ERROR ESTIMATES FOR DISCONTINUOUS GALERKIN TIME-STEPPING SCHEMES FOR ROBIN BOUNDARY CONTROL PROBLEMS CONSTRAINED TO PARABOLIC PDES

KONSTANTINOS CHRYSAFINOS* AND EFTHIMIOS N. KARATZAS*

Abstract. We consider fully discrete finite element approximations of a Robin optimal boundary control problem, constrained by linear parabolic PDEs with rough initial data. Conforming finite element methods for spatial discretization combined with discontinuous time-stepping Galerkin schemes are being used for the space-time discretization. Error estimates are proved under weak regularity hypotheses for the state, adjoint and control variables. Computational examples validating our expected rates of convergence are also provided.

Keywords: Discontinuous Time-Stepping Schemes, Finite Element Approximations, Robin Boundary Control, Parabolic equations, Error Estimates.

Mathematics Subject Classification: 65M60, 49J20.

1. Introduction. We consider an optimal control problem associated to the minimization of the tracking functional subject to linear parabolic PDEs with rough initial data. In particular, given a target function y_d we seek state variable y and Robin boundary control variable g such that the functional

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

is minimized subject to the constraints,

$$y_t - \eta\Delta y = f \text{ in } (0,T]\times\Omega, \quad y + \frac{\eta}{\lambda}\frac{\partial y}{\partial \mathbf{n}} = g \text{ on } (0,T]\times\Gamma, \quad y(0,x) = y_0 \text{ in } \Omega. \tag{1.2}$$

Here, $\Omega \subset \mathbb{R}^2$ denotes an open bounded polygonal and convex domain, with Lipschitz boundary Γ. The control g is applied on the boundary Γ and it is of Robin type. Our analysis and results will be primarily focused on the case of low regularity assumptions, i.e., initial data $y_0 \in L^2(\Omega)$, but our analysis will be also applicable in other cases where the solution possesses additional regularity. Furthermore, we are also interested in case of pointwise control constraints in the sense that $g_a \le g(t,x) \le g_b$ for a.e. $(t,x) \in (0,T]\times\Gamma$, where $g_a, g_b \in \mathbb{R}$. A precise formulation will be given in the next section. The forcing term f and the parameters $\lambda > 0$, $\eta > 0$ are given data, while $\alpha > 0$ denotes a penalty parameter which limits the size of the control and it is comparable to the discretization parameters. The case of rough initial data is very important within the context of such boundary optimal control problems and great care is exercised in order to include this case into our analysis.

The main goal is to show that the error estimates of the corresponding optimality system have the same structure to the estimates of the uncontrolled linear parabolic equation with Robin boundary data. The key -but not the only- structural difficulty associated to boundary optimal control problems with rough initial data stems from the lack of sufficient regularity of the state, adjoint and control variables. In

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

particular, if $y_0 \in L^2(\Omega)$ then the regularity of the state variable is limited to $L^2[0,T;H^1(\Omega)] \cap H^1[0,T;H^1(\Omega)^*]$. Hence, classical boot-strap arguments for the uncontrolled parabolic pdes which rely on standard Ritz-Galerkin elliptic projections, typically fail due to the lack of regularity. As a consequence, error estimates for space-time approximations of parabolic optimal control problems with rough initial data $y_0 \in L^2(\Omega)$ in Lipschitz domains have not been treated before.

To overcome the lack of regularity, we analyze a scheme which is based on a discontinuous time-stepping approach, which is suitable for problems without regular enough solutions. The analysis showcases the favorable behavior of such schemes even in presence of essential Robin boundary controls. The key feature of our discrete schemes is that they exhibit the same regularity properties to the continuous weak problem. Our results can be summarized as follows:

1. We develop a symmetric error estimate under minimal regularity assumptions on the natural norm $W(0,T) \equiv L^\infty[0,T;L^2(\Omega)] \cap L^2[0,T;H^1(\Omega)] \times L^2[0,T;L^2(\Gamma)]$ associated to our discontinuous time-stepping scheme, i.e.,
$$\|\text{error}\|_{W(0,T)} \leq C\|\text{best approximation error}\|_{W(0,T)},$$
which states that the error is as good as the regularity and approximation theory allows it to be.
2. We define a new generalized space-time projection that exhibits best approximation properties in $L^2[0,T;L^2(\Omega)]$, and which is also applicable for $y_t \in L^2[0,T;H^1(\Omega)^*]$. Using the above projection, and an appropriate duality argument for an auxiliary system, we obtain a rate of $\mathcal{O}(h)$ for the $L^2[0,T;L^2(\Omega)]$ norm, when $\tau \leq Ch^2$.
3. In case of bounded controls, we demonstrate the applicability of our estimates within the variational discretization concept of [25]. This approach allows to overcome the lack of the enhanced regularity for the state variable due to the failure of classical "boot-strap" arguments for the control, state and adjoint variables.

To our best knowledge our estimates are new, and optimal in terms of the prescribed regularity of the solutions, and the presence of essential boundary conditions. In addition, even in presence of additional regularity on the data, i.e., $y_0 \in H^1(\Omega)$, and despite the use of L^2 projections which exhibit best approximation properties, the rate $\mathcal{O}(h^{3/2})$ (when $\tau \leq Ch^2$) appears to be optimal since there is no possibility to obtain a better estimate at least when polygonal and convex domains are involved. We also point out that the Robin boundary control can be viewed as a penalization approach for Dirichlet boundary control problems (see for instance the works of [3, 7, 27] and references within). For this reason the dependence upon the parameters λ, α, η of various constants appearing in our estimates is carefully tracked.

1.1. Related results. Previous related results regarding discontinuous time-stepping approaches are almost exclusively related to distributed controls. For instance, the discontinuous Galerkin framework is explored in the works of [34] and [33] where a-priori estimates are developed for distributed optimal control problems with and without control constraints respectively for the heat equation. In [8, 9] a priori error estimates in terms of suitable space-time projections, are derived for unconstrained distributed optimal control problems related to parabolic and implicit parabolic pdes with general and possibly time-dependent coefficients in the elliptic part. Error estimates related to distributed optimal control problems for semi-linear parabolic pdes are proved in the work of [37], with control constraints and $H_0^1(\Omega) \cap L^\infty(\Omega)$ initial data, while a priori error estimates of symmetric type for problems without control constraints are analyzed in [13]. A-priori error estimates for the velocity tracking problem with control constraints are analyzed in the works of [5, 6]. A convergence result for discontinuous time-stepping schemes for Robin optimal control problems (without control constraints) related to semi-linear parabolic pdes, under $L^2(\Omega)$ data is recently considered in [10]. Finally, in [32] fully-discrete approximations of a Neumann boundary control problem related to homogeneous linear parabolic pdes are analyzed, for the implicit Euler scheme, for smooth domains and for regular enough data.

Several results regarding the analysis of optimal boundary control problems can be found in [22, 30, 31, 36, 39] (see also references within). Various boundary control problems related to time-dependent pdes were studied in the previous works of [1, 2, 11, 23, 26, 28, 29, 40, 39, 41].

2. Background.

2.1. Notation. We use the standard notation for the Sobolev spaces $H^s(\Omega)$, and $H^s(\Gamma)$, with $s \in \mathbb{R}$ with norms denoted by $\|\cdot\|_{H^s(\Omega)}$ and $\|\cdot\|_{H^s(\Gamma)}$ respectively. The dual space of $H^1(\Omega)$ is denoted by $H^1(\Omega)^*$, and the corresponding duality pairing by $\langle\cdot,\cdot\rangle_{H^1(\Omega)^*,H^1(\Omega)} \equiv \langle\cdot,\cdot\rangle$. We will also use the space $H^{1/2}(\Gamma)$, its dual denoted by $H^{-1/2}(\Gamma)$, and their duality pairing denoted by $\langle.,.\rangle_{H^{-1/2}(\Gamma),H^{1/2}(\Gamma)} \equiv \langle\cdot,\cdot\rangle_\Gamma$. Finally, the standard notation $(\cdot,\cdot)$, $(\cdot,\cdot)_\Gamma$ will be used for the $L^2(\Omega)$ and $L^2(\Gamma)$ inner products respectively. For any of the above Sobolev spaces, we define the space-time spaces $L^p[0,T;X]$, $L^\infty[0,T;X]$, $C[0,T;X]$ and $H^1[0,T;X]$ in a standard fashion (see e.g. [18, Chapter 5]). We will frequently use the space $W(0,T) := L^\infty[0,T;L^2(\Omega)] \cap L^2[0,T;H^1(\Omega)] \times L^2[0,T;L^2(\Gamma)]$ endowed with the standard "graph" norm. For any $\gamma \geq 0$, we also define the space $H^\gamma[0,T;X]$ in a standard way (see e.g. [18, Chapter 5]).The bilinear form associated to our operator is given by

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

and satisfies the following properties:

$$a(y,y) = \eta \|\nabla y\|^2_{L^2(\Omega)}, \ \alpha(y,v) \leq C\eta \|y\|_{H^1(\Omega)} \|v\|_{H^1(\Omega)} \ \forall y,v \in H^1(\Omega).$$

Finally we recall some useful inequalities which will be used subsequently.
Sobolev's Boundary Inequality (see e.g. [4, Theorem 1.6.6]): If Ω has a Lipschitz boundary then there exists $C > 0$, such that: $\|v\|_{L^2(\Gamma)} \leq C\|v\|^{1/2}_{L^2(\Omega)}\|v\|^{1/2}_{H^1(\Omega)}$, $\forall v \in H^1(\Omega)$.
Generalized Friedrichs' Inequality (see e.g. [35, Theorem 1.9]): There exists $C_F > 0$ (depending only on Ω) such that: $\|\nabla v\|^2_{L^2(\Omega)} + \|v\|^2_{L^2(\Gamma)} \geq C_F\|v\|^2_{H^1(\Omega)}$.

2.2. The continuous control problem. We begin by stating the weak formulation of the state equation. Given $f \in L^2\left[0,T;H^1(\Omega)^*\right]$, $g \in L^2\left[0,T;H^{-1/2}(\Gamma)\right]$, and $y_0 \in L^2(\Omega)$ we seek $y \in L^2[0,T;H^1(\Omega)] \cap H^1[0,T;H^1(\Omega)^*]$ such that for a.e. $t \in (0,T]$, and for all $v \in H^1(\Omega)$,

$$\langle y_t, v\rangle + a(y,v) + \lambda \langle y,v\rangle_\Gamma = \langle f,v\rangle + \lambda \langle g,v\rangle_\Gamma \qquad \text{and} \qquad (y(0),v) = (y_0,v)\,. \quad (2.1)$$

An equivalent weak formulation of (2.1) suitable for the analysis of dG schemes, is to seek $y \in W(0,T)$ such that for all $v \in L^2[0,T;H^1(\Omega)] \cap H^1[0,T;H^1(\Omega)^*]$,

$$\begin{aligned}(y(T),v(T)) &+ \int_0^T (-\langle y, v_t\rangle + a(y,v) + \lambda \langle y,v\rangle_\Gamma)dt \\ &= (y_0,v(0)) + \int_0^T (\langle f,v\rangle + \lambda \langle g,v\rangle_\Gamma)dt. \qquad (2.2)\end{aligned}$$

The basic existence, uniqueness and regularity result of (2.2) follows (see e.g. [11]).

THEOREM 2.1. *Suppose* $g \in L^2[0,T;H^{-1/2+\theta}(\Gamma)] \cap H^\theta[0,T;H^{-1/2}(\Gamma)]$, $y_0 \in H^\theta(\Omega)$, *and* $f \in L^2[0,T;H^{1-\theta}(\Omega)^*]$ *for some* $\theta \in [0,1]$. *Then, there exists a unique* $y \in L^2[0,T;H^{1+\theta}(\Omega)] \cap H^1[0,T;H^{1-\theta}(\Omega)^*]$ *satisfying* (2.2)*, and*

$$\begin{aligned}&\|y\|_{L^2[0,T;H^{1+\theta}(\Omega)]} + \|y_t\|_{L^2[0,T;H^{1-\theta}(\Omega)^*]} \\ &\le C\Big(\|f\|_{L^2[0,T;H^{1-\theta}(\Omega)^*]} + \|u_0\|_{H^\theta(\Omega)} + \|g\|_{L^2[0,T;H^{-1/2+\theta}(\Gamma)]} + \|g\|_{H^\theta[0,T;H^{-1/2}(\Gamma)]}\Big).\end{aligned}$$

Thus, the control to state mapping $G : L^2[0,T;L^2(\Gamma)] \to W(0,T)$, which associates to each control g the state $G(g) = y_g \equiv y(g)$ via (2.2) is well defined, and continuous. Hence, the cost functional, frequently denoted to by its reduced form, $J(y,g) \equiv J(y(g)) \equiv J(g) : L^2[0,T;L^2(\Gamma)] \to \mathbb{R}$ is also well defined and continuous.

DEFINITION 2.2. *Let* $f \in L^2[0,T;H^1(\Omega)^*]$, $y_0 \in L^2(\Omega)$, *and* $y_d \in L^2[0,T;L^2(\Omega)]$ *be given data. Then, the set of admissible controls (denoted by* $\mathcal{A}_{ad}$*), is defined by:*

1. *Unconstrained Controls:* $\mathcal{A}_{ad} \equiv L^2[0,T;L^2(\Gamma)]$.
2. *Constrained Controls:* $\mathcal{A}_{ad} \equiv \{g \in L^2[0,T;L^2(\Gamma)] : g_a \le g(t,x) \le g_b$ *for a.e.* $(t,x) \in (0,T) \times \Gamma\}$.

The pair $(y(g),g) \in W(0,T) \times \mathcal{A}_{ad}$*, is said to be an optimal solution if* $J(y(g),g) \le J(w(h),h)\ \forall (w(h),h) \in W(0,T) \times \mathcal{A}_{ad}$.

We will occasionally abbreviate the notation $y \equiv y_g \equiv y(g)$. Below, we state the main result concerning the existence of an optimal solution (see for instance [39]).

THEOREM 2.3. *Let* $y_0 \in L^2(\Omega)$, $f \in L^2[0,T;H^1(\Omega)^*]$, $y_d \in L^2[0,T;L^2(\Omega)]$ *be given. Then, the boundary control problem has unique solution* $(\bar{y}(\bar{g}),\bar{g}) \in W(0,T) \times \mathcal{A}_{ad}$.

2.3. The optimality system. An optimality system of equations can be derived by using standard techniques; see for instance [39] or [11, Section 2]. We first state the basic differentiability property of the cost functional.

LEMMA 2.4. *The cost functional* $J : L^2[0,T;L^2(\Gamma)] \to \mathbb{R}$ *is of class* C^∞ *and for every* $g,u \in L^2[0,T;L^2(\Gamma)]$,

$$J'(g)u = \int_0^T \int_\Gamma (\mu(g) + \alpha g)u\,dx\,dt,$$

where $\mu(g) \equiv \mu_g \in W(0,T)$ *is the unique solution of the following problem: For all* $v \in L^2[0,T;H^1(\Omega)] \cap H^1[0,T;H^1(\Omega)^*]$,

$$\int_0^T \big(\langle \mu_g, v_t\rangle + a(\mu_g,v) + \lambda \langle \mu_g, v\rangle_\Gamma\big)\,dt = -(\mu_g(0),v(0)) + \int_0^T (y_g - y_d, v)dt \quad (2.3)$$

where $\mu_g(T) = 0$*, and* y_g *is defined by (2.2). In addition,* $(\mu_g)_t \in L^2[0,T;H^1(\Omega)^*]$.

Therefore the optimality system which consists of the state and adjoint equations, and the optimality condition takes the form:

LEMMA 2.5. *Let $(\bar{y}_{\bar{g}}, \bar{g}) \equiv (\bar{y}, \bar{g}) \in W(0,T) \times \mathcal{A}_{ad}$ denote the unique optimal pair of Definition 2.2. Then, there exists an adjoint $\bar{\mu} \in W(0,T)$ satisfying, $\bar{\mu}(T) = 0$ such that for all $v \in L^2[0,T; H^1(\Omega)] \cap H^1[0,T; H^1(\Omega)^*]$,*

$$(\bar{y}(T), v(T)) + \int_0^T \left(-\langle \bar{y}, v_t \rangle + a(\bar{y}, v) + \lambda \langle \bar{y}, v \rangle_\Gamma \right) dt \qquad (2.4)$$

$$= (\bar{y}_0, v(0)) + \int_0^T \left(\langle f, v \rangle + \lambda \langle \bar{g}, v \rangle_\Gamma \right) dt,$$

$$\int_0^T \left(\langle \bar{\mu}, v_t \rangle + a(v, \bar{\mu}) + \lambda \langle \bar{\mu}, v \rangle_\Gamma \right) dt = -(\bar{\mu}(0), v(0)) + \int_0^T (\bar{y} - y_d, v)\, dt, \quad (2.5)$$

1) Unconstrained Controls: $$\int_0^T (\alpha \bar{g} + \lambda \bar{\mu}, u)_\Gamma \, dt = 0 \quad \forall u \in \mathcal{A}_{ad}, \qquad (2.6)$$

2) Constrained Controls: $$\int_0^T \int_\Gamma (\alpha \bar{g} + \lambda \bar{\mu})(u - \bar{g})\, dx dt \geq 0 \quad \forall u \in \mathcal{A}_{ad}. \qquad (2.7)$$

In addition, $\bar{y}_t \in L^2[0,T; H^1(\Omega)^*]$, $\bar{\mu} \in L^2[0,T; H^2(\Omega)] \cap H^1[0,T; L^2(\Omega)]$, and (2.7) is equivalent to $\bar{g}(t,x) = Proj_{[g_a, g_b]} \left(-\frac{\lambda}{\alpha} \bar{\mu}(t,x) \right)$ for a.e. $(t,x) \in (0,T] \times \Gamma$.

Proof. The derivation of the optimality system is standard (see e.g. [39]). For the enhanced regularity on $\bar{\mu}$, we note that $\bar{y} - y_d \in L^2[0,T; L^2(\Omega)]$ and apply the analogue of Theorem 2.1 for (2.5) to get that $\bar{\mu} \in L^2[0,T; H^2(\Omega)] \cap H^1[0,T; L^2(\Omega)]$. For the projection formula we refer the reader to [7, Lemma 4.2] and we note that (2.5) corresponds to backwards in time problem with zero Robin data. □

REMARK 2.6. *We point out that for smooth boundary and for any $v \in H^2(\Omega)$ we obtain that the normal derivative $\frac{\partial v}{\partial \mathbf{n}}$ is well defined and belongs to $H^{1/2}(\Gamma)$. This is not the case when Γ is polygonal domain (say only Lipschitz continuous), despite the fact that on each straight component (denoted by Γ_i) we clearly obtain $\frac{\partial v}{\partial \mathbf{n}}|_{\Gamma_i} \in H^{1/2}(\Gamma_i)$. We refer the reader to [20] for related regularity results for general polygonal domains. If the boundary is smooth, e.g. of class C^2 then $\bar{\mu}|_\Gamma \in L^2[0,T; H^{3/2}(\Gamma)] \cap H^{3/4}[0,T; L^2(\Gamma)]$. Hence, a bootstrap argument can be applied in order to improve the regularity of $\bar{g}, \bar{y}$ (see e.g. [32]). For example, in case of unconstrained controls, $\bar{g} \in L^2[0,T; H^{3/2}(\Gamma)] \cap H^{3/4}[0,T; L^2(\Gamma)]$ too, which results $\bar{y} \in L^2[0,T; H^2(\Omega)] \cap H^1[0,T; L^2(\Omega)]$, when $y_0 \in H^1(\Omega)$.*

3. The discrete optimal control problem.

3.1. Preliminaries. We consider a family of triangulations (say $\{\mathcal{T}_h\}_{h>0}$) of Ω, defined in the standard way, (see e.g. [17]). To every element $T \in \mathcal{T}_h$, we associate two parameters h_T and ρ_T, denoting the diameter of the set T, and the diameter of the largest ball contained in T respectively. The size of the mesh is denoted by $h = \max_{T \in \mathcal{T}_h} h_T$. The following standard properties of the mesh will be assumed:
(i) – There exist two positive constants $\rho_\mathcal{T}$ and $\delta_\mathcal{T}$ such that $\frac{h_T}{\rho_T} \leq \rho_\mathcal{T}$ and $\frac{h}{h_T} \leq \delta_\mathcal{T}$ $\forall T \in \mathcal{T}_h$ and $\forall h > 0$.
(ii) – Given h, let $\{T_j\}_{j=1}^{N_h}$ denote the family of triangles belonging to $\mathcal{T}_h$ and having one side included on the boundary Γ. Thus, if the vertices of $T_j \cap \Gamma$ are denoted by

$x_{j,\Gamma}$, $x_{j+1,\Gamma}$ then the straight line $[x_{j,\Gamma}, x_{j+1,\Gamma}] \equiv T_j \cap \Gamma$. Here, we also assume that $x_{1,\Gamma} = x_{N_h+1,\Gamma}$.

On the mesh $\mathcal{T}_h$ we consider finite dimensional spaces $U_h \subset H^1(\Omega)$ constructed by piecewise polynomials in Ω. Standard approximation theory assumptions are assumed on these spaces. In particular, for any $v \in H^{l+1}(\Omega)$, there exists an integer $\ell \geq 1$, and a constant $C > 0$ (independent of h) such that:

$$\inf_{v_h \in U_h} \|v - v_h\|_{H^s(\Omega)} \leq C h^{l+1-s} \|v\|_{H^{l+1}(\Omega)}, \quad \text{for } 0 \leq l \leq \ell \text{ and } s = -1, 0, 1.$$

We also use inverse inequalities on quasi-uniform triangulations, i.e., there exist constants $C \geq 0$, such that $\|v_h\|_{H^1(\Omega)} \leq C/h \|v_h\|_{L^2(\Omega)}$, and $\|v_h\|_{L^2(\Omega)} \leq C/h \|v_h\|_{H^1(\Omega)^*}$, etc. Approximations will be constructed on a (quasi-uniform) partition $0 = t^0 < t^1 < \ldots < t^N = T$ of $[0, T]$, i.e., there exists a constant $0 < \theta < 1$ such that $\min_{n=1,..,N}(t^n - t^{n-1}) \geq \theta \max_{n=1,...,N}(t^n - t^{n-1})$. We also use the notation $\tau^n = t^n - t^{n-1}, \tau = \max_{n=1,...,N} \tau^n$ and we denote by $\mathcal{P}_k[t^{n-1}, t^n; U_h]$ the space of polynomials of degree k or less having values in U_h. We seek approximate solutions who belong to the space

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

By convention, the functions of $\mathcal{U}_h$ are left continuous with right limits and hence will write $y_h^n \equiv y_{h-}^n$ for $y_h(t^n) = y_h(t_-^n)$, and y_{h+}^n for $y_h(t_+^n)$, while the jump at t^n, is denoted by $[y_h^n] = y_{h+}^n - y_h^n$. In the above definitions, we have used the following notational abbreviation, $y_{h,\tau} \equiv y_h$, $\mathcal{U}_{h,\tau} \equiv \mathcal{U}_h$ etc. For the time-discretization, our main focus will be the lowest order scheme ($k = 0$) which corresponds to the discontinuous Galerkin variant of the implicit Euler. We emphasize that other schemes (including schemes of arbitrary order in time and space) can be included in our proofs. However, the limited regularity will be acting as a barrier in terms of developing estimates of higher order.

For the control variable, we have two separate choices for the constrained and the unconstrained case respectively. In both cases our discretization is motivated by the optimality condition (see also [14]).

Case 1: Unconstrained Controls: We employ a discretization which allows the presence of discontinuities (in time), i.e., we define,

$$\mathcal{G}_h = \{g_h \in L^2[0, T; L^2(\Gamma)] : g_h|_{(t^{n-1}, t^n]} \in \mathcal{P}_k[t^{n-1}, t^n; G_h]\}.$$

Here, a conforming subspace $G_h \subset L^2(\Gamma)$ is specified at each time interval $(t^{n-1}, t^n]$, which satisfy standard approximation properties. Even though various choices of G_h are possible, here we focus our attention to the natural choice, $G_h = U_h|_\Gamma$ and we refer the reader to [19, 21] (see also references within) for a detailed analysis. Only $L^2[0, T; L^2(\Gamma)]$ regularity will be needed in the error estimates. To summarize, for the choice of piecewise linears (in space), we choose:

$$U_h = \{v_h \in C(\bar{\Omega}) : v_h|_T \in \mathcal{P}_1, \text{ for all } T \in \mathcal{T}_h\},$$
$$G_h = \{u_h \in C(\Gamma) : u_h|_{[x_{i,\Gamma}, x_{i+1,\Gamma}]} \in \mathcal{P}_1, \text{ for } i = 1, ..., N_h\}.$$

Case 2: Constrained Controls: Analogously to the previous case, we employ the variational discretization concept (see e.g. [25]) which allows the natural discretization of the controls via the adjoint variable. In this case, we do not explicitly discretize the control variable, i.e., $\mathcal{G}_h \equiv L^2[0, T; L^2(\Gamma)]$.

3.2. The fully-discrete optimal control problem. The discontinuous time-stepping fully-discrete scheme for the control to state mapping $G_h : L^2[0,T;L^2(\Gamma)] \to \mathcal{U}_h$, which associates to each control g its state $G_h(g) = y_{g,h} \equiv y_h(g)$ is defined as follows: For any boundary data $g \in L^2[0,T;L^2(\Gamma)]$, for given data $y_0 \in L^2(\Omega)$, $f \in L^2[0,T;H^1(\Omega)^*]$, and target $y_d \in L^2[0,T;L^2(\Omega)]$ we seek $y_h \in \mathcal{U}_h$ such that for $n = 1, ..., N$, and for all $v_h \in \mathcal{P}_k[t^{n-1}, t^n; U_h]$,

$$(y_h^n, v_h^n) + \int_{t^{n-1}}^{t^n} \Big(- \langle y_h, v_{ht} \rangle + a(y_h, v_h) + \lambda \langle y_h, v_h \rangle_\Gamma \Big) dt \tag{3.1}$$
$$= (y_h^{n-1}, v_{h+}^{n-1}) + \int_{t^{n-1}}^{t^n} \Big(\langle f, v_h \rangle + \lambda \langle g, v_h \rangle_\Gamma \Big) dt.$$

Here, $y_h^0 = P_h y_0$, where P_h denotes the standard $L^2(\Omega)$ projection, i.e., $(P_h y_0 - y_0, v_h) = 0, \quad \forall v_h \in U_h$. We note that only $g \in L^2[0,T;L^2(\Gamma)]$ regularity is needed to validate the fully-discrete formulation. Stability estimates at partition time-points as well as in $L^2[0,T;H^1(\Omega)]$ and $L^2[0,T;L^2(\Gamma)]$ norms easily follow by setting $v_h = y_h$ into (3.1). For the estimate at arbitrary time-points, we may apply the techniques which were developed in [15, Section 2] for general linear parabolic PDEs, (see also [10, Section 3] for stability estimate for semilinear parabolic PDEs with Robin data). Similar to the continuous case, the control to fully-discrete state mapping $G_h : L^2[0,T;L^2(\Gamma)] \to \mathcal{U}_h$, is well defined, and continuous. The definition of the discrete Robin boundary control problem, now follows:

Definition 3.1. *Let $f \in L^2[0,T;H^1(\Omega)^*]$, $y_0 \in L^2(\Omega)$, $y_d \in L^2[0,T;L^2(\Omega)]$ be given data. Suppose that the set of discrete admissible controls is denoted by $\mathcal{A}_{ad}^d \equiv \mathcal{G}_h \cap \mathcal{A}_{ad}$ (see Section 3.1), and let $J_h(y_h, g_h) \equiv \frac{1}{2}\int_0^T \int_\Omega |y_h - y_d|^2 dxdt + \frac{\alpha}{2}\int_0^T \int_\Gamma |g_h|^2 dxdt$. Here the pair $(y_h, g_h) \in \mathcal{U}_h \times \mathcal{A}_{ad}^d$ satisfy (3.1). Then, the pair $(\bar{y}_h, \bar{g}_h) \in \mathcal{U}_h \times \mathcal{A}_{ad}^d$, is said to be an optimal solution if $J_h(\bar{y}_h, \bar{g}_h) \le J_h(w_h, u_h)$, $\forall (w_h, u_h) \in \mathcal{U}_h \times \mathcal{A}_{ad}^d$.*

The existence of the discrete optimal control problem can be proved by standard techniques while uniqueness follows from the structure of the functional, and the linearity of the equation. The basic stability estimates in terms of the optimal pair $(\bar{y}_h, \bar{g}_h) \in W(0,T) \times L^2[0,T;L^2(\Gamma)]$ can be easily obtained. We close this subsection by quoting the estimate at arbitrary time-points, for schemes of arbitrary order under minimal regularity assumptions, adapted to our case from [10, Section 3]. The estimate highlights the fact that the natural choice of the discrete energy norm for the state variable associated to discontinuous time-stepping schemes is $\|\cdot\|_{W(0,T)} = \|\cdot\|_{L^\infty[0,T;L^2(\Omega)]} + \|\cdot\|_{L^2[0,T;H^1(\Omega)]} + \|\cdot\|_{L^2[0,T;L^2(\Gamma)]}$.

Lemma 3.2. *Let $y_0 \in L^2(\Omega)$, $f \in L^2[0,T;H^1(\Omega)^*]$. If $(\bar{y}_h, \bar{g}_h) \in \mathcal{U}_h \times \mathcal{A}_{ad}^d$ denotes the solution pair of the discrete optimal control problem, then,*

$$\|\bar{y}_h\|_{L^\infty[0,T;L^2(\Omega)]} \le C \max\Big\{1, \Big(\frac{\lambda}{\alpha}\Big)^{1/2}\Big\}\big(\|y_0\|_{L^2(\Omega)} + \|f\|_{L^2[0,T;H^1(\Omega)^*]} \big).$$

Here, $C \ge 0$ depends on $1/C_F \min\{\eta, \lambda\}$, C_k and Ω but not on α, τ, h. We note that the above estimate remains valid for the control constrained case assuming that $0 \in \mathcal{A}_{ad}^d$. Otherwise, the constant C of Lemma 3.2 also depends upon $\max\{|g_a|, |g_b|\}$.

3.3. The discrete optimality system. Using well known techniques and the stability estimates in $W(0,T)$, it is easy to show the differentiability of the relation

$g \to y_h(g)$, for any $g \in L^2[0,T;L^2(\Gamma)]$. Hence, the discrete analogue of Lemma 2.4, takes the following form:

LEMMA 3.3. *The cost functional $J_h : L^2[0,T;L^2(\Gamma)] \to \mathbb{R}$ is well defined, differentiable, and for every $g, u \in L^2[0,T;L^2(\Gamma)]$,*

$$J_h^{'}(g)u = \int_0^T \int_\Gamma (\mu_h(g) + \alpha g) u dx dt,$$

where $\mu_h(g) \equiv \mu_{g,h} \in W(0,T)$ is the unique solution of following problem: For all $n = 1, ..., N$, and for all $v_h \in P_k[t^{n-1}, t^n; U_h]$,

$$\begin{aligned} -(\mu^n_{g,h+}, v^n_h) + \int_{t^{n-1}}^{t^n} \Big(\langle \mu_{g,h}, v_{ht} \rangle + a(v_h, \mu_{g,h}) + \lambda \langle \mu_{g,h}, v_h \rangle_\Gamma \Big) dt \\ = -(\mu^{n-1}_{g,h+}, v^{n-1}_{h+}) + \int_{t^{n-1}}^{t^n} \langle y_{g,h} - y_d, v_h \rangle dt, \end{aligned} \tag{3.2}$$

where $\mu^N_{g,h+} = 0$. Here, $y_{h,g} \equiv y_h(g)$ is the solution of (3.1).

Thus, the fully-discrete optimality system takes the following form.

LEMMA 3.4. *Let $(\bar{y}_h(\bar{g}_h), \bar{g}_h) \equiv (\bar{y}_h, \bar{g}_h) \in \mathcal{U}_h \times \mathcal{A}^d_{ad}$ denote the unique optimal pair of Definition 3.1. Then, there exists an adjoint $\bar{\mu}_h \in \mathcal{U}_h$ satisfying $\bar{\mu}^N_+ = 0$, such that for all $v_h \in \mathcal{P}_k[t^{n-1}, t^n; U_h]\}$, and for all $n = 1, ..., N$*

$$\begin{aligned} (\bar{y}^n_h, v^n_h) + \int_{t^{n-1}}^{t^n} \left(-\langle \bar{y}_h, v_{ht} \rangle + a(\bar{y}_h, v_h) + \lambda \langle \bar{y}_h, v_h \rangle_\Gamma \right) dt \\ = (\bar{y}^{n-1}_h, v^{n-1}_{h+}) + \int_{t^{n-1}}^{t^n} \left(\langle f, v_h \rangle + \lambda \langle \bar{g}_h, v_h \rangle_\Gamma \right) dt, \end{aligned} \tag{3.3}$$

$$\begin{aligned} -(\bar{\mu}^n_{h+}, v^n_h) + \int_{t^{n-1}}^{t^n} \left(\langle \bar{\mu}_h, v_{ht} \rangle + a(\bar{\mu}_h, v_h) + \lambda \langle \bar{\mu}_h, v_h \rangle_\Gamma \right) dt \\ = -(\bar{\mu}^{n-1}_{h+}, v^{n-1}_{h+}) + \int_{t^{n-1}}^{t^n} \left(\bar{y}_h - y_d, v_h \right) dt, \end{aligned} \tag{3.4}$$

and the following optimality condition holds: For all $u_h \in \mathcal{A}^d_{ad}$,

$$\text{1) Unconstrained Controls: } \int_0^T (\alpha \bar{g}_h + \lambda \bar{\mu}_h, u_h)_\Gamma dt = 0, \tag{3.5}$$

$$\text{2) Constrained Controls: } \int_0^T \int_\Gamma \left(\alpha \bar{g}_h + \lambda \bar{\mu}_h \right) \left(u_h - \bar{g}_h \right) dx dt \geq 0. \tag{3.6}$$

Estimates for the adjoint variable at partition points and in $L^2[0,T;H^1(\Omega)]$ can be derived easily, while for an estimate in $L^\infty[0,T;L^2(\Omega)]$ we refer the reader to [10]. The following estimate highlights the fact that the discrete solutions produced by discontinuous time-stepping schemes possess the same regularity properties of the continuous problem.

LEMMA 3.5. *Let $(\bar{y}_h, \bar{g}_h)$ denote the discrete optimal solution and $(\bar{y}_h, \bar{\mu}_h, \bar{g}_h)$ satisfy the system (3.3)-(3.4)-(3.5) or (3.6). Then,*

$$\|\bar{\mu_h}\|_{L^\infty[0,T;H^1(\Omega)]} + \lambda^{1/2} \|\bar{\mu}_h\|_{L^\infty[0,T;L^2(\Gamma)]} \leq C \|\bar{y}_h - y_d\|_{L^2[0,T;L^2(\Omega)]},$$

where C does not depend on α, τ, h but only on $1/\eta$, C_k, Ω.

Proof. The proof follows based on the techniques of [16, Theorem 4.10], modified in order to handle the Robin boundary data, and the backward in time nature of our pde. First, we note that $\mu(T) = 0$, and $y_h - y_d \in L^\infty[0,T;L^2(\Omega)]$. Hence, at each time $t \in (t^{n-1}, t^n]$ let $a_p(t) \in U_h$ denote the following discrete approximation of the Laplacian (with Robin boundary data),

$$(a_p(t), v_h) = (1/\eta)a(\bar{\mu}_h(t), v_h) + (\lambda/\eta)(\bar{\mu}_h(t), v_h)_\Gamma, \quad \forall v_h \in U_h.$$

Thus, $a_p \in \mathcal{P}_k[t^{n-1}, t^n; U_h]$, and hence setting $v_h(t) = \bar{\mu}_{ht}(t) \in U_h$, and $v_h(t) = a_p(t) \in U_h$, we obtain

$$(1/2)\frac{d}{dt}(\|\nabla\bar{\mu}_h(t)\|^2_{L^2(\Omega)} + (\lambda/\eta)\|\bar{\mu}_h(t)\|_{L^2(\Gamma)}) = (a_p(t), \bar{\mu}_{ht}(t)),$$

and

$$a(\bar{\mu}_h(t), a_p(t)) + \lambda\langle\bar{\mu}_h(t), a_p(t)\rangle_\Gamma = \eta(a_p(t), a_p(t)).$$

Integrating by parts in time (3.4), setting $v_h = a_p$ into the resulting equality, using the last two equalities, the definition of $a_p(t^n)$, i.e., $(a_p(t^n), \bar{\mu}^n_{h+} - \bar{\mu}^n_h) = (\nabla\bar{\mu}^n_h, \nabla(\bar{\mu}^n_{h+} - \bar{\mu}^n_h)) + (\lambda/\eta)(\bar{\mu}^n_h, \bar{\mu}^n_{h+} - \bar{\mu}^n_h)_\Gamma$, and standard algebra, we obtain,

$$\begin{aligned}
&(1/2)\|\nabla\bar{\mu}^{n-1}_{h+}\|_{L^2(\Omega)} + (\lambda/2\eta)\|\bar{\mu}^{n-1}_{h+}\|_{L^2(\Gamma)} + \eta\int_{t^{n-1}}^{t^n}\|a_p\|^2_{L^2(\Omega)}dt \\
&\le (1/2)\|\nabla\bar{\mu}^n_{h+}\|_{L^2(\Omega)} + (\lambda/2\eta)\|\bar{\mu}^{n-1}_{h+}\|_{L^2(\Gamma)} + \int_{t^{n-1}}^{t^n}(\bar{y}_h - y_d, a_p)dt.
\end{aligned}$$

The above inequality implies bounds at the partition points, and hence bounds in $L^\infty[0,T;H^1(\Omega)]$, when $k = 0,1$ after inserting the stability bound on $\bar{y}_h$. For high-order (in time) schemes, we directly follow the approach of [16, Theorem 4.10]. □

4. Error estimates. The key ingredient of the proof is the definition of a suitable generalized space-time dG projection capable of handling the low regularity of $y_t \in L^2[0,T;H^1(\Omega)^*]$, and an auxiliary optimality system which plays the role of a global space-time projection and exhibits best approximation properties.

4.1. The fully-discrete projection. Let $w_h, z_h \in \mathcal{U}_h$ be defined as the solutions of the following system. Given data f, y_0, initial conditions $w^0_h = y^0_h$, where $y^0_h \equiv P_h y_0$, (recall that P_h denotes the standard L^2 projection, i.e., $(P_h y_0 - y_0, v_h) = 0, \quad \forall\, v_h \in U_h$), and terminal condition $z^N_+ = 0$, we seek $w_h, z_h \in \mathcal{U}_h$ such that for $n = 1, ..., N$ and for all $v_h \in \mathcal{P}_k[t^{n-1}, t^n; U_h]$,

$$\begin{aligned}
&(w^n_h, v^n_h) + \int_{t^{n-1}}^{t^n}\Big(-\langle w_h, v_{ht}\rangle + a(w_h, v_h) + \lambda\langle w_h, v_h\rangle_\Gamma\Big)dt \qquad (4.1)\\
&\qquad = (w^{n-1}_h, v^{n-1}_{h+}) + \int_{t^{n-1}}^{t^n}\Big(\langle f, v_h\rangle + \lambda\langle\bar{g}, v_h\rangle_\Gamma\Big)dt,
\end{aligned}$$

$$\begin{aligned}
&-(z^n_{h+}, v^n_h) + \int_{t^{n-1}}^{t^n}\Big(\langle z_h, v_{ht}\rangle + a(z_h, v_h) + \lambda\langle z_h, v_h\rangle_\Gamma\Big)dt \qquad (4.2)\\
&\qquad = -(z^{n-1}_{h+}, v^{n-1}_{h+}) + \int_{t^{n-1}}^{t^n}(w_h - y_d, v_h)dt.
\end{aligned}$$

The solutions $w_h, z_h \in \mathcal{U}_h$ exist due to the regularity of $\bar{y}, \bar{\mu} \in W(0,T)$. 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 $\bar{y} - w_h, \bar{\mu} - z_h$ will be derived in terms of local L^2 projections associated to discontinuous time-stepping methods (see e.g. [38]).

DEFINITION 4.1. *(1) The projection* $P_n^{loc} : C[t^{n-1}, t^n; L^2(\Omega)] \to \mathcal{P}_k[t^{n-1}, t^n; U_h]$ *satisfies* $(P_n^{loc} v)^n = P_h v(t^n)$, *and*

$$\int_{t^{n-1}}^{t^n} (v - P_n^{loc} v, v_h) = 0, \qquad \forall\, v_h \in \mathcal{P}_{k-1}[t^{n-1}, t^n; U_h]. \tag{4.3}$$

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

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

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

Due to the lack of regularity, and in particular the fact that $\bar{y} \in L^2[0,T; H^1(\Omega)] \cap H^1[0,T; H^1(\Omega)^*]$, we construct a space-time generalized L^2 projection which combines the standard dG time stepping projection, and the spacial generalized L^2 projection $Q_h : H^1(\Omega)^* \to U_h$. Recall that the definition of Q_h states that $\langle v - Q_h v, v_h \rangle = 0$, for all $v \in H^1(\Omega)^*$ and $v_h \in U_h$ (see for instance [12, Section 2]).

DEFINITION 4.2. *(1) The projection* $Q_n^{loc} : C[t^{n-1}, t^n; H^1(\Omega)^*] \to \mathcal{P}_k[t^{n-1}, t^n; U_h]$ *satisfies* $(Q_n^{loc} v)^n = Q_h v(t^n)$, *and*

$$\int_{t^{n-1}}^{t^n} \langle v - Q_n^{loc} v, v_h \rangle = 0, \qquad \forall\, v_h \in \mathcal{P}_{k-1}[t^{n-1}, t^n; U_h].$$

Here we also use the convention $(Q_n^{loc} v)^n \equiv (Q_n^{loc} v)(t^n)$ *and* $Q_h : H^1(\Omega)^* \to U_h$ *is the generalized orthogonal projection operator onto* $U_h \subset H^1(\Omega)$.

(2) The projection $Q_h^{loc} : C[0,T; H^1(\Omega)^*] \to \mathcal{U}_h$ *satisfies*

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

For $k = 0$, *the projection* $Q_h^{loc} : C[0,T; H^1(\Omega)^*] \to \mathcal{U}_h$ *reduces to* $Q_h^{loc} v(t) = Q_h v(t^n)$ *for all* $t \in (t^{n-1}, t^n]$, $n = 1, ..., N$.

The key feature of Q_h^{loc} is that it coincides to P_h^{loc}, when $v \in L^2[0,T; L^2(\Omega)]$ i.e., $P_h^{loc} v = Q_h^{loc} v$ when $v \in L^2[0,T; L^2(\Omega)]$, and hence exhibits best approximation properties, but is also applicable for $v \equiv \bar{y}_t \in L^2[0,T; H^1(\Omega)^*]$. For the backwards in time problem a modification of the above projections (still denoted by P_n^{loc}, Q_n^{loc} respectively) will be needed. For example, in addition to relation (4.3), we need to impose the "matching condition" on the left, i.e., $(P_n^{loc} v)_+^{n-1} = P_h v(t_+^{n-1})$ instead of imposing the condition on the right. In the following Lemma, we collect several results regarding (optimal) rates of convergence for the above projection. Here, the emphasis is placed on the approximation properties of the generalized projection Q_h^{loc}, under minimal regularity assumptions, i.e., for $v \in L^2[0,T; H^1(\Omega)] \cap H^1[0,T; H^1(\Omega)^*]$ for the lowest order scheme.

LEMMA 4.3. *Let* $U_h \subset H^1(\Omega)$, *and* P_h^{loc}, Q_h^{loc} *defined in Definitions 4.1 and 4.2 respectively. Then, for all* $v \in L^2[0,T; H^{l+1}(\Omega)] \cap H^{k+1}[0,T; L^2(\Omega)]$ *there exists constant* $C \geq 0$ *independent of* h, τ *such that*

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

If in addition, $k = 0, l = 1$, and $v \in L^2[0,T;H^1(\Omega)] \cap H^1[0,T;H^1(\Omega)^*]$ then there exists a constant $C \geq 0$ independent of h, τ such that

$$\begin{aligned}
&\|v - Q_h^{loc} v\|_{L^2[0,T;L^2(\Omega)]} \leq C\big(h\|v\|_{L^2[0,T;H^1(\Omega)]} \\
&\quad +\tau^{1/2}(\|v\|_{L^2[0,T;H^1(\Omega)]} + \|v_t\|_{L^2[0,T;H^1(\Omega)^*]})\big), \\
&\|v - Q_h^{loc} v\|_{L^2[0,T;H^1(\Omega)]} \leq C\big(\|v\|_{L^2[0,T;H^1(\Omega)]} + (\tau/h^2)\|v_t\|_{L^2[0,T;H^1(\Omega)^*]}\big).
\end{aligned}$$

Let $k = 0, l = 1$, and $v \in L^2[0,T;H^2(\Omega)] \cap H^1[0,T;L^2(\Omega)]$. Then there exists constant $C \geq 0$ independent of h, τ such that,

$$\begin{aligned}
&\|v - Q_h^{loc} v\|_{L^2[0,T;H^1(\Omega)]} \leq C\big(h\|v\|_{L^2[0,T;H^2(\Omega)]} \\
&\quad +\tau^{1/2}(\|v_t\|_{L^2[0,T;L^2(\Omega)]} + \|v\|_{L^2[0,T;H^2(\Omega)]})\big).
\end{aligned}$$

Proof. The first estimate is given in [16, Theorem 4.3, and Corollary 4.8]. For the second estimate, for any $t \in (t^{n-1}, t^n]$, adding and subtracting appropriate terms, and using the definition of Q_h^{loc}, we obtain,

$$\|v - Q_h^{loc} v\|^2_{L^2[0,T;L^2(\Omega)]} \leq \sum_{n=1}^{N} \int_{t^{n-1}}^{t^n} (\|v(t) - v(t^n)\|^2_{L^2(\Omega)} + \|v(t^n) - Q_h v(t^n)\|^2_{L^2(\Omega)}) dt.$$

For the first term,

$$\int_{t^{n-1}}^{t^n} \|v(t) - v(t^n)\|^2_{L^2(\Omega)} dt \leq C\tau \int_{t^{n-1}}^{t^n} \big(\|v_t\|^2_{H^1(\Omega)^*} + \|v\|^2_{H^1(\Omega)}\big) dt.$$

The second term can be approximated by triangle inequality, the approximation property $\|v(t) - Q_h v(t)\|_{L^2(\Omega)} \leq Ch\|v(t)\|_{H^1(\Omega)}$, and the bound on $\|v(t) - v(t^n)\|_{L^2(\Omega)}$. For the third estimate, we first note that the generalized orthogonal projection $Q_h : H^1(\Omega)^* \to U_h$ is stable in $\|.\|_{H^1(\Omega)^*}$ norm. Indeed, for all $v \in H^1(\Omega)^*$, $w \in H^1(\Omega)$, by the definition of projections Q_h and P_h,

$$\begin{aligned}
\|Q_h v\|_{H^1(\Omega)^*} &= \sup_{0 \neq w \in H^1(\Omega)} \frac{|\langle Q_h v, w\rangle|}{\|w\|_{H^1(\Omega)}} \leq \sup_{0 \neq w \in H^1(\Omega)} \left(\frac{|\langle Q_h v - v, w\rangle|}{\|w\|_{H^1(\Omega)}} + \frac{|\langle v, w\rangle|}{\|w\|_{H^1(\Omega)}} \right) \\
&\leq \sup_{0 \neq w \in H^1(\Omega)} \frac{|\langle Q_h v - v, w - P_h w\rangle|}{\|w\|_{H^1(\Omega)}} + \|v\|_{H^1(\Omega)^*}
\end{aligned}$$

where at the last inequality we have used the fact that $\langle Q_h v - v, P_h w\rangle = 0$. Note also that by the definition of projection P_h, we deduce that $\langle Q_h v - v, w - P_h w\rangle = \langle -v, w - P_h w\rangle$. Hence, the $H^1(\Omega)$ stability of the P_h projection implies,

$$\begin{aligned}
\|Q_h v\|_{H^1(\Omega)^*} &\leq \sup_{0 \neq w \in H^1(\Omega)} \frac{|\langle v, w - P_h w\rangle|}{\|w\|_{H^1(\Omega)}} + \|v\|_{H^1(\Omega)^*} \\
&\leq C \frac{\|v\|_{H^1(\Omega)^*} \|w - P_h w\|_{H^1(\Omega)}}{\|w\|_{H^1(\Omega)}} + \|v\|_{H^1(\Omega)^*} \leq C\|v\|_{H^1(\Omega)^*}.
\end{aligned}$$

Thus, the definition of Q_h^{loc} for $k = 0$, $l = 1$, the inverse estimate $\|Q_h v\|_{L^2(\Omega)} \leq$

$C/h\|Q_hv\|_{H^1(\Omega)^*}$, and the stability of Q_h in $H^1(\Omega)^*$ norm, imply

$$
\begin{aligned}
&\|v - Q_h^{loc}v\|_{L^2[0,T;H^1(\Omega)]} = \left(\sum_{n=1}^{N}\int_{t^{n-1}}^{t^n}\|v(t) - Q_hv(t^n)\|^2_{H^1(\Omega)}dt\right)^{1/2}\\
&= \left(\sum_{n=1}^{N}\int_{t^{n-1}}^{t^n}\|v(t) - Q_hv(t)\|^2_{H^1(\Omega)}dt\right)^{1/2} + \left(\sum_{n=1}^{N}\int_{t^{n-1}}^{t^n}\|Q_hv(t) - Q_hv(t^n)\|^2_{H^1(\Omega)}dt\right)^{1/2}\\
&\leq C\|v\|_{L^2[0,T;H^1(\Omega)]} + \frac{C}{h^2}\left(\sum_{n=1}^{N}\int_{t^{n-1}}^{t^n}\|Q_hv(t) - Q_hv(t^n)\|^2_{H^1(\Omega)^*}dt\right)^{1/2}\\
&\leq C\|v\|_{L^2[0,T;H^1(\Omega)]} + \frac{C}{h^2}\left(\sum_{n=1}^{N}\int_{t^{n-1}}^{t^n}\|v(t) - v(t^n)\|^2_{H^1(\Omega)^*}dt\right)^{1/2}\\
&\leq C\|v\|_{L^2[0,T;H^1(\Omega)]} + \frac{C}{h^2}\left(\sum_{n=1}^{N}\int_{t^{n-1}}^{t^n}(t^n - t)\int_{t^{n-1}}^{t^n}\|v_t\|^2_{H^1(\Omega)^*}dsdt\right)^{1/2}\\
&\leq C\|v\|_{L^2[0,T;H^1(\Omega)]} + C\frac{\tau}{h^2}\|v_t\|_{L^2[0,T;H^1(\Omega)^*]},
\end{aligned}
$$

for all $v \in L^2[0,T;H^1(\Omega)] \cap H^1[0,T;H^1(\Omega)^*]$, which completes the proof of the third estimate. The last estimate is standard, and can be derived similar to the second one, after noting that $Q_h^{loc} \equiv P_h^{loc}$, and $\frac{d}{dt}\|\nabla v(t)\|^2_{L^2(\Omega)} = 2\langle \nabla v_t, \nabla v\rangle$. ☐

REMARK 4.4. *The stability estimate in $L^2[0,T;H^1(\Omega)]$ requires the time-step restriction of $\tau \leq Ch^2$ due to the lack of regularity with respect to time. If $v \in L^2[0,T;H^{l+1}(\Omega)] \cap H^{k+1}[0,T;L^2(\Omega)]$ then the first estimate of Lemma 4.3 implies that,*

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

Indeed, using [16, Theorem 4.3, Corollary 4.8], we obtain the following (local in time) estimates:

$$
\begin{aligned}
&\|v - P_n^{loc}v\|_{L^2[t^{n-1},t^n;H^1(\Omega)]}\\
&\leq C\big(\|v - P_hv\|_{L^2[t^{n-1},t^n;H^1(\Omega)]} + \tau^{k+1}\|P_hv^{(k+1)}\|_{L^2[t^{n-1},t^n;H^1(\Omega)]}\big)\\
&\leq C\big(h^l\|v\|_{L^2[t^{n-1},t^n;H^{l+1}(\Omega)]} + (\tau^{k+1}/h)\|v^{(k+1)}\|_{L^2[t^{n-1},t^n;L^2(\Omega)]}\big),
\end{aligned}
$$

where at the last estimate we have used an inverse estimate. We note that if more regularity is available, the inverse estimate is not necessary. In particular if $v^{(k+1)} \in L^2[0,T;H^1(\Omega)]$, then the improved rate of $\mathcal{O}(h^l + \tau^{k+1})$ holds in $\|.\|_{L^2[0,T;H^1(\Omega)]}$ norm. However, we note that for our boundary optimal control problem the increased regularity $v_t \in L^2[0,T;H^1(\Omega)]$ is not available. Hence, we emphasize that the lack of regularity acts as a barrier for developing a truly higher order scheme. Working similarly we also obtain an estimate at arbitrary time-points, i.e.,

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

Below, we state the main result for related to the auxiliary problem, which acts as the global space-time dG projection. Our goal is to state that the projection error is as good as the local dG projection error allows it to be, and hence it is optimal in the sense of the available regularity.

THEOREM 4.5. *Let $f \in L^2[0,T;H^1(\Omega)^*]$ and $y_0 \in L^2(\Omega)$ be given, and $\bar{y},\bar{\mu} \in W(0,T)$ be the solutions of (2.4)-(2.5)-(2.6) or (2.7), and $w_h, z_h \in \mathcal{U}_h$ be the solutions of (4.1)-(4.2). Denote by $e_1 = \bar{y} - w_h$, $r_1 = \bar{\mu} - z_h$ and let $e_p \equiv \bar{y} - Q_h^{loc}\bar{y}$, $r_p = \bar{\mu} - P_h^{loc}\bar{\mu}$, where P_h^{loc}, Q_h^{loc} are defined in Definitions 4.1 and 4.2. Then, there exists an algebraic constant $C > 0$ depending only on Ω such that,*

$$1) C_F \min\{\eta,\lambda\}\|e_1\|^2_{L^2[0,T;H^1(\Omega)]} + \sum_{i=0}^{N-1}\|[e_1^i]\|^2_{L^2(\Omega)} + \lambda\|e_1\|^2_{L^2[0,T;L^2(\Gamma)]}$$
$$\leq C\big(\|e_1^0\|^2_{L^2(\Omega)} + (1/C_F\min\{\eta,\lambda\})\big(\|e_p\|^2_{L^2[0,T;H^1(\Omega)]} + \lambda\|e_p\|^2_{L^2[0,T;L^2(\Gamma)]}\big),$$

$$2) C_F \min\{\eta,\lambda\}\|r_1\|^2_{L^2[0,T;H^1(\Omega)]} + \sum_{i=1}^{N}\|[r_1^i]\|^2_{L^2(\Omega)} + \lambda\|r_1\|^2_{L^2[0,T;L^2(\Gamma)]}$$
$$\leq C\Big((1/C_F\min\{\eta,\lambda\})\big(\|e_1\|^2_{L^2[0,T;L^2(\Omega)]} + \|r_p\|^2_{L^2[0,T;H^1(\Omega)]}\big) + \lambda\|r_p\|^2_{L^2[0,T;L^2(\Gamma)]}\Big),$$

$$3) \|e_1\|_{L^2[0,T;L^2(\Omega)]} \leq C\big(\eta\|e_p\|_{L^2[0,T;L^2(\Omega)]} + \tau^{1/2}(\|e_p\|_{L^2[0,T;H^1(\Omega)]} + \|e_p\|_{L^2[0,T;L^2(\Gamma)]})\big),$$

$$4) \|r_1\|_{L^2[0,T;L^2(\Omega)]} \leq C\big(\eta\|e_1\|_{L^2[0,T;L^2(\Omega)]} + \|r_p\|_{L^2[0,T;L^2(\Omega)]}$$
$$+\tau^{1/2}(\|r_p\|_{L^2[0,T;H^1(\Omega)]} + \|r_p\|_{L^2[0,T;L^2(\Gamma)]})\big).$$

Here, $w_h^0 = y_h^0$, where $y_h^0 = P_h y_0$, and C a constant depending upon on the domain Ω.

Proof. Step 1: Preliminary estimates. Throughout this proof, we denote by $e_1 = \bar{y} - w_h, r_1 = \bar{\mu} - z_h$ and we split e_1, r_1 to $e_1 \equiv e_{1h} + e_p \equiv (Q_h^{loc}\bar{y} - w_h) + (\bar{y} - Q_h^{loc}\bar{y})$, $r_1 \equiv r_{1h} + r_p \equiv (P_h^{loc}\bar{\mu} - z_h) + (\bar{\mu} - P_h^{loc}\bar{\mu})$, where P_h^{loc}, Q_h^{loc} are defined in Definitions 4.1 and 4.2. Subtracting (4.1) from (2.4), and (4.2) from (2.5) we obtain the orthogonality condition: For $n = 1,...,N$, and for all $v_h \in \mathcal{P}_k[t^{n-1},t^n;U_h]$,

$$(e_1^n, v_h^n) + \int_{t^{n-1}}^{t^n}\Big(-\langle e_1, v_{ht}\rangle + a(e_1,v_h) + \lambda\langle e_1, v_h\rangle_\Gamma\Big)dt = (e_1^{n-1}, v_{h+}^{n-1}), \quad (4.4)$$

$$-(r_{1+}^n, v_h^n) + \int_{t^{n-1}}^{t^n}\Big(\langle r_1, v_{ht}\rangle + a(r_1,v_h) + \lambda\langle r_1, v_h\rangle_\Gamma\Big)dt$$
$$= -(r_{1+}^{n-1}, v_{h+}^{n-1}) + \int_{t^{n-1}}^{t^n}(e_1, v_h)dt. \quad (4.5)$$

Note that the orthogonality condition (4.4) is essentially uncoupled and identical to the orthogonality condition of [15, Relation (2.6)]. Hence applying [15, Theorem 2.2], we derive the first estimate. In a similar way, the orthogonality condition (4.5) is equivalent to: For $n = 1,...,N$, and for all $v_h \in \mathcal{P}_k[t^{n-1},t^n;U_h]$,

$$-(r_{1h+}^n, v_h^n) + \int_{t^{n-1}}^{t^n}\Big(\langle r_{1h}, v_{ht}\rangle + a(r_{1h},v_h) + \lambda\langle r_{1h}, v_h\rangle_\Gamma\Big)dt \quad (4.6)$$
$$= -(r_{1h+}^{n-1}, v_{h+}^{n-1}) + \int_{t^{n-1}}^{t^n}\Big((e_1,v_h) - a(r_p,v_h) - \lambda(r_p,v_h)_\Gamma\Big)dt.$$

Here, we have used the Definition 4.1 of the projection P_h^{loc}, which implies that $\int_{t^{n-1}}^{t^n}\langle r_p, v_{ht}\rangle dt = 0$, $(r_{p+}^n, v_h^n) = 0$, and $(r_{p+}^{n-1}, v_{h+}^{n-1}) = 0$. Setting $v_h = r_{1h}$ into (4.6), using the Friedrichs' inequality to bound the second and the third term on the left, Young's inequality to bound the terms on the right, and standard algebra, we obtain

$$-(1/2)\|r^n_{1h+}\|^2_{L^2(\Omega)} + (1/2)\|[r^n_{1h}]\|^2_{L^2(\Omega)} + (1/2)\|r^{n-1}_{1h+}\|^2_{L^2(\Omega)} + (\lambda/4)\int_{t^{n-1}}^{t^n} \|r_{1h}\|^2_{L^2(\Gamma)} dt$$
$$+(C_F \min\{\lambda,\eta\}/4)\int_{t^{n-1}}^{t^n} \|r_{1h}\|^2_{H^1(\Omega)} dt + (\eta/2)\int_{t^{n-1}}^{t^n} \|\nabla r_{1h}\|^2_{L^2(\Omega)} dt$$
$$\leq C\int_{t^{n-1}}^{t^n} \Big((1/C_F \min\{\lambda,\eta\})\|e_1\|^2_{L^2(\Omega)} + (1/C_F \min\{\lambda,\eta\})\|r_p\|^2_{H^1(\Omega)} + \lambda\|r_p\|^2_{L^2(\Gamma)}\Big) dt.$$

The second estimate now follows upon summation.
Step 2: Duality arguments. We turn our attention to the last two estimates. In order to obtain the improved rate for the $L^2[0,T;L^2(\Omega)]$ norm we employ a duality argument to derive a better bound for the quantity $\|e_{1h}\|^2_{L^2[0,T;L^2(\Omega)]}$. For this purpose, we define a backwards in time parabolic problem with right hand side $e_{1h} \in L^2[0,T;L^2(\Omega)]$, and zero Robin and terminal data, i.e., $\lambda\phi + \eta\frac{\partial\phi}{\partial n}|_\Gamma = 0$, and $\phi(T) = 0$. For $n = 1, ..., N$ and for all $v \in L^2[0,T;H^1(\Omega)] \cap H^1[0,T;H^1(\Omega)^*]$, we seek $\phi \in W(0,T)$ such that

$$\int_{t^{n-1}}^{t^n} \big(\langle \phi, v_t\rangle + a(v,\phi) + \lambda\langle \phi, v\rangle_\Gamma\big) dt + (\phi(t^{n-1}, v(t^{n-1})) = \int_{t^{n-1}}^{t^n} (e_{1h}, v) dt. \tag{4.7}$$

Note that since $e_{1h} \in L^\infty[0,T;L^2(\Omega)]$, then $\phi \in L^2[0,T;H^2(\Omega)] \cap H^1[0,T;L^2(\Omega)]$ (see Theorem 2.1). In particular, the following estimate hold:

$$\|\phi\|_{L^2[0,T;H^2(\Omega)]} + \|\phi_t\|_{L^2[0,T;L^2(\Omega)]} + \lambda\|\phi\|_{L^2[0,T;L^2(\Gamma)]} \leq C\|e_{1h}\|_{L^2[0,T;L^2(\Omega)]}. \tag{4.8}$$

The lack of regularity of the right hand side of (4.7) due to the presence of discontinuities, implies that we can not improve regularity of ϕ in $[0,T]$. The associated discontinuous time-stepping scheme can be defined as follows: Given, terminal data $\phi^N_{h+} = 0$, we seek $\phi_h \in \mathcal{U}_h$ such that for all $v_h \in \mathcal{P}_k[t^{n-1}, t^n; U_h]$,

$$-(\phi^n_{h+}, v^n_{h-}) + \int_{t^{n-1}}^{t^n} \big((\phi_h, v_{ht}) + a(\phi_h, v_h) + \lambda\langle \phi_h, v_h\rangle_\Gamma\big) dt$$
$$+(\phi^{n-1}_{h+}, v^{n-1}_{h+}) = \int_{t^{n-1}}^{t^n} (e_{1h}, v_h) dt. \tag{4.9}$$

Hence using Lemma 3.5, the following stability estimate holds:

$$\|\phi_h\|_{L^\infty[0,T;H^1(\Omega)]} + \lambda\|\phi_h\|_{L^\infty[0,T;L^2(\Gamma)]} \leq C_k\|e_{1h}\|_{L^2[0,T;L^2(\Omega)]}. \tag{4.10}$$

It is now clear that we have the following estimate for $\phi - \phi_h$, which is a straightforward application of the previous estimates in $L^2[0,T;H^1(\Omega)]$, the approximation properties of Lemma 4.3, of projections P^{loc}_h, Q^{loc}_h, and the boundary Sobolev inequality,

$$\|\phi - \phi_h\|_{L^2[0,T;H^1(\Omega)]} + \lambda\|\phi - \phi_h\|_{L^2[0,T;L^2(\Gamma)]} \tag{4.11}$$
$$\leq C\big(h + \tau^{1/2}\big)\big(\|\phi\|_{L^2[0,T;H^2(\Omega)]} + \|\phi_t\|_{L^2[0,T;L^2(\Omega)]}\big) \leq C\big(h + \tau^{1/2}\big)\|e_{1h}\|_{L^2[0,T;L^2(\Omega)]}.$$

We note that the lack of regularity on the right hand side, restricts the rate of convergence to the rate given by the lowest order scheme $l = 1$, $k = 0$, even if high order schemes (in time) are chosen. Setting $v_h = e_{1h}$, into (4.9), we obtain,

$$-(\phi_{h+}^n, e_{1h-}^n) + \int_{t^{n-1}}^{t^n} \big((\phi_h, e_{1ht}) + a(e_{1h}, \phi_h) + \lambda\langle\phi_h, e_{1h}\rangle_\Gamma\big)dt + (\phi_{h+}^{n-1}, e_{1h+}^{n-1})$$
$$= \int_{t^{n-1}}^{t^n} \|e_{1h}\|_{L^2(\Omega)}^2 dt.$$

Integrating by parts in time, we deduce,

$$-(\phi_{h+}^n, e_{1h-}^n) + (\phi_{h-}^n, e_{1h-}^n) + \int_{t^{n-1}}^{t^n} \big(-(\phi_{ht}, e_{1h}) + a(\phi_h, e_{1h}) + \lambda\langle\phi_h, e_{1h}\rangle_\Gamma\big)dt$$
$$= \int_{t^{n-1}}^{t^n} \|e_{1h}\|_{L^2(\Omega)}^2 dt. \tag{4.12}$$

Setting $v_h = \phi_h$ into (4.4) and using $e_1 = e_p + e_{1h}$, and the definition of projection Q_h^{loc} of Definition 4.2 we obtain,

$$(e_{1h-}^n, \phi_{h-}^n) + \int_{t^{n-1}}^{t^n} \big(-(e_{1h}, \phi_{ht}) + a(e_{1h}, \phi_h) + \lambda\langle e_{1h}, \phi_h\rangle_\Gamma\big)dt - (e_{1h-}^{n-1}, \phi_{h+}^{n-1})$$
$$= -\int_{t^{n-1}}^{t^n} \big(a(e_p, \phi_h) + \lambda\langle e_p, \phi_h\rangle\big)dt. \tag{4.13}$$

Here, we have used the fact that the definition of projection Q_h^{loc} of Definition 4.2, implies that $(e_p^n, \phi_{h-}^n) = 0$, $\int_{t^{n-1}}^{t^n} (e_p, v_{ht})dt = 0$ and $(e_{p-}^{n-1}, \phi_{h+}^{n-1}) = 0$. Using (4.12) to replace the first four terms of (4.13) we arrive to

$$(\phi_{h+}^n, e_{1h-}^n) - (e_{1h-}^{n-1}, \phi_{h+}^{n-1}) + \int_{t^{n-1}}^{t^n} \|e_{1h}\|_{L^2(\Omega)}^2 dt$$
$$= -\int_{t^{n-1}}^{t^n} \big(a(e_p, \phi_h) + \lambda\langle e_p, \phi_h\rangle_\Gamma\big)dt$$
$$= -\int_{t^{n-1}}^{t^n} \big(a(e_p, \phi_h - \phi) + a(e_p, \phi) + \lambda\langle e_p, \phi_h - \phi\rangle_\Gamma + \lambda\langle e_p, \phi\rangle_\Gamma\big)dt$$
$$= -\int_{t^{n-1}}^{t^n} \Big(a(e_p, \phi_h - \phi) + \lambda\langle e_p, \phi_h - \phi\rangle_\Gamma - \eta(e_p, \Delta\phi) + \eta\langle e_p, \frac{\partial\phi}{\partial n}\rangle_\Gamma + \lambda\langle e_p, \phi\rangle_\Gamma\Big)dt$$
$$= -\int_{t^{n-1}}^{t^n} \big(a(e_p, \phi_h - \phi) + \lambda\langle e_p, \phi_h - \phi\rangle_\Gamma - \eta(e_p, \Delta\phi)\big)dt$$

where at the last two equalities we have used integration by parts (in space), and the definition of ϕ as a dual problem with zero Robin boundary data respectively. Therefore,

$$\int_{t^{n-1}}^{t^n} \|e_{1h}\|_{L^2(\Omega)}^2 dt + (\phi_{h+}^n, e_{1h-}^n) - (e_{1h-}^{n-1}, \phi_{h+}^{n-1}) \leq \int_{t^{n-1}}^{t^n} \eta\|\phi_h - \phi\|_{H^1(\Omega)}\|e_p\|_{H^1(\Omega)} dt$$
$$+ \int_{t^{n-1}}^{t^n} \big(\eta\|e_p\|_{L^2(\Omega)}\|\Delta\phi\|_{L^2(\Omega)} + \lambda\|e_p\|_{L^2(\Gamma)}\|\phi_h - \phi\|_{L^2(\Gamma)}\big)dt.$$

Then summing the above inequalities and using the fact that $\phi_+^N \equiv 0$ and $e_{1h-}^0 = 0$

(by definition) and rearranging terms, we obtain

$$
\begin{aligned}
&(1/2)\|e_{1h}\|^2_{L^2[0,T;L^2(\Omega)]} \leq C\eta \int_0^T \|e_p\|_{L^2(\Omega)}\|\phi\|_{H^2(\Omega)}dt \\
&\quad +C\int_0^T \big(\eta\|\phi_h-\phi\|_{H^1(\Omega)}\|e_p\|_{H^1(\Omega)} + \lambda\|e_p\|_{L^2(\Gamma)}\|\phi_h-\phi\|_{L^2(\Gamma)}\big)dt \\
&\leq C\Big(\eta\|e_p\|_{L^2[0,T;L^2(\Omega)]}\|\phi\|_{L^2[0,T;H^2(\Omega)]} + \eta\|\phi_h-\phi\|_{L^2[0,T;H^1(\Omega)]}\|e_p\|_{L^2[0,T;H^1(\Omega)]} \\
&\quad +\lambda\|e_p\|_{L^2[0,T;L^2(\Gamma)]}\|\phi_h-\phi\|_{L^2[0,T;L^2(\Gamma)]}\Big) \\
&\leq C\Big(\eta\|e_p\|_{L^2[0,T;L^2(\Omega)]}\|e_{1h}\|_{L^2[0,T;L^2(\Omega)]} \\
&\quad +(h+\tau^{1/2})\|e_{1h}\|_{L^2[0,T;L^2(\Omega)]}(\|e_p\|_{L^2[0,T;H^1(\Omega)]} + \lambda\|e_p\|_{L^2[0,T;L^2(\Gamma)]})\Big).
\end{aligned}
$$

Here, we have used the Cauchy-Schwarz inequality, the stability bounds of dual equation (4.8), i.e., and the error estimates (4.11) on $\phi_h - \phi$. Finally, the estimate on $\|r_1\|_{L^2[0,T;L^2(\Omega)]}$ follows by using a similar duality argument. ☐

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 [15, Section 2] can be applied directly to derive an estimate in $L^\infty[0,T;L^2(\Omega)]$ (see also Proposition 4.10).

THEOREM 4.6. *Let $w_h, z_h \in \mathcal{U}_h$ be the solutions of (4.1)-(4.2). Denote by $e_1 = \bar{y}-w_h$, $r_1 = \bar{\mu} - z_h$ and suppose that the assumptions of Theorem 4.5 hold. Then there exists a constant C depending on C_k, Ω such that*

$$
\begin{aligned}
\|e_1\|_{L^\infty[0,T;L^2(\Omega)]} &\leq C\big(\|e_p\|_{L^\infty[0,T;L^2(\Omega)]} + \|e_1^0\|_{L^2(\Omega)} \\
&\quad +\|e_p\|_{L^2[0,T;H^1(\Omega)]} + \lambda\|e_p\|_{L^2[0,T;L^2(\Gamma)]}\big), \\
\|r_1\|_{L^\infty[0,T;L^2(\Omega)]} &\leq C\big(\|r_p\|_{L^\infty[0,T;L^2(\Omega)]} + \|e_1\|_{L^2[0,T;L^2(\Omega)]} \\
&\quad +\|r_p\|_{L^2[0,T;H^1(\Omega)]} + \lambda\|r_p\|_{L^2[0,T;L^2(\Gamma)]}\big).
\end{aligned}
$$

Here $e_p = \bar{y} - Q_h^{loc}\bar{y}$, $r_p = \bar{\mu} - P_h^{loc}\bar{\mu}$.

Proof. Splitting the error as in the previous theorem, i.e., $e_1 = e_{1h} + e_p$ it suffices to bound the term $\sup_{t^{n-1}<t\leq t^n}\|e_{1h}(t)\|^2_{L^2(\Omega)}$. This is done in [15, Theorem 2.5] (note that the orthogonality condition is uncoupled). The estimate for the adjoint variable can be derived similarly. ☐

REMARK 4.7. *The combination of the last two Theorems implies the "symmetric, regularity free" structure of our estimate. In particular, suppose that the initial data $y_0 \in L^2(\Omega)$, and the forcing term $f \in L^2[0,T;H^1(\Omega)^*]$. Then, define the natural energy norm $\|(.,.)\|_X$ endowed by the weak formulation under minimal regularity assumptions as follows:*

$$\|(e_1, r_1)\|_X \equiv \|e_1\|_{W(0,T)} + \|r_1\|_{W(0,T)}.$$

Then, using Theorems 4.5, 4.6 we obtain an estimate of the form

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

The above estimate indicates that the error is as good as the approximation properties enables it to be, under the natural parabolic regularity assumptions, and it can be *viewed as the fully-discrete analogue of Céa's Lemma (see e.g. [17]). Hence, the rates of convergence for e_1, r_1 depend only on the approximation and regularity results, via the projection error e_p, r_p as indicated in Lemma 4.3 and Remark 4.4. If $y_0 \in L^2(\Omega)$, i.e, $\bar{y} \in L^2[0,T;H^1(\Omega)] \cap H^1[0,T;H^1(\Omega)^*]$, and $\bar{\mu} \in L^2[0,T;H^2(\Omega)] \cap H^1[0,T;L^2(\Omega)]$ then for $l=1, k=0$, and for $\tau \le Ch^2$ we obtain that*

1. $\|e_p\|_{L^2[0,T;H^1(\Omega)]} \le C$,
2. $\|r_p\|_{L^2[0,T;H^1(\Omega)]} \le C(h\|\bar{\mu}\|_{L^2[0,T;H^2(\Omega)]} + \tau^{1/2}\|\bar{\mu}_t\|_{L^2[0,T;L^2(\Omega)]})$,
3. $\|e_p\|_{L^2[0,T;L^2(\Omega)]} \le C(h\|\bar{y}\|_{L^2[0,T;H^1(\Omega)]} + \tau^{1/2}\|\bar{y}_t\|_{L^2[0,T;H^1(\Omega)^*]})$,
4. $\|r_p\|_{L^2[0,T;L^2(\Omega)]} \le C(h^2\|\bar{\mu}\|_{L^2[0,T;H^2(\Omega)]} + \tau\|\bar{\mu}_t\|_{L^2[0,T;L^2(\Omega)]})$,
5. $\|e_p\|_{L^2[0,T;L^2(\Gamma)]} \le C\|e_p\|^{1/2}_{L^2[0,T;L^2(\Omega)]}\|e_p\|^{1/2}_{L^2[0,T;H^1(\Omega)]} \le C(h+\tau^{1/2})^{1/2}$.

Therefore, the above estimates, and Theorem 4.5, imply for $\tau \le Ch^2$ the following rates: $\|e_1\|_{L^2[0,T;L^2(\Omega)]} \approx \mathcal{O}(h)$, and $\|r_1\|_{L^2[0,T;L^2(\Gamma)]} \approx \mathcal{O}(h)$.

The estimate is applicable even in case of more regular solutions. For example, if in addition both $\bar{y}, \bar{\mu} \in L^2[0,T;H^2(\Omega)] \cap H^1[0,T;L^2(\Omega)]$ (here $l=1$, and $k=0$),

1. $\|e_p\|_{L^2[0,T;H^1(\Omega)]} \le Ch\|\bar{y}\|_{L^2[0,T;H^2(\Omega)]} + \tau^{1/2}\|\bar{y}_t\|_{L^2[0,T;L^2(\Omega)]}$.
2. $\|e_p\|_{L^2[0,T;L^2(\Omega)]} \le Ch^2\|\bar{y}\|_{L^2[0,T;H^2(\Omega)]} + \tau\|\bar{y}_t\|_{L^2[0,T;L^2(\Omega)]}$.
3. $\|e_p\|_{L^2[0,T;L^2(\Gamma)]} \le C(h^2+\tau)^{1/2}(h+\tau^{1/2})^{1/2}$.

For the boundary norm we have used Sobolev's boundary inequality. Same rates hold also the related norms of r_p. Therefore, from Theorem 4.5, we obtain that $\|e_1\|_{L^2[0,T;H^1(\Omega)]} \approx \mathcal{O}(h)$, $\|r_1\|_{L^2[0,T;H^1(\Omega)]} \approx \mathcal{O}(h)$, $\|e_1\|_{L^2[0,T;L^2(\Omega)]} \approx \mathcal{O}(h^{3/2})$ and $\|r_1\|_{L^2[0,T;L^2(\Omega)]} \approx \mathcal{O}(h^{3/2})$ when $\tau \le Ch^2$.

4.2. Unconstrained Controls: Preliminary estimates for the optimality system. It remains to compare the discrete optimality system (3.3)-(3.4)-(3.5) to the auxiliary system (4.1)-(4.2).

LEMMA 4.8. *Let $\bar{y}_h, \bar{\mu}_h, w_h, z_h \in \mathcal{U}_h$ be the solutions the discrete optimality system (3.3)-(3.4)-(3.5) and of the auxiliary system (4.1)-(4.2) respectively. Denote by $e_1 \equiv \bar{y} - w_h$, $r_1 \equiv \bar{\mu} - z_h$, and let $e_{2h} \equiv w_h - \bar{y}_h$, $r_{2h} \equiv z_h - \bar{\mu}_h$. Then there exists algebraic constant $C > 0$ such that:*

$$\|e_{2h}\|_{L^2[0,T;L^2(\Omega)]} + (\lambda/\alpha^{1/2})\|r_{2h}\|_{L^2[0,T;L^2(\Gamma)]} \le C\lambda/\alpha^{1/2}\|r_1\|_{L^2[0,T;L^2(\Gamma)]}.$$

Proof. Subtracting (3.4) from (4.2) we obtain the equation: For $n = 1, ..., N$,

$$\begin{aligned} &-(r^n_{2h+}, v^n_h) + \int_{t^{n-1}}^{t^n} \Big(\langle r_{2h}, v_{ht}\rangle + a(r_{2h}, v_h) + \lambda\langle r_{2h}, v_h\rangle_\Gamma\Big)dt \\ &= -(r^{n-1}_{2h+}, v^{n-1}_{h+}) + \int_{t^{n-1}}^{t^n} (e_{2h}, v_h)dt \quad \forall v_h \in \mathcal{P}_k[t^{n-1}, t^n; U_h]. \end{aligned} \tag{4.14}$$

Subtracting (3.3) from (4.1) and using (2.6)-(3.5), we obtain: For $n = 1, ..., N$,

$$\begin{aligned} &(e^n_{2h}, v^n_h) + \int_{t^{n-1}}^{t^n} \Big(- \langle e_{2h}, v_{ht}\rangle + a(e_{2h}, v_h) + \lambda\langle e_{2h}, v_h\rangle_\Gamma\Big)dt \\ &= (e^{n-1}_{2h}, v^{n-1}_{h+}) + \int_{t^{n-1}}^{t^n} -(\lambda^2/\alpha)(\bar{\mu} - \bar{\mu}_h, v_h)_\Gamma dt \quad \forall v_h \in \mathcal{P}_k[t^{n-1}, t^n; U_h]. \end{aligned} \tag{4.15}$$

We set $v_h = e_{2h}$ into (4.14) to obtain

$$-(r^n_{2h+}, e^n_{2h}) + \int_{t^{n-1}}^{t^n} \Big(\langle r_{2h}, e_{2ht} \rangle + a(r_{2h}, e_{2h}) + \lambda \langle r_{2h}, e_{2h} \rangle_\Gamma \Big) dt \tag{4.16}$$
$$+(r^{n-1}_{2h+}, e^{n-1}_{2h+}) = \int_{t^{n-1}}^{t^n} \|e_{2h}\|^2_{L^2(\Omega)} dt.$$

Similarly, setting $v_h = r_{2h}$ into (4.15) we deduce,

$$(e^n_{2h}, r^n_{2h}) + \int_{t^{n-1}}^{t^n} \Big(-\langle e_{2h}, r_{2ht} \rangle + a(e_{2h}, r_{2h}) + \lambda \langle e_{2h}, r_{2h} \rangle_\Gamma \Big) dt \tag{4.17}$$
$$-(e^{n-1}_{2h}, r^{n-1}_{2h+}) = \int_{t^{n-1}}^{t^n} \Big(-(\lambda^2/\alpha) \langle r_1, r_{2h} \rangle_\Gamma - (\lambda^2/\alpha) \|r_{2h}\|^2_{L^2(\Gamma)} \Big) dt.$$

Integrating by parts with respect to time in (4.17), and subtracting the resulting equation from (4.16), we arrive to

$$(r^n_{2h+}, e^n_{2h}) - (e^{n-1}_{2h}, r^{n-1}_{2h+}) + \int_{t^{n-1}}^{t^n} \Big(\|e_{2h}\|^2_{L^2(\Omega)} + (\lambda^2/\alpha) \|r_{2h}\|^2_{L^2(\Gamma)} \Big) dt \tag{4.18}$$
$$= -(\lambda^2/\alpha) \int_{t^{n-1}}^{t^n} \langle r_1, r_{2h} \rangle_\Gamma dt.$$

Using Young's inequality to bound the right hand side, adding the resulting inequalities (4.18) from 1 to N, and noting that $\sum_{n=1}^N \big((r^n_{2h+}, e^n_{2h}) - (e^{n-1}_{2h}, r^{n-1}_{2h+}) \big) = 0$ (since $e^0_{2h} \equiv 0$, $r^N_{2h+} = 0$) we obtain the desired estimate. □

Estimates easily follow by the previous Lemma and the estimates on the projections e_1 and r_1 together with a classical "boot-strap" argument.

PROPOSITION 4.9. *Let $\bar{y}_h, \bar{\mu}_h, w_h, z_h \in \mathcal{U}_h$ be the solutions of the optimality system (3.3)-(3.4)-(3.5) and of the auxiliary system (4.1)-(4.2) respectively. Denote by $e_1 \equiv \bar{y} - w_h$, $r_1 \equiv \bar{\mu} - z_h$, and let $e_{2h} \equiv w_h - \bar{y}_h$, $r_{2h} \equiv z_h - \bar{\mu}_h$. Then, the following estimate holds:*

$$\|e^N_{2h}\|^2_{L^2(\Omega)} + \sum_{i=0}^{N-1} \|[e^i_{2h}]\|^2_{L^2(\Omega)} + C_F \min\{\eta, \lambda\} \int_0^T \|e_{2h}\|^2_{H^1(\Omega)} dt$$
$$+\lambda \int_0^T \|e_{2h}\|^2_{L^2(\Gamma)} dt \leq (C/\lambda\alpha^2) \int_{t^{n-1}}^{t^n} \|r_1\|^2_{L^2(\Gamma)} dt$$
$$\|r^0_{2h+}\|^2_{L^2(\Omega)} + \sum_{i=1}^{N} \|[r^i_{2h}]\|^2_{L^2(\Omega)} + C_F \min\{\eta, \lambda\} \int_0^T \|r_{2h}\|^2_{H^1(\Omega)} dt$$
$$+\lambda \int_0^T \|r_{2h}\|^2_{L^2(\Gamma)} dt \leq \big(C\lambda^2/\alpha C_F \min\{\eta, \lambda\} \big) \int_0^T \|r_1\|^2_{L^2(\Gamma)} dt,$$

where C is constant depending only upon Ω.

Proof. Step 1: Estimates for the state: Setting $v_h = e_{2h}$ into (4.15) and noting that $\mu - \mu_h = r_1 + r_{2h}$ we obtain

$$(1/2)\|e^n_{2h}\|^2_{L^2(\Omega)} + (1/2)\|[e^{n-1}_{2h}]\|^2_{L^2(\Omega)} - (1/2)\|e^{n-1}_{2h}\|^2_{L^2(\Omega)} + \eta \int_{t^{n-1}}^{t^n} \|\nabla e_{2h}\|^2_{L^2(\Omega)} dt$$
$$+\lambda \int_{t^{n-1}}^{t^n} \|e_{2h}\|^2_{L^2(\Gamma)} dt \leq -(\lambda^2/\alpha) \int_{t^{n-1}}^{t^n} (r_1 + r_{2h}, e_{2h})_\Gamma dt. \tag{4.19}$$

Using Young's inequality for the first term on the right hand side, (4.19) gives,

$$(1/2)\|e_{2h}^n\|^2_{L^2(\Omega)} + (1/2)\|[e_{2h}^{n-1}]\|^2_{L^2(\Omega)} - (1/2)\|e_{2h}^{n-1}\|^2_{L^2(\Omega)} + \eta \int_{t^{n-1}}^{t^n} \|\nabla e_{2h}\|^2_{L^2(\Omega)} dt$$

$$+(\lambda/2)\int_{t^{n-1}}^{t^n} \|e_{2h}\|^2_{L^2(\Gamma)} dt \leq (1/\lambda\alpha^2)\int_{t^{n-1}}^{t^n} (\|r_1\|^2_{L^2(\Gamma)} + \|r_{2h}\|^2_{L^2(\Gamma)})dt. \qquad (4.20)$$

Using Friedrichs' inequality, and standard algebra we obtain the estimate upon summation by using the estimate on $\|r_{2h}\|_{L^2[0,T;L^2(\Gamma)]}$ of Lemma 4.8.
Step 2: Estimates for the adjoint: Setting $v_h = r_{2h}$ into (4.14), and using Friedrichs' and Young's inequalities, and Lemma 4.8 to bound the norm of $\|e_{2h}\|_{L^2[0,T;L^2(\Omega)]}$ we obtain the desired estimate. □

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 [15] into the Robin boundary linear case. Here, the stability estimate at arbitrary time-points will be also needed.

Proposition 4.10. *Suppose that the assumptions of Theorem 4.5, and Proposition 4.9 hold. Then there exists a constant C depending only upon constant C_k, and the domain such that,*

$$\begin{aligned}\|e_{2h}\|_{L^\infty[0,T;L^2(\Omega)]} &\leq C\big(\eta\|e_{2h}\|_{L^2[0,T;H^1(\Omega)]} + \lambda\|e_{2h}\|_{L^2[0,T;L^2(\Gamma)]} \\ &\quad +(\lambda^{3/2}/\alpha)\|r_1\|_{L^2[0,T;L^2(\Gamma)]}\big), \\ \|r_{2h}\|_{L^\infty[0,T;L^2(\Omega)]} &\leq C\big(\eta\|r_{2h}\|_{L^2[0,T;H^1(\Omega)]} + (\lambda/\alpha^{1/2})\|r_1\|_{L^2[0,T;L^2(\Gamma)]}\big).\end{aligned}$$

Proof. The proof closely follows the techniques of [15, Section 2], adjusted to the Robin boundary data case. For completeness, we state the proof for the first estimate, while the second one can be treated similarly. First, we briefly recall the main tool of approximations of the discrete characteristic function. For any polynomial $s \in \mathcal{P}_k(t^{n-1}, t^n)$, we denote the discrete approximation of $\chi_{[t^{n-1},t)}s$ by the polynomial $\hat{s} \in \{\hat{s} \in \mathcal{P}_k(t^{n-1}, t^n), \hat{s}(t^{n-1}) = s(t^{n-1})\}$ which satisfies

$$\int_{t^{n-1}}^{t^n} \hat{s}q = \int_{t^{n-1}}^{t} sq \qquad \forall\, q \in \mathcal{P}_{k-1}(t^{n-1}, t^n).$$

The motivation for the above construction stems from the elementary observation that for $q = s'$ we obtain $\int_{t^{n-1}}^{t^n} s'\hat{s} = \int_{t^{n-1}}^{t} ss' = \frac{1}{2}(s^2(t) - s^2(t^{n-1}))$. The construction can be extended to approximations of $\chi_{[t^{n-1},t)}v$ for $v \in \mathcal{P}_k[t^{n-1}, t^n; V]$ where V is a linear space. The discrete approximation of $\chi_{[t^{n-1},t)}v$ in $\mathcal{P}_k[t^{n-1}, t^n; V]$ 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,

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

Then, [15, Lemma 2.4] states various continuity properties, and in particular that

$$\|\hat{v}\|_{L^2[t^{n-1},t^n;V]} \leq C_k\|v\|_{L^2[t^{n-1},t^n;V]}, \|\hat{v} - \chi_{[t^{n-1},t)}v\|_{L^2[t^{n-1},t^n;V]} \leq C_k\|v\|_{L^2[t^{n-1},t^n;V]}$$

where C_k is a constant depending on k. We begin by integrating by parts with respect to time in (4.15), 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 above. The definition of the $\hat{e}_{2h}$ and the fact that $e_{2ht} \in \mathcal{P}_{k-1}[t^{n-1},t^n;U_h]$ implies that $\int_{t^{n-1}}^{t^n}(e_{2ht},\hat{e}_{2h})dt = \int_{t^{n-1}}^{t}(e_{2ht},e_{2h})dt$ and hence,

$$(1/2)\|e_{2h}(t)\|^2_{L^2(\Omega)} + (1/2)\|[e_{2h}^{n-1}]\|^2_{L^2(\Omega)} + \int_{t^{n-1}}^{t^n} a(e_{2h},\hat{e}_{2h})dt \tag{4.21}$$
$$= (1/2)\|e_{2h}^{n-1}\|^2_{L^2(\Omega)} - \lambda\int_{t^{n-1}}^{t^n}(e_{2h},\hat{e}_{2h})_\Gamma dt - \int_{t^{n-1}}^{t^n}(\lambda^2/\alpha)(r_1 + r_{2h},\hat{e}_{2h})_\Gamma dt.$$

Recall also that the continuity property on $a(.,.)$, imply

$$\Big|\int_{t^{n-1}}^{t^n}\big(a(e_{2h},\hat{e}_{2h}) + \lambda(e_{2h},\hat{e}_{2h})_\Gamma\big)dt\Big| \le C_k\int_{t^{n-1}}^{t^n}(\eta\|e_{2h}\|^2_{H^1(\Omega)} + \lambda\|e_{2h}\|^2_{L^2(\Gamma)})dt$$

while the coupling term can be bounded as:

$$\Big|\frac{\lambda^2}{\alpha}\int_{t^{n-1}}^{t^n}(r_1 + r_{2h},\hat{e}_{2h})_\Gamma dt\Big| \le (C_k\lambda^3/\alpha^2)\int_{t^{n-1}}^{t^n}\Big(\|r_{2h}\|^2_{L^2(\Gamma)} + \|r_1\|^2_{L^2(\Gamma)}\Big)dt$$
$$+C_k\lambda\int_{t^{n-1}}^{t^n}\|e_{2h}\|^2_{L^2(\Gamma)}dt.$$

Here we have used Young's inequality with appropriate $\delta > 0$, and in various instances of the continuity property of the approximation of the discrete characteristic. Hence, substituting the above estimates into (4.21), we obtain an inequality of the form, $(1 - C\tau)a^n \le a^{n-1} + f^n$, 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 $\|e_{2h}^{n-1}\|^2_{L^2(\Omega)} \le a_{n-1}$. Hence the desired estimate follows by summation and by Lemma 4.8. □

4.3. Unconstrained Controls: Symmetric error estimates and estimates for rough initial data. Various estimates can be derived, using results of previous subsections and standard approximation theory results. We begin by stating symmetric error estimates which can be viewed as the analogue of the classical Céa's Lemma.

THEOREM 4.11. *Let $\bar{y}_h, \bar{\mu}_h \in \mathcal{U}_h$ and $(\bar{y},\bar{\mu}) \in W(0,T)$ denote the approximate solutions of the discrete and continuous optimality systems (3.3)-(3.4)-(3.5) and (2.4)-(2.5)-(2.6) respectively. Let $e_p = \bar{y} - Q_h^{loc}\bar{y}$, $r_p = \bar{\mu} - P_h^{loc}\bar{\mu}$ denote the projection error, where P_h^{loc}, Q_h^{loc} defined in Definition of 4.1, and 4.2 respectively. Then, the following estimate holds for the error $e = \bar{y} - \bar{y}_h$ and $r = \bar{\mu} - \bar{\mu}_h$:*

$$\|(e,r)\|_X \le \tilde{\mathbf{C}}(1/\alpha)\|(e_p,r_p)\|_X$$

where $\tilde{\mathbf{C}}$ depends upon constants of Theorems 4.5, 4.6, and Proposition 4.9, 4.10, and is independent of τ, h, α.

Proof. The first estimate follows by using triangle inequality and previous estimates of Theorem 4.5 and 4.6 and Propositions 4.9 and 4.10. □

An improved estimate for the $L^2[0,T;L^2(\Omega)]$ norm for the state, and in $L^2[0,T;L^2(\Gamma)]$ for the adjoint follow by combining the estimates of Theorem 4.5, and Lemma 4.8.

THEOREM 4.12. *Suppose that $y_0 \in L^2(\Omega)$, $f \in L^2[0,T;H^1(\Omega)^*]$, and the assumptions of Theorem 4.5 and Lemma 4.8 hold. Let $e_p = \bar{y} - Q_h^{loc}\bar{y}$, $r_p = \bar{\mu} - P_h^{loc}\bar{\mu}$ denote the* *projection error, where P_h^{loc}, Q_h^{loc} defined in Definition of 4.1, and 4.2 respectively. Then, there exists a constant C independent of h, τ, α such that,*

$$\|e\|_{L^2[0,T;L^2(\Omega)]} \leq C\Big(\|e_p\|_{L^2[0,T;L^2(\Omega)]} + \tau^{1/2}(\|e_p\|_{L^2[0,T;H^1(\Omega)]} + \|e_p\|_{L^2[0,T;L^2(\Gamma)]}) + (\lambda/\alpha^{1/2})\|r_1\|_{L^2[0,T;L^2(\Gamma)]}\Big)$$

$$\|r\|_{L^2[0,T;L^2(\Gamma)]} \leq C\|r_1\|^{1/2}_{L^2[0,T;L^2(\Omega)]}\|r_1\|^{1/2}_{L^2[0,T;H^1(\Omega)]},$$

where r_1 is estimated in terms of projection errors e_p, r_p by Theorem 4.5.

Proof. The first estimate follows by using triangle inequality and previous estimates of Theorem 4.5, Lemma 4.8. The second estimate follows by triangle inequality, the estimate of Lemma 4.8 to bound r_{2h} and Sobolev's boundary inequality. □

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 4.13. *Suppose that the assumptions of Theorem 4.5 and Lemma 4.8 hold, and let $y_0 \in L^2(\Omega)$, $f \in L^2[0,T;H^1(\Omega)^*]$. Assume that piecewise linear polynomials are being used to construct the subspaces $U_h \subset H^1(\Omega)$ in each time step, and piecewise constants polynomials $k = 0$ for the temporal discretization. Then, for $\tau \leq Ch^2$ we obtain,*

$$\|e\|_{L^2[0,T;L^2(\Omega)]} \leq Ch \text{ and } \|r\|_{L^2[0,T;L^2(\Gamma)]} \leq Ch.$$

If in addition, $\bar{y}, \bar{\mu} \in L^2[0,T;H^2(\Omega)] \cap H^1[0,T;L^2(\Omega)]$ then,

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

$$\|e\|_{L^2[0,T;L^2(\Omega)]} \leq C(1/\alpha^{1/2})\big(h^2 + \tau + (h^2+\tau)^{1/2}(h+\tau^{1/2})^{1/2} + (h+\tau^{1/2})^2\big),$$

$$\|r\|_{L^2[0,T;L^2(\Gamma)]} \leq C(h^2+\tau)^{1/2}(h+\tau^{1/2})^{1/2},$$

which imply for $\tau \approx h^2$, the rates $\|(e,r)\|_X \approx \mathcal{O}(h)$, $\|e\|_{L^2[0,T;L^2(\Omega)]} \approx \mathcal{O}(h^{3/2})$, and $\|r\|_{L^2[0,T;L^2(\Gamma)]} \approx \mathcal{O}(h^{3/2})$.

Proof. The rates directly follow from Theorem 4.11, Theorem 4.12, Lemma 4.3 and Remark 4.7. □

4.4. Control Constraints: The variational discretization approach. It is worth noting that our estimates are also applicable in case of point-wise control constraints, when using the variational discretization approach of Hinze ([25]). The variational discretization approach implies that $\mathcal{A}_{ad}^d \equiv \mathcal{A}_{ad}$, i.e., the control is not discretized explicitly, but only implicitly via the adjoint variable. Thus, our discrete optimal control problem now coincides to: Minimize functional $J_h(y_h(g), g) \equiv (1/2)\int_0^T \|y_h(g) - y_d\|^2_{L^2(\Omega)}dt + (\alpha/2)\int_0^T \|g\|^2_{L^2(\Gamma)}dt$ subject to (3.1), where $y_h(g) \in \mathcal{U}_h$ denotes the solution of (3.1) with right hand side given control $g \in L^2[0,T;L^2(\Gamma)]$. Then, the optimal control (abusing the notation, denoted again by $\bar{g}_h$) satisfies the following first order optimality condition,

$$J_h'(\bar{g}_h)(u - \bar{g}_h) \geq 0, \qquad \text{for all } u \in L^2[0,T;L^2(\Gamma)],$$

where $\bar{g}_h$ can take the form $\bar{g}_h(t,x) = Proj_{[g_a,g_b]}(-\frac{\lambda}{\alpha}\bar{\mu}_h(\bar{g}_h(t,x)))$, for a.e. $(t,x) \in (0,T] \times \Gamma$ similar to continuous case. We note that the $\bar{g}_h$ is not in general a finite element function corresponding to our finite element mesh, hence its algorithmic construction requires extra care (see e.g. [25]). However, in most practical situations, the main goal is to minimize and compute the state variable, and not necessarily the control that is used to achieve our goal. For the second derivative we easily obtain an estimate independent of $\bar{g}, \bar{g}_h$, and in particular,

$$J_h''(u)(\tilde{u},\tilde{u}) \geq \alpha\|\tilde{u}\|^2_{L^2[0,T;L^2(\Gamma)]}, \qquad \text{for all } \tilde{u} \in L^2[0,T;L^2(\Gamma)].$$

THEOREM 4.14. *Let $y_0 \in L^2(\Omega)$, $f \in L^2[0,T;H^1(\Omega)^*]$, and $y_d \in L^2[0,T;L^2(\Omega)]$. Suppose that $\mathcal{A}^d_{ad} \equiv \mathcal{A}_{ad}$ and let $\bar{g}, \bar{g}_h$ denote the solutions of the corresponding continuous and discrete optimal control problems. Then, the following estimate hold:*

$$\|\bar{g}-\bar{g}_h\|_{L^2[0,T;L^2(\Gamma)]} \leq C(1/\alpha)\|\mu(\bar{g})-\mu_h(\bar{g})\|_{L^2[0,T;L^2(\Gamma)]},$$

where $\mu_h(\bar{g})$ and $\mu(\bar{g})$ denote the solutions of (3.2) and (2.3) respectively. Furthermore, if $\tau \leq Ch^2$, $\|\bar{g}-\bar{g}_h\|_{L^2[0,T;L^2(\Gamma)]} \approx \mathcal{O}(h)$.

Proof. We note that $\mathcal{A}^d_{ad} \equiv \mathcal{A}_{ad}$, and hence the optimality conditions imply that

$$J_h'(\bar{g}_h)(\bar{g}-\bar{g}_h) \geq 0 \qquad \text{and} \qquad J'(\bar{g})(\bar{g}-\bar{g}_h) \leq 0. \tag{4.22}$$

Therefore, using the second order condition and the mean value theorem, we obtain for any $u \in L^2[0,T;L^2(\Gamma)]$, (and hence for the one resulting from the mean value theorem) and inequalities (4.22),

$$\begin{aligned}
&\alpha\|\bar{g}-\bar{g}_h\|^2_{L^2[0,T;L^2(\Gamma)]} \leq J_h''(u)(\bar{g}-\bar{g}_h,\bar{g}-\bar{g}_h)\\
&= J_h'(\bar{g})(\bar{g}-\bar{g}_h) - J_h'(\bar{g}_h)(\bar{g}-\bar{g}_h) \leq J_h'(\bar{g})(\bar{g}-\bar{g}_h) - J'(\bar{g})(\bar{g}-\bar{g}_h)\\
&= \int_0^T\int_\Gamma (\mu_h(\bar{g})-\mu(\bar{g}))(\bar{g}-\bar{g}_h)dxdt \leq C\|\mu(\bar{g})-\mu_h(\bar{g})\|_{L^2[0,T;L^2(\Gamma)]}\|\bar{g}-\bar{g}_h\|_{L^2[0,T;L^2(\Gamma)]},
\end{aligned}$$

which clearly implies the first estimate. Now, a rate of convergence can be obtained using similar arguments to Theorem 4.5. Indeed, note that subtracting (3.2) from (2.3) and setting $\bar{r} = \mu_h(\bar{g}) - \mu(\bar{g})$, and $\bar{e} = y_h(\bar{g}) - y(\bar{g})$, we obtain the analog of orthogonality condition (4.4)-(4.5), i.e., for all $n = 1, ..., N$ and for all $v_h \in \mathcal{P}_k[t^{n-1}, t^n; U_h]$,

$$(\bar{e}_1^n, v_h^n) + \int_{t^{n-1}}^{t^n} \Big(-\langle \bar{e}_1, v_{ht}\rangle + a(\bar{e}_1, v_h) + \lambda\langle \bar{e}_1, v_h\rangle_\Gamma \Big)dt = (\bar{e}_1^{n-1}, v_{h+}^{n-1}),$$

$$\begin{aligned}
-(\bar{r}_+^n, v_h^n) &+ \int_{t^{n-1}}^{t^n} \Big(\langle \bar{r}, v_{ht}\rangle + a(\bar{r}, v_h) + \lambda\langle \bar{r}, v_h\rangle_\Gamma \Big)dt\\
&= -(\bar{r}_{1+}^{n-1}, v_{h+}^{n-1}) + \int_{t^{n-1}}^{t^n} (\bar{e}, v_h)dt,
\end{aligned}$$

Using Sobolev's boundary inequality, the estimates of Theorem 4.5, and the rates of Proposition 4.13, we obtain the desired estimate, after noting the reduced regularity of $\bar{e}$. ☐

4.5. Numerical experiments. We consider three numerical examples for the model problem on $\Omega \times [0,T] = [0,1]^2 \times [0, 0.1]$. The first two examples are computed based on piecewise constants (in time) ($k = 0$), and standard piecewise linear elements ($l = 1$) in space, while the third one with piecewise linear polynomials in both time and space ($k = 1$, $l = 1$). More specifically, we consider three examples with:

1. Regular initial condition for the state variable (with known exact solution for the state variable).
2. Discontinuous initial data $y_0 \in L^2(\Omega)$: In this case we assume that exact solution is the solution with time-space mesh $dt = 2.71267e{-}05$, $h = 5.20833e{-}03$ (3687 and 37249 ndof respectively).
3. A coarse time stepping approach for a problem with known exact solution for the state variable.

We note that the boundary control does not possesses continuous derivatives at some points. The examples are based on the one presented in [13]. In all examples we fix the regularization parameter in the functional as $\alpha = \pi^{-4}$. The optimal control problem is solved by the finite element toolkit FreeFem++ (see [24]) using a gradient algorithm method in a 4 Six-Core AMD Opteron(tm) Processor 8431, 96 GB RAM computer.

EXAMPLE 1. Let $a = -\sqrt{5}$. We choose right-hand side

$$f(t, x_1, x_2) = \pi^2 e^{a\pi^2 t} \left\{ -(2x_2^2 - 2x_2 + 2x_1^2 + a + 1) \sin(\pi x_1 x_2) \sin(\pi x_1(x_2 - 1)) + 2(x_2^2 - x_2 + x_1^2) \cos(\pi x_1 x_2) \cos(\pi x_1(x_2 - 1)) \right\},$$

initial condition $y_0(x_1, x_2) = \sin(\pi(1 + x_1 x_2)) \sin(\pi x_1(x_2 - 1))$, and target function $y_d(t, x_1, x_2) = 0.5$, in a way to guarantee that the optimal solution $(\bar{y}, \bar{g})$ of the optimal control problem is given by

$$\bar{y}(t, x_1, x_2) = \exp(a\pi^2 t) \sin(\pi(1 + x_1 x_2)) \sin(\pi x_1(x_2 - 1)),$$

while $\bar{g}$ has been computed by using the Robin condition at each component of the boundary. For this choice of data the corresponding errors for the state and the control variable for different meshes are shown in the Table 4.1. In this case, the predicted (theoretical) rates for the $L^2[0, T; L^2(\Omega)]$ and $L^2[0, T; H^1(\Omega)]$ norms are $\mathcal{O}(\tau + h^{3/2})$, and $\mathcal{O}(\tau + h)$ respectively, and they are both verified numerically.

TABLE 4.1
Experiment 1-Rates of convergence for the 2d solution with $\tau = h^2/2$ and regular initial data.

Discretization	Error		
$\tau = h^2/2$	$\|e\|_{L^2[0,T;L^2(\Omega)]}$	$\|e\|_{L^2[0,T;H^1(\Omega)]}$	$J(y, g)$
$h = 0.2357022$	0.018310605	0.070340370	0.002395820
$h = 0.1178511$	0.004085497	0.031958661	0.001857961
$h = 0.0589255$	0.001335615	0.016375314	0.001738954
$h = 0.0294627$	0.000766443	0.008819160	0.001711876
$h = 0.0147313$	0.000676697	0.005626214	0.001705198
Conv. rate	1.526118558	0.998546583	-

EXAMPLE 2. In this test problem Ω, and T, are the same as in the Example 1. The difference is that in this example the initial data y_0 is a discontinuous function, defined as follows:

$$y_0 = \begin{cases} \sin(\pi(1 + x_1 x_2)) \sin(\pi x_1(x_2 - 1)) & \text{if } x_1, x_2 \geq 0.5, \\ 10 + \sin(\pi(1 + x_1 x_2)) \sin(\pi x_1(x_2 - 1)) & \text{otherwise.} \end{cases}$$

The results related to the errors are demonstrated in Table 4.2, where the rate of $\mathcal{O}(h)$ when $\tau \leq Ch^2$, for the $L^2[0, T; L^2(\Omega)]$ norm is verified for the state and adjoint

Table 4.2
Experiment 2-Rates of convergence for the 2d solution with $\tau = h^2/2$ and rough initial data.

Discretization	Error		
$\tau = h^2/2$	$\|e\|_{L^2[0,T;\mathrm{L}^2(\Omega)]}$	$\|r\|_{L^2[0,T;\mathrm{L}^2(\Omega)]}$	$J(y,g)$
$h = 0.2357022$	0.4093275092	0.008552165422	0.9411555956
$h = 0.1178511$	0.1555909764	0.005056762072	0.8225865966
$h = 0.0589255$	0.0714820269	0.002440981965	0.7424795375
$h = 0.0294627$	0.0302970740	0.001179518135	0.7066657202
$h = 0.01473139$	0.0100448501	0.001097951813	0.6883517113
Conv. rate	1.2520017243	0.952697386266	-

variable. The results give a little bit better rate of convergence due to the constructive way of the state variable. Obviously the error norm $L^2[0,T;H^1(\Omega)]$ doesn't give a rate, since the data $y_0 \in L^2(\Omega)$ and the initial discontinuity is disseminated through characteristics in the whole exact solution.

EXAMPLE 3. To illustrate the potential applicability of higher order time stepping schemes, we consider a coarse time-stepping approach based on the $k = 1$ time stepping scheme. Here, we return to the example 1, with the known smooth solution $\bar{y}$ given by $\bar{y}(t, x_1, x_2) = \exp(a\pi^2 t)\sin(\pi(1 + x_1x_2))\sin(\pi x_1(x_2 - 1))$ for $k = 1$, $l = 1$. Note that despite the fact that we have chosen smooth state variable, the presence of a Robin boundary control limits the regularity at least near by the boundary for the time derivative of the adjoint and control variables. However overall, we expect that the parabolic regularity will appear as time progresses. Our best approximation type estimates for "smooth" state, adjoint and control variables yield a convergence rate with respect to $L^2[0,T;H^1(\Omega)]$ norm of order $\mathcal{O}(\tau^2 + h)$, when piecewise linears are considered for both time and space i.e., $k = 1$, $l = 1$. In the following experiments we present the rate based on a coarse time stepping approach. In particular, for $\tau = h^{1/2}$, which corresponds to very few time steps compared to the standard approaches, the Table 4.3, clearly indicates that we still obtain a rate, of almost $\mathcal{O}(h)$. Of course, it is expected that the rate is suboptimal due to the lack of smoothness near the boundary.

Table 4.3
Experiment 3-Rates of convergence for the 2d solution with $k = 1$, $\tau = h^{1/2}$, regular initial data.

Discretization	Error	
$\tau = h^{1/2}$	$\|e\|_{L^2[0,T;\mathrm{H}^1(\Omega)]}$	$J(y,g)$
$h = 0.2357022$	0.068070558	0.002676642
$h = 0.1178511$	0.040332082	0.002579619
$h = 0.0589255$	0.019010050	0.002468955
$h = 0.0294627$	0.012117836	0.002384007
$h = 0.0147313$	0.008222888	0.002322462
$h = 0.0073656$	0.005980212	0.002276926
Conv. rate	0.762328463	-

Acknowledgements: The first author would like to thank Prof. Boris Vexler for many useful discussions regarding the Lemma 4.3. The second author is supported by Papakyriakopoulos Scholarship.

REFERENCES

[1] W. Alt and U. Mackenroth, Convergence of Finite Element Approximations to State Constrained Convex Parabolic Boundary Control Problems, *SIAM J. on Control and Optim.*, **27(4)** (1989), pp. 718-736.

[2] W. Barthel, C. John and F. Tröltzsch, Optimal boundary control of a system of reaction diffusion equations, *ZAMM.*, **90(12),** (2010), pp. 966982.

[3] F. Ben Belgacem, H. El Fekih, and J.-P. Raymond, A penalized approach for solving a parabolic equation with nonsmooth Dirichlet boundary conditions, *Asympt. Anal.*, **34** (2003), pp 121-136.

[4] S. Brenner and L. Scott, The Mathematical Theory of Finite Element Methods, Springer-Verlag, 1996.

[5] E. Casas, and K. Chrysafinos, A discontinuous Galerkin time stepping scheme for the velocity tracking problem, *SIAM. J. Numer. Anal.*, **50** (2012), pp. 2281-2306.

[6] E. Casas, and K. Chrysafinos, Error estimates for the discretization of the velocity tracking problem, *submitted, available at http://www.math.ntua.gr/ chrysafinos.*

[7] E. Casas and M. Mateos, and J.-P. Raymond, Penalization of dirichlet optimal control problems, *ESAIM COCV*, **15** (2009), pp 782–809.

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

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

[10] K. Chrysafinos, Convergence of discontinuous time-stepping schemes for a Robin boundary control problem under minimal regularity assumptions, *Int. J. Numer. Anal. and Model.*, **10** (2013), pp. 673-696 .

[11] 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.

[12] K. Chrysafinos, and L.S. Hou, Error estimates for semidiscrete finite element approximations for linear and semilinear parabolic equations under minimal regularity assumptions, *SIAM J. Numer. Anal.*, **40** (2002), pp. 282-306.

[13] K. Chrysafinos and E. Karatzas, Symmetric error estimates for discontinuous Galerkin approximations for an optimal control problem associated to semilinear parabolic PDEs, *Disc. and Contin. Dynam. Syst. - Ser. B*, **17** (2012), pp.: 1473 - 1506.

[14] K. Chrysafinos and E. Karatzas, Discontinuous Galerkin time-stepping schemes for Robin boundary control problems constrained to parabolic PDEs, *IFAC Workshop on Control of Systems Governed by PDEs*, (2013).

[15] 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.

[16] K. Chrysafinos and N. J. Walkington, Discontinuous Galerkin approximations of the Stokes and Navier-Stokes equations, Math. Comp., **79** (2010), 2135–2167.

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

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

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

[20] P. Grisvard, Elliptic problems in nonsmooth domains, Pitman, Boston, 1985.

[21] M.D.Gunzburger and L.S.Hou, Treating inhomogeneous essential boundary conditions in finite element methods and the calculation of boundary stresses, *SIAM J.Numer. Anal.*, 29 (**2**), (1992) pp. 390-424.

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

[23] M.D. Gunzburger and S. Manservisi, The velocity tracking problem for Navier-Stokes flow with boundary control, *SIAM J. on Control and Optim.*, **39** (2000), pp. 594634.

[24] F. Hecht, FreeFem++, Third edition, Version 3.13, 2011. Available from: http://www.freefem.org/ff++.

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

[26] M.Hinze and K.Kunisch, Second order methods for boundary control of the instationary Navier-Stokes system, *ZAMMZ. Angew. Math. Mech.*, **84** (2004), pp. 171-187.

[27] L.S. Hou and S. Ravindran, A penalized Neumann control approach for solving an optimal Dirichlet control problem for the Navier-Stokes equations, *SIAM J. Control Optim.*, **36** (1998), pp. 1795-1814.

[28] K.Kunisch and B.Vexler, Constrained Dirichlet boundary control in L^2 for a class of evolution equations, *SIAM J. on Control and Optim.*, **46 (5)** (2007), pp. 1726-1753.

[29] 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.

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

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

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

[33] 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.

[34] D. Meidner and B. Vexler, A priori error estimates for space-time finite element discretization of parabolic optimal control problems. Part II: Problems with control constraints, *SIAM J. Control and Optim.*, **47** (2008), pp. 1301-1329.

[35] J. Necas, Les méthodes directes en Théorie des equations elliptiques, Masson et Cue. Paris 1967.

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

[37] I. Neitzel and B. Vexler, A priori error estimates for space-time finite element discretization of semilinear parabolic optimal control problems, *Numer. Math.*, **120** (2012), pp. 345-386.

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

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

[40] 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.

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