estimate, which states that,

$$\|\text{error}\|_X \leq C\Big(\|\text{in. data error}\|_{L^2(\Omega)} + \|\text{best approx. error}\|_X + \|\text{subsp. error}\|_{X_1}\Big).$$

Here, $\|.\|_X = \|.\|_{L^\infty[0,T;L^2(\Omega)]} + \|.\|_{L^2[0,T;H^1(\Omega)]}$, and $\|.\|_{X_1}$ denotes a norm related to a possible change of finite element subspaces every other (or every few) time steps and can be omitted when the same subspaces are being used in every time step. The term $\|\text{best approx. error}\|_X$ is posed in terms of suitable local L^2 projections and allows optimal rates of convergence when the solution is sufficiently smooth. The constant C does not depend exponentially on quantities of the form $1/\alpha$. The dependance upon α of various constants appearing in these estimates is essential to the underlying optimal control problem and hence it should be carefully tracked. In particular, in most computational and practical engineering examples, we are interested for small values of the parameter α, and in certain cases even comparable to the discretization parameter h.

The structure of the estimate is similar to the work of [11] which concerns the DG approximations of linear (uncontrolled) parabolic PDE's, and it leads to optimal error estimates in terms of the regularity of the solutions and the approximation theory of the chosen subspaces.

The proof of the main estimate, is based on estimates of an auxiliary and essentially uncoupled system together with a "boot-strap" argument and stability estimates at arbitrary time-points under minimal regularity assumptions. The key element of the proposed methodology is the use of a "duality" type of argument for discontinuous time-stepping schemes, to facilitate the decoupling of the optimality system. In particular, using the adjoint variable as test function in the primal equation, and the primal variable as test function in the adjoint equation, we first show that

$$\|\text{error}\|^2_{L^2[0,T;L^2(\Omega)]} \leq \|\text{best approx. error}\|^2_X + \alpha^{1/2}\|\text{error}\|^2_{L^2[0,T;H^1(\Omega)]}.$$

Then, for α suitably small, we apply a "boot-strap" argument to obtain the desired symmetric estimate. To our best knowledge the above symmetric estimates and their particular structure are new within our optimal control setting.

The motivation for using a DG approach, stems from its performance in a vast area of problems where the given data satisfy low regularity properties, such as optimal control problems. Furthermore, the concept of symmetric error estimates can be effectively capture the interplay between regularity of solutions and approximation properties of the subspaces. Such estimates are also recently applicable to a variety of problems such as error analysis of moving meshes, Lagrangian moving mesh methodologies (see e.g. [18, 42]) and can be viewed as generalization of the classical Céa Lemma (see e.g. [15]). In addition, discontinuous (in time) schemes accommodate the use of different subspaces in each time step, and hence basic adaptivity ideas, in a natural way. In the recent works of [7, 8, 40, 41, 44, 45] discontinuous Galerkin schemes were analyzed for distributed optimal control problems constrained to linear parabolic PDE's, while the case of semi-linear constraints is analyzed in [9, 48]. In [9], convergence of discontinuous time-stepping schemes for optimal control problems (without control constraints) related to semi-linear parabolic PDE's is studied, under minimal regularity assumptions on the data and growth assumptions on the semi-linear term. In the very recent work of [48] first order (in time) error estimates for the controls are presented for an optimal control problem related to semi-linear parabolic PDE, with control constraints, in case that the initial data belong to $H_0^1(\Omega)\cap L^\infty(\Omega)$ under weak hypothesis on semi-linear term. The controls are discretized by piecewise constants in time and space, however the analysis is also applicable when piecewise constants (in time) piecewise linears (in space) are being used. For the state equation, the lowest order ($k = 0$) discontinuous Galerkin (in time) combined with standard conforming finite elements (in space), are being used. The first-order (in time) estimates presented in [48] successfully address a variety of difficulties due to the presence of control constraints, and the corresponding nonconvexity. The estimates and the analysis of [48] are different compared to the ones presented here. Our work primarily focuses on the development of estimates that possess the symmetric structure (and their advantageous features described above) for the associated optimality system.

Below, we give a brief description of other related results.

1.1. Related results. Several problems with distributed controls have been studied before analytically in [26, 29, 35, 38, 39, 47, 53] (see also references within). Several results related to the analysis of numerical algorithms for optimal control problems were studied in [4, 5, 6, 14, 16, 17, 25, 28, 30, 31, 33, 34, 36, 37, 43, 49, 51, 52, 53, 55, 56].

A posteriori estimates for DG schemes were studied in [40, 41] for optimal control problems related to linear parabolic PDE's, while in [44] an adaptive space-time finite element algorithm is constructed and analyzed. A priori error estimates for an optimal control problem of distributed type, having states constrained to the heat equation are presented in [45] while in [7, 8], a priori error estimates for DG schemes for the tracking problem related to linear parabolic PDE's and implicit parabolic PDE's respectively, with non-selfadjoint possibly time dependent coefficients are established. In [46] a Petrov-Galerkin Crank-Nicolson scheme is applied to an optimal control problem with control constraints related to linear parabolic PDE's, while in [3] a Crank-Nicolson formulation is analyzed. In both papers, second order rates of convergence are obtained.

There is an abundant literature concerning DG schemes for the solution of parabolic equations without applying controls (see e.g. [50] and references therein). The relation of the discontinuous Galerkin method to adaptive techniques was studied in detail in [20, 21, 50]. Some results related to finite element approximation of semi-linear and general nonlinear parabolic problems are presented in [1, 19, 22, 23].

1.2. Synopsis. An outline of this paper follows. After introducing the necessary notation in section 2, the optimal control problem and its corresponding optimality system of equations are described in section 3. In section 4, we formulate the discrete optimal control problem and state the key stability estimates at arbitrary time points under minimal regularity assumptions for the state and adjoint variables. In section 5, error estimates on the energy norm and at arbitrary time-points are obtained, for an auxiliary (and essentially uncoupled) system of parabolic PDE's, using L^2 projection techniques. Then, utilizing the estimates on the auxiliary system, the stability estimates of section 4, and a "boot-strap" argument, we obtain estimates for the nonlinear optimality system. This technique allow us to derive fully-discrete error estimates of arbitrary order provided that the natural parabolic regularity is present.

Our work is concluded by presenting a simple numerical experiment, which validates our basic estimate.

2. Preliminaries. We use standard notation for Hilbert spaces $L^2(\Omega)$, $H^s(\Omega)$, $0 < s \in \mathbb{R}$, $H_0^1(\Omega) \equiv \{v \in H^1(\Omega) : v|_\Gamma = 0\}$, related norms and inner products (see e.g. [24, Chapter 5]). We denote by $H^{-1}(\Omega)$ the dual of $H_0^1(\Omega)$ and the corresponding duality pairing by $\langle .,. \rangle$. For any Banach space X, we denote by $L^p[0,T;X], L^\infty[0,T;X]$ the standard time-space spaces, endowed with norms:

$$\|v\|_{L^p[0,T;X]} = \left(\int_0^T \|v\|_X^p dt \right)^{\frac{1}{p}}, \quad \|v\|_{L^\infty[0,T;X]} = \operatorname{esssup}_{t\in[0,T]} \|v\|_X.$$

The set of all continuous functions $v : [0,T] \to X$, is denoted by $C[0,T;X]$, with norm defined by $\|v\|_{C[0,T;X]} = \max_{t\in[0,T]} \|v(t)\|_X$. Finally, we denote by $H^1[0,T;X]$,

$$\|v\|_{H^1[0,T;X]} = \left(\int_0^T \|v\|_X^2 dt \right)^{\frac{1}{2}} + \left(\int_0^T \|v_t\|_X^2 dt \right)^{\frac{1}{2}} \le C < \infty,$$

and by $W(0,T)$ the solution space $W(0,T) = L^2[0,T;H_0^1(\Omega)] \cap H^1[0,T;H^{-1}(\Omega)]$ with norm

$$\|v\|_{W(0,T)}^2 = \|v\|_{L^2[0,T;H^1(\Omega)]}^2 + \|v_t\|_{L^2[0,T;H^{-1}(\Omega)]}^2.$$

The bilinear form associated to our operator, is defined by

$$a(y,v) = \int_\Omega A(x) \nabla y \nabla v dx \qquad \forall\, y, v \in H^1(\Omega),$$

and satisfies the standard coercivity and continuity conditions

$$a(y,y) \ge \eta \|y\|_{H^1(\Omega)}^2, \qquad a(y,v) \le C_c \|y\|_{H^1(\Omega)} \|v\|_{H^1(\Omega)} \qquad \forall\, y, v \in H_0^1(\Omega).$$

A weak formulation of (1.2) is then defined as follows: Given f, g, y_0 we seek $y \in W(0,T)$ such that

$$\begin{aligned} (y(T), v(T)) + \int_0^T \Big(-\langle y, v_t \rangle + a(y,v) + \langle \phi(y), v \rangle \Big) dt \\ = (y_0, v(0)) + \int_0^T \Big(\langle f, v \rangle + (g, v) \Big) dt, \end{aligned} \tag{2.1}$$

for all $v \in W(0,T)$. The data satisfy the minimal regularity assumptions which guarantee the existence of a weak solution $y \in W(0,T)$, i.e.,

$$f \in L^2[0,T;H^{-1}(\Omega)], \qquad y_0 \in L^2(\Omega)$$

while the distributed control g will be sought in $L^2[0,T;L^2(\Omega)]$. Note that under the above regularity assumptions one can only show convergence of the discrete schemes see [9, Section 3] (even in the uncontrolled case). For error estimates, additional regularity assumptions are needed in order to guarantee rates of convergence. In particular, we will assume that $y \in L^\infty[0,T;L^4(\Omega)]$, which typically requires that $y_0 \in H_0^1(\Omega)$, $f \in L^2[0,T;L^2(\Omega)]$. The choice of the control space significantly simplifies the implementation of the finite element algorithm, since it leads to an algebraic optimality condition. Hence, it avoids the use of spaces of fractional order, or the solution of an extra PDE which typically occurs when other norms of g are included in the functional (see e.g. [29]).

For the subsequent analysis it suffices that the target $U \in L^2[0,T;L^2(\Omega)]$. However in most cases U is actually smoother, since the target typically corresponds to the solution a parabolic PDE, and hence it can be assumed that $U \in W(0,T)$. For the analysis of our discrete schemes, the semi-linear term is required to fulfill the following structural assumptions.

ASSUMPTION 2.1. *(a) For convergence of the state variable: The semi-linear term $\phi \in \mathcal{C}^1(\mathbb{R};\mathbb{R})$ satisfies,*

$$\phi'(s) \geq 0, \quad \big|\phi(s)\big| \leq C|s|^p, \quad \big|\phi'(s)\big| \leq C\big|s\big|^{p-1}, \quad s\phi(s) \geq C\big|s\big|^{p+1}, \quad \textit{for } 1 < p \leq 3.$$

(b) For convergence of the state and adjoint variable: In addition to (a), ϕ' be Lipschitz continuous, with Lipschitz constant C_L, or $\phi \in C^2(\mathbb{R};\mathbb{R})$ with $|\phi''(s)| \leq C|s|^{p-2}$ for $2 < p \leq 3$.
(c) If the semi-linear also includes time-space coefficients, i.e., $\phi(s) \equiv \phi(t,x,s) : [0,T]\times\Omega\times\mathbb{R} \to \mathbb{R}$ then, in addition to (a)-(b), $\phi(0), \phi'(0)$ are required to be uniformly bounded.

REMARK 2.2. *Convergence can be shown by simply assuming growth and monotonicity conditions of Assumption 2.1 (a)-(b) on ϕ, ϕ' (see [9, Section 3]). The Lipschitz continuity assumption on ϕ' is imposed only to minimize technicalities. Most of the results presented here, are still valid under the weaker assumptions of [48], provided that the initial data belong to $H^1_0(\Omega) \cap L^\infty(\Omega)$. We refer the reader to [53] (see also references within) for a detailed analysis of possible assumptions on the semi-linear term and on the regularity of the data. Here, we have chosen to impose the minimal regularity assumptions that guarantee the existence on the corresponding discrete solution on the space $L^\infty[0,T;L^2(\Omega)] \cap L^2[0,T;H^1_0(\Omega)]$.*

We close this preliminary section, by recalling generalized Hölder's and Young's inequalities which will be used subsequently.
Generalized Hölder's Inequality: For any measurable set E, of any dimension and for $(1/s_1) + (1/s_2) + (1/s_3) = 1$, $s_i \geq 1$,

$$\int_E f_1 f_2 f_3 dE \leq \|f_1\|_{L^{s_1}(E)} \|f_2\|_{L^{s_2}(E)} \|f_3\|_{L^{s_3}(E)}.$$

Young's Inequality: For any $a, b \geq 0$, $\delta > 0$, $ab \leq \delta a^2 + (1/4\delta)b^2$

3. The continuous optimal control problem.

In this section, we formulate the optimal control problem and state results regarding the existence of (an) optimal solution(s) and of the corresponding optimality system (first order necessary conditions). We refer the reader to [53] (see also references within) for an excellent overview regarding existence / uniqueness and issues related to first and second order necessary and sufficient conditions.

3.1. Existence of optimal solution.

First, we quote a result regarding the solvability of the weak problem (2.1) on the natural energy space under minimal regularity assumptions.

THEOREM 3.1. *Let $f \in L^2[0,T;H^{-1}(\Omega)]$, $y_0 \in L^2(\Omega)$, $g \in L^2[0,T;L^2(\Omega)]$. Then, there exists a unique solution $y \in W(0,T)$ which satisfies the following energy estimate*

$$\|y\|_{W(0,T)} \leq C\Big(\|f\|_{L^2[0,T;H^{-1}(\Omega)]} + \|y_0\|_{L^2(\Omega)} + \|g\|_{L^2[0,T;L^2(\Omega)]}\Big).$$

Proof. The proof is standard (see e.g. [14, 24, 57]). □

Next, we state the definition of the set of admissible solutions $\mathcal{A}_{ad}$ and of the (local) optimal pair.

DEFINITION 3.2.

1. *Given data $f \in L^2[0,T;H^{-1}(\Omega)]$, $y_0 \in L^2(\Omega)$, and target $U \in L^2[0,T;L^2(\Omega)]$, then (y,g) is said to be an admissible element (pair) if $y \in W(0,T)$, $g \in L^2[0,T;L^2(\Omega)]$ satisfy (2.1). (Note that $J(y,g)$ is bounded, due to Theorem 3.1).*

2. *Given data $f \in L^2[0,T;H^{-1}(\Omega)]$, $y_0 \in L^2(\Omega)$, and target $U \in L^2[0,T;L^2(\Omega)]$ we seek pair $(y,g) \in \mathcal{A}_{ad}$ such that $J(y,g) \leq J(w,h)\,\forall\,(w,h) \in \mathcal{A}_{ad}$, when $\|y-w\|_{W(0,T)} + \|g-h\|_{L^2[0,T;L^2(\Omega)]} \leq \delta$ for $\delta > 0$ appropriately chosen.*

Below, we state the main result concerning the existence of an optimal solution for the minimization of the functional (1.1).

THEOREM 3.3. *Suppose $y_0 \in L^2(\Omega)$, $f \in L^2[0,T;H^{-1}(\Omega)]$, $U \in L^2[0,T;L^2(\Omega)]$. Then, the optimal control problem has solution $(y,g) \in W(0,T) \times L^2[0,T;L^2(\Omega)]$.*

Proof. Similar to [14, 26, 38, 53]. □

REMARK 3.4. *The solution to optimal control problems having states constrained to nonlinear parabolic PDE's is in general not unique. However, we note that under additional assumptions on the data of the control problem and the structure of the semi-linear term it is possible to prove that there exists a unique optimal control g (see e.g. [39, Chapter 3, pp 43]), and that the corresponding optimality system admits a unique solution.*

3.2. The continuous optimality system. Suppose now that $(y,g) \in \mathcal{A}_{ad}$ is a (local) optimal solution in the sense of Definition 3.2. Then, an optimality system corresponding to the optimal control problem of Definition 3.2 can be easily derived based on well known Lagrange multiplier techniques (see e.g. [14, 26, 38, 47]). Given f, y_0, U satisfying the assumptions of Definition 3.2, we seek $y, \mu \in W(0,T)$ such that for all $v \in W(0,T)$,

$$\begin{cases} (y(T),v(T)) + \displaystyle\int_0^T \Big(-\langle y, v_t\rangle + a(y,v) + \langle \phi(y), v\rangle\Big)dt \\ \qquad = (y_0, v(0)) + \displaystyle\int_0^T \Big(\langle f, v\rangle + (g,v)\Big)dt \\ y(0,x) = y_0 \end{cases} \tag{3.1}$$

$$\begin{cases} \displaystyle\int_0^T \Big(\langle \mu, v_t\rangle + a(\mu,v) + \langle \phi'(y)\mu, v\rangle\Big)dt = -(\mu(0),v(0)) + \int_0^T (y-U,v)dt \\ \mu(T,x) = 0 \end{cases} \tag{3.2}$$

$$\int_0^T (\alpha g + \mu, u)dt = 0 \qquad \forall u \in L^2[0,T;L^2(\Omega)]. \tag{3.3}$$

REMARK 3.5. *Note that due to optimality condition we obtain that the control g is actually smoother, i.e., $g = -(1/\alpha)\mu \in W(0,T)$. The later can be used to obtain improved regularity results for the primal and adjoint variables via "boot-strap" argument, when additional regularity on U, f, y_0 is available.*

4. The discrete optimal control problem.

4.1. The fully-discrete optimal control problem. The fully-discrete approximations are constructed on a partition $0 = t^0 < t^1 < \ldots < t^N = T$ of $[0,T]$. On each time interval $(t^{n-1}, t^n]$, of length $\tau_n \equiv t^n - t^{n-1}$, a subspace U_h^n of $H_0^1(\Omega)$ is specified, and it is assumed that each U_h^n satisfies the classical approximation theory results (see e.g. [15]). We also assume that the time-steps are quasi-uniform, i.e., there exists $0 \leq \theta \leq 1$, such that $\min_{n=1,\ldots,N} \tau_n \geq \theta \max_{n=1,\ldots,N} \tau_n$. Now, we seek approximate solutions who belong to the space

$$\mathcal{U}_h = \{y_h \in L^2[0,T;H_0^1(\Omega)] : y_h|_{(t^{n-1},t^n]} \in \mathcal{P}_k[t^{n-1}, t^n; U_h^n]\}.$$

Here $\mathcal{P}_k[t^{n-1}, t^n; U_h^n]$ denotes the space of polynomials of degree k or less having values in U_h^n. We also use the following notational abbreviation, $y_{h,\tau} \equiv y_h$, $\mathcal{U}_{h,\tau} \equiv \mathcal{U}_h$ etc. The discretization of the control can be effectively achieved through the discretization of the adjoint variable μ. However, we point out that the only regularity assumption on the discrete control is $g_h \in L^2[0,T;L^2(\Omega)]$.

By convention, the functions of $\mathcal{U}_h$ are left continuous with right limits and hence will subsequently write (abusing the notation) y^n for $y_h(t^n) = y_h(t^n_-)$, and y^n_+ for $y(t^n_+)$. The above notation will also be used for the error $e = y - y_h$ function. Due to a well known embedding result $W(0,T) \subset C[0,T;L^2(\Omega)]$ (see e.g. [24, Chapter 5]), the exact solution y is in $C[0,T;L^2(\Omega)]$, so that the jump in the error at t^n, denoted by $[e^n]$, is $[e^n] = [y^n] = y^n_+ - y^n$.

The discrete state equation can be defined as follows: Under the assumptions of Definition 3.2, we seek state $y_h \in \mathcal{U}_h$, such that for any $g_h \in L^2[0,T;L^2(\Omega)]$,

$$\begin{aligned}(y^n, v^n) &+ \int_{t^{n-1}}^{t^n} \Big(-\langle y_h, v_{ht}\rangle + a(y_h, v_h) + (\phi(y_h), v_h)\Big) dt \\ &= (y^{n-1}, v_+^{n-1}) + \int_{t^{n-1}}^{t^n} \Big(\langle f, v_h\rangle + (g_h, v_h)\Big) dt \qquad \forall\, v_h \in \mathcal{P}_k[t^{n-1}, t^n; U_h^n],\end{aligned} \tag{4.1}$$

for $n = 1, ..., N$. The discrete admissible set $\mathcal{A}_{ad}^d$ and the discrete (local) optimal control problem is now defined analogously to the continuous problem.

DEFINITION 4.1. *Suppose that the assumptions of Section 2 hold.*

1. $\mathcal{A}_{ad}^d \equiv \{(y_h, g_h) \in \mathcal{U}_h \times L^2[0,T;U_h^n]$ *such that (4.1) holds*$\}$.
2. *Discrete (local) Optimal Pair: We seek pair* $(y_h, g_h) \in \mathcal{A}_{ad}^d$ *such that* $J(y_h, g_h) \leq J(w_h, u_h)$ *for all* $(w_h, u_h) \in \mathcal{A}_{ad}^d$ *when* $\|y_h - w_h\|_{L^2[0,T;H_0^1(\Omega)]} + \|y_h - w_h\|_{L^\infty[0,T;L^2(\Omega)]} + \|g_h - u_h\|_{L^2[0,T;L^2(\Omega)]} \leq \delta'$ *for* $\delta' > 0$ *appropriately chosen.*

Let $\tilde{y}_h$ be the solution of 4.1 without control. Without loss of generality, it is understood that the pair $(\tilde{y}_h, 0) \in \mathcal{A}_{ad}^d$, and δ' are chosen in a way to guarantee that

$J(y_h, g_h) \leq J(\tilde{y}_h, 0)$. The proof of existence of optimal solution of the discrete problem and its corresponding discrete optimality system of equations (first order necessary conditions) require stability estimates for the solution of (4.1), under minimal regularity assumptions (see e.g. [9, Section 3]). These stability estimates are also needed for the derivation of error estimates. The $\|y_h\|_X \equiv \|y_h\|_{L^\infty[0,T;L^2(\Omega)]} + \|y_h\|_{L^2[0,T;H^1(\Omega)]}$ norm is used as the natural energy norm associated to the DG formulation, since the discrete time-derivative does not possesses any meaningful regularity due to the presence of discontinuities.

4.2. Stability estimates. Now we are ready to state stability estimates for the discrete optimal control problem. Under an additional assumption on the semi-linear term, we derive a stability bound, which improves the dependence of τ upon the penalty parameter α compared to the result of [9, Lemma 3.6].

ASSUMPTION 4.2. *Suppose that $\{t^n\}_{n=0}^N$ denotes a quasi-uniform partition of $[0,T]$. In addition to Assumption 2.1, we assume that ϕ satisfies the following assumption: For all $n = 1, ..., N$ and $s_1, s_2 \in L^2[t^{n-1}, t^n; L^2(\Omega)]$, with $\|s_1 - s_2\|_{L^2[t^{n-1},t^n;L^2(\Omega)]} \leq \epsilon$, for some $\epsilon > 0$, there exists $C_L > 0$ (algebraic constant) such that*

$$\|\phi(s_1) - \phi(s_2)\|_{L^2[t^{n-1},t^n;L^2(\Omega)]} \leq C_L \|s_1 - s_2\|_{L^2[t^{n-1},t^n;L^2(\Omega)]}.$$

REMARK 4.3. *In the remaining of this paper, we will denote by C_L constants that depend only upon Lipschitz constants of Assumptions 2.1 and 4.2, and by C_k constants that depend upon k. Both constants can be different in different appearances.*

LEMMA 4.4. *Suppose that $y_0 \in L^2(\Omega)$, $U \in L^2[0,T;L^2(\Omega)]$, $f \in L^2[0,T;H^{-1}(\Omega)]$ are given functions, and let ϕ satisfy Assumptions 2.1, and 4.2. If $(y_h, g_h) \in \mathcal{U}_h \times L^2[0,T;U_h^n]$ denotes a solution pair of the discrete (local) optimal control problem, then*

$$\int_0^T \|y_h - U\|^2_{L^2(\Omega)} dt + (\alpha/2) \int_0^T \|g_h\|^2_{L^2(\Omega)} dt$$
$$\leq C\Big(\|y^0\|^2_{L^2(\Omega)} + (1/\eta) \int_0^T \|f\|^2_{H^{-1}(\Omega)} dt + \int_0^T \|U\|^2_{L^2(\Omega)} dt\Big) \equiv C_{st}$$

where C is a constant depending only on Ω. In addition, for all $n = 1, ..., N$

$$\|y^n\|^2_{L^2(\Omega)} + \sum_{i=0}^{n-1} \|[y^i]\|^2_{L^2(\Omega)} + \int_0^{t^n} \eta \|y_h\|^2_{H^1(\Omega)} dt \leq D_{yst},$$

with $D_{yst} \equiv C_{st} \max\{1, 1/\alpha^{1/2}\}$. Let $\tau \equiv \max_{i=1,..,n} \tau_i$, with $\tau_i = t^i - t^{i-1}$. If $\tau \leq \min\{C_k/8C_L C_{st}^{1/2}, C_k \alpha^{1/2}/8\}$, then the following estimate holds:

$$\|y_h\|^2_{L^\infty[0,T;L^2(\Omega)]} \leq C D_{yst},$$

where C depends on $(C_c/\eta), C_k$ and Ω but not on α, τ, h.

Proof. For the first two estimates we simply note that $J(y_h, g_h) \leq J(\tilde{y}_h, 0) \equiv (1/2)\int_0^T \|\tilde{y}_h - U\|^2_{L^2(\Omega)} dt$, where $\tilde{y}_h$ corresponds to the solution of (4.1) without control. The estimate on $\tilde{y}_h$ follows from [11, Section 2]. For the second estimate, we set

$v_h = y_h$ into (4.1), and use Young's inequality to obtain

$$(1/2)\|y^n\|^2_{L^2(\Omega)} + (1/2)\|[y^{n-1}]\|^2_{L^2(\Omega)} + \eta \int_{t^{n-1}}^{t^n} \|y_h\|^2_{H^1(\Omega)} dt \leq (1/2)\|y^{n-1}\|^2_{L^2(\Omega)}$$
$$+ (1/(4\alpha^{1/2})) \int_{t^{n-1}}^{t^n} \|y_h\|^2_{L^2(\Omega)} dt + \alpha^{1/2} \int_{t^{n-1}}^{t^n} \|g_h\|^2_{L^2(\Omega)} dt.$$

The estimate now follows by adding the above inequalities and using the first estimate. For the estimate at arbitrary points, the proof uses ideas of [12]. For completeness, we sketch the proof. Set $v_h = \bar{y}_h$ into (4.1), where $\bar{y}_h$ is the exponential interpolant of $e^{-\rho(t-t^{n-1})}y_h$ of y_h (for some $\rho > 0$) and defined as in Appendix A.1. Then, the definition of $\bar{y}_h$ allows to obtain

$$\int_{t^{n-1}}^{t^n} (y_{ht}, \bar{y}_h) dt = \int_{t^{n-1}}^{t^n} (y_{ht}, y_h) e^{-\rho(t-t^{n-1})} dt \tag{4.2}$$
$$= (1/2)\|y^n\|^2_{L^2(\Omega)} e^{-\rho(t^n-t^{n-1})} - (1/2)\|y^{n-1}\|^2_{L^2(\Omega)} + (\rho/2) \int_{t^{n-1}}^{t^n} \|y_h(t)\|^2_{L^2(\Omega)} e^{-\rho(t-t^{n-1})} dt.$$

Hence, integrate by parts with respect to time in (4.1), and using (4.2), we obtain

$$(1/2)\|y^n\|^2_{L^2(\Omega)} e^{-\rho(t-t^{n-1})} + (1/2)\|[y^{n-1}]\|^2_{L^2(\Omega)} - (1/2)\|y^{n-1}\|^2_{L^2(\Omega)}$$
$$+ (\rho/2) \int_{t^{n-1}}^{t^n} \|y_h(t)\|^2_{L^2(\Omega)} e^{-\rho(t-t^{n-1})} dt + \int_{t^{n-1}}^{t^n} \langle \phi(y_h), \bar{y}_h \rangle dt$$
$$\leq \int_{t^{n-1}}^{t^n} |a(y_h, \bar{y}_h)| dt + \int_{t^{n-1}}^{t^n} |\langle f, \bar{y}_h \rangle| dt + \int_{t^{n-1}}^{t^n} |(g_h, \bar{y}_h)| dt.$$

Using Lemma A.2, we may bound $\bar{y}_h$ in terms of y_h in various norms. In particular, using Young's inequalities, we obtain

$$(1/2)\|y^n\|^2_{L^2(\Omega)} e^{-\rho(t-t^{n-1})} + (1/2)\|[y^{n-1}]\|^2_{L^2(\Omega)} - (1/2)\|y^{n-1}\|^2_{L^2(\Omega)} \tag{4.3}$$
$$+ (\rho/2) \int_{t^{n-1}}^{t^n} \|y_h(t)\|^2_{L^2(\Omega)} e^{-\rho(t-t^{n-1})} dt + \int_{t^{n-1}}^{t^n} \langle \phi(y_h), \bar{y}_h \rangle dt$$
$$\leq C_k \int_{t^{n-1}}^{t^n} \Big(\|f\|^2_{H^{-1}(\Omega)} + (C_c + \eta)\|y_h\|^2_{H^1(\Omega)} + \alpha^{1/2}\|g_h\|^2_{L^2(\Omega)} + (1/\alpha^{1/2})\|y_h\|^2_{L^2(\Omega)} \Big) dt.$$

It remains to bound the semi-linear term. For this purpose, using Assumption 2.1, we obtain,

$$\int_{t^{n-1}}^{t^n} \langle \phi(y_h), \bar{y}_h \rangle dt \geq \int_{t^{n-1}}^{t^n} \langle \phi(y_h) - \phi(\bar{y}_h), \bar{y}_h \rangle dt.$$

Moving the last integral on the right hand side of (4.3) we obtain a bound as follows: Lemma A.2, implies that the difference $y_h - \bar{y}_h$ remains small. In particular, using the previously derived estimate on $\|y_h\|_{L^2[0,T;L^2(\Omega)]}$ we may bound $\|y_h - \bar{y}_h\|_{L^2[t^{n-1},t^n;L^2(\Omega)]} \leq C_k \rho \tau_n \|y_h\|_{L^2[t^{n-1},t^n;L^2(\Omega)]} \leq C_k \rho \tau C_{st}^{1/2}$. Therefore, we deduce from Assumption 4.2, and Hölder's inequality

$$\int_{t^{n-1}}^{t^n} \langle \phi(y_h) - \phi(\bar{y}_h), \bar{y}_h \rangle dt \leq C_L \|y_h - \bar{y}_h\|_{L^2[t^{n-1},t^n;L^2(\Omega)]} \|\bar{y}_h\|_{L^2[t^{n-1},t^n;L^2(\Omega)]}$$
$$\leq C_k C_L \rho \tau_n C_{st}^{1/2} \int_{t^{n-1}}^{t^n} \|y_h\|^2_{L^2(\Omega)} dt.$$

Collecting the above inequalities into (4.3), we obtain

$$\begin{aligned}
&(1/2)\|y^n\|^2_{L^2(\Omega)}e^{-\rho(t-t^{n-1})} + (1/2)\|[y^{n-1}]\|^2_{L^2(\Omega)} - (1/2)\|y^{n-1}\|^2_{L^2(\Omega)} \\
&+(\rho/2)\int_{t^{n-1}}^{t^n}\|y_h(t)\|^2_{L^2(\Omega)}e^{-\rho(t-t^{n-1})}dt \\
&\leq C_k\int_{t^{n-1}}^{t^n}\Big(\|f\|^2_{H^{-1}(\Omega)} + (C_c+\eta)\|y_h\|^2_{H^1(\Omega)} + \alpha^{1/2}\|g_h\|^2_{L^2(\Omega)}\Big)dt \\
&+\Big((1/\alpha^{1/2}) + C_kC_L\rho\tau_nC_{st}^{1/2}\Big)\int_{t^{n-1}}^{t^n}\|y_h\|^2_{L^2(\Omega)}dt \\
&\leq C_k\int_{t^{n-1}}^{t^n}\Big(\|f\|^2_{H^{-1}(\Omega)} + (C_c+\eta)\|y_h\|^2_{H^1(\Omega)} + \alpha^{1/2}\|g_h\|^2_{L^2(\Omega)}\Big)dt \\
&+\tau_n\Big((1/\alpha^{1/2}) + C_kC_L\rho\tau_nC_{st}^{1/2}\Big)\|y_h\|^2_{L^\infty[t^{n-1},t^n;L^2(\Omega)]}.
\end{aligned}$$

Hence, selecting $\rho = 1/\tau_n$ and using the inverse estimate $\|y_h\|^2_{L^\infty[t^{n-1},t^n;L^2(\Omega)]} \leq C_k/\tau_n\int_{t^{n-1}}^{t^n}\|y_h(t)\|^2_{L^2(\Omega)}$, we observe that the last term on the left hand side can be bounded from below by,

$$\begin{aligned}
&(\rho/2)\int_{t^{n-1}}^{t^n}\|y_h(t)\|^2_{L^2(\Omega)}e^{-\rho(t-t^{n-1})}dt \geq (e^{-1}/2\tau_n)\int_{t^{n-1}}^{t^n}\|y_h(t)\|^2_{L^2(\Omega)}dt \\
&\geq C_k\|y_h\|^2_{L^\infty[t^{n-1},t^n;L^2(\Omega)]}.
\end{aligned}$$

It remains to bound the last term at the right hand side. Choosing $\tau_n > 0$ in a way hide this term on the left hand side, at the right hand side, i.e., $C_{st}^{1/2}C_kC_L\tau_n \leq (C_k/8)$ and $(\tau_n/\alpha^{1/2}) \leq (C_k/8)$, i.e., for $\tau_n \leq \min\{\big(C_k/8C_LC_{st}^{1/2}, (\alpha^{1/2}C_k/8)\}$ we obtain,

$$\begin{aligned}
&(1/4)\|y_h\|^2_{L^\infty[t^{n-1},t^n;L^2(\Omega)]} \leq \|y^{n-1}\|^2_{L^2(\Omega)} \\
&\quad +C_k\int_{t^{n-1}}^{t^n}\Big(\|f\|^2_{H^{-1}(\Omega)} + (C_c+\eta)\|y_h\|^2_{H^1(\Omega)} + \alpha^{1/2}\|g_h\|^2_{L^2(\Omega)}\Big)dt.
\end{aligned}$$

The estimate now follows by using the previously derived estimates at the energy norm and at partition points. □

Remark 4.5. *The Assumption 4.2 is also helpful in order to minimize technicalities in the subsequent derivation of symmetric error estimates. However, we note that if the growth condition is satisfied with exponent $1 \leq p \leq 2$, it can be easily shown that $\|\phi(y_h) - \phi(\bar{y}_h)\|_{L^2[t^{n-1},t^n;L^2(\Omega)]} \leq C(C_{st}, C_k)\|y_h - \bar{y}_h\|_{L^2[t^{n-1},t^n;L^2(\Omega)]}$.*

Similar to the case of [9, Theorem 3.8], (where ϕ satisfies growth and monotonicity conditions) the following convergence result can be established when the same subspaces are being used at every time interval, i.e., $U_h^n = U_h \subset H_0^1(\Omega)$, for $n = 1, ..., N$, under minimal regularity assumptions.

Theorem 4.6. *Given fixed h and partition $0 = t^0 < t^1 < ... < t^N = T$ of $[0,T]$, with $\tau = \max_{i=1,...,N}\tau_i$, satisfying the assumptions of Lemma 4.4, and let the Assumption 2.1 hold. Suppose also that $f \in L^2[0,T;H^{-1}(\Omega)]$, $y_0 \in L^2(\Omega)$, $U \in L^2[0,T;L^2(\Omega)]$ and let $\alpha > 0$. Then, for $U_h^n \equiv U_h \subset H_0^1(\Omega)$ and for quasi-uniform time-steps, we obtain,*

- *There exist $y_h \in \mathcal{U}_h$ and $g_h \in L^2[0,T;L^2(\Omega)]$ such that the pair (y_h, g_h) satisfies the discrete equation (4.1) and the functional $J(y_h, g_h)$ is minimized.*

- *The discrete pair (y_h, g_h) converges as $\tau, h \to 0$ to solution (y, g) of the continuous optimal control problem, in the following sense:*

$$y_h \to y \quad \text{weakly in } L^2[0,T;H_0^1(\Omega)] \quad y_h \to y \quad \text{weakly-* in } L^\infty[0,T;L^2(\Omega)]$$
$$y_h \to y \quad \text{strongly in } L^2[0,T;L^2(\Omega)] \quad g_h \to g \quad \text{weakly in } L^2[0,T;L^2(\Omega)].$$

Proof. See [9, Theorem 3.8]. ☐

REMARK 4.7. *The stability estimates under minimal regularity assumptions are valid even when different subspaces are being used at every time interval. The convergence result of [9, Theorem 3.8] is based on a discrete compactness argument of Walkington (see [54, Theorem 3.1]) for discontinuous time-stepping schemes which is established when $U_h^n \equiv U_h$. However it is possible to extend the main result even in case of different subspaces. We note also that the proof of Theorem 4.6 requires only the growth and monotonicity assumptions of Assumption 2.1.*

4.3. The fully-discrete optimality system. The fully-discrete optimality system is defined as follows: We seek $y_h, \mu_h \in \mathcal{U}_h$ such that for $n = 1, ..., N$ and for every $v_h \in \mathcal{P}_k[t^{n-1}, t^n; U_h^n]$,

$$(y^n, v^n) + \int_{t^{n-1}}^{t^n} \Big(- \langle y_h, v_{ht} \rangle + a(y_h, v_h) + (\phi(y_h), v_h) \Big) dt \tag{4.4}$$
$$= (y^{n-1}, v_+^{n-1}) + \int_{t^{n-1}}^{t^n} \Big(\langle f, v_h \rangle + (g_h, v_h) \Big) dt$$

$$-(\mu_+^n, v^n) + \int_{t^{n-1}}^{t^n} \Big(\langle \mu_h, v_{ht} \rangle + a(v_h, \mu_h) + (\phi'(y_h)\mu_h, v_h) \Big) dt \tag{4.5}$$
$$= -(\mu_+^{n-1}, v_+^{n-1}) + \int_{t^{n-1}}^{t^n} (y_h - U, v_h) dt$$

$$\int_0^T (\alpha g_h + \mu_h, u_h) dt = 0 \qquad \forall\, u_h \in L^2[0,T;U_h^n]. \tag{4.6}$$

Here, $y^0 = y_{h0}, \mu_+^N = 0, f, U$ are given data, and y_{h0} denotes an approximation of y_0.

REMARK 4.8. *For low order schemes ($k = 0$, or $k = 1$) the proof of existence of the discrete optimality system can be derived by standard techniques. For high order schemes, we refer the reader to [10, Section 4].*

REMARK 4.9. *Note that testing the optimality condition (4.6) with functions of polynomial in time structure, we may easily see that (4.6) is equivalent to $\int_{t^{n-1}}^{t^n} (\alpha g_h + \mu_h, v_h) = 0$ for all $v_h \in \mathcal{P}_k[t^{n-1}, t^n; U_h^n]$, and $n = 1, ..., N$.*

The remaining of this section is devoted to stability estimates on the adjoint variable μ_h. These estimates will play a crucial role in the subsequent analysis of error estimates for the fully-discrete optimality system.

LEMMA 4.10. *Suppose that $y_0 \in L^2(\Omega)$, $U \in L^2[0,T;L^2(\Omega)]$, $f \in L^2[0,T;H^{-1}(\Omega)]$ are given functions, let ϕ satisfy Assumptions 2.1-4.2. If (y_h, μ_h) satisfy (4.4)-(4.5)-(4.6) then*

$$\int_0^T \|\mu_h\|_{L^2(\Omega)}^2 dt \le C_{st}\alpha, \quad \|\mu_+^0\|_{L^2(\Omega)}^2 + \sum_{i=1}^N \|[\mu^i]\|_{L^2(\Omega)}^2 + \eta \int_0^T \|\mu_h\|_{H^1(\Omega)}^2 dt \le C_{st}\alpha^{1/2}$$

and for $n = 1, ..., N$, $\|\mu_+^{n-1}\|^2_{L^2(\Omega)} \le C_{st}\alpha^{1/2}$, where C_{st} is defined in Lemma 4.4. Suppose that in addition to the assumptions of Lemma 4.4, $\tau \equiv \max_{i=1,..,n} \tau_i$, satisfies $\big((D_{yst}C_L^2C_k^2/4\eta) + (C_k/4\alpha^{1/2})\big)\tau \le (1/4)$. Then, we obtain,

$$\|\mu_h\|^2_{L^\infty[0,T;L^2(\Omega)]} \le CC_{st}\alpha^{1/2} \equiv D_{\mu st}$$

where C does not depend on α, τ, h, but only on $C_c/\eta, C_k$ and Ω, and D_{yst} denotes the stability constant of Lemma 4.4.

Proof. (*Sketch:*) The first two estimates are identical to [9, Lemma 3.8]. For the estimate at arbitrary time, we proceed as follows. Similar to [9, Section 4], we set $v_h = \bar{\mu}_h$ where $\bar{\mu}_h$ is the exponential interpolant $e^{-\rho(t^n-t)}\mu_h$ of μ_h (for some $\rho > 0$) and defined as in Appendix A.1 (suitably modified to handle the backwards in time problem). Then, the analog of (4.3), takes the form

$$\begin{aligned}
&(1/2)\|\mu_+^{n-1}\|^2_{L^2(\Omega)}e^{-\rho(t-t^{n-1})} + (1/2)\|[\mu^n]\|^2_{L^2(\Omega)} - (1/2)\|\mu_+^n\|^2_{L^2(\Omega)} \\
&+(\rho/2)\int_{t^{n-1}}^{t^n} \|\mu_h(t)\|^2_{L^2(\Omega)}e^{-\rho(t-t^{n-1})}dt + \int_{t^{n-1}}^{t^n} \langle \phi'(y_h)\mu_h, \bar{\mu}_h\rangle dt \\
&\le C_k \int_{t^{n-1}}^{t^n} \Big(\|\mu_h\|^2_{H^1(\Omega)} + (C_k/\alpha^{1/2})\|\mu_h\|^2_{L^2(\Omega)} + \alpha^{1/2}\|y_h - U\|^2_{L^2(\Omega)}\Big)dt.
\end{aligned} \tag{4.7}$$

It remains to treat the semi-linear term. Note that adding and subtracting μ_h, the semi-linear term takes the form,

$$\int_{t^{n-1}}^{t^n} \langle \phi'(y_h)\mu_h, \bar{\mu}_h\rangle dt = \int_{t^{n-1}}^{t^n} \langle \phi'(y_h)\mu_h, \bar{\mu}_h - \mu_h\rangle dt + \int_{t^{n-1}}^{t^n} \langle \phi'(y_h)\mu_h, \mu_h\rangle dt.$$

Hence, we may drop the last term due to the monotonicity of ϕ, and move the first term at the right hand side. Then, using the Lipschitz continuity of ϕ', the interpolation inequality $\|.\|^2_{L^4(\Omega)} \le C\|.\|_{L^2(\Omega)}\|.\|_{H^1(\Omega)}$, Hölder's inequality, and Appendix A.1, we obtain,

$$\begin{aligned}
&\int_{t^{n-1}}^{t^n} |\langle \phi'(y_h)\mu_h, \bar{\mu}_h - \mu_h\rangle| dt \le C_L \int_{t^{n-1}}^{t^n} \|y_h\|_{L^2(\Omega)}\|\mu_h\|_{L^4(\Omega)}\|\bar{\mu}_h - \mu_h\|_{L^4(\Omega)}dt \\
&\le CC_LD_{yst}^{1/2} \int_{t^{n-1}}^{t^n} \|\mu_h\|^{1/2}_{L^2(\Omega)}\|\mu_h\|^{1/2}_{H^1(\Omega)}\|\bar{\mu}_h - \mu_h\|^{1/2}_{L^2(\Omega)}\|\bar{\mu}_h - \mu_h\|^{1/2}_{H^1(\Omega)}dt \\
&\le C_kC_LD_{yst}^{1/2}\rho\tau_n\|\mu_h\|_{L^2[t^{n-1},t^n;L^2(\Omega)]}\|\mu_h\|_{L^2[t^{n-1},t^n;H^1(\Omega)]}.
\end{aligned}$$

Therefore, using Young's inequality with $\delta > 0$, we deduce that

$$\int_{t^{n-1}}^{t^n} |\langle \phi'(y_h)\mu_h, \bar{\mu}_h - \mu_h\rangle| dt \le (C_k^2C_L^2D_{yst}\rho^2\tau_n^2/4\eta)\|\mu_h\|^2_{L^2[t^{n-1},t^n;L^2(\Omega)]} + \eta\int_{t^{n-1}}^{t^n} \|\mu_h\|^2_{H^1(\Omega)}dt.$$

Then, combining the last three relations into (4.7) and selecting $\rho = (1/\tau_n)$, we obtain the desired estimate working identical to Lemma 4.4. ◻

Remark 4.11. *We close this section by noting that the discrete stability bounds for the adjoint variable scale better in terms of the parameter α compared to stability constant of the state variable, as expected.*

The rest of the paper is devoted in proving that the DG approximations of the optimality system exhibit the same rate of convergence to the related (uncontrolled) linear parabolic PDE, for appropriate data f, u_0, U and the parameter α.

5. Error estimates for the optimality system. The key ingredient of the proof will be the stability estimate at arbitrary time-points of section 4, along with estimates for an auxiliary optimality system (based on suitable L^2 projection techniques), and a "duality argument" in order to treat the nonlinear terms. In order to obtain an actual rate of convergence more regularity is needed.

ASSUMPTION 5.1. *Let (y, g) be an optimal pair in the sense of Definition 3.2. In addition, let $y_0 \in H_0^1(\Omega)$, $f \in L^2[0, T; L^2(\Omega)]$ and assume that $\alpha^{1/2}\|y\|^2_{L^\infty[0,T;L^4(\Omega)]} \le C_d$, where C_d is constant depending only upon data f, U, y_0, the constants C_c, η and the domain Ω.*

REMARK 5.2. *The above assumption implies a mild restriction on the size of y, in terms of the penalty parameter α and the given data. We refer the reader to [57] for a detailed analysis of regularity results for semi-linear parabolic PDEs. Analogous $L^\infty[0, T; H^1(\Omega)]$ stability results for the discrete optimal control problem, and for the optimality system (4.4)-(4.5)-(4.6) will be studied in detail elsewhere.*

5.1. An auxiliary optimality system. First, we define an auxiliary optimality system which will help uncoupling the discrete optimality system. Let $w_h, z_h \in \mathcal{U}_h$ be defined as the solutions of the following system. Given data f, U, y_0, and initial conditions $w_{h0} = y_{h0}$, where y_{h0} denote the initial approximation of y_0, $z_+^N = 0$, we seek $w_h, z_h \in \mathcal{U}_h$ such that for $n = 1, ..., N$ and for all $v_h \in \mathcal{P}_k[t^{n-1}, t^n; U_h^n]$,

$$(w^n, v^n) + \int_{t^{n-1}}^{t^n} \Big(- \langle w_h, v_{ht}\rangle + a(w_h, v_h) + \langle \phi(y), v_h\rangle \Big) dt \tag{5.1}$$
$$= (w^{n-1}, v_+^{n-1}) + \int_{t^{n-1}}^{t^n} \Big(\langle f, v_h\rangle - (1/\alpha)(\mu, v_h) \Big) dt,$$

$$-(z_+^n, v^n) + \int_{t^{n-1}}^{t^n} \Big(\langle z_h, v_{ht}\rangle + a(z_h, v_h) + \langle \phi'(y)\mu, v_h\rangle \Big) dt \tag{5.2}$$
$$= -(z_+^{n-1}, v_+^{n-1}) + \int_{t^{n-1}}^{t^n} (w_h - U, v_h) dt.$$

The solutions $w_h, z_h \in \mathcal{U}_h$ exist since $\phi(y), \phi'(y)\mu$ belong at least to $L^2[0, T; H^{-1}(\Omega)]$, due to Assumptions 2.1-4.2 and the regularity of $y, \mu \in W(0, T)$. The solutions of the auxiliary optimality system play the role of "global projections" onto $\mathcal{U}_h$. The basic estimate on the energy norm of $y - w_h, \mu - z_h$ will be derived in terms of local L^2 projections using techniques of [11, Section 2] into the auxiliary system (3.1)-(3.2),(5.1)-(5.2). A key feature of these estimates is that they are valid under minimal regularity assumptions. The following standard projections associated to DG method (see e.g. [50]) are needed.

DEFINITION 5.3. *(1) The projection $\mathbb{P}_n^{loc} : C[t^{n-1}, t^n; L^2(\Omega)] \to \mathcal{P}_k[t^{n-1}, t^n; U_h^n]$ satisfies $(\mathbb{P}_n^{loc} v)^n = P_n v(t^n)$, and*

$$\int_{t^{n-1}}^{t^n} (v - \mathbb{P}_n^{loc} v, v_h) = 0, \qquad \forall\, v_h \in \mathcal{P}_{k-1}[t^{n-1}, t^n; U_h^n]. \tag{5.3}$$

Here we have used the convention $(\mathbb{P}_n^{loc} v)^n \equiv (\mathbb{P}_n^{loc} v)(t^n)$ and $P_n : L^2(\Omega) \to U_h^n$ is the orthogonal projection operator onto $U_h^n \subset H_0^1(\Omega)$.

(2) The projection $\mathbb{P}_h^{loc} : C[0,T; L^2(\Omega)] \to \mathcal{U}_h$ *satisfies*

$$\mathbb{P}_h^{loc} v \in \mathcal{U}_h \text{ and } (\mathbb{P}_h^{loc} v)|_{(t^{n-1},t^n]} = \mathbb{P}_n^{loc}(v|_{[t^{n-1},t^n]}).$$

For the backwards in time problem a modification of the above projection (still denoted by $\mathbb{P}_n^{loc}$) will be needed. In particular, in addition to relation (5.3), we need to impose the "matching condition" on the left, i.e., $(\mathbb{P}_n^{loc} v)_+^{n-1} = P_n v(t_+^{n-1})$ instead of imposing the condition on the right. Note that the projection of Definition 5.3 can be viewed as the one step DG approximation of $v_t = f$ on the interval $(t^{n-1}, t^n]$ with exact initial data $v(t^{n-1})$ and $f = v_t$ specified, while the modified projection for the backwards in time stems from the one step DG approximation of the backwards in time ODE, with given terminal data. Recall that due to [50, Theorem 12.1] or [13] these projections satisfy the expected approximation properties. Below, we state the main result for the auxiliary problem.

THEOREM 5.4. *Let* $f \in L^2[0,T; H^{-1}(\Omega)], y_0 \in L^2(\Omega)$, *and* $U \in L^2[0,T; L^2(\Omega)]$ *be given, and let Assumption 2.1 hold. Let* $y, \mu \in W(0,T)$ *be the solutions of (3.1)-(3.2) and* $w_h, z_h \in \mathcal{U}_h$ *be the solutions of (5.1)-(5.2) computed using the DG scheme. Denote by* $e_1 = y - w_h$, $r_1 = \mu - z_h$ *and let* $e_p \equiv y - \mathbb{P}_h^{loc} y$, $r_p = \mu - \mathbb{P}_h^{loc}\mu$, *where* $\mathbb{P}_h^{loc}$ *is defined in Definition 5.3. Then, there exists an algebraic constant* $C > 0$ *depending only on* Ω *such that,*

$$\eta \|e_1\|^2_{L^2[0,T;H^1(\Omega)]} + \sum_{i=0}^{N-1} \|[e_1^i]\|^2_{L^2(\Omega)} \leq C\Big(\|e_1^0\|^2_{L^2(\Omega)} + (C_c^2/\eta)\|e_p\|^2_{L^2[0,T;H^1(\Omega)]} \Big)$$
$$+ \sum_{i=0}^{N-1} 2\min\Big(\|(I-P_i)y(t^i)\|^2_{L^2(\Omega)}, (1/(\tau_{i+1}\eta))\|P_{i+1}(I-P_i)y(t^i)\|^2_{H^{-1}(\Omega)} \Big),$$
$$\eta \|r_1\|^2_{L^2[0,T;H^1(\Omega)]} + \sum_{i=1}^{N} \|[r_1^i]\|^2_{L^2(\Omega)} \leq C\Big((1/\eta)\|e_1\|^2_{L^2[0,T;L^2(\Omega)]} + (C_c^2/\eta)\|r_p\|^2_{L^2[0,T;H^1(\Omega)]} \Big)$$
$$+ \sum_{i=1}^{N} 2\min\Big(\|(I-P_{i+1})\mu(t^i)\|^2_{L^2(\Omega)}, (1/(\tau_i\eta))\|P_i(I-P_{i+1})\mu(t^i)\|^2_{H^{-1}(\Omega)} \Big).$$

Here, $w_{0h} = y_{0h}$, *where* y_{0h} *denotes an approximation of* y_0, $\tau_i = t^i - t^{i-1}$, P_n *denotes the* L^2 *projection on* U_h^n *and we have used the convention* $P_0 \equiv P_1$, $P_{N+1} \equiv P_N$.

Proof. Throughout this proof, we denote by $e_1 = y - w_h, r_1 = \mu - z_h$ and we split e_1, r_1 to $e_1 \equiv e_{1h} + e_p \equiv (\mathbb{P}_h^{loc} y - w_h) + (y - \mathbb{P}_h^{loc} y)$, $r_1 \equiv r_{1h} + r_p \equiv (\mathbb{P}_h^{loc}\mu - z_h) + (\mu - \mathbb{P}_h^{loc}\mu)$, where $\mathbb{P}_h^{loc}$ is defined in Definition 5.3. Using the above notation, and subtracting (5.1) from (3.1), and (5.2) from (3.2) we obtain the orthogonality condition: for $n = 1, ..., N$

$$(e_1^n, v^n) + \int_{t^{n-1}}^{t^n} \Big(-\langle e_1, v_{ht}\rangle + a(e_1, v_h) \Big) dt = (e_1^{n-1}, v_+^{n-1}), \tag{5.4}$$

$$-(r_{1+}^n, v^n) + \int_{t^{n-1}}^{t^n} \Big(\langle r_1, v_{ht}\rangle + a(r_1, v_h) \Big) dt = -(r_{1+}^{n-1}, v_+^{n-1}) + \int_{t^{n-1}}^{t^n} (e_1, v_h) dt, \tag{5.5}$$

for all $v_h \in \mathcal{P}_k[t^{n-1}, t^n; U_h^n]$. Note that the orthogonality condition (5.4) is essentially uncoupled and identical to the orthogonality condition of [11, Relation (2.6)].

Hence applying [11, Theorem 2.2], we derive the first estimate. In a similar way, the orthogonality condition (5.5) is equivalent to:

$$
\begin{aligned}
-(r^n_{1h+}, v^n) + \int_{t^{n-1}}^{t^n} \Big(\langle r_{1h}, v_{ht}\rangle + a(r_{1h}, v_h)\Big) dt = -(r^{n-1}_{1h+}, v^{n-1}_+) \qquad (5.6)\\
+ \int_{t^{n-1}}^{t^n} \Big((e_1, v_h) - a(r_p, v_h)\Big) dt + (r^n_{p+}, v^n) \qquad \forall\, v_h \in \mathcal{P}_k[t^{n-1}, t^n; U^n_h].
\end{aligned}
$$

Here, we have used the definition of the projection. Setting $v_h = r_{1h}$ into (5.6), using the bounds,

$$
\int_{t^{n-1}}^{t^n} \big|(e_1, r_{1h})\big| dt \leq \int_{t^{n-1}}^{t^n} \Big((\eta/4)\|r_{1h}\|^2_{H^1(\Omega)} + (C/\eta)\|e_1\|^2_{L^2(\Omega)}\Big) dt,
$$

$$
\int_{t^{n-1}}^{t^n} \big|a(r_{1h}, r_p)\big| dt \leq (\eta/4) \int_{t^{n-1}}^{t^n} \|r_{1h}\|^2_{H^1(\Omega)} dt + (C_c^2/\eta) \int_{t^{n-1}}^{t^n} \|r_p\|^2_{H^1(\Omega)} dt,
$$

and standard algebra, we obtain

$$
\begin{aligned}
&-(1/2)\|r^n_{1h+}\|^2_{L^2(\Omega)} + (1/2)\|[r^n_{1h}]\|^2_{L^2(\Omega)} + (1/2)\|r^{n-1}_{1h+}\|^2_{L^2(\Omega)} + (\eta/2)\int_{t^{n-1}}^{t^n} \|r_{1h}\|^2_{H^1(\Omega)} dt\\
&\leq C \int_{t^{n-1}}^{t^n} \Big((C_c^2/\eta)\|r_p\|^2_{H^1(\Omega)} + (1/\eta)\|e_1\|^2_{L^2(\Omega)}\Big) dt + \big|((I - P_{n+1})\mu(t^n_+), r^n_{1h})\big|. \qquad (5.7)
\end{aligned}
$$

Finally for the last term, observe that $r^n_{h+} \in U^{n+1}_h$ and hence,

$$
\begin{aligned}
((I - P_{n+1})\mu(t^n_+), r^n_{1h}) &= ((I - P_{n+1})\mu(t^n_+), r^n_{1h} - r^n_{1h+})\\
&\leq \|(I - P_{n+1})\mu(t^n_+)\|^2_{L^2(\Omega)} + (1/4)\|r^n_{1h+} - r^n_{1h}\|^2_{L^2(\Omega)}.
\end{aligned}
$$

An alternative bound can be obtained by using the inverse estimate $\|r^n_{1h}\|^2_{H^1(\Omega)} \leq (C_k/\tau_n)\int_{t^{n-1}}^{t^n} \|r_{1h}\|^2_{H^1(\Omega)} dt$, and noting that $r^n_{1h} \in U^n_h$,

$$
\begin{aligned}
((I - P_{n+1})\mu(t^n_+), r^n_{1h}) &= (P_n(I - P_{n+1})\mu(t^n_+), r^n_{1h})\\
&\leq \|P_n(I - P_{n+1})\mu(t^n_+)\|_{H^{-1}(\Omega)}\|r^n_{1h}\|_{H^1(\Omega)}\\
&\leq (C_k^2/(\tau_n\eta))\|P_n(I - P_{n+1})\mu(t^n_+)\|^2_{H^{-1}(\Omega)} + (\eta/4)\int_{t^{n-1}}^{t^n} \|r_{1h}\|^2_{H^1(\Omega)} dt.
\end{aligned}
$$

where at the last step we have also used Young's inequality. Collecting the last two estimates and equation (5.7) we obtain the desired estimate upon summation and standard algebra. □

Remark 5.5. *If the same subspaces are being used every time step, i.e., $U^n_h \equiv U_h \subset H^1_0(\Omega)$ then we observe that there is no contribution from the summation term in Theorem 5.4. Indeed, inspecting the above proof, we note that for $i = 1, ..., N$ the local $L^2(\Omega)$ projection $P_i \equiv P_{i+1} \equiv P_{L^2} : L^2(\Omega) \to U_h$ is the same at each time step. Therefore, $r^n_{h+} \in U_h$ implies that*

$$
\big((I - P_{n+1})\mu(t^n_+), r^n_{1h}\big) \equiv \big((I - P_{L^2})\mu(t^n_+), r^n_{1h}\big) \equiv 0.
$$

Hence, (5.7) takes the form

$$-(1/2)\|r_{1h+}^n\|_{L^2(\Omega)}^2 + (1/2)\|[r_{1h}^n]\|_{L^2(\Omega)}^2 + (1/2)\|r_{1h+}^{n-1}\|_{L^2(\Omega)}^2 + (\eta/2)\int_{t^{n-1}}^{t^n} \|r_{1h}\|_{H^1(\Omega)}^2 dt$$
$$\leq C\int_{t^{n-1}}^{t^n} \Big((C_c^2/\eta)\|r_p\|_{H^1(\Omega)}^2 + (1/\eta)\|e_1\|_{L^2(\Omega)}^2\Big)dt.$$

Working similarly for the forward (in time) problem, we obtain th following estimates:

$$\eta\|e_1\|_{L^2[0,T;H^1(\Omega)]}^2 + \sum_{i=0}^{N-1}\|[e_1^i]\|_{L^2(\Omega)}^2 \leq C\Big(\|e_1^0\|_{L^2(\Omega)}^2 + (C_c^2/\eta)\|e_p\|_{L^2[0,T;H^1(\Omega)]}^2\Big),$$
$$\eta\|r_1\|_{L^2[0,T;H^1(\Omega)]}^2 + \sum_{i=1}^{N}\|[r_1^i]\|_{L^2(\Omega)}^2 \leq C\Big((1/\eta)\|e_1\|_{L^2[0,T;L^2(\Omega)]}^2 + (C_c^2/\eta)\|r_p\|_{L^2[0,T;H^1(\Omega)]}^2\Big).$$

Subsequently, an estimate on the $L^\infty[0,T;L^2(\Omega)]$ norm is derived, using the approximation of the discrete characteristic (see Appendix B, and the subsequent Theorem 5.12). Since, an estimate on the $L^2[0,T;H^1(\Omega)]$ norm is already obtained, and the auxiliary optimality system is now essentially uncoupled, the techniques of [11, Section 2] can be applied directly.

THEOREM 5.6. *Let $w_h, z_h \in \mathcal{U}_h$ be the solutions of (5.1)-(5.2) computed using the DG scheme. Denote by $e_1 = y - w_h$, $r_1 = \mu - z_h$ and suppose that the assumptions of Theorem 5.4 hold. Then there exists a constant C depending on C_k, Ω such that*

$$\|e_1\|_{L^\infty[0,T;L^2(\Omega)]}^2 \leq C\Big[\|e_p\|_{L^\infty[0,T;L^2(\Omega)]}^2 + \|e_1^0\|_{L^2(\Omega)}^2 + (C_c^2/\eta)\|e_p\|_{L^2[0,T;H^1(\Omega)]}^2$$
$$+\sum_{i=0}^{N-1} 2\min\Big(\|(I-P_i)y(t^i)\|_{L^2(\Omega)}^2, (1/(\tau_{i+1}\eta))\|P_{i+1}(I-P_i)y(t^i)\|_{H^{-1}(\Omega)}^2\Big)\Big],$$
$$\|r_1\|_{L^\infty[0,T;L^2(\Omega)]}^2 \leq C\Big[\|r_p\|_{L^\infty[0,T;L^2(\Omega)]}^2$$
$$+(1/\eta)\|e_1\|_{L^2[0,T;L^2(\Omega)]}^2 + (C_c^2/\eta)\|r_p\|_{L^2[0,T;H^1(\Omega)]}^2$$
$$+\sum_{i=1}^{N} 2\min\Big(\|(I-P_{i+1})\mu(t^i)\|_{L^2(\Omega)}^2, (1/(\tau_i\eta))\|P_i(I-P_{i+1})\mu(t^i)\|_{H^{-1}(\Omega)}^2\Big)\Big].$$

Proof. Splitting the error as in the previous theorem, i.e., $e_1 = e_{1h} + e_p$ it suffices to bound the term $\sup_{t^{n-1}<t\leq t^n}\|e_{1h}(t)\|_{L^2(\Omega)}^2$. This is done in [11, Theorem 2.5] (note that the orthogonality condition is uncoupled).

The estimate for the adjoint variable can be derived similarly starting from orthogonality condition (5.5), and using a suitable approximation for the discrete characteristic for the backwards in time problem. □

REMARK 5.7. *Similar to Remark 5.5 an improved bound holds when $U_h^n = U_h$, $n = 1, ..., N$. In particular,*

$$\|e_1\|_{L^\infty[0,T;L^2(\Omega)]}^2 \leq C\Big(\|e_p\|_{L^\infty[0,T;L^2(\Omega)]}^2 + \|e_1^0\|_{L^2(\Omega)}^2 + (C_c^2/\eta)\|e_p\|_{L^2[0,T;H^1(\Omega)]}^2\Big)$$
$$\|r_1\|_{L^\infty[0,T;L^2(\Omega)]}^2 \leq C\Big(\|r_p\|_{L^\infty[0,T;L^2(\Omega)]}^2 + (1/\eta)\|e_1\|_{L^2[0,T;L^2(\Omega)]}^2$$
$$+(C_c^2/\eta)\|r_p\|_{L^2[0,T;H^1(\Omega)]}^2\Big).$$

Remark 5.8. *The combination of the last two Theorems implies the "symmetric" structure of our estimate. In particular, let* $\|(.,.)\|_X$, $\|(.,.)\|_{X_1}$ *be defined by*

$$\|(e_1,r_1)\|_X^2 \equiv \|e_1\|_X^2 + \|r_1\|_X^2 \equiv \|e_1\|^2_{L^2[0,T;H^1(\Omega)]} + \|r_1\|^2_{L^2[0,T;H^1(\Omega)]} \\ +\|e_1\|^2_{L^\infty[0,T;L^2(\Omega)]} + \|r_1\|^2_{L^\infty[0,T;L^2(\Omega)]},$$

and

$$\begin{aligned}&\|(e_1,r_1)\|_{X_1}^2 \equiv \|e_1\|_{X_1}^2 + \|r_1\|_{X_1}^2 \\ &\equiv \sum_{i=0}^{N-1} 2\min\Big(\|(I-P_i)y(t^i)\|^2_{L^2(\Omega)}, (1/(\tau_{i+1}\eta))\|P_{i+1}(I-P_i)y(t^i)\|^2_{H^{-1}(\Omega)}\Big) \\ &+\sum_{i=1}^{N} 2\min\Big(\|(I-P_{i+1})\mu(t^i)\|^2_{L^2(\Omega)}, (1/(\tau_i\eta))\|P_i(I-P_{i+1})\mu(t^i)\|^2_{H^{-1}(\Omega)}\Big).\end{aligned}$$

Then, using Theorems 5.4, 5.6 we obtain an estimate of the form

$$\|error\|_X \le C\Big(\|in.\ data\ error\|_{L^2(\Omega)} + \|best\ approx.\ error\|_X + \|subsp.error\|_{X_1}\Big).$$

The above estimate indicates that the error is as good as the approximation properties enables it to be, and it is applicable for higher order elements under the natural parabolic regularity assumptions. If $U_h^n \equiv U_h$ *for* $n = 1, ..., N$ *then the subspace error can be dropped, and thus we obtain symmetric estimate of the form*

$$\|error\|_X \le C\Big(\|in.\ data\ error\|_{L^2(\Omega)} + \|best\ approx.\ error\|_X\Big), \tag{5.8}$$

which can be viewed as the fully-discrete analogue of Céa's Lemma (see e.g. [15]).

5.2. The nonlinear optimality system. It remains to compare the discrete optimality system (4.4)-(4.5) to the auxiliary system (5.1)-(5.2). In the remaining of this work, we denote by $e_{2h} \equiv w_h - y_h$, and by $r_{2h} \equiv z_h - \mu_h$. We begin by establishing an auxiliary bound for $\|e_{2h}\|^2_{L^2[0,T;L^2(\Omega)]}$ and $(1/\alpha)\|r_{2h}\|^2_{L^2[0,T;L^2(\Omega)]}$ in terms of $\alpha^{1/2}\|e_{2h}\|^2_{L^2[0,T;H^1(\Omega)]}$ and projection terms e_1, r_1. Here, we note that without loss of generality we assume $\alpha < 1$, which corresponds to the physical case.

Lemma 5.9. *Suppose that Assuptions 2.1-4.2-5.1 hold. Let* $y_h, \mu_h, w_h, z_h \in \mathcal{U}_h$ *be the solutions the optimality system (4.4)-(4.5) and of the auxiliary system (5.1)-(5.2) respectively, computed using the discontinuous Galerkin scheme. Denote by* $e_1 \equiv y - w_h$, $r_1 \equiv \mu - z_h$, *and let* $e_{2h} \equiv w_h - y_h$, $r_{2h} \equiv z_h - \mu_h$. *Then, there exists constant* $\mathbf{C}$ *depending on* η, C_L, C_c *and the constants* C_d, C_{st} *of Assumption 5.1 and Lemma 4.4 respectively such that for* τ *satisfying the Assumptions of Lemmas 4.4, and 4.10, and for* $\alpha < CC_L$ *the following estimate holds:*

$$\begin{aligned}&\int_0^T \|e_{2h}\|^2_{L^2(\Omega)}dt + (1/\alpha)\int_0^T \|r_{2h}\|^2_{L^2(\Omega)}dt \\ &\le C\int_0^T \Big((1/\alpha)\|e_1\|^2_{H^1(\Omega)} + \|r_1\|^2_{H^1(\Omega)}\Big)dt + \mathbf{C}\alpha^{1/2}\int_0^T \|e_{2h}\|^2_{H^1(\Omega)}dt.\end{aligned}$$

Proof. Subtracting (4.5) from (5.2) we obtain the equation,

$$-(r^n_{2h+}, v^n) + \int_{t^{n-1}}^{t^n} \Big(\langle r_{2h}, v_{ht} \rangle + a(r_{2h}, v_h) + \langle \phi'(y)\mu - \phi'(y_h)\mu_h, v_h \rangle \Big) dt$$
$$= -(r^{n-1}_{2h+}, v^{n-1}_+) + \int_{t^{n-1}}^{t^n} (e_{2h}, v_h) dt \qquad \forall\, v_h \in \mathcal{P}_k[t^{n-1}, t^n; U^n_h]. \tag{5.9}$$

Subtracting (4.4) from (5.1) we obtain the equation:

$$(e^n_{2h}, v^n) + \int_{t^{n-1}}^{t^n} \Big(-\langle e_{2h}, v_{ht} \rangle + a(e_{2h}, v_h) + \langle \phi(y) - \phi(y_h), v_h \rangle \Big) dt$$
$$= (e^{n-1}_{2h}, v^{n-1}_+) + \int_{t^{n-1}}^{t^n} -(1/\alpha)(\mu - \mu_h, v_h) dt \qquad \forall\, v_h \in \mathcal{P}_k[t^{n-1}, t^n; U^n_h]. \tag{5.10}$$

We will obtain an auxiliary bound for $\|e_{2h}\|^2_{L^2[0,T;L^2(\Omega)]}$ and $(1/\alpha)\|r_{2h}\|^2_{L^2[0,T;L^2(\Omega)]}$ in terms of $\alpha^{1/2}\|e_{2h}\|^2_{L^2[0,T;H^1(\Omega)]}$ and projection terms. For this purpose we set $v_h = e_{2h}$ into (5.9) to obtain

$$-(r^n_{2h+}, e^n_{2h}) + \int_{t^{n-1}}^{t^n} \Big(\langle r_{2h}, e_{2ht} \rangle + a(r_{2h}, e_{2h}) + \langle \phi'(y)\mu - \phi'(y_h)\mu_h, e_{2h} \rangle \Big) dt$$
$$+(r^{n-1}_{2h+}, e^{n-1}_{2h+}) = \int_{t^{n-1}}^{t^n} \|e_{2h}\|^2_{L^2(\Omega)} dt. \tag{5.11}$$

and $v_h = r_{2h}$ into (5.10)

$$(e^n_{2h}, r^n_{2h}) + \int_{t^{n-1}}^{t^n} \Big(-\langle e_{2h}, r_{2ht} \rangle + a(e_{2h}, r_{2h}) + \langle \phi(y) - \phi(y_h), v_h \rangle \Big) dt$$
$$-(e^{n-1}_{2h}, r^{n-1}_{2h+}) = \int_{t^{n-1}}^{t^n} -(1/\alpha)(r_1, r_{2h}) - (1/\alpha)\|r_{2h}\|^2_{L^2(\Omega)} dt. \tag{5.12}$$

Integrating by parts with respect to time in (5.12), and subtracting the resulting equation from (5.11), we arrive to

$$(r^n_{2h+}, e^n_{2h}) - (e^{n-1}_{2h}, r^{n-1}_{2h+}) + \int_{t^{n-1}}^{t^n} \Big(\|e_{2h}\|^2_{L^2(\Omega)} + (1/\alpha)\|r_{2h}\|^2_{L^2(\Omega)} \Big) dt \tag{5.13}$$
$$= \int_{t^{n-1}}^{t^n} \left(\langle \phi'(y)\mu - \phi'(y_h)\mu_h, e_{2h} \rangle - \langle \phi(y) - \phi(y_h), r_{2h} \rangle \right) dt - (1/\alpha) \int_{t^{n-1}}^{t^n} (r_1, r_{2h}) dt.$$

We need to bound the three terms of the right hand side. We begin by estimating the last two terms. For this purpose, note that,

$$\left| (1/\alpha) \int_{t^{n-1}}^{t^n} (r_1, r_{2h}) dt \right| \le (1/4\alpha) \int_{t^{n-1}}^{t^n} \|r_{2h}\|^2_{L^2(\Omega)} dt + (1/\alpha) \int_{t^{n-1}}^{t^n} \|r_1\|^2_{L^2(\Omega)} dt,$$

while Assumption 4.2 (note that there exists $\epsilon > 0$ such that $\|y_h - y\|_{L^2[t^{n-1}, t^n; L^2(\Omega)]} \le \epsilon$ due to Theorem 4.6) and Young's inequality imply that

$$\int_{t^{n-1}}^{t^n} |\langle \phi(y) - \phi(y_h), r_{2h} \rangle| \, dt \le C_L^2 \alpha \int_{t^{n-1}}^{t^n} \Big(\|e_{2h}\|^2_{L^2(\Omega)} + \|e_1\|^2_{L^2(\Omega)} \Big) dt$$
$$+(1/4\alpha) \int_{t^{n-1}}^{t^n} \|r_{2h}\|^2_{L^2(\Omega)} dt.$$

Furthermore, for the final term, we may bound

$$\begin{aligned}
\mathbf{I}_{nl} &\equiv \int_{t^{n-1}}^{t^n} |\langle \phi'(y)\mu - \phi'(y_h)\mu_h, e_{2h}\rangle| \, dt \\
&\leq \int_{t^{n-1}}^{t^n} |\langle \phi'(y)(\mu - \mu_h), e_{2h}\rangle| \, dt + \int_{t^{n-1}}^{t^n} |\langle (\phi'(y) - \phi'(y_h))\mu_h, e_{2h}\rangle| \, dt \\
&\equiv \mathbf{I}_{nl}^1 + \mathbf{I}_{nl}^2.
\end{aligned}$$

For integral $\mathbf{I}_{nl}^1$, adding and subtracting $\phi'(0)$,

$$\begin{aligned}
\mathbf{I}_{nl}^1 &= \int_{t^{n-1}}^{t^n} |\langle \phi'(y)(\mu - \mu_h), e_{2h}\rangle| \, dt \\
&\leq \int_{t^{n-1}}^{t^n} \left(|\langle (\phi'(y) - \phi'(0))(\mu - \mu_h), e_{2h}\rangle| + |\langle \phi'(0)(\mu - \mu_h), e_{2h}\rangle| \right) dt.
\end{aligned}$$

Hence, using the Lipschitz continuity of ϕ', the uniform bound on $\phi'(0)$, the embedding $H^1(\Omega) \subset L^4(\Omega)$, and Young's inequality with suitable $\delta > 0$, we obtain

$$\begin{aligned}
\mathbf{I}_{nl}^1 &\leq CC_L \int_{t^{n-1}}^{t^n} \|y\|_{L^4(\Omega)} \|r_{2h} + r_1\|_{L^2(\Omega)} \|e_{2h}\|_{L^4(\Omega)} dt + C \int_{t^{n-1}}^{t^n} \|r_{2h} + r_1\|_{L^2(\Omega)} \|e_{2h}\|_{L^2(\Omega)} dt \\
&\leq (1/\alpha) \int_{t^{n-1}}^{t^n} \|r_1\|_{L^2(\Omega)}^2 dt + (1/4\alpha) \int_{t^{n-1}}^{t^n} \|r_{2h}\|_{L^2(\Omega)}^2 dt \\
&\quad + \alpha C(C_L) \|y\|_{L^\infty[0,T;L^4(\Omega)]}^2 \int_{t^{n-1}}^{t^n} \|e_{2h}\|_{H^1(\Omega)}^2 dt + C\alpha \int_{t^{n-1}}^{t^n} \|e_{2h}\|_{L^2(\Omega)}^2 dt \\
&\leq (1/\alpha) \int_{t^{n-1}}^{t^n} \|r_1\|_{L^2(\Omega)}^2 dt + (1/4\alpha) \int_{t^{n-1}}^{t^n} \|r_{2h}\|_{L^2(\Omega)}^2 dt \\
&\quad + \alpha^{1/2} C(C_L, C_d) \int_{t^{n-1}}^{t^n} \|e_{2h}\|_{H^1(\Omega)}^2 dt + C\alpha \int_{t^{n-1}}^{t^n} \|e_{2h}\|_{L^2(\Omega)}^2 dt
\end{aligned}$$

where at the last inequality we have used Assumption 5.1. Here $C(C_L, C_d)$ denote constant depending upon C_L, the data f, y_0, U, η and Ω. In addition, the Lipschitz continuity of ϕ' and the generalized Hölder's inequality, imply that

$$\begin{aligned}
\mathbf{I}_{nl}^2 = \int_{t^{n-1}}^{t^n} |\langle (\phi'(y) - \phi'(y_h))\mu_h, e_{2h}\rangle| \, dt &\leq C_L \int_{t^{n-1}}^{t^n} \|e_1\|_{L^4(\Omega)} \|\mu_h\|_{L^2(\Omega)} \|e_{2h}\|_{L^4(\Omega)} dt \\
&\quad + \int_{t^{n-1}}^{t^n} \|e_{2h}\|_{L^4(\Omega)} \|\mu_h\|_{L^2(\Omega)} \|e_{2h}\|_{L^4(\Omega)} dt.
\end{aligned}$$

The first part of $\mathbf{I}_{nl}^2$ can be bounded by using the embedding $H^1(\Omega) \subset L^4(\Omega)$ and Young's inequality,

$$\begin{aligned}
\int_{t^{n-1}}^{t^n} \|e_1\|_{H^1(\Omega)} \|\mu_h\|_{L^2(\Omega)} \|e_{2h}\|_{H^1(\Omega)} dt &\leq (CD_{\mu st}/\alpha^{1/2}) \int_{t^{n-1}}^{t^n} \|e_1\|_{H^1(\Omega)}^2 dt \\
&\quad + \alpha^{1/2} \int_{t^{n-1}}^{t^n} \|e_{2h}\|_{H^1(\Omega)}^2 dt,
\end{aligned}$$

where here we denote by $D_{\mu st}$ the stability constant of Lemma 4.10. Finally, observe that interpolation inequality $\|.\|_{L^4(\Omega)}^2 \leq C\|.\|_{L^2(\Omega)}\|.\|_{H^1(\Omega)}$, the stability inequality of

μ_h of Lemma 4.10 and Young's inequality with appropriate δ, imply that

$$\int_{t^{n-1}}^{t^n} \|e_{2h}\|_{L^4(\Omega)} \|e_{2h}\|_{L^4(\Omega)} \|\mu_h\|_{L^2(\Omega)} dt \leq \|\mu_h\|_{L^\infty[t^{n-1},t^n;L^2(\Omega)]} \int_{t^{n-1}}^{t^n} \|e_{2h}\|_{L^2(\Omega)} \|e_{2h}\|_{H^1(\Omega)} dt$$
$$\leq (1/4) \int_{t^{n-1}}^{t^n} \|e_{2h}\|^2_{L^2(\Omega)} dt + C\|\mu_h\|^2_{L^\infty[t^{n-1},t^n;L^2(\Omega)]} \int_{t^{n-1}}^{t^n} \|e_{2h}\|^2_{H^1(\Omega)} dt$$
$$\leq (1/4) \int_{t^{n-1}}^{t^n} \|e_{2h}\|^2_{L^2(\Omega)} dt + CC_{st}\alpha^{1/2} \int_{t^{n-1}}^{t^n} \|e_{2h}\|^2_{H^1(\Omega)} dt$$

Substituting the above bounds into (5.13) and adding the resulting inequalities from 1 to N, noting that $\sum_{n=1}^{N} \big((r^n_{2h+}, e^n_{2h}) - (e^{n-1}_{2h}, r^{n-1}_{2h+})\big) = 0$ (since $e^0_{2h} \equiv 0$, $r^N_{2h+} = 0$), and choosing $\alpha < C(C_L)$ to hide $\int_{t^{n-1}}^{t^n} \|e_{2h}\|^2_{L^2(\Omega)} dt$, we obtain the desired estimate. ∎

Remark 5.10. *In the above proof we have use the Lipschitz continuity of ϕ' to avoid any additional technicalities. The assumption that $y \in L^\infty[0,T;L^4(\Omega)]$, will require to impose additional regularity assumptions on the data, in particular, $y_0 \in H^1_0(\Omega)$, $f \in L^2[0,T;L^2(\Omega)]$, but not additional regularity on the control and the target.*

Estimates follow using projection techniques of Theorem 5.4 which allow to treat the forward and backward (in time) coupled PDE's together with a "boot-strap" argument.

Theorem 5.11. *Let Assumptions 2.1-4.2-5.1 hold. Let $y_h, \mu_h, w_h, z_h \in \mathcal{U}_h$ be the solutions of the optimality system (4.4)-(4.5) and of the auxiliary system (5.1)-(5.2) respectively, computed using the discontinuous Galerkin scheme. Denote by $e_1 \equiv y - w_h$, $r_1 \equiv \mu - z_h$, and let $e_{2h} \equiv w_h - y_h$, $r_{2h} \equiv z_h - \mu_h$. Then, there exists constant $\mathbf{D}$, depending on $\|y\|_{L^\infty[0,T;L^2(\Omega)]}/\eta$, the constant $\mathbf{C}$ of Lemma 5.9, and $\rho \equiv \frac{CC^2_{st}/\eta+\beta\mathbf{C}}{\eta/4+CC^2_{st}/\eta+\beta\mathbf{C}} < 1$ (for $\beta > 0$) such that for τ satisfying the assumptions of Lemmas 4.4, and 4.10, the following estimate holds:*

$$\|e^N_{2h}\|^2_{L^2(\Omega)} + \eta \int_0^T \|e_{2h}\|^2_{H^1(\Omega)} dt + \sum_{i=0}^{N-1} \|[e^i_{2h}]\|^2_{L^2(\Omega)}$$
$$+(\eta/\alpha) \int_0^T \|r_{2h}\|^2_{H^1(\Omega)} dt + (1/\alpha)\|r^0_{2h+}\|^2_{L^2(\Omega)} + (1/\alpha) \sum_{i=1}^{N} \|[r^i_{2h}]\|^2_{L^2(\Omega)}$$
$$\leq \mathbf{D}(1/\alpha^2) \int_0^T \Big(\|e_1\|^2_{H^1(\Omega)} + \|r_1\|^2_{H^1(\Omega)}\Big) dt.$$

Here the constant $\mathbf{D}$ is independent of τ, h, α.

Remark 5.12. *We note that we are interested in the case where the values of α are small, and possibly comparable to h, which guarantee fast convergence to the target U. Hence, great care is exercised to avoid the use of Grönwall's type arguments which typically lead to constants of the form $exp(1/\alpha)$.*

Proof. Step 1: Preliminary estimates for the state: Setting $v_h = e_{2h}$ into (5.10) and noting that $\mu - \mu_h = r_1 + r_{2h}$ we obtain

$$(1/2)\|e_{2h}^n\|^2_{L^2(\Omega)} + (1/2)\|[e_{2h}^{n-1}]\|^2_{L^2(\Omega)} - (1/2)\|e_{2h}^{n-1}\|^2_{L^2(\Omega)} + \eta \int_{t^{n-1}}^{t^n} \|e_{2h}\|^2_{H^1(\Omega)} dt$$
$$+ \int_{t^{n-1}}^{t^n} \langle \phi(y) - \phi(y_h), e_{2h}\rangle dt \leq -(1/\alpha) \int_{t^{n-1}}^{t^n} (r_1 + r_{2h}, e_{2h}) dt, \tag{5.14}$$

For the first term on the right hand side, note that

$$\Big|(1/\alpha) \int_{t^{n-1}}^{t^n} (r_1, e_{2h}) dt\Big| \leq (\eta/4) \int_{t^{n-1}}^{t^n} \|e_{2h}\|^2_{H^1(\Omega)} dt + (C/\eta\alpha^2) \int_{t^{n-1}}^{t^n} \|r_1\|^2_{L^2(\Omega)} dt.$$

Next we focus on the nonlinear terms. Notice that the monotonicity of ϕ implies that

$$\mathbf{I}_{nl} \equiv \int_{t^{n-1}}^{t^n} \langle \phi(y) - \phi(y_h), e_{2h}\rangle dt \geq \int_{t^{n-1}}^{t^n} \langle \phi(y) - \phi(w_h), e_{2h}\rangle dt$$

and hence we moving the above term on the right hand side, we may bound the term by using Assumption 4.2, Poincaré inequality, and Young's inequality, as follows:

$$\Big|\mathbf{I}_{nl}\Big| \leq C_L \int_{t^{n-1}}^{t^n} \|e_1\|_{L^2(\Omega)} \|e_{2h}\|_{L^2(\Omega)} dt$$
$$\leq (\eta/4) \int_{t^{n-1}}^{t^n} \|e_{2h}\|^2_{H^1(\Omega)} dt + (CC_L/\eta) \int_{t^{n-1}}^{t^n} \|e_1\|^2_{H^1(\Omega)} dt.$$

Therefore collecting the above bounds into (5.14) and multiplying by $\alpha^{1/2}$ we obtain:

$$\alpha^{1/2}\Big(\|e_{2h}^n\|^2_{L^2(\Omega)} + \|[e_{2h}^{n-1}]\|^2_{L^2(\Omega)} - \|e_{2h}^{n-1}\|^2_{L^2(\Omega)} + (\eta/4) \int_{t^{n-1}}^{t^n} \|e_{2h}\|^2_{H^1(\Omega)} dt\Big) \tag{5.15}$$
$$\leq \int_{t^{n-1}}^{t^n} \Big((C/\eta\alpha^{3/2})\|r_1\|^2_{H^1(\Omega)} + (CC_L\alpha^{1/2}/\eta)\|e_1\|^2_{H^1(\Omega)}\Big) dt - (1/\alpha^{1/2}) \int_{t^{n-1}}^{t^n} (r_{2h}, e_{2h}) dt.$$

Step 2: Preliminary estimates for the adjoint: Setting $v_h = r_{2h}$ into (5.9), we obtain

$$-(1/2)\|r_{2h+}^n\|^2_{L^2(\Omega)} + (1/2)\|[r_{2h}^n]\|^2_{L^2(\Omega)} + (1/2)\|r_{2h+}^{n-1}\|^2_{L^2(\Omega)} + \eta \int_{t^{n-1}}^{t^n} \|r_{2h}\|^2_{H^1(\Omega)} dt$$
$$+ \int_{t^{n-1}}^{t^n} \langle \phi'(y)\mu - \phi'(y_h)\mu_h, r_{2h}\rangle dt \leq \int_{t^{n-1}}^{t^n} (e_{2h}, r_{2h}) dt. \tag{5.16}$$

Using the monotonicity of ϕ, and noting that $\mu - \mu_h = r_1 + r_{2h}$, the nonlinearity of the adjoint equation can be written as:

$$\int_{t^{n-1}}^{t^n} \langle \phi'(y)\mu - \phi'(y_h)\mu_h, r_{2h}\rangle dt$$
$$= \int_{t^{n-1}}^{t^n} \langle \phi'(y)\mu - \phi'(y)\mu_h, r_{2h}\rangle dt + \int_{t^{n-1}}^{t^n} \langle \phi'(y)\mu_h - \phi'(y_h)\mu_h, r_{2h}\rangle dt$$
$$\geq \int_{t^{n-1}}^{t^n} \langle \phi'(y) r_1, r_{2h}\rangle dt + \int_{t^{n-1}}^{t^n} \langle \phi'(y)\mu_h - \phi'(y_h)\mu_h, r_{2h}\rangle dt.$$

Moving the last two integrals on the right hand side, we derive appropriate bounds. For the first integral, using the Lipschitz continuity of ϕ', the uniform bound on $\phi'(0)$, the generalized Hölder's inequality and the embedding $H^1(\Omega) \subset L^4(\Omega)$, we easily obtain

$$\Big| \int_{t^{n-1}}^{t^n} \langle \phi'(y) r_1, r_{2h} \rangle dt \Big| \leq \Big| \int_{t^{n-1}}^{t^n} \langle (\phi'(y) - \phi'(0)) r_1, r_{2h} \rangle dt \Big| + \Big| \int_{t^{n-1}}^{t^n} \langle \phi'(0) r_1, r_{2h} \rangle dt \Big|$$
$$\leq (\eta/4) \int_{t^{n-1}}^{t^n} \|r_{2h}\|^2_{H^1(\Omega)} dt + (C_y/\eta) \int_{t^{n-1}}^{t^n} \|r_1\|^2_{H^1(\Omega)} dt,$$

where C_y depends only on $\|y\|_{L^\infty[0,T;L^2(\Omega)]}$ and the domain. Similarly, for the second integral, the Lipschitz continuity of ϕ', the generalized Hölder inequality and the fact that $y - y_h = e_1 + e_{2h}$ imply,

$$\Big| \int_{t^{n-1}}^{t^n} \langle (\phi'(y) - \phi'(y_h)) \mu_h, r_{2h} \rangle dt \Big| \leq C_L \int_{t^{n-1}}^{t^n} \|\mu_h\|_{L^2(\Omega)} \|e_1 + e_{2h}\|_{L^4(\Omega)} \|r_{2h}\|_{L^4(\Omega)} dt$$
$$\leq \mathbf{II}^1_{nl} + \mathbf{II}^2_{nl}.$$

It remains to bound the last two integrals. Starting from $\mathbf{II}^2_{nl}$, using the interpolation inequality $\|.\|^2_{L^4(\Omega)} \leq C \|.\|_{L^2(\Omega)} \|.\|_{H^1(\Omega)}$ and stability estimates on μ_h, we obtain:

$$\mathbf{II}^2_{nl} \leq C_L \int_{t^{n-1}}^{t^n} \|\mu_h\|_{L^2(\Omega)} \|e_{2h}\|_{L^4(\Omega)} \|r_{2h}\|_{L^4(\Omega)} dt$$
$$\leq (\eta/4) \int_{t^{n-1}}^{t^n} \|\mu_h\|^2_{L^2(\Omega)} \|r_{2h}\|_{H^1(\Omega)} \|e_{2h}\|_{H^1(\Omega)} dt + (CC_L/\eta) \int_{t^{n-1}}^{t^n} \|e_{2h}\|_{L^2(\Omega)} \|r_{2h}\|_{L^2(\Omega)} dt$$
$$\leq (\eta/4) \int_{t^{n-1}}^{t^n} \|r_{2h}\|^2_{H^1(\Omega)} dt + (\|\mu_h\|^4_{L^\infty[t^{n-1},t^n;L^2(\Omega)]} \eta/16) \int_{t^{n-1}}^{t^n} \|e_{2h}\|^2_{H^1(\Omega)} dt$$
$$+ (CC_L/\eta) \int_{t^{n-1}}^{t^n} \left(\alpha^{1/2} \|e_{2h}\|^2_{L^2(\Omega)} + (1/\alpha^{1/2}) \|r_{2h}\|^2_{L^2(\Omega)} \right) dt,$$
$$\leq (\eta/4) \int_{t^{n-1}}^{t^n} \|r_{2h}\|^2_{H^1(\Omega)} dt + (CC^2_{st} \alpha \eta/16) \int_{t^{n-1}}^{t^n} \|e_{2h}\|^2_{H^1(\Omega)} dt$$
$$+ (CC_L/\eta) \int_{t^{n-1}}^{t^n} \left(\alpha^{1/2} \|e_{2h}\|^2_{L^2(\Omega)} + (1/\alpha^{1/2}) \|r_{2h}\|^2_{L^2(\Omega)} \right) dt,$$

where we have used the stability bound of Lemma 4.10. For $\mathbf{II}^1_{nl}$, using the Hölder's inequality and the embedding $H^1(\Omega) \subset L^4(\Omega)$, we obtain,

$$\mathbf{II}^1_{nl} \leq C \int_{t^{n-1}}^{t^n} \|\mu_h\|_{L^2(\Omega)} \|e_1\|_{H^1(\Omega)} \|r_{2h}\|_{H^1(\Omega)} dt$$
$$\leq (\eta/4) \int_{t^{n-1}}^{t^n} \|r_{2h}\|^2_{H^1(\Omega)} dt + (CC_L C_{st} \alpha^{1/2}/\eta) \int_{t^{n-1}} \|e_1\|^2_{H^1(\Omega)} dt.$$

Inserting the bounds on $\mathbf{II}^1_{nl}, \mathbf{II}^2_{nl}$ into (5.16), and multiplying by $(1/\alpha^{1/2})$, we obtain

$$\begin{aligned}
&-(1/2\alpha^{1/2})\|r^n_{2h+}\|^2_{L^2(\Omega)} + (1/2\alpha^{1/2})\|[r^n_{2h}]\|^2_{L^2(\Omega)} + (1/2\alpha^{1/2})\|r^{n-1}_{2h+}\|^2_{L^2(\Omega)} \\
&\quad +(\eta/2\alpha^{1/2})\int_{t^{n-1}}^{t^n}\|r_{2h}\|^2_{H^1(\Omega)}dt \qquad (5.17)\\
&\leq \mathbf{D}\int_{t^{n-1}}^{t^n}\Big(\|e_1\|^2_{H^1(\Omega)} + (1/\alpha^{1/2})\|r_1\|^2_{H^1(\Omega)}\Big)dt + (1/\alpha^{1/2})\int_{t^{n-1}}^{t^n}(e_{2h}, r_{2h})dt \\
&\quad +CC^2_{st}\alpha^{1/2}\int_{t^{n-1}}^{t^n}\|e_{2h}\|^2_{H^1(\Omega)}dt + CC_L/\eta\int_{t^{n-1}}^{t^n}\Big(\|e_{2h}\|^2_{L^2(\Omega)} + (1/\alpha)\|r_{2h}\|^2_{L^2(\Omega)}\Big)\,dt,
\end{aligned}$$

where $\mathbf{D}$ depends upon CC_LC_{st}/η, and C_y/η.

Step 3: Combination of (5.15)-(5.17): Next we will form the convex combination of (5.15)-(5.17) by multiplying $1-\rho$ equation (5.17) and by ρ equation (5.15), $0<\rho<1$, (ρ to be determined later), and we add the resulting equations:

$$\begin{aligned}
&\rho\alpha^{1/2}\|e^n_{2h}\|^2_{L^2(\Omega)} + \rho\alpha^{1/2}\|[e^{n-1}_{2h}]\|^2_{L^2(\Omega)} - \rho\alpha^{1/2}\|e^{n-1}_{2h}\|^2_{L^2(\Omega)} + (\rho\eta\alpha^{1/2}/4)\int_{t^{n-1}}^{t^n}\|e_{2h}\|^2_{H^1(\Omega)}dt \\
&-((1-\rho)/2\alpha^{1/2})\|r^n_{2h+}\|^2_{L^2(\Omega)} + ((1-\rho)/2\alpha^{1/2})\|[r^n_{2h}]\|^2_{L^2(\Omega)} + ((1-\rho)/2\alpha^{1/2})\|r^{n-1}_{2h+}\|^2_{L^2(\Omega)} \\
&+((1-\rho)\eta/4\alpha^{1/2})\int_{t^{n-1}}^{t^n}\|r_{2h}\|^2_{H^1(\Omega)}dt \\
&\leq \mathbf{D}(1/\alpha^{3/2})\int_{t^{n-1}}^{t^n}\Big(\|r_1\|^2_{H^1(\Omega)} + \|e_1\|^2_{H^1(\Omega)}\Big)\,dt + (1-\rho)CC^2_{st}\alpha^{1/2}\int_{t^{n-1}}^{t^n}\|e_{2h}\|^2_{H^1(\Omega)}dt \\
&\quad +(1-\rho)(CC_L/\eta)\int_{t^{n-1}}^{t^n}\Big(\|e_{2h}\|^2_{L^2(\Omega)} + (1/\alpha)\|r_{2h}\|^2_{L^2(\Omega)}\Big)\,dt \\
&\quad +(1-\rho)/\alpha^{1/2}\int_{t^{n-1}}^{t^n}(e_{2h}, r_{2h})dt - (\rho/\alpha^{1/2})\int_{t^{n-1}}^{t^n}(e_{2h}, r_{2h})dt. \qquad (5.18)
\end{aligned}$$

There are two distinct cases. If $0<\rho\leq(1/2)$, then $\rho\leq(1-\rho)$ and we may bound the last two terms, by $2(1-\rho)/\alpha^{1/2}\int_{t^{n-1}}^{t^n}|(e_{2h}, r_{2h})|dt$, and hence using Young's inequality,

$$2(1-\rho)/\alpha^{1/2}\int_{t^{n-1}}^{t^n}\Big(\alpha^{1/2}\|e_{2h}\|^2_{L^2(\Omega)} + (1/\alpha^{1/2})\|r_{2h}\|^2_{L^2(\Omega)}\Big)\,dt.$$

Substituting the last inequality into (5.18), and summing from 1 to N we deduce

$$
\begin{aligned}
&\rho\alpha^{1/2}\|e_{2h}^N\|^2_{L^2(\Omega)} + \rho\alpha^{1/2}\sum_{i=1}^N \|[e_{2h}^{i-1}]\|^2_{L^2(\Omega)} + (\rho\eta\alpha^{1/2}/4)\int_0^T \|e_{2h}\|^2_{H^1(\Omega)}dt \\
&+((1-\rho)/2\alpha^{1/2})\sum_{i=1}^N \|[r_{2h}^i]\|^2_{L^2(\Omega)} + ((1-\rho)/2\alpha^{1/2})\|r^0_{2h+}\|^2_{L^2(\Omega)} \\
&+((1-\rho)\eta/4\alpha^{1/2})\int_0^T \|r_{2h}\|^2_{H^1(\Omega)}dt \\
&\leq \mathbf{D}(1/\alpha^{3/2})\int_0^T \Big(\|r_1\|^2_{H^1(\Omega)} + \|e_1\|^2_{H^1(\Omega)}\Big)\,dt + (1-\rho)CC_{st}^2\alpha^{1/2}\int_0^T \|e_{2h}\|^2_{H^1(\Omega)}dt \\
&\quad +(1-\rho)CC_L/\eta\int_0^T \Big(\|e_{2h}\|^2_{L^2(\Omega)} + (1/\alpha)\|r_{2h}\|^2_{L^2(\Omega)}\Big)\,dt \\
&\quad +2(1-\rho)\int_0^T \Big(\|e_{2h}\|^2_{L^2(\Omega)} + (1/\alpha)\|r_{2h}\|^2_{L^2(\Omega)}\Big)\,dt. \qquad (5.19)
\end{aligned}
$$

where $\mathbf{D}$ depends only upon the stability constant C_{st}, η, C_L. Note that we may use Lemma 5.9 to replace, the last two integrals, by projection terms e_1, r_1 and $\alpha^{1/2}\|e_{2h}\|^2_{L^2[0,T;H^1(\Omega)]}$. Thus,

$$
\begin{aligned}
&\rho\alpha^{1/2}\|e_{2h}^N\|^2_{L^2(\Omega)} + \rho\alpha^{1/2}\sum_{i=1}^N \|[e_{2h}^{i-1}]\|^2_{L^2(\Omega)} + (\rho\eta\alpha^{1/2}/4)\int_0^T \|e_{2h}\|^2_{H^1(\Omega)}dt \\
&+((1-\rho)/2\alpha^{1/2})\sum_{i=1}^N \|[r_{2h}^i]\|^2_{L^2(\Omega)} + ((1-\rho)/2\alpha^{1/2})\|r^0_{2h+}\|^2_{L^2(\Omega)} \\
&+((1-\rho)\eta/4\alpha^{1/2})\int_0^T \|r_{2h}\|^2_{H^1(\Omega)}dt \\
&\leq \mathbf{D}(\rho)(1/\alpha^{3/2})\int_0^T \Big(\|r_1\|^2_{H^1(\Omega)} + \|e_1\|^2_{H^1(\Omega)}\Big)\,dt \qquad (5.20) \\
&\quad +(1-\rho)CC_{st}^2\alpha^{1/2}/\eta\int_0^T \|e_{2h}\|^2_{H^1(\Omega)}dt + (1-\rho)3\mathbf{C}\alpha^{1/2}\int_0^T \|e_{2h}\|^2_{H^1(\Omega)}dt.
\end{aligned}
$$

Here, $\mathbf{C}$ denotes the constant of Lemma 5.9. Then, choosing ρ in order to hide the term $\|e_{2h}\|_{L^2[0,T;H^1(\Omega)]}$ on the left, i.e.,

$$
(1-\rho)(CC_{st}^2/\eta + 3\mathbf{C})\alpha^{1/2} = \rho\eta\alpha^{1/2}/4, \qquad \rho \equiv \frac{CC_{st}^2/\eta + 3\mathbf{C}}{\eta/4 + CC_{st}^2/\eta + 3\mathbf{C}} < 1,
$$

(noting that ρ is independent of α) we arrive at the desired estimate. We also note that so far we have treated the case $0 < \rho \leq (1/2)$, which implies an assumption on the size of data, and in particular, $CC_{st}^2/\eta + 3\mathbf{C} < \eta/4$. It remains to treat the case where $(1/2) < \rho < 1$. Again, we are interested in treating the last two terms of (5.18). For this purpose, note that

$$
\begin{aligned}
&(1-\rho)/\alpha^{1/2}\int_{t^{n-1}}^{t^n}(e_{2h}, r_{2h})dt - (\rho/\alpha^{1/2})\int_{t^{n-1}}^{t^n}(e_{2h}, r_{2h})dt \\
&\leq |(1-2\rho)|/\alpha^{1/2}\int_{t^{n-1}}^{t^n}|(e_{2h}, r_{2h})|dt.
\end{aligned}
$$

Since, $(1/2) < \rho < 1$, we deduce $|(1-2\rho)| = (2\rho - 1) \leq \beta(1-\rho)$, for some $\beta > 0$. Indeed, we note that if $\beta > 0$ big enough, then $\rho \approx< 1$ since $\rho \leq (1+\beta)/(2+\beta) \approx< 1$. The remaining of the proof remains the same. The analog of (5.20) takes the form,

$$\begin{aligned}
&\rho\alpha^{1/2}\|e_{2h}^N\|^2_{L^2(\Omega)} + \rho\alpha^{1/2}\sum_{i=1}^N \|[e_{2h}^{i-1}]\|^2_{L^2(\Omega)} + (\rho\eta\alpha^{1/2}/4)\int_0^T \|e_{2h}\|^2_{H^1(\Omega)}dt \\
&+((1-\rho)/2\alpha^{1/2})\sum_{i=1}^N \|[r_{2h}^i]\|^2_{L^2(\Omega)} + ((1-\rho)/2\alpha^{1/2})\|r^0_{2h+}\|^2_{L^2(\Omega)} \\
&+((1-\rho)\eta/4\alpha^{1/2})\int_0^T \|r_{2h}\|^2_{H^1(\Omega)}dt \\
&\leq \mathbf{D}(\rho)(1/\alpha^{3/2})\int_0^T \left(\|r_1\|^2_{H^1(\Omega)} + \|e_1\|^2_{H^1(\Omega)}\right)dt \\
&\qquad +(1-\rho)CC_{st}^2\alpha^{1/2}/\eta\int_0^T \|e_{2h}\|^2_{H^1(\Omega)}dt + \beta(1-\rho)\mathbf{C}\alpha^{1/2}\int_0^T \|e_{2h}\|^2_{H^1(\Omega)}dt.
\end{aligned}$$

Then, choosing ρ (independent of α) in order to hide the last two terms on the left hand side, i.e, for

$$(1-\rho)(CC_{st}^2/\eta + \beta\mathbf{C})\alpha^{1/2} = \rho\eta\alpha^{1/2}/4, \qquad \rho \equiv \frac{CC_{st}^2/\eta + \beta\mathbf{C}}{\eta/4 + CC_{st}^2/\eta + \beta\mathbf{C}} < 1,$$

we obtain the desired estimate. □

REMARK 5.13. *In most practical situations, such as short time-setting or not very large data C_{st}, we note that the values of the parameters ρ or $1-\rho$ are not comparable to $a^{1/2} << 1$, hence the dependence of the estimate upon α does not deteriorate further.*

Based on the estimates at the energy norms, we proceed to derive estimates at arbitrary times. Since, an estimate on the energy norm $\|r_1\|_{L^2[0,T;H^1(\Omega)]}$ is already obtained in Theorem 5.11, the optimality system is now essentially uncoupled. An estimate at arbitrary time points for the forward in time equation can be derived by applying the approximation of the discrete characteristic technique of [11] into the semi-linear case. Here, the stability estimate at arbitrary time-points will be also needed.

THEOREM 5.14. *Let $y_h, \mu_h \in \mathcal{U}_h$ be the solutions of (4.4)-(4.5). If in addition to the assumptions of Theorems 5.4, 5.11, τ satisfies $\tau \leq C_k/\eta$, then there exists a constant $\tilde{\mathbf{D}}$ depending on the ratios $(C_y/\eta), (C_c/\eta)$, $e^{TC_k/\eta}$ and the constant $\mathbf{D}$ of Theorem 5.11, such that*

$$\|e_{2h}\|^2_{L^\infty[0,T;L^2(\Omega)]} \leq \tilde{\mathbf{D}}(1/\alpha^2)\int_0^T \left(\|e_1\|^2_{H^1(\Omega)} + \|r_1\|^2_{H^1(\Omega)}\right)dt.$$

Here, $\tilde{\mathbf{D}}$ is also independent of τ, h, α.

Proof. We begin by integrating by parts with respect to time in (5.10), and substituting $v_h = \hat{e}_{2h}$, where $\hat{e}_{2h}$ denotes the approximation of the discrete characteristic function $\chi_{[t^{n-1},t)}e_{2h}$ (for any fixed $t \in [t^{n-1}, t^n)$), as constructed in Appendix B. The definition of the $\hat{e}_{2h}$ (see Appendix B) and the fact that $e_{2ht} \in \mathcal{P}_{k-1}[t^{n-1}, t^n; U_h^n]$

implies that $\int_{t^{n-1}}^{t^n}(e_{2ht},\hat{e}_{2h})dt = \int_{t^{n-1}}^{t}(e_{2ht},e_{eh})dt$ which implies,

$$(1/2)\|e_{2h}(t)\|^2_{L^2(\Omega)} + (1/2)\|[e_{2h}^{n-1}]\|^2_{L^2(\Omega)} + \int_{t^{n-1}}^{t^n}\Big(a(e_{2h},\hat{e}_{2h}) + \langle\phi(y)-\phi(y_h),\hat{e}_{2h}\rangle\Big)dt$$
$$= (1/2)\|e_{2h}^{n-1}\|^2_{L^2(\Omega)} - \int_{t^{n-1}}^{t^n}(1/\alpha)(r_1+r_{2h},\hat{e}_{2h})dt. \qquad (5.21)$$

Recall also that the continuity properties on $a(.,.)$, ϕ and Proposition B.1, imply

$$\Big|\int_{t^{n-1}}^{t^n} a(e_{2h},\hat{e}_{2h})dt\Big| \le C(C_k,C_c)\int_{t^{n-1}}^{t^n}\|e_{2h}\|^2_{H^1(\Omega)}dt$$

while the coupling term can be bounded as:

$$\Big|\frac{1}{\alpha}\int_{t^{n-1}}^{t^n}(r_1+r_{2h},\hat{e}_{2h})dt\Big| \le (C_k/\alpha^2)\int_{t^{n-1}}^{t^n}\Big(\|r_{2h}\|^2_{L^2(\Omega)}+\|r_1\|^2_{L^2(\Omega)}\Big)dt$$
$$+C_k\int_{t^{n-1}}^{t^n}\|e_{2h}\|^2_{L^2(\Omega)}dt.$$

Here we have used Young's inequality with appropriate $\delta > 0$ and Proposition B.1. For the semilinear term, recall that the growth condition, and generalized Hölder inequality, the embedding $H^1(\Omega)\subset L^4(\Omega)$ imply

$$\int_{t^{n-1}}^{t^n}\langle\phi(y)-\phi(y_h),\hat{e}_{2h}\rangle dt \le C_L\int_{t^{n-1}}^{t^n}\|y-y_h\|_{H^1(\Omega)}\|\hat{e}_{2h}\|_{H^1(\Omega)}dt.$$

Using Young's inequality, we finally arrive at:

$$\int_{t^{n-1}}^{t^n}\langle\phi(y)-\phi(y_h),\hat{e}_{2h}\rangle dt \le C_k(C_y+C_L)\int_{t^{n-1}}^{t^n}\Big(\|e_1\|^2_{H^1(\Omega)}+\|e_{2h}\|^2_{H^1(\Omega)}\Big)dt$$

where C_y depends only upon $\|y\|_{L^\infty[0,T;L^2(\Omega)]}$. Hence, substituting the above estimates into (5.21), we obtain an inequality of the form, $(1-C\tau_n)a^n \le a^{n-1}+f^n$, where $a^n = \sup_{s\in(t^{n-1},t^n]}\|e_{2h}(s)\|^2_{L^2(\Omega)}$. Indeed, let $t\in(t^{n-1},t^n]$ to be chosen as $a^n \equiv \|e_{2h}(t)\|^2_{L^2(\Omega)}$ and note that $C_k\int_{t^{n-1}}^{t^n}\|e_{2h}\|^2_{L^2(\Omega)}dt \le C_k\tau_n a^n$, for τ_n satisfying $\tau_n C_k < (1/4)$ the desired estimate follows by the discrete Grönwall Lemma, upon using the previous bounds of Lemma 5.9, Theorems 5.4, 5.11, and standard algebra. ∎

Estimate on the adjoint variable μ, follow using similar techniques and the previously derived estimates on the primal variable. Below, we state the relevant estimate.

THEOREM 5.15. *Let $y_h,\mu_h\in\mathcal{U}_h$ be the solutions of (4.4)-(4.5). Suppose that the Assumptions of Theorems 5.11-5.14 hold. Then there exists a constant* $\tilde{\mathbf{D}} > 0$ *(similar to Theorem 5.14) such that*

$$\|r_{2h}\|^2_{L^\infty[0,T;L^2(\Omega)]} \le \tilde{\mathbf{D}}\int_0^T\Big(\|e_1\|^2_{H^1(\Omega)}+\|r_1\|^2_{H^1(\Omega)}\Big)dt.$$

5.3. Symmetric Error Estimates. Various estimates can be derived, using results of Section 5, and standard approximation theory results. We begin by stating symmetric error estimates.

THEOREM 5.16. *Suppose that Assumptions 2.1-4.2-5.1 hold. Let $y_h, \mu_h \in \mathcal{U}_h$ denote the approximate solutions of the optimality system (4.4)-(4.5) computed using the discontinuous Galerkin scheme. Suppose that $\tau = \max_{i=1,\dots,n} \tau_n$, h, satisfy the conditions of Lemmas 4.4,4.10 and Theorem 5.14. Then, the following estimate holds:*

$$\|e\|_X^2 + (1/\alpha)\|r\|_X^2 \le \tilde{\mathbf{C}}(1/\alpha^2)\Big(\|e_0\|_{L^2(\Omega)}^2 + \|e_p\|_X^2 + \|r_p\|_X^2\Big)$$
$$+ \sum_{i=0}^{N-1} 2\min\Big(\|(I-P_i)y(t^i)\|_{L^2(\Omega)}^2, (1/\tau^{i+1}\eta)\|P_{i+1}(I-P_i)y(t^i)\|_{H^{-1}(\Omega)}^2\Big)$$
$$+ \sum_{i=1}^{N} 2\min\Big(\|(I-P_{i+1})\mu(t^i)\|_{L^2(\Omega)}^2, (1/\tau^i\eta)\|P_i(I-P_{i+1})\mu(t^i)\|_{H^{-1}(\Omega)}^2\Big)$$

where $\tilde{\mathbf{C}}$ depends upon the stability constants of Lemmas 4.4, 4.10, and the constants $\mathbf{C}$, $\mathbf{D}$, $\tilde{\mathbf{D}}$ of Lemma 5.9 and Theorems 5.11, 5.14 respectively, but is independent of τ, h, α. In addition, suppose that the same subspaces are being used, i.e., $U_h^n = U_h$. Then,

$$\|e\|_X^2 + (1/\alpha)\|r\|_X^2 \le \tilde{\mathbf{C}}(1/\alpha^2)\Big(\|e_0\|_{L^2(\Omega)}^2 + \|e_p\|_X^2 + \|r_p\|_X^2\Big).$$

Proof. The first estimate follows by using triangle inequality and previous estimates of Theorems 5.4-5.11. The second estimate follows by Remark 5.5. □

Using now standard regularity and approximation theory results we obtain convergence rates. Below, we state convergence rates in two distinct cases, depending on the available regularity.

PROPOSITION 5.17. *Suppose that the assumptions of Theorems 5.4-5.11 hold. Suppose also that y, μ satisfy,*

$$(y,\mu) \in L^\infty[0,T; H^{l+1} \cap H_0^1(\Omega)] \qquad (y^{(k+1)}, \mu^{(k+1)}) \in L^\infty[0,T; H^1(\Omega)].$$

Assume that piecewise polynomials of degree l are being used to construct the subspaces $U_h^n \subset H^1(\Omega)$ in each time step, where h denotes the spacial discretization parameter. Then the following estimate holds:

$$\|e\|_X^2 + (1/\alpha)\|r\|_X^2 \le \tilde{\mathbf{C}}(1/\alpha^2)\Big(h^{2l} + \tau^{2(k+1)} + h^{2l}\min\{h^4/(\tau^2\eta), h^2/\tau\}\Big).$$

Here the constant $\tilde{\mathbf{C}}$ denotes the constant of Theorem 5.16. In case that $U_h^n = U_h$ then the following estimate is valid

$$\|e\|_X^2 + (1/\alpha)\|r\|_X^2 \le \tilde{\mathbf{C}}(1/\alpha^2)\Big(h^{2l} + \tau^{2(k+1)}\Big).$$

Proof. It remains to estimate e_p, r_p. Using [13, Corollary 4.8], and the standard approximation properties of P_n, we obtain,

$$\|y - \mathbb{P}_n^{loc} y\|_{L^2[t^{n-1},t^n;H^1(\Omega)]} \le C\big(\|y - P_n y\|_{L^2[t^{n-1},t^n;H^1(\Omega)]} + \tau^{k+1}\|P_n y^{(k+1)}\|_{L^2[t^{n-1},t^n;H^1(\Omega)]}\big)$$
$$\le C\big(h^l\|y\|_{L^2[t^{n-1},t^n;H^{l+1}(\Omega)]} + \tau^{k+1}\|y^{(k+1)}\|_{L^2[t^{n-1},t^n;H^1(\Omega)]}\big).$$

Therefore,

$$\|y - \mathbb{P}_h^{loc} y\|_{L^2[0,T;H^1(\Omega)]} \leq C\big(h^l \|y\|_{L^2[0,T;H^{l+1}(\Omega)]} + \tau^{k+1}\|y^{(k+1)}\|_{L^2[0,T;H^1(\Omega)]}\big).$$

Working similarly, we also obtain that

$$\|y - \mathbb{P}_h^{loc} y\|_{L^\infty[0,T;L^2(\Omega)]} \leq C\big(h^{l+1} \|y\|_{L^\infty[0,T;H^{l+1}(\Omega)]} + \tau^{k+1}\|y^{(k+1)}\|_{L^\infty[0,T;H^1(\Omega)]}\big).$$

Similar estimates also hold for r_p. It remains to bound the error terms due to the change of subspaces. For that purpose, it is easy to see that

$$\sum_{i=0}^{N-1} 2\min\Big(\|(I-P_i)y(t^i)\|^2_{L^2(\Omega)}, (1/\tau^{i+1}\eta)\|P_{i+1}(I-P_i)y(t^i)\|^2_{H^{-1}(\Omega)}\Big)$$
$$\leq C\|y\|^2_{C[0,T;H^{l+1}(\Omega)]}\min\{h^{2l+4}/(\tau^2\eta), h^{2+2l}/\tau\}$$

while a similar estimates also holds for the terms involving the adjoint variable.
□

Our last result concerns error estimates under more restrictive regularity assumptions on the solution, and in particular on the time-derivative.

PROPOSITION 5.18. *Suppose that the assumptions of 5.4-5.11 hold. Suppose also that y, μ satisfy,*

$$(y, \mu) \in L^\infty[0,T; H^{l+1}\cap H^1_0(\Omega)] \qquad (y^{(k+1)}, \mu^{(k+1)}) \in L^\infty[0,T;L^2(\Omega)].$$

Assume that the same subspaces are being used in every time-step $U_h^n = U_h$ and piecewise polynomials of degree l are being used to construct the subspace $U_h \subset H^1(\Omega)$, where h denotes the spacial discretization parameter. Suppose that the assumptions of Theorem 5.16 hold. Then, we obtain,

$$\|e\|_X^2 + (1/\alpha)\|r\|^2_X \leq \tilde{\mathbf{C}}(1/\alpha^2)\Big(h^{2l} + (\tau^{2k+2}/h^2)\Big),$$

where $\tilde{\mathbf{C}}$ denote the constant of Theorem 5.16.

Proof. Working similar to the previous theorem, and an inverse estimate lead to

$$\|y - \mathbb{P}_n^{loc} y\|^2_{L^2[t^{n-1},t^n;H^1(\Omega)]} \leq C\|y - P_n y\|^2_{L^2[t^{n-1},t^n;H^1(\Omega)]} + C_k\tau^{2(k+1)}\|P_n y^{(k+1)}\|^2_{L^2[t^{n-1},t^n;H^1(\Omega)]}$$
$$\leq C\left(\|y-P_n y\|^2_{L^2[t^{n-1},t^n;H^1(\Omega)]} + (\tau^{2(k+1)}/h^2)\|P_n y^{(k+1)}\|^2_{L^2[t^{n-1},t^n;L^2(\Omega)]}\right).$$

The projection error in $L^\infty[t^{n-1},t^n;L^2(\Omega)]$ can be treated similarly. The adjoint variable can be treated similarly. Thus, using the stability of the orthogonal projection, we obtain the desired estimate. □

REMARK 5.19. *It is clear from the proofs of Propositions 5.17 and 5.18 that the enhanced regularity assumptions on (y, μ) is only needed to obtain (optimal) rates with respect to $\|.\|_{L^\infty[0,T;L^2(\Omega)]}$ part of the corresponding $\|.\|_X$ norm. Indeed, if we choose the same subspaces in each time step $U_h^n = U_h$ then there is no contribution from the jump-terms, and hence we may combine the results or Remark 5.5, and Theorem 5.11, to relate the errors $\|e\|_{L^2[0,T;H^1(\Omega)]}$ and $\|r\|_{L^2[0,T;H^1(\Omega)]}$ with projection errors $y - \mathbb{P}_h^{loc} y$ and $\mu - \mathbb{P}_h^{loc}\mu$ at the same norms. As a consequence, the rates of convergence of Proposition 5.18, with respect to $\|.\|_{L^2[0,T;H^1(\Omega)]}$ norms only require $(y,\mu) \in L^2[0,T;H^{l+1}(\Omega)] \cap H^{k+1}[0,T;L^2(\Omega)]$ regularity.*

REMARK 5.20. *Due to the absence of control constraints, an estimate on the controls $g - g_h$ follow directly from the estimate on the adjoint $\|\mu - \mu_h\|_X$ using the optimality condition. However, as it is indicated in the subsequent numerical experiments, an improved rate of convergence in the $L^2[0,T;L^2(\Omega)]$ norm is expected for the controls. This issue will be investigated elsewhere.*

5.4. Numerical Experiments. In this section, we are going to validate numerically the proven a priori error estimates for $k = 0$, $l = 1$, in the cases $\tau = h^2$ and $\tau = h$ for the error in the control, state, and adjoint state.

We consider the following numerical example for the model problem with known analytical exact solution on $\Omega \times (0,T) = (0,1)^2 \times (0,0.1)$ and homogenous Dirichlet boundary conditions, similar to the one presented in [48]. In particular, we minimize the functional

$$J(y,g) = \frac{1}{2}\int_0^T \|y - U\|^2_{L^2(\Omega)} dt + \frac{\alpha}{2}\int_0^T \|g\|^2_{L^2(\Omega)} dt$$

subject to the constraints,

$$\begin{cases} y_t - \operatorname{div}[A(x)\nabla y] + (1/3)y^3 &= f + g \quad \text{in } (0,T)\times\Omega \\ y &= 0 \quad \text{on } (0,T)\times\Gamma \\ y(0,x) &= y_0 \quad \text{in } \Omega. \end{cases}$$

We will chose regularization parameter $\alpha = \pi^{-4}$, right-hand side

$$\begin{aligned} f(t,x_1,x_2) = &-\pi^4 e^{-\sqrt{5}\pi^2 T} sin(\pi x_1) sin(\pi x_2) \\ &+\frac{1}{3}(\frac{-1}{2-\sqrt{5}}\pi^2 e^{-\sqrt{5}\pi^2 t} sin(\pi x_1) sin(\pi x_2))^3, \end{aligned}$$

target function

$$\begin{aligned} U(t,x_1,x_2) = &\left(2\pi^2 e^{-\sqrt{5}\pi^2 T} - \frac{\pi^4}{(2-\sqrt{5})^2}\left(e^{-\sqrt{5}\pi^2 t} sin(\pi x_1) sin(\pi x_2)\right)^2 \left(e^{-\sqrt{5}\pi^2 t} - e^{-\sqrt{5}\pi^2 T}\right)\right) \\ &\times sin(\pi x_1) sin(\pi x_2), \end{aligned}$$

and initial condition $y_0(x_1,x_2) = \frac{-1}{2-\sqrt{5}}\pi^2 sin(\pi x_1) sin(\pi x_2)$, in a way to guarantee that the optimal solution triple (y,μ,g) of the above problem is given by

$$y(t,x_1,x_2) = \frac{-1}{2-\sqrt{5}}\pi^2 e^{-\sqrt{5}\pi^2 t} sin(\pi x_1) sin(\pi x_2),$$

$$\mu(t,x_1,x_2) = (e^{-\sqrt{5}\pi^2 t} - e^{-\sqrt{5}\pi^2 T}) sin(\pi x_1) sin(\pi x_2),$$

$$g(t,x_1,x_2) = -\pi^4 (e^{-\sqrt{5}\pi^2 t} - e^{-\sqrt{5}\pi^2 T}) sin(\pi x_1) sin(\pi x_2).$$

The optimal control problem is solved by the finite element toolkit FreeFem++ (see [2, 32]) using a conjugate gradient algorithm method. The mesh-generator and the linear algebra solver is the standard one provided by the toolkit. Two different experiments are being performed, with modest values of the discretization parameters τ, h. The first one, for $\tau \approx h$, while the second one requires a more restrictive time-step approach $\tau \approx h^2$, which is typically more standard. In both cases the expected rates of convergence are being computed for the $L^2[0,T;H^1(\Omega)]$ norms for the the state and adjoint variables, i.e., $\mathcal{O}(h)$, while in the second experiment we also recover the expected quadratic rate of convergence in $L^2[0,T;L^2(\Omega)]$ norm for the control.

TABLE 5.1
Rates of convergence for the 2d solution with $k = 0$, $l = 1$ ($h = \tau$).

Discretization	Error		
$h = \tau$	$\|e\|_{L^2[0,T;\mathrm{H}_0^1(\Omega)]}$	$\|r\|_{L^2[0,T;\mathrm{H}_0^1(\Omega)]}$	$\|g - g_h\|_{L^2[0,T;\mathrm{L}^2(\Omega)]}$
$h = 0.02946280$	3.631050	0.05551130	0.02498330
$h = 0.01473140$	1.508560	0.02618430	0.01082740
$h = 0.00736570$	0.772711	0.01454260	0.00561528
$h = 0.00368285$	0.391391	0.00758848	0.00281426
Convergence rate	1.071233	0.95696566	1.05004366

TABLE 5.2
Rates of convergence for the 2d solution with $k = 0$, $l = 1$ ($h^2 = \tau$).

Discretization	Error		
$h^2 = \tau$	$\|e\|_{L^2[0,T;\mathrm{H}_0^1(\Omega)]}$	$\|r\|_{L^2[0,T;\mathrm{H}_0^1(\Omega)]}$	$\|g - g_h\|_{L^2[0,T;\mathrm{L}^2(\Omega)]}$
$h = 0.1178510$	2.254550	0.04141390	0.07661170
$h = 0.0589256$	1.003230	0.01943350	0.02208320
$h = 0.0294628$	0.470049	0.00914215	0.00546600
$h = 0.0147314$	0.229416	0.00445367	0.00135706
Convergence rate	1.051790	1.06430666	1.89617666

6. Conclusion. We conclude this work by noting that the above symmetric estimates imply that the error will be as good as the approximation theory of the subspaces and the regularity theory of the underlying control problem will allow it to be. A key feature of the analysis is that there is no exponential dependence upon the parameter $(1/\alpha)$ which captures the information about the size of the control. The estimates are still applicable for time-steps that are chosen independent of the spacial discretization parameter h. Other type of controls and some computational results will be studied elsewhere.

Appendix A. Quotation of results related to an exponential interpolant.

The polynomial interpolant of functions $e^{-\rho(t-t^{n-1})}v$, where $v \in \mathcal{P}_k[t^{n-1}, t^n; V]$ and V is any linear space, is needed in the proof of the main stability estimate. Here, we quote the definition and the main results from [12].

DEFINITION A.1. *Let V be a linear space, and $\rho > 0$ be given. If $v = \sum_{i=0}^{k} r_i(t) v_i \in \mathcal{P}_k[t^{n-1}, t^n; V]$, with $r_i \in \mathcal{P}_k[t^{n-1}, t^n]$ and $v_i \in V$, we define the exponential interpolant of v by*

$$\bar{v} = \sum_{i=0}^{k} \bar{r}_i(t) v_i$$

where $\bar{r}_i \in \mathcal{P}_k[t^{n-1}, t^n]$ is the approximation of $r_i(t)e^{-\rho(t-t^{n-1})}$ satisfying $r_i(t^{n-1}) = \bar{r}_i(t^{n-1})$ and

$$\int_{t^{n-1}}^{t^n} \bar{r}_i(t) q(t) dt = \int_{t^{n-1}}^{t^n} r_i(t) q(t) e^{-\rho(t-t^{n-1})} dt, \qquad q \in \mathcal{P}_{k-1}[t^{n-1}, t^n].$$

The following Lemma (see [12, Lemma 3.4]) asserts that the difference $v - \bar{v}$ remains small in various norms.

Lemma A.2. *Let V and Q be linear spaces and $v \to \bar{v}$ be the map constructed in Definition A.1, for given $\rho > 0$. If $L(.,.) : V \times Q \to \mathbb{R}$ denotes a bilinear mapping and $v \in \mathcal{P}_k[t^{n-1}, t^n; V]$ then*

$$\int_{t^{n-1}}^{t^n} L(\bar{v}(t), q(t))dt = \int_{t^{n-1}}^{t^n} L(v(t), q(t))e^{-\rho(t-t^{n-1})}dt, \qquad \forall\, q \in \mathcal{P}_{k-1}[t^{n-1}, t^n; Q].$$

If $(.,.)_V$ is a (semi) inner product on V, then there exists a constant C_k independent of $\rho > 0$, such that

$$\|v - \bar{v}\|_{L^2[t^{n-1},t^n;V]} \leq C_k \rho(t^n - t^{n-1})\|v\|_{L^2[t^{n-1},t^n;V]}.$$

Appendix B. Quotation of results related to the discrete characteristic function.

Note that the computation of the error at arbitrary times $t \in [t^{n-1}, t^n)$ can be facilitated by substituting $v_h = \chi_{[t^{n-1},t)} y_h$ into the discrete equations. However, this choice is not available since $\chi_{[t^{n-1},t)} y_h$ is not a member of $\mathcal{U}_h$, unless t is actually a partition point. Therefore approximations of such functions need to be constructed. This is done in [11, Section 2.3]. For completeness we state the main results. The approximations are constructed on the interval $[0, \tau)$, where $\tau = t^n - t^{n-1}$ and they are invariant under translations.

Let $t \in (0, \tau)$. We consider polynomials $s \in \mathcal{P}_k(0, \tau)$, and we denote the discrete approximation of $\chi_{[0,t)} s$ by the polynomial $\hat{s} \in \{\hat{s} \in \mathcal{P}_k(0, \tau), \hat{s}(0) = s(0)\}$ which satisfies

$$\int_0^\tau \hat{s} q = \int_0^t s q \qquad \forall\, q \in \mathcal{P}_{k-1}(0, \tau).$$

The motivation for the above construction stems from the elementary observation that for $q = s'$ we obtain $\int_0^\tau s' \hat{s} = \int_0^t s s' = \frac{1}{2}(s^2(t) - s^2(0))$.

The construction can be extended to approximations of $\chi_{[0,t)} v$ for $v \in \mathcal{P}_k[0, \tau; V]$ where V is a linear space. The discrete approximation of $\chi_{[0,t)} v$ in $\mathcal{P}_k[0, \tau; V]$ is defined by $\hat{v} = \sum_{i=0}^k \hat{s}_i(t) v_i$ and if V is a semi-inner product space then,

$$\hat{v}(0) = v(0), \qquad \text{and} \int_0^\tau (\hat{v}, w)_V = \int_0^t (v, w)_V \quad \forall w \in \mathcal{P}_{k-1}[0, \tau; V].$$

Finally, we quote the main result from [11].

Proposition B.1. *Suppose that V is a (semi) inner product space. Then the mapping $\sum_{i=0}^k s_i(t) v_i \to \sum_{i=0}^k \hat{s}_i(t) v_i$ on $\mathcal{P}_k[0, \tau; V]$ is continuous in $\|.\|_{L^2[0,\tau;V]}$. In particular,*

$$\|\hat{v}\|_{L^2[0,\tau;V]} \leq C_k \|v\|_{L^2[0,\tau;V]}, \qquad \|\hat{v} - \chi_{[0,t)} v\|_{L^2[0,\tau;V]} \leq C_k \|v\|_{L^2[0,\tau;V]}$$

where C_k is a constant depending on k.

Proof. See [11, Lemma 2.4]. □

Remark B.2. *Combining the above estimate with standard scaling arguments and the finite dimensionality of $\mathcal{P}_k[0, \tau]$ we also obtain an estimate of the form*

$$\|\hat{v}\|_{L^\infty[0,\tau;L^2(\Omega)]} \leq C_k \|v\|_{L^\infty[0,\tau,L^2(\Omega)]}.$$

Acknowledgements: The authors would like to thank the referees for their detailed reports, which help the improvement of the paper. The second author acknowledges the financial support of a *Papakyriakopoulos* grant.

REFERENCES

[1] G. Akrivis and C. Makridakis, Galerkin time-stepping methods for nonlinear parabolic equations, *ESAIM: Math. Model. and Numer. Anal.*,**38** (2004), pp. 261-289.

[2] G. Allaire and O. Pantz , Structural optimization with FreeFem++. *Struct. Multidiscip. Optim.* , **32** (2), 2006, pp. 173-181.

[3] T. Apel, and T. Flaig, Crank-Nicolson schemes for optimal control problems with evolution equations, *submitted, available at http://www.unibw.de/bauv1/personen/apel/papers*

[4] A. Borzi, and R. Griesse, Distributed optimal control for lambda-omega systems, *J. Numer. Math.*, **14** (2006), pp 17-40.

[5] E. Casas, and J.-P. Raymond, Error estimates for the numerical approximation of Dirichlet boundary control for semilinear elliptic equation, *SIAM J. Control and Optim.*, **45** (No 5) (2006), pp. 1586-1611.

[6] E. Casas, M. Mateos and F. Tröltzsch, Error estimates for the numerical approximation of boundary semilinear elliptic control problem, *Comput. Optim. and Appl.*, **31** (2005), pp. 193-219.

[7] K. Chrysafinos, Discontinous Galerkin approximations for distributed optimal control problems constrained to linear parabolic PDE's., *Int. J. Numer. Anal. and Mod.*, **4**, (No 3-4) (2007), pp 690-712.

[8] K. Chrysafinos, Analysis and finite element approximations for distributed optimal control problems for implicit parabolic equations, *J. Comput. Appl. and Math.*, **231** (2009), pp 327-348.

[9] K. Chrysafinos, Convergence of discontinuous Galerkin approximations of an optimal control problem associated to semilinear parabolic PDE's., *ESAIM M^2AN*, **44** (No 1) (2010), pp 189-206.

[10] K. Chrysafinos, Convergence of discontinuous time-stepping schemes for a Robin boundary control problem under minimal regularity assumptions, *submitted, available at http://www.math.ntua.gr/∼chrysafinos.*

[11] K. Chrysafinos and N.J. Walkington, Error estimates for the discontinuous Galerkin methods for parabolic equations, *SIAM J. Numer. Anal.*, **44** (No 1) (2006), pp 349-366.

[12] K. Chrysafinos and N.J. Walkington, Discontinuous Galerkin approximations of the Stokes and Navier-Stokes equations, *Math. Comp*, **79** (No 272) (2010), pp 2135-2167.

[13] K. Chrysafinos and N.J. Walkington, Lagrangian and moving mesh methods for the convection diffusion equation, *ESAIM M^2AN.*, **42** (No 1) (2008), pp 25-55.

[14] 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.

[15] P.G. Ciarlet, The finite element method for elliptic problems, *SIAM Classics in Applied Math*, 2002.

[16] K. Deckelnick and M. Hinze, Semidiscretization and error estimates for distributed control of the instationary Navier-Stokes equations, *Numer. Math.*, **97** (2004), pp. 297-320.

[17] K. Deckelnick and M. Hinze, Variational discretization of parabolic control problems in the presence of pointwise state constraints, *J. Comput. Math.* **29** No 1 (2011), pp 1-16.

[18] 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.

[19] D. Estep and S. Larsson, The discontinuous Galerkin method for semilinear parabolic equations, *RAIRO Modél. Math. Anal. Numér.*, **27** (1993), pp 35-54.

[20] K. Eriksson and C. Johnson, Adaptive finite element methods for parabolic problems. I. A linear model problem, *SIAM J. Numer. Anal.*, **28** (1991), pp. 43-77.

[21] K. Eriksson and C. Johnson, Adaptive finite element methods for parabolic problems. II. Optimal error estimates in $L_\infty(L^2)$ and $L_\infty(L_\infty)$, *SIAM J. Numer. Anal.*, **32** (1995), pp. 706-740.

[22] K. Ericksson and C. Johnson, Adaptive finite element methods for parabolic problems IV: Nonlinear problems, *SIAM J. Numer. Anal.*, **32** (6) (1995), pp. 1729-1749.

[23] K. Eriksson, C. Johnson and V. Thomée, Time discretization of parabolic problems by the discontinuous Galerkin method, *RAIRO Modél. Math. Anal. Numér.*, **29** (1985), pp 611-643.

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

[25] R. Falk, Approximation of a class of otimal control problems with order of convergence estimates, *J. Math. Anal. Appl.*, **44** (1973), pp. 28-47.

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

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

[28] W. Gong, M. Hinze and Z. Zhou, A priori analysis for finite element approximation of parabolic optimal control problems with pointwise control, *submitted, available at http://preprint.math.uni-hamburg.de/public/papers/hbaum/hbaum2011-07.pdf*

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

[30] 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.

[31] 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.

[32] F. Hecht, FreeFem++ Third Edition - Version 3.13, *available at http://www.freefem.org/ff++*, 2011.

[33] M. Hinze, A variational discretization concept in control constrained optimization: The linear-quadratic case, *Comput. Optim. Appl.*, **30** (2005), pp. 45-61.

[34] M. Hinze and K. Kunisch, Second order methods for optimal control of time-dependent fluid flow, *SIAM J. Control and Optim.* **40** 2001, pp. 925-946.

[35] K. Ito and K. Kunisch, Lagrange multiplier approach to variational problems and applications, *Advances in Deisign and Control*, **15** 2008, SIAM publications.

[36] G. Knowles, Finite element approximation of parabolic time optimal control problems, *SIAM J.Control and Optim.*, **20** (1982), pp. 414-427.

[37] 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.

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

[39] J.-L. Lions, Some aspects of the control of distributed parameter systems, Conference Board of the Mathematical Sciences, SIAM, 1972.

[40] 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.

[41] 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** no 3 (2004), pp. 1032-1061.

[42] 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.

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

[44] D. Meidner and B. Vexler, Adaptive space-time finite element methods for parabolic optimization problems, *SIAM J. Control and Optim.*, **46** (2007), pp. 116-142.

[45] 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.

[46] D. Meidner and B. Vexler, A-priori error analysis of the Petrov-Galerkin Crank-Nicolson scheme for parabolic optimal control problems, *accepted in SIAM J. Control and Optim. Preprint available at http://www-m1.ma.tum.de/bin/view/Lehrstuhl/BorisVexlerPublic*

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

[48] I. Neitzel and B. Vexler, A priori error estimates for space-time finite element discretization of semilinear parabolic optimal control problems, *published on line in Numer. Math.*.

[49] A. Rösch, Error estimates for parabolic optimal control problems with control constraints, *Zeitschrift für Analysis und ihre Anwendunges, ZAA*, **23** (2004), pp. 353-376.

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

[51] F. Tröltzsch, Semidiscrete Ritz-Galerkin approximation of nonlinear parabolic boundary control problems, *International Series of Numerical Mathematics*, **111** (1993), pp 57-68.

[52] F. Tröltzsch, Semidiscrete Ritz-Galerkin approximation of nonlinear parabolic boundary control problems- Strong convergence of optimal controls, *Appl. Math. Optim.*, **29** (1994), pp 309-329.

[53] F. Tröltzsch, Optimal control of partial differential equations: Theory, methods and applications, *Graduate Studies in Mathematics, AMS*, **112**, Providence 2010.

[54] N. J. Walkington, Compactness properties of the DG and CG time stepping schemes for parabolic equations, *SIAM J. Numer. Anal*, **47** (2010), pp. 4680-4710.

[55] R. Winther, Error estimates for a Galerkin approximation of a parabolic control problem, *Ann. Math. Pura Appl.* **117** (1978), pp 173-206.

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

[57] E. Zeidler, Nonlinear functional analysis and its applications, II/B Nonlinear monotone operators, Springer-Verlag, New York, 1990.