SYMMETRIC ERROR ESTIMATES FOR DISCONTINUOUS GALERKIN TIME-STEPPING SCHEMES FOR OPTIMAL CONTROL PROBLEMS CONSTRAINED TO EVOLUTIONARY STOKES EQUATIONS

KONSTANTINOS CHRYSAFINOS,* AND EFTHIMIOS N. KARATZAS†

Abstract. We consider fully discrete finite element approximations of a distributed optimal control problem, constrained by the evolutionary Stokes equations. 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. The estimates are also applicable when high order schemes are being used. Computational examples validating our expected rates of convergence are also provided.

Keywords: Discontinuous Time-Stepping Schemes, Finite Element Approximations, Stokes Equations, Velocity Tracking Problem, Distributed Controls, 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 the evolutionary Stokes equations. In particular, given a target function y_d we seek velocity y and control variable g such that the functional

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

is minimized subject to the constrains,

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

Here, $\Omega \subset \mathbb{R}^d$, $d = 2,3$ denotes an open bounded and convex domain. Our main estimates are valid under the general assumption of a Lipschitz boundary Γ, and the domain should allow $\mathbf{H}^2$ regularity for the velocity solution of the corresponding steady Stokes equations (see Remark 2.6 and Section 3.1 for a more detailed discussion). The control g is of distributed type, the forcing term f, and the viscosity constant $\nu > 0$, are given data, while $\alpha > 0$ denotes a small penalty parameter which limits the size of the control, and in many instances can be of comparable size to the time and spacial discretization parameters (denoted by τ, and h respectively). Special emphasis is placed in the case of rough initial data, i.e., $y_0 \in \mathbf{W}(\Omega) \equiv \{v \in \mathbf{L}^2(\Omega) : \operatorname{div} v = 0, \quad v\cdot\mathbf{n} = 0\}$, however our analysis also includes the possibility of using high order schemes. Furthermore, we are also interested in

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

†Department of Mathematics, National Technical University of Athnens, Zografou Campus, Athens 15780, Greece, karmakis@math.ntua.gr

The second author is supported by Papakyriakopoulos Scholarship

case of pointwise control constraints in the sense that $g_a \leq g(t,x) \leq g_b$ for a.e. $(t,x) \in (0,T) \times \Omega$, where $g_a, g_b \in \mathbb{R}$.

The main goal is to show that the error estimates of the corresponding optimality system have the same structure to those of the uncontrolled evolutionary Stokes equations. In particular, we develop an almost symmetric error estimate under minimal regularity assumptions on the natural energy norm $\|.\|_{W(0,T)} \equiv \|.\|_{L^\infty[0,T;\mathbf{L}^2(\Omega)]} + \|.\|_{L^2[0,T;\mathbf{H}^1(\Omega)]}$ associated to our discontinuous time-stepping scheme, i.e., an estimate of the form,

$$\begin{aligned} \|\text{error}\|_{W(0,T)} &\leq C \|\text{best approximation error}\|_{W(0,T)} \\ &\quad + \|\text{best approx. error pressure}\|_{L^2[0,T;L^2(\Omega)]}, \end{aligned}$$

which states that the error is as good as the regularity and approximation theory allows it to be. The above "best approximation" framework naturally includes high order schemes, since it separates the issue of regularity of the optimal pair from the choice of the approximation scheme. As a consequence, estimates of high order can be also included similar to the uncontrolled case, at least in case of unbounded controls, when classical boot-strap arguments imply enhanced regularity.

The key feature of our estimate is that it is valid under low regularity assumptions on the given data. More precisely, the symmetric error estimate, only requires velocity $y \in L^2[0,T;\mathbf{V}(\Omega)] \cap H^1[0,T;\mathbf{H}^{-1}(\Omega)]$ and pressure $p \in L^2[0,T;L^2_0(\Omega)]$, where $\mathbf{V}(\Omega) = \{v \in \mathbf{H}^1_0(\Omega) : \operatorname{div} v = 0\}$, and $L^2_0(\Omega) = \{p \in L^2(\Omega) : \int_\Omega p dx = 0\}$.

Note that if $y_0 \in \mathbf{W}(\Omega)$ then the regularity of the state variable is limited to $L^2[0,T;\mathbf{V}(\Omega)] \cap H^1[0,T;\mathbf{V}(\Omega)^*]$, where $\mathbf{V}(\Omega)^*$ denotes the dual of $\mathbf{V}(\Omega)$. Furthermore, despite the fact that $y_t + \nabla p \in L^2[0,T;\mathbf{H}^{-1}(\Omega)]$ it is not known whether $p \in L^2[0,T;L^2_0(\Omega)]$ and $y_t \in L^2[0,T;\mathbf{H}^{-1}(\Omega)]$. As a consequence the pressure p satisfies (1.2) in a distributional sense. Hence, the assumption $p \in L^2[0,T;L^2_0(\Omega)]$ is the minimal one, to guarantee the decomposition between y_t and p and hence to validate a suitable weak formulation for rough initial data from the numerical analysis viewpoint. We emphasize that classical approaches within or without the discontinuous Galerkin framework for related parabolic control problems, typically fail to include the case of rough initial data since they demand more regularity with respect to the time-derivative y_t. As a result, error estimates for space-time approximations of the velocity tracking problem with rough initial data $y_0 \in \mathbf{W}(\Omega)$ have not been treated before, despite the fact that the case of rough initial data is of extreme importance within the context of controlling fluid flows (see e.g. [20]). Another important difference, between the error analysis of parabolic optimal control problems, and to the one of the velocity tracking problem is the presence of the incompressibility constraint which significantly complicates the analysis of discrete schemes.

In our work, we analyze a scheme which is based on a discontinuous time-stepping framework, which is suitable for problems without regular enough solutions, and we prove an estimate of symmetric type. The analysis showcases the favorable behavior of such schemes since it allows a unified treatment for a broad category of schemes for optimal control problems for the evolutionary Stokes equations. The key ingredient, that distinguishes our analysis, is the definition of a generalized divergence free space-time projection which exhibits best approximation properties in $L^2[0,T;\mathbf{L}^2(\Omega)]$, but is also applicable when $y_t \in L^2[0,T;\mathbf{H}^{-1}(\Omega)]$ only. Thus, constructing a global space-time projection, and using an appropriate duality argument, we obtain a rate of $\mathcal{O}(h)$ for the $L^2[0,T;\mathbf{L}^2(\Omega)]$ norm, when $\tau \leq Ch^2$. We note that we use a direct fully-discretized approach, instead of the typical two step approach which also includes a semi-discretized (in time) optimal control problem as an intermediate step. Similarly, in case of bounded controls, we demonstrate the applicability of our estimates within the variational discretization concept of [26]. This approach allows to overcome the lack of the enhanced regularity resulting from a “boot-strap” argument for the control and state variable.

To our best knowledge our estimates are new, and optimal in terms of the prescribed regularity of the solutions and the presence of the incompressibility constraint.

1.1. Related results. Several results regarding the analysis of related control problems were presented in [1, 3, 16, 20, 27, 28, 30, 41, 45] (see also references within), where various aspects, including first and second order necessary conditions are developed and analyzed. To the contrary the literature regarding numerical analysis for optimal control problems related to evolutionary Navier-Stokes equations is very limited and concentrated to the lowest order (in time) scheme with initial data (at least) in $\mathbf{V}(\Omega)$. In [21, 22, 23] convergence of a gradient algorithm is proven, in case of distributed controls, of bounded distributed controls, and Dirichlet boundary controls. Error estimates for the semi-discrete (in space) discretization are derived in [15] in case of distributed controls without control constraints by using a variational discretization approach, while in [14] fully-discrete error estimates for the implicit Euler scheme are presented for the velocity tracking problem (without control constraints) for the homogeneous Stokes equations using the variational discretization approach, for smooth data and for smooth solutions.

Recently, a-priori error estimates for the velocity tracking problem for Navier-Stokes flows with control constraints were analyzed in the works of [4, 5] with initial data belonging to $\mathbf{V}(\Omega)$. The lowest order (piecewise constants) discontinuous Galerkin scheme in time, combined with conforming elements in space for the velocity and the pressure was analyzed, and estimates for the state, adjoint, and control variables were derived for three separate choices of control discretization (piecewise constants, linears, and the variational discretization). Our work, is motivated by the results of [4, 5] and it can be viewed as an attempt to extend these results to include the cases of rough data, and high order schemes via the derivation of a symmetric estimate.

Other results concerning discontinuous time-stepping approaches are mainly related to distributed controls for linear and semilinear parabolic pdes. In case of distributed optimal control problems for the heat equation, a-priori error estimates for discontinuous time stepping schemes were previously established in [35, 36], with and without control constraints, as well as in [6, 7] in case of distributed optimal control problems without control constraints for general parabolic equations with time-dependent coefficients in the elliptic part of the operator. In [35, 36] optimal a-priori estimates are presented in $L^2[0,T;L^2(\Omega)]$ norm for the control, state and adjoint variables, for $H^1_0(\Omega)$ initial data using an auxiliary semi-discrete (in time) optimal control problem before proceeding to the fully-discrete problem. To the contrary, in [6, 7] symmetric estimates in the natural energy norm are developed, under low regularity on the data, using fully-discrete projection techniques. Recently in [9], a Robin boundary control problem related to the heat equation with rough initial data has been studied.

Estimates related to distributed optimal control problems for semi-linear parabolic pdes with control constraints are developed in the work of [38], for $H^1_0(\Omega) \cap L^\infty(\Omega)$ initial data, and in [8] for estimates of symmetric type for problems without control constraints. We also mention several related works [2, 39, 40] regarding parabolic optimal control problems with and and without control constraints which involve high order discrete schemes.

Various results regarding the analysis of optimal control problems can be found in [20, 31, 32, 37, 44] (see also references within). For general results related to discontinuous Galerkin methods for parabolic pdes (without applying controls) we refer the reader to [43] (see also references therein). A posteriori estimates and related adaptivity issues within the discontinuous Galerkin framework for optimal control problems were also explored in the works of [33, 34] (see also references within).

2. Background.

2.1. Notation.

We use the standard notation for the Sobolev spaces $H^s(\Omega)$ and their vector valued counterparts $\mathbf{H}^s(\Omega)$ with $s \in \mathbb{R}$ with norms denoted by $\|\cdot\|_{H^s(\Omega)}$ and $\|\cdot\|_{\mathbf{H}^s(\Omega)}$ respectively. Furthermore, let $H^1_0(\Omega) = \{v \in H^1(\Omega) : v|_\Gamma = 0\}$, $\mathbf{H}^1_0 = \{v \in \mathbf{H}^1(\Omega) : v|_\Gamma = 0\}$. We also denote by $(.,.)$ the standard $\mathbf{L}^2(\Omega)$ inner product, by $\mathbf{H}^{-1}(\Omega)$ the dual of $\mathbf{H}^1_0(\Omega)$, and their duality pairing by $\langle\cdot,\cdot\rangle_{\mathbf{H}^{-1}(\Omega),\mathbf{H}^1(\Omega)} \equiv \langle\cdot,\cdot\rangle$. We will also denote the corresponding divergence free spaces by $\mathbf{V}(\Omega) = \{v \in \mathbf{H}^1_0(\Omega) : \operatorname{div} v = 0\}$, $\mathbf{W}(\Omega) = \{v \in \mathbf{L}^2(\Omega) : \operatorname{div} v = 0, v \cdot \mathbf{n} = 0\}$, endowed with $\mathbf{H}^1(\Omega)$ and $\mathbf{L}^2(\Omega)$ norms respectively, and by $\mathbf{V}(\Omega)^*$ the dual of $\mathbf{V}(\Omega)$. Finally, for the pressure we will also need the space, $L^2_0(\Omega) = \{p \in L^2(\Omega) : \int_\Omega p = 0\}$ endowed with the standard $L^2(\Omega)$ norm. 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. [17, Chapter 5]).

We will frequently use the space $W(0,T) := L^\infty[0,T;\mathbf{W}(\Omega)] \cap L^2[0,T;\mathbf{V}(\Omega)]$ endowed with the norm $\|v\|^2_{W(0,T)} \equiv \|v\|^2_{L^\infty[0,T;\mathbf{L}^2(\Omega)]} + \|v\|^2_{L^2[0,T;\mathbf{H}^1(\Omega)]}$. For any $\gamma \geq 0$, we also define the space $H^\gamma[a,b;X]$ in a standard way (see e.g. [17, Chapter 5]). The bilinear form associated to our operator is given by

$$a(y,v) = \nu \int_\Omega \nabla y \nabla v dx, \ \forall y, v \in \mathbf{H}^1_0(\Omega),$$

and satisfies the standard coercivity and continuity conditions

$$a(y,y) \geq \nu \|\nabla y\|^2_{\mathbf{L}^2(\Omega)}, \ \alpha(y,v) \leq C\nu \|y\|_{\mathbf{H}^1(\Omega)} \|v\|_{\mathbf{H}^1(\Omega)} \ \forall y, v \in \mathbf{H}^1_0(\Omega).$$

Finally the bilinear form associated to the pressure is given by

$$b(v,q) = \int_\Omega -q\nabla.v dx, \qquad \forall v \in \mathbf{H}^1_0(\Omega), q \in L^2(\Omega),$$

which satisfies the standard continuity and inf-sup conditions (see e.g [18, 42]), i.e.,

$$b(v,q) \leq C\|v\|_{\mathbf{H}^1(\Omega)}\|q\|_{L^2(\Omega)}, \quad \text{and} \inf_{q \in L^2_0(\Omega)} \sup_{v \in \mathbf{H}^1_0(\Omega)} \frac{b(v,q)}{\|v\|_{H^1(\Omega)}\|q\|_{L^2(\Omega)}} \geq c > 0,$$

respectively.

2.2. The continuous control problem.

A weak formulation of (1.2), suitable for the case of rough initial data, is defined by using divergence-free test functions and can be written as follows: Given $f \in L^2[0,T;\mathbf{V}(\Omega)^*]$, $g \in L^2[0,T;\mathbf{L}^2(\Omega)]$, and $y_0 \in \mathbf{W}(\Omega)$ we seek $y \in L^2[0,T;\mathbf{V}(\Omega)] \cap H^1[0,T;\mathbf{V}(\Omega)^*]$ such that for a.e. $t \in (0,T]$,

$$\langle y_t, v\rangle + a(y,v) = \langle f, v\rangle + (g,v) \ \forall v \in \mathbf{V}(\Omega), \quad (y(0),v) = (y_0,v) \ \forall v \in \mathbf{W}(\Omega). \tag{2.1}$$

To the contrary, from the numerical analysis viewpoint, a desirable weak formulation suitable for the analysis of dG schemes, is to seek $y \in L^\infty[0,T;\mathbf{L}^2(\Omega)] \cap L^2[0,T;\mathbf{H}^1(\Omega)]$, and $p \in L^2[0,T;L^2_0(\Omega)]$ such that for all $v \in L^2[0,T;\mathbf{H}^1(\Omega)] \cap H^1[0,T;\mathbf{H}^{-1}(\Omega)]$, and for all $q \in L^2[0,T;L^2_0(\Omega)]$,

$$\begin{cases} (y(T),v(T)) + \displaystyle\int_0^T \left(-\langle y, v_t\rangle + a(y,v) + b(v,p)\right) dt \\ \qquad = (y_0, v(0)) + \displaystyle\int_0^T \left(\langle f,v\rangle + (g,v)\right) dt, \\ \displaystyle\int_0^T b(y,q) dt = 0. \end{cases} \tag{2.2}$$

Some comments regarding the existence and uniqueness of weak solutions of the evolutionary Stokes and the equivalence of formulations (2.1) and (2.2) follow.

Remark 2.1. *Recall that standard regularity theorems in [12, 42] show that if $f, g \in L^2[0,T;\mathbf{W}(\Omega)]$ and $y_0 \in \mathbf{V}(\Omega)$ then the solution (y,p) of equations (2.1) satisfies*

$$(y,p) \in L^2[0,T;\mathbf{H}^2(\Omega) \cap \mathbf{V}(\Omega)] \cap H^1[0,T;\mathbf{W}(\Omega)] \times L^2[0,T;H^1(\Omega) \cap L^2_0(\Omega)].$$

In this case weak formulations (2.1), and (2.2), are essentially equivalent. If the data $f \in L^2[0,T;\mathbf{V}^*(\Omega)], y_0 \in \mathbf{W}(\Omega)$ then there exists a unique weak solution that satisfies $y \in L^2[0,T;\mathbf{V}(\Omega)] \cap H^{-1}[0,T;\mathbf{V}^*(\Omega)]$, while the pressure p satisfies (1.2) in a distributional sense, and $y_t + \nabla p \in L^2[0,T;\mathbf{H}^{-1}(\Omega)]$. In the above case, we note that it is not evident whether the pressure belongs to $L^2[0,T;L^2_0(\Omega)]$ under minimal regularity assumptions (see e.g. [12, 42]), and hence formulation (2.2) is not necessarily valid, unless the existence of a pressure $p \in L^2[0,T;L^2_0(\Omega)]$ is assumed.

The control to state mapping $G : L^2[0,T;\mathbf{L}^2(\Omega))] \to W(0,T)$, which associates to each control g the state $G(g) = y_g \equiv y(g)$ via (2.1) is well defined, and continuous. Furthermore, we note that if more regularity is available to data, i.e., if $y_0 \in \mathbf{V}(\Omega)$, and $f \in L^2[0,T;\mathbf{L}^2(\Omega)]$, then $y(g) \in L^2[0,T;\mathbf{H}^2(\Omega) \cap \mathbf{V}(\Omega)] \cap H^1[0,T;\mathbf{L}^2(\Omega)]$ and $p \in L^2[0,T;H^1(\Omega) \cap L^2_0(\Omega)]$. Hence, the cost functional is frequently denoted by its reduced form, $J(y,g) \equiv J(y(g)) \equiv J(g) : L^2[0,T;\mathbf{L}^2(\Omega)] \to \mathbb{R}$ where $J(g) \equiv \frac{1}{2}\int_0^T \|y_g - y_d\|^2_{\mathbf{L}^2(\Omega)} dt + \frac{\alpha}{2}\int_0^T \|g\|^2_{\mathbf{L}^2(\Omega)} dt$, and $y_g \equiv y(g)$ is defined by (2.1).

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

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

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

The main result concerning the existence of an optimal solution follows directly from the setting of our problem (see for e.g. [44]), since $\mathcal{A}_{ad} \neq \emptyset$ (note that $(y(0),0) \in W(0,T) \times \mathcal{A}_{ad}$ for instance, without loss of generality).

THEOREM 2.3. *Let* $y_0 \in \mathbf{W}(\Omega)$, $f \in L^2[0,T;\mathbf{V}(\Omega)^*]$, $y_d \in L^2[0,T;\mathbf{W}(\Omega)]$ *be given data. Then, the optimal control problem has unique solution* $(\bar{y}(\bar{g}),\bar{g}) \in W(0,T) \times L^2[0,T;\mathbf{L}^2(\Omega)]$*. In addition there exists a pressure* $\bar{p}$ *that satisfies (1.2) in a distributional sense. If in addition,* $y_0 \in \mathbf{V}(\Omega)$, $f \in L^2[0,T;\mathbf{W}(\Omega)]$*, then* $\bar{p} \in L^2[0,T;H^1(\Omega)\cap L^2_0(\Omega)]$ *and the pair* $(\bar{y},\bar{p})$ *also satisfies (2.2).*

2.3. The optimality system. An optimality system of equations can be derived by using standard techniques; see for instance [20, 44] or [4, Section 3]. We first state the basic differentiability property of the cost functional.

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

$$J^{'}(g)u = \int_0^T (\mu(g) + \alpha g, u) dt,$$

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

$$\int_0^T \left(\langle \mu_g, v_t\rangle + a(\mu_g, v)\right) dt = -(\mu_g(0), v(0)) + \int_0^T (y_g - y_d, v) dt, \qquad (2.3)$$

where $\mu_g(T) = 0$*, and* $y_g = y(g)$ *satisfies (2.1). In addition,* $(\mu_g)_t \in L^2[0,T;\mathbf{L}^2(\Omega)]$*, and there exists pressure* $\phi \in L^2[0,T;H^1(\Omega)\cap L^2_0(\Omega)]$ *such that the backwards in time Stokes equation is satisfied in the sense of weak formulation (2.2).*

Therefore the optimality system which consists of the state and adjoint equations, and the optimality condition takes the following 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* $v \in L^2[0,T;\mathbf{V}(\Omega)] \cap H^1[0,T;\mathbf{V}(\Omega)^*]$,

$$(\bar{y}(T), v(T)) + \int_0^T \left(-\langle \bar{y}, v_t\rangle + a(\bar{y}, v)\right) dt = (\bar{y}_0, v(0)) + \int_0^T \left(\langle f, v\rangle + (\bar{g}, v)\right) dt, \quad (2.4)$$

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

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

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

In addition, $\bar{y}_t \in L^2[0,T;\mathbf{V}(\Omega)^*]$, $\bar{\mu} \in L^2[0,T;\mathbf{H}^2(\Omega)\cap\mathbf{V}(\Omega)] \cap H^1[0,T;\mathbf{L}^2(\Omega)]$*, and (2.7), is equivalent to* $\bar{g}(t,x) = Proj_{[g_a,g_b]}\left(-\frac{1}{\alpha}\bar{\mu}(t,x)\right)$ *for a.e.* $(t,x) \in (0,T]\times\Omega$*. In addition, there exists a pressure* $\bar{\phi} \in L^2[0,T;H^1(\Omega)\cap L^2_0(\Omega)]$ *associated to the adjoint variable* $\bar{\mu}$ *satisfying the backwards' in time evolutionary Stokes, in the sense of formulation (2.2).*

Proof. The derivation of the optimality system is standard (see e.g. [44]). For the enhanced regularity on $\bar{\mu}$, we note that $\bar{\mu}(T) = 0$, and $\bar{y} - y_d \in L^2[0,T;\mathbf{W}(\Omega)]$ and hence (2.5) implies that to get that $\bar{\mu} \in L^2[0,T;\mathbf{H}^2(\Omega)\cap\mathbf{V}(\Omega)]\cap H^1[0,T;\mathbf{L}^2(\Omega)]$. For the regularity of the corresponding pressure $\bar{\phi}$ we refer to Remark 2.1. ∎

Remark 2.6. *For the numerical analysis it is preferable that (2.4) and (2.5) take the following form: For all* $v \in L^2[0,T;\mathbf{H}_0^1(\Omega)] \cap H^1[0,T;\mathbf{H}^{-1}(\Omega)]$, $q \in L^2[0,T;L_0^2(\Omega)]$, *we seek* $\bar{y}, \bar{\mu} \in L^\infty[0,T;\mathbf{L}^2(\Omega)] \cap L^2[0,T;\mathbf{H}_0^1(\Omega)]$ *such that,*

$$\begin{cases} (\bar{y}(T), v(T)) + \int_0^T (-\langle \bar{y}, v_t\rangle + a(\bar{y}, v) + b(v, \bar{p}))\, dt \\ \qquad = (\bar{y}_0, v(0)) + \int_0^T (\langle f, v\rangle + (\bar{g}, v))\, dt, \\ \int_0^T b(\bar{y}, q) dt = 0, \end{cases} \tag{2.8}$$

$$\begin{cases} \int_0^T \left(\langle \bar{\mu}, v_t\rangle + a(v, \bar{\mu}) + b(v, \bar{\phi})\right) dt = -(\bar{\mu}(0), v(0)) + \int_0^T (\bar{y} - y_d, v) dt, \\ \int_0^T b(\bar{\mu}, q) dt = 0. \end{cases} \tag{2.9}$$

It is clear that $\bar{\mu} \in L^2[0,T;\mathbf{H}^2(\Omega) \cap \mathbf{V}(\Omega)] \cap H^1[0,T;\mathbf{L}^2(\Omega)]$ *and* $\bar{\phi} \in L^2[0,T;H^1(\Omega) \cap L_0^2(\Omega)]$, *and hence equation (2.9) is equivalent to (2.5). In addition, when* $y_0 \in \mathbf{V}(\Omega)$, *and* $f \in L^2[0,T;\mathbf{W}(\Omega)]$, *then (2.8) is also equivalent to (2.4). However, we point out that for rough data,* $y_0 \in \mathbf{W}(\Omega)$ *the regularity of the corresponding pressure* $\bar{p} \in L^2[0,T;L_0^2(\Omega)]$ *has to be assumed. In any case, for the numerical analysis we will consider the optimality system (2.8)-(2.9) and one of the optimality conditions (2.6) or (2.7). When control constraints are not present then a bootstrap argument can be applied in order to improve the regularity of* $\bar{g}, \bar{\mu}, \bar{y}$ *in a straightforward manner, provided natural regularity assumptions on the given data, the smoothness of the domain, and appropriate compatibility conditions (see for instance [12]). We refer the reader to [4, 5] for enhanced related regularity results when control constraints are involved. Let* $\Omega_T = \Omega \times (0,T]$. *If the boundary is of class* C^2, *then* $\bar{g} \in \mathbf{H}^1(\Omega_T) \cap C[0,T;\mathbf{H}_0^1(\Omega)] \cap L^2[0,T;\mathbf{W}^{1,p}]$ *when* $y_0 \in \mathbf{V}(\Omega)$ *and* $f \in L^2[0,T;\mathbf{L}^2(\Omega)]$.

3. The discrete optimal control problem.

3.1. Preliminaries. A family of triangulations (denoted by $\{\mathcal{T}_h\}_{h>0}$) of Ω, is defined in the standard way, (see e.g. [11]). We assume that to every element $T \in \mathcal{T}_h$, 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 are assigned, and the associated size of the mesh is denoted by $h \equiv \max_{T\in\mathcal{T}_h} h_T$. 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) – Define $\bar{\Omega}_h = \cup_{T\in\mathcal{T}_h} T$ and denote by Ω_h, and Γ_h its interior and boundary respectively. We also assume that the boundary vertices of $\mathcal{T}_h$ are points of Γ.

On the mesh $\mathcal{T}_h$ we consider two finite dimensional spaces $\mathbf{Y}_h \subset \mathbf{H}_0^1(\Omega)$ and $Q_h \subset L_0^2(\Omega)$ constructed by piecewise polynomials in Ω_h, and vanishing in $\Omega \setminus \Omega_h$. We note that under the above structural assumptions, if Ω is convex, then Ω_h is convex, and $|\Omega \setminus \Omega_h| \leq Ch^2$. The above assumptions are enough in order to obtain best-approximation estimates for the cases where the initial data belong to $\mathbf{W}(\Omega)$ or $\mathbf{V}(\Omega)$.

Alternatively, the assumption on the domain to be convex and polygonal in $\mathbb{R}^2$ is also enough to guarantee the $\mathbf{H}^2$ regularity of the steady Navier-Stokes. For convex and polyhedral domains in $\mathbb{R}^3$, the $\mathbf{H}^2$ regularity is also believed to hold. We note that it is not known if convexity is enough to guarantee the $\mathbf{H}^2$ elliptic regularity of the stationary Stokes equations in $\mathbb{R}^3$ (see for instance [25]). Furthermore, more regularity on the boundary Γ (say C^2), implies $\mathbf{H}^2$ regularity of the stationary Stokes, while when dealing with higher order schemes, we emphasize that additional smoothness on Γ should be assumed (see for instance [42, Remark 3.7], or [13, pp 34]), together with compatibility conditions in order to guarantee the appropriate regularity of the solutions. For example let $\Omega_T = \Omega \times [0,T]$, where $\Omega \subset \mathbb{R}^3$ is an open bounded domain, with boundary of type C^{2r+2}, $r = \max\{k+2, 2\}$. If data $f \in \mathbf{H}^{2k,k}(\Omega_T)$, $y_0 \in \mathbf{H}^{2k+1} \cap \mathbf{W}(\Omega)$ (together with appropriate compatibility conditions on the data that we omit) then $y \in \mathbf{H}^{2k+2,k+1}(\Omega_T) \cap C[0,T;\mathbf{V}(\Omega)]$, $\nabla p \in H^{2k,k}(\Omega_T)$. Here, we denote by $\mathbf{H}^{s_1,s_2}(\Omega_T)$ the space of all functions with all partial derivatives up to order $s_1 \geq 0$ in space, and up to order $s_2 \geq 0$ in time, bounded in $L^2(\Omega_T)$. Under the previous hypotheses, and for notational consistency, in various instances even if the domain is not polygonal (polyhedral), we will still denote $(.,.)_{\mathbf{L}^2(\Omega_h)}$ by $(.,.)_{\mathbf{L}^2(\Omega)}$, $\|.\|_{\mathbf{H}^s(\Omega_h)}$ by $\|.\|_{\mathbf{H}^s(\Omega)}$ etc, and we will assume that it can be approximated by a suitable polygonal (polyhedral) domain.

Standard approximation theory assumptions are assumed on spaces $\mathbf{Y}_h$ and Q_h. In particular, for any $v \in \mathbf{H}^{l+1}(\Omega) \cap \mathbf{H}^1_0(\Omega)$, there exists an integer $\ell \geq 1$, and a constant $C > 0$ (independent of h) such that:

$$\inf_{v_h \in \mathbf{Y}_h} \|v - v_h\|_{\mathbf{H}^s(\Omega)} \leq C h^{l+1-s} \|v\|_{\mathbf{H}^{l+1}(\Omega)}, \quad \text{for } 0 \leq l \leq \ell \text{ and } s = -1, 0, 1. \tag{3.1}$$

Also for any $q \in H^l(\Omega) \cap L^2_0(\Omega)$, then

$$\inf_{q_h \in Q_h} \|q - q_h\|_{L^2(\Omega)} \leq C h^l \|q\|_{H^l(\Omega)}, \quad \text{for } 0 \leq l \leq \ell. \tag{3.2}$$

In addition, the spaces $\mathbf{Y_h}$ and Q_h must satisfy the inf-sup condition, i.e., there exists $C > 0$ (independent of h) such that

$$\inf_{q_h \in Q_h} \sup_{v_h \in \mathbf{Y}_h} \frac{b(v_h, q_h)}{\|v_h\|_{\mathbf{H}^1(\Omega_h)} \|q_h\|_{L^2(\Omega_h)}} > C. \tag{3.3}$$

We also consider the discrete divergence free analog of $\mathbf{Y}_h$ denoted by

$$\mathbf{U}_h = \{v_h \in \mathbf{Y}_h : b(v_h, q_h) = 0 \quad \forall\, q_h \in Q_h\}.$$

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 $0 < \theta \leq 1$ such that $\min_{n=1,..,N}(t^n - t^{n-1}) \geq \theta \max_{n=1,...,N}(t^n - t^{n-1})$. We denote by $\tau^n = t^n - t^{n-1}$, $\tau = \max_{n=1,...,N} \tau^n$ and by $\mathcal{P}_k[t^{n-1}, t^n; \mathbf{Y}_h]$, $\mathcal{P}_k[t^{n-1}, t^n; \mathbf{U}_h]$, and $\mathcal{P}_k[t^{n-1}, t^n; Q_h]$ the spaces of polynomials of degree k or less having values in $\mathbf{Y}_h$, $\mathbf{U}_h$ and Q_h respectively. We seek approximate solutions for the velocity and the pressure who belong to the spaces:

$$\begin{aligned}
\mathcal{Y}_h &= \{y_h \in L^2[0,T;\mathbf{H}^1_0(\Omega)] : y_h|_{(t^{n-1},t^n]} \in \mathcal{P}_k[t^{n-1}, t^n; \mathbf{Y}_h]\}, \\
\mathcal{U}_h &= \{y_h \in L^2[0,T;\mathbf{H}^1_0(\Omega)] : y_h|_{(t^{n-1},t^n]} \in \mathcal{P}_k[t^{n-1}, t^n; \mathbf{U}_h]\}, \\
\mathcal{Q}_h &= \{y_h \in L^2[0,T;L^2_0(\Omega)] : y_h|_{(t^{n-1},t^n]} \in \mathcal{P}_k[t^{n-1}, t^n; Q_h]\}.
\end{aligned}$$

The following remark highlights why the use of same degree of polynomials with respect to time is the natural choice for the discretization (in time) of the pressure.

Remark 3.1. *It is obvious that the analogue for the discrete divergence free subspace of* $\mathcal{P}_k[t^{n-1},t^n;\mathbf{Y}_h]$ *is,* $\mathcal{Z}_h^n = \{v_h \in \mathcal{P}_k[t^{n-1},t^n;\mathbf{Y}_h] : \int_{t^{n-1}}^{t^n} b(v_h,q_h) = 0, \quad \forall\, q_h \in \mathcal{P}_k[t^{n-1},t^n;Q_h]\}$*. Then, [10, Lemma 2.3] states that* $\mathcal{Z}_h^n \equiv \mathcal{P}_k[t^{n-1},t^n;\mathbf{U}_h]$*. Therefore, we may write that*

$$
\begin{aligned}
\mathcal{Z}_h &\equiv \{v_h \in \mathcal{Y}_h : \int_0^T b(v_h,q_h) = 0, \quad \forall\, q_h \in Q_h\} \\
&= \{v_h \in \mathcal{Y}_h : v_h|_{(t^{n-1},t^n]} \in \mathcal{Z}_h^n\} \\
&= \{v_h \in \mathcal{Y}_h : v_h|_{(t^{n-1},t^n]} \in \mathcal{P}_k[t^{n-1},t^n;\mathbf{U}_h]\} \equiv \mathcal{U}_h.
\end{aligned}
$$

We refer the reader to [10, Section 2] for more details.

By convention, the functions of $\mathcal{U}_h$ are left continuous with right limits. Thus, we will write y^n for $y(t^n) \equiv y(t^n_-)$, y^{n-1}_+ for $y(t^{n-1}_+)$, y^n_h for $y_h(t^n) \equiv y_h(t^n_-)$, and y^n_{h+} for $y(t^n_+)$, while the jump at t^n, is denoted by $[y_h^n] = y^n_{h+} - y^n_h$. In the above definitions, we have also used the following notational abbreviation, $y_{h,\tau} \equiv y_h$, $\mathcal{Y}_{h,\tau} \equiv \mathcal{Y}_h$, $\mathcal{U}_{h,\tau} \equiv \mathcal{U}_h$ etc. This is due to the fact that the time-discretization parameter τ can be chosen independent of h.

We emphasize that schemes of arbitrary order in time and space will be included in our proofs. However, the limited regularity will be acting as a barrier in terms of developing estimates of high order, at least in presence of control constraints. The case of the lowest order scheme, in time and space, has been treated in detail in the recent works of [4] and [5] for the velocity tracking problem of Navier-Stokes flows, with control constraints, when data $y_0 \in \mathbf{V}(\Omega)$, $f \in L^2[0,T;\mathbf{L}^2(\Omega)]$.

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.
Case 1: Unconstrained Controls: We employ the natural space-time discretization which allows the presence of discontinuities (in time). In particular, we define by $\mathcal{G}_h \equiv \mathcal{Y}_h$. Only $L^2[0,T;\mathbf{L}^2(\Omega)]$ regularity will be needed in the error estimates.
Case 2: Constrained Controls: Analogously to the previous case, we employ the variational discretization concept (see e.g. [26]) 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., $\mathcal{G}_h \equiv L^2[0,T;\mathbf{L}^2(\Omega)]$.

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

$$
\begin{aligned}
&(y_h^n, v_h^n) + \int_{t^{n-1}}^{t^n} \Big(-\langle y_h, v_{ht}\rangle + a(y_h,v_h)\Big) dt \qquad (3.4)\\
&= (y_h^{n-1}, v_{h+}^{n-1}) + \int_{t^{n-1}}^{t^n} \Big(\langle f, v_h\rangle + (g, v_h)\Big) dt.
\end{aligned}
$$

Here y_h^0 denotes a suitable approximation of the initial data y_0. Stability estimates at partition points as well as in $L^2[0,T;\mathbf{H}^1(\Omega)]$ norm easily follow by setting $v_h = y_h$ into (3.4). For the estimate at arbitrary time-points, we refer the reader to [10, Appendix

A]. Thus, stability estimates imply that the control to fully-discrete state mapping $G_h : L^2[0,T;\mathbf{L}^2(\Omega)] \to \mathcal{U}_h$, is well defined, and continuous. Due to Remark 3.1, we will primarily focus on the equivalent weak formulation: We seek $(y_h, p_h) \in \mathcal{Y}_h \times \mathcal{Q}_h$ such that the following formulation is satisfied: For $n = 1, ..., N$, and for all $v_h \in \mathcal{P}_k[t^{n-1}, t^n; \mathbf{Y}_h]$, $q_h \in L^2[t^{n-1}, t^n; Q_h]$,

$$\begin{cases} (y_h^n, v_h^n) + \displaystyle\int_{t^{n-1}}^{t^n} \Big(-\langle y_h, v_{ht}\rangle + a(y_h, v_h) + b(v_h, p_h)\Big) dt \\ \qquad = (y_h^{n-1}, v_{h+}^{n-1}) + \displaystyle\int_{t^{n-1}}^{t^n} \Big(\langle f, v_h\rangle + (g, v_h)\Big) dt, \\ \displaystyle\int_{t^{n-1}}^{t^n} b(y_h, q_h) dt = 0. \end{cases} \tag{3.5}$$

The key advantage of the above formulation is that it clearly justifies the computation of the pressure within the discontinuous Galerkin (in time) framework (see Remark 3.1), and showcases the importance of the inf-sup condition when constructing finite element schemes (see e.g. [18, 25, 42]) for the Stokes and Navier-Stokes equations. To this end, we note that the algorithmic interpretation of the resulting schemes typically uses the discrete formulation (3.5) while the pressure is frequently computed through penalization approaches. In any case, and similar to the elliptic case (see for instance [18, Chapter II]) the spatial approximation properties of spaces $\mathcal{U}_h$ are obtained through spatial approximation properties of $\mathcal{Y}_h$ and $\mathcal{Q}_h$ and the inf-sup condition, while the temporal approximation properties within the discontinuous Galerkin framework are established by the local (in time) L^2 projection techniques into polynomial spaces (see for instance [43, Chapter 14] for parabolic problems and / or [10, Sections 2, and 3] for the Stokes equations, and the subsequent Lemma 4.3, for the Stokes equations under low regularity assumptions).

The fully-discrete optimal control problem can be defined as follows:

Definition 3.2. *Let $f \in L^2[0,T;\mathbf{V}(\Omega)^*]$, $y_0 \in \mathbf{W}(\Omega)$, $y_d \in L^2[0,T;\mathbf{W}(\Omega)]$ 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 $J(y_h, g_h) \equiv \frac{1}{2}\int_0^T \|y_h - y_d\|^2_{\mathbf{L}^2(\Omega)} dt + \frac{\alpha}{2}\int_0^T \|g_h\|^2_{\mathbf{L}^2(\Omega)} dt$. Here the pair $(y_h, g_h) \in \mathcal{Y}_h \times \mathcal{A}_{ad}^d$ and the associated pressure $p_h \in \mathcal{Q}_h$ satisfy (3.5). Then, the pair $(\bar{y}_h, \bar{g}_h) \in \mathcal{Y}_h \times \mathcal{A}_{ad}^d$, is said to be an optimal solution if $J(\bar{y}_h, \bar{g}_h) \leq J(w_h, u_h)$ $\forall (w_h, u_h) \in \mathcal{Y}_h \times \mathcal{A}_{ad}^d$.*

The existence and uniqueness of the discrete optimal control problem can be proved by standard techniques. 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 4]. The estimate highlights the fact that the natural discrete energy norm for the state variable associated to discontinuous time-stepping schemes is $\|.\|_{W(0,T)} = \|.\|_{L^\infty[0,T;\mathbf{L}^2(\Omega)]} + \|.\|_{L^2[0,T;\mathbf{H}^1(\Omega)]}$.

Lemma 3.3. *Suppose that $y_0 \in \mathbf{W}(\Omega)$, $f \in L^2[0,T;\mathbf{V}(\Omega)^*]$. If $(\bar{y}_h, \bar{g}_h) \in \mathcal{Y}_h \times \mathcal{A}_{ad}^d$ denotes the solution pair of the discrete optimal control problem, then there exists constant $C > 0$ depending on $1/\nu$, C_k and Ω but not on α, τ, h, such that,*

$$\|\bar{y}_h\|^2_{L^\infty[0,T;\mathbf{L}^2(\Omega)]} \leq C(1/\alpha)\big(\|y_0\|^2_{\mathbf{L}^2(\Omega)} + \|f\|^2_{L^2[0,T;\mathbf{V}(\Omega)^*]}\big).$$

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 $g \in L^2[0,T;\mathbf{L}^2(\Omega)]$. Here, we note that $y_h(g)$ is defined by (3.5).

LEMMA 3.4. *The cost functional $J_h : L^2[0,T;\mathbf{L}^2(\Omega)] \to \mathbb{R}$, with $J_h(g) := J(y_h(g), g)$, is well defined, differentiable, and for every $g, u \in L^2[0,T;\mathbf{L}^2(\Omega)]$,*

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

where $\mu_h(g) \equiv \mu_{g,h} \in \mathcal{Y}_h$ and its associated pressure $\phi_{g,h} \in \mathcal{Q}_h$ is the unique solution pair of following problem: For all $n = 1, ..., N$ and for all $v_h \in P_k[t^{n-1}, t^n; \mathbf{Y}_h]$, $q_h \in L^2[t^{n-1}, t^n; Q_h]$,

$$\begin{cases} -(\mu^n_{g,h+}, v^n_h) + \displaystyle\int_{t^{n-1}}^{t^n} \Big(\langle \mu_{g,h}, v_{ht}\rangle + a(v_h, \mu_{g,h}) + b(v_h, \phi_{g,h})\Big)dt \\ = -(\mu^{n-1}_{g,h+}, v^{n-1}_{h+}) + \displaystyle\int_{t^{n-1}}^{t^n} \langle y_{g,h} - y_d, v_h\rangle dt, \\ \displaystyle\int_{t^{n-1}}^{t^n} b(\mu_{g,h}, q_h)dt = 0, \end{cases}$$

where $\mu^N_{g,+} = 0$. Here, $y_{g,h} \equiv y_h(g)$ is the solution of (3.5).

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

LEMMA 3.5. *Let $(\bar{y}_h(\bar{g}_h), \bar{g}_h) \equiv (\bar{y}_h, \bar{g}_h) \in \mathcal{Y}_h \times \mathcal{A}^d_{ad}$ denote the unique optimal pair of Definition 3.2, and let $\bar{p}_h \in \mathcal{Q}_h$ the associated pressure. Then, there exists an adjoint $\bar{\mu}_h \in \mathcal{Y}_h$ and an associated pressure $\bar{\phi}_h \in \mathcal{Q}_h$ satisfying, $\bar{\mu}^N_+ = 0$ such that for all $v_h \in \mathcal{P}_k[t^{n-1}, t^n; \mathbf{U}_h]$, $q_h \in L^2[t^{n-1}, t^n; Q_h]$, and for all $n = 1, ..., N$*

$$\begin{cases} (\bar{y}^n_h, v^n_h) + \displaystyle\int_{t^{n-1}}^{t^n} (-\langle \bar{y}_h, v_{ht}\rangle + a(\bar{y}_h, v_h) + b(v_h, \bar{p}_h))\, dt \\ \qquad = (\bar{y}^{n-1}_h, v^{n-1}_{h+}) + \displaystyle\int_{t^{n-1}}^{t^n} (\langle f, v_h\rangle + (\bar{g}_h, v_h))\, dt, \\ \displaystyle\int_{t^{n-1}}^{t^n} b(\bar{y}_h, q_h)dt = 0, \end{cases} \tag{3.6}$$

$$\begin{cases} -(\bar{\mu}^n_{h+}, v^n_h) + \displaystyle\int_{t^{n-1}}^{t^n} \big(\langle \bar{\mu}_h, v_{ht}\rangle + a(\bar{\mu}_h, v_h) + b(v_h, \bar{\phi}_h)\big)\, dt \\ \qquad = -(\bar{\mu}^{n-1}_{h+}, v^{n-1}_{h+}) + \displaystyle\int_{t^{n-1}}^{t^n} (\bar{y}_h - y_d, v_h)\, dt, \\ \displaystyle\int_{t^{n-1}}^{t^n} b(\bar{\mu}_h, q_h)dt = 0, \end{cases} \tag{3.7}$$

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

$$\text{1) Unconstrained Controls:} \int_0^T (\alpha \bar{g}_h + \bar{\mu}_h, u_h)dt = 0, \tag{3.8}$$

$$\text{2) Constrained Controls:} \int_0^T (\alpha \bar{g}_h + \bar{\mu}_h, u_h - \bar{g}_h)\, dt \geq 0. \tag{3.9}$$

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

LEMMA 3.6. *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 (3.6)-(3.7)-(3.8) or (3.9). Then,*

$$\|\bar{\mu}_h\|_{L^\infty[0,T;\mathbf{H}^1(\Omega)]} \le C\|\bar{y}_h - y_d\|_{L^2[0,T;\mathbf{L}^2(\Omega)]},$$

where C does not depend on α, τ, h but only on $1/\nu$, C_k, Ω. If in addition, $y_0 \in \mathbf{V}(\Omega)$, $f \in L^2[0,T;\mathbf{L}^2(\Omega)]$ then the solution $\bar{y}_h$ of (3.6) also satisfies,

$$\|\bar{y}_h\|_{L^\infty[0,T;\mathbf{H}^1(\Omega)]} \le C.$$

Proof. The proof is given for the forward in time evolutionary Stokes equations in [10, Theorem 4.10]. For the backwards in time problem, we simply note that $\bar{y}_h - y_d \in L^2[0,T;\mathbf{W}(\Omega)]$, and hence by a simple modification of the technique, we obtain the desired result. ∎

4. Symmetric error estimates. First, an auxiliary system which plays the role of a global space-time dG projection is defined. Throughout the remaining of our paper, we will work with weak formulations that assume the existence of a pressure $\bar{p} \in L^2[0,T;L^2_0(\Omega)]$ (and hence of $\bar{y}_t \in L^2[0,T;\mathbf{H}^{-1}(\Omega)]$). Hence, the continuous optimality system consists of the state and adjoint equations (2.8)-(2.9) and the optimality condition (2.6) or (2.7), and the discrete optimality system by (3.6)-(3.7) and (3.8) or (3.9).

4.1. The fully-discrete projection. Given data f, y_0, initial condition $w_h^0 = y_h^0 \equiv P_h y_0$, where $P_h y_0$ denotes the L^2 projection onto the discrete divergence free space $\mathbf{U}_h$, and $z_+^N = 0$, we seek (w_h, p_{1h}),$(z_h, \phi_{1h}) \in \mathcal{Y}_h \times \mathcal{Q}_h$ such that for $n = 1, ..., N$ and for all $v_h \in \mathcal{P}_k[t^{n-1}, t^n; \mathbf{Y}_h]$, $q_h \in \mathcal{P}_k[t^{n-1}, t^n; Q_h]$,

$$\begin{cases} (w_h^n, v_h^n) + \displaystyle\int_{t^{n-1}}^{t^n} \Big(-\langle w_h, v_{ht}\rangle + a(w_h, v_h) + b(v_h, p_{1h})\Big) dt \\ \qquad = (w_h^{n-1}, v_{h+}^{n-1}) + \displaystyle\int_{t^{n-1}}^{t^n} \Big(\langle f, v_h\rangle + (\bar{g}, v_h)\Big) dt, \\ \displaystyle\int_{t^{n-1}}^{t^n} b(w_h, q_h) dt = 0, \end{cases} \tag{4.1}$$

$$\begin{cases} -(z_{h+}^n, v_h^n) + \displaystyle\int_{t^{n-1}}^{t^n} \Big(\langle z_h, v_{ht}\rangle + a(z_h, v_h) + b(v_h, \phi_{1h})\Big) dt \\ \qquad = -(z_{h+}^{n-1}, v_{h+}^{n-1}) + \displaystyle\int_{t^{n-1}}^{t^n} (w_h - y_d, v_h) dt, \\ \displaystyle\int_{t^{n-1}}^{t^n} b(z_h, q_h) dt = 0. \end{cases} \tag{4.2}$$

The solutions $w_h, z_h \in \mathcal{Y}_h$ exist for any given data $f \in L^2[0,T;\mathbf{V}(\Omega)^*]$, $y_0 \in \mathbf{W}(\Omega)$, and $y_d \in L^2[0,T;\mathbf{L}^2(\Omega)]$. In particular, the stability estimates imply that $w_h, z_h \in W(0,T)$. In addition, due to the enhanced regularity of $w_h - y_d$, we also obtain the stability of z_h in $L^\infty[0,T;\mathbf{H}^1(\Omega)]$ norm.

The solutions of the auxiliary optimality system play the role of "global projections" onto $\mathcal{Y}_h$. 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 projection techniques into the auxiliary system. The following standard projection associated to discontinuous time-stepping methods for the Navier-Stokes equations (see e.g. [10, Definitions 4.1, 4.2]) is needed.

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

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

Here we have used the convention $(P_n^{loc} v)^n \equiv (P_n^{loc} v)(t^n)$ *and* $P_h : L^2(\Omega) \to \mathbf{U}_h$ *is the orthogonal projection operator onto discrete divergence free subspace* $\mathbf{U}_h$.

(2) The projection $P_h^{loc} : C[0, T; \mathbf{L}^2(\Omega)] \to \mathcal{U}_h$ *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 the coupling between the time-derivative and the pressure, we will also need the following generalized dG projection, which will be applicable when $\bar{p} \in L^2[0, T; L_0^2(\Omega)]$, $\bar{y}_t \in L^2[0, T; \mathbf{H}^{-1}(\Omega)]$. In particular, motivated by a similar construction for linear parabolic problems with rough initial data [9, Section 4], we construct a space-time generalized $\mathbf{L}^2$ divergence free projection, which combines the standard dG time stepping projection, and the generalized $\mathbf{L}^2$ projection $\Xi_h : \mathbf{H}^{-1}(\Omega) \to \mathbf{U}_h$. For various properties of Ξ_h see for instance [29, Section 2]. Recall that the definition of Ξ_h states that $\langle v - \Xi_h v, v_h \rangle = 0$, for all $v \in \mathbf{H}^{-1}(\Omega)$ and $v_h \in \mathbf{U}_h$. The projection is well defined in $\mathbf{H}^{-1}(\Omega)$, and coincides to P_h for $v \in \mathbf{L}^2(\Omega)$.

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

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

Here we also use the convention $(\Xi_n^{loc} v)^n \equiv (\Xi_n^{loc} v)(t^n)$, *and* $\Xi_h : \mathbf{H}^{-1}(\Omega) \to \mathbf{U}_h$ *is the generalized orthogonal projection operator onto* $\mathbf{U}_h$.

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

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

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

By definition, Ξ_h^{loc} coincides to P_h^{loc}, when $v \in L^2[0, T; \mathbf{L}^2(\Omega)]$ i.e., $P_h^{loc} v = \Xi_h^{loc} v$ when $v \in L^2[0, T; \mathbf{L}^2(\Omega)]$, and hence exhibits best approximation properties. However, we emphasize that is also applicable for $v \equiv y_t \in L^2[0, T; \mathbf{H}^{-1}(\Omega)]$. For the backwards in time problem a modification of the above projections (still denoted by P_n^{loc}, Ξ_n^{loc} respectively) is defined in a similar manner. 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 Ξ_h^{loc}, under minimal regularity assumptions, i.e., for $v \in L^2[0,T;\mathbf{V}(\Omega)] \cap H^1[0,T;\mathbf{H}^{-1}(\Omega)]$ for the lowest order scheme.

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

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

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

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

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

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

Proof. The first estimate is given in [10, Theorem 4.3, and Corollary 4.8]. For the second one, using [10, Theorem 4.3, Corollary 4.8], and standard approximation properties of P_h, we obtain for every $v \in L^2[t^{n-1}, t^n; \mathbf{H}^{l+1}(\Omega)]$, with $v^{(k+1)} \in L^2[t^{n-1}, t^n; \mathbf{L}^2(\Omega)]$, the following estimates:

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

where at the last estimate we have used an inverse estimate. Thus the second estimate is proved. The third estimate is standard, and we omit the proof.

The fourth estimate, follows by well known arguments after simple modifications to handle the divergence free nature of the projection. For completeness, we state the main arguments. For any $t \in (t^{n-1}, t^n]$, adding and subtracting appropriate terms, and using the definition of Ξ_h^{loc}, we obtain,

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

For the first term, we define $e(t) = v(t^n) - v(t)$, and note that $(1/2)\frac{d}{dt}\|e(t)\|^2_{\mathbf{L}^2(\Omega)} = \langle e_t, e\rangle = -\langle v_t(t), v(t^n) - v(t)\rangle$. Hence, integrating with respect to time in $(s, t^n]$, we obtain $(1/2)\big(\|e(t^n)\|^2_{\mathbf{L}^2(\Omega)} - \|e(s)\|^2_{\mathbf{L}^2(\Omega)}\big) = \int_s^{t^n} -\langle v_t(t), v(t^n) - v(t)\rangle dt$. Note that

$e(t^n) = 0$, and hence we obtain after integration by parts in time, $(1/2)\|e(s)\|^2_{\mathbf{L}^2(\Omega)} = -\langle v(s), v(t^n) - v(s)\rangle - \int_s^{t^n} \langle v_t(t), v(t)\rangle dt$. Thus, using Young's inequality, we obtain, $(1/4)\|e(s)\|^2_{\mathbf{L}^2(\Omega)} \leq \|v(s)\|^2_{\mathbf{L}^2(\Omega)} + \int_s^{t^n} \|v_t\|_{\mathbf{H}^{-1}(\Omega)}\|v\|_{\mathbf{H}^1(\Omega)} dt$. Using the embedding $L^2[s, t^n; \mathbf{H}^1_0(\Omega)] \cap H^1[s, t^n; \mathbf{H}^{-1}(\Omega)] \subset L^\infty[s, t^n; \mathbf{L}^2(\Omega)]$, Hölder's inequality, and integrating in time from t^{n-1} to t^n, we finally arrive to

$$(1/4)\int_{t^{n-1}}^{t^n} \|e(s)\|^2_{\mathbf{L}^2(\Omega)} ds \leq C\tau \int_{t^{n-1}}^{t^n} \big(\|v_t\|^2_{\mathbf{H}^{-1}(\Omega)} + \|v\|^2_{\mathbf{H}^1(\Omega)}\big) ds$$

which implies the desired estimate for the first term. The second term, can be proven similarly using triangle inequality, and the approximation property $\|v(t) - \Xi_h v(t)\|_{\mathbf{L}^2(\Omega)} \leq Ch\|v\|_{\mathbf{H}^1(\Omega)}$, (note that $v \in L^2[0, T; \mathbf{V}(\Omega)]$). For the last estimate, we will use the previous estimate. Thus, the definition of Ξ_h^{loc} for $k = 0$, $l \geq 1$, the inverse estimate $\|\Xi_h v\|_{\mathbf{H}^1(\Omega)} \leq C/h\|\Xi_h v\|_{\mathbf{L}^2(\Omega)}$, imply

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

for all $v \in L^2[0, T; \mathbf{V}(\Omega)] \cap H^1[0, T; \mathbf{H}^{-1}(\Omega)]$, which completes the proof of the fourth estimate. □

Remark 4.4. *(1) The last estimate of the above Lemma in $L^2[0, T; \mathbf{H}^1\Omega)]$ norm, requires the time-step restriction of $\tau \leq Ch^2$ due to the lack of regularity with respect to time. For the second estimate, we also note that if more regularity is available, the inverse estimate is not necessary. In particular if $v^{(k+1)} \in L^2[0, T; \mathbf{H}^1(\Omega)]$, then the improved rate of $\mathcal{O}(h^l + \tau^{k+1})$ holds in $\|.\|_{L^2[0,T;\mathbf{H}^1(\Omega)]}$ norm. However, we note that for the lowest order scheme $k = 0, l \geq 1$, the increased regularity $v_t \in L^2[0, T; \mathbf{H}^1(\Omega)]$ is not available at least in presence of control constraints. Hence, we emphasize that the lack of regularity acts as a barrier for developing a true higher order scheme. Working similarly we also obtain an estimate at arbitrary time-points, (see for instance [10]).*
(2) It is worth noting that approximation properties of Ξ_h^{loc} in $\|.\|_{L^2[0,T;\mathbf{H}^{-1}(\Omega)]}$ norm (see for instance [29, Proposition 2.12]) hold only on the divergence free subspace, $\mathbf{V}^{-1} \equiv \{v \in \mathbf{H}^{-1}(\Omega) : divv = 0\}$ endowed with the norm $\|.\|_{\mathbf{V}^{-1}} = \|.\|_{\mathbf{H}^{-1}}$. Here, the divergence free condition is understood as follows:

$$\langle v, \nabla\phi\rangle = 0 \qquad \forall\, \phi \in H^2_0(\Omega) \equiv \{\phi \in H^2(\Omega) \cap H^1_0(\Omega) : (\nabla\phi)|_\Gamma = 0\},$$

where $\langle .,. \rangle \equiv \langle .,. \rangle_{\mathbf{H}^{-1},\mathbf{H}^1_0}$. We refer the reader to [29, Section 2.3] for a detailed analysis of the projection and its properties, but we point out that in the subsequent analysis the use of $\|.\|_{L^2[0,T;\mathbf{H}^{-1}(\Omega)]}$ projection estimates is not needed.

The next result states that the error related to the auxiliary projection 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 $f \in L^2[0,T;\mathbf{H}^{-1}(\Omega)]$ and $y_0 \in \mathbf{W}(\Omega)$ be given, and $(\bar{y},\bar{p}),(\bar{\mu},\bar{\phi}) \in W(0,T) \times L^2[0,T;L^2_0(\Omega)]$ be the solutions of (2.8)-(2.9) and optimality conditions (2.6) or (2.7), and $w_h, z_h \in \mathcal{U}_h$ be the solutions of (4.1)-(4.2). Denote by $\bar{e} = \bar{y} - w_h$, $\bar{r} = \bar{\mu} - z_h$ and let $e_p \equiv \bar{y} - \Xi_h^{loc}\bar{y}$, $r_p = \bar{\mu} - P_h^{loc}\bar{\mu}$, where P_h^{loc}, Ξ_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, for any $q_h \in L^2[0,T;L^2_0(\Omega)]$,*

$$(1) \qquad \|\bar{e}\|^2_{W(0,T)} + \sum_{i=0}^{N-1} \|[\bar{e}^i]\|^2_{\mathbf{L}^2(\Omega)} \leq C\big(\|\bar{e}^0\|^2_{\mathbf{L}^2(\Omega)} + (1/\nu)\big(\|e_p\|^2_{W(0,T)} + \|\bar{p} - q_h\|^2_{L^2[0,T;L^2(\Omega)]}\big)\big),$$

$$(2) \qquad \|\bar{r}\|^2_{W(0,T)} + \sum_{i=1}^{N} \|[\bar{r}^i]\|^2_{\mathbf{L}^2(\Omega)} \leq C(1/\nu)\big(\|\bar{e}\|^2_{L^2[0,T;\mathbf{L}^2(\Omega)]} + \|r_p\|^2_{W(0,T)} + \|\bar{\phi} - q_h\|^2_{L^2[0,T;L^2(\Omega)]}\big),$$

$$(3) \qquad \|\bar{e}\|_{L^2[0,T;\mathbf{L}^2(\Omega)]} \leq C(1/\nu)\big(\nu\|e_p\|_{L^2[0,T;\mathbf{L}^2(\Omega)]} + \tau^{1/2}(\|e_p\|_{L^2[0,T;\mathbf{H}^1(\Omega)]} + \|\bar{p} - q_h\|_{L^2[0,T;L^2(\Omega)]})\big),$$

$$(4) \qquad \|\bar{r}\|_{L^2[0,T;\mathbf{L}^2(\Omega)]} \leq C\big(\nu\|\bar{e}\|_{L^2[0,T;\mathbf{L}^2(\Omega)]} + \|r_p\|_{L^2[0,T;\mathbf{L}^2(\Omega)]} + \tau^{1/2}(\|r_p\|_{L^2[0,T;\mathbf{H}^1(\Omega)]} + \|\bar{\phi} - q_h\|_{L^2[0,T;L^2_0(\Omega)]})\big).$$

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

Proof. Estimates (1)-(2): Throughout this proof, we denote by $\bar{e} = \bar{y} - w_h, \bar{r} = \bar{\mu} - z_h$ and we split $\bar{e}, \bar{r}$ to $\bar{e} \equiv e_{1h} + e_p \equiv (\Xi_h^{loc}\bar{y} - w_h) + (\bar{y} - \Xi_h^{loc}\bar{y})$, $\bar{r} \equiv r_{1h} + r_p \equiv (P_h^{loc}\bar{\mu} - z_h) + (\bar{\mu} - P_h^{loc}\bar{\mu})$, where P_h^{loc}, Ξ_h^{loc} are defined in Definitions 4.1 and 4.2. Subtracting (4.1) from (2.8), and (4.2) from (2.9) we obtain the orthogonality condition: For $n = 1, ..., N$, and for all $v_h \in \mathcal{P}_k[t^{n-1}, t^n; \mathbf{Y}_h]$, $q_h \in \mathcal{P}_k[t^{n-1}, t^n; Q_h]$,

$$\begin{cases} (\bar{e}^n, v_h^n) + \displaystyle\int_{t^{n-1}}^{t^n} \Big(- \langle \bar{e}, v_{ht} \rangle + a(\bar{e}, v_h) + b(v_h, \bar{p} - p_{1h}) \Big) dt = (\bar{e}^{n-1}, v_{h+}^{n-1}), \\ \displaystyle\int_{t^{n-1}}^{t^n} b(\bar{y} - w_h, q_h) dt = 0, \end{cases} \tag{4.4}$$

$$\begin{cases} -(\bar{r}_+^n, v_h^n) + \displaystyle\int_{t^{n-1}}^{t^n} \Big(\langle \bar{r}, v_{ht} \rangle + a(\bar{r}, v_h) + b(v_h, \bar{\phi} - \phi_{1h}) \Big) dt \\ \qquad = -(\bar{r}_+^{n-1}, v_{h+}^{n-1}) + \displaystyle\int_{t^{n-1}}^{t^n} (\bar{e}, v_h) dt, \\ \displaystyle\int_{t^{n-1}}^{t^n} b(\bar{\mu} - z_h, q_h) dt = 0. \end{cases} \tag{4.5}$$

Note that the orthogonality condition (4.4) is essentially uncoupled and identical to the orthogonality condition of [10, Equation (4.4)]. Hence applying [10, Theorems

4.6 and 4.7], we derive the first estimate. For the second estimate, we note that the orthogonality condition (4.5) is equivalent to: For $n = 1, ..., N$, and for all $v_h \in \mathcal{P}_k[t^{n-1}, t^n; \mathbf{Y}_h]$, $q_h \in \mathcal{P}_k[t^{n-1}, t^n; Q_h]$,

$$\begin{cases} -(r^n_{1h+}, v^n_h) + \int_{t^{n-1}}^{t^n} \Big(\langle r_{1h}, v_{ht}\rangle + a(r_{1h}, v_h) + \ b(v_h, \bar{\phi} - \phi_{1h})\Big)dt \\ \qquad = -(r^{n-1}_{1h+}, v^{n-1}_{h+}) + \int_{t^{n-1}}^{t^n} \Big((\bar{e}, v_h) - a(r_p, v_h)\Big)dt, \\ \int_{t^{n-1}}^{t^n} b(\bar{\mu} - z_h, q_h)dt = 0. \end{cases} \tag{4.6}$$

Here, we have used the Definition 4.1 of the projection P_h^{loc}, which implies that $\int_{t^{n-1}}^{t^n} \langle r_p, v_{ht}\rangle dt = 0$ and $(r^n_{p+}, v^n) = 0$. Setting $v_h = r_{1h} \in \mathcal{U}_h$ into (4.6), using the incompressibility constraint to write, $\int_{t^{n-1}}^{t^n} b(r_{1h}, \bar{\phi} - \phi_{1h}) = \int_{t^{n-1}}^{t^n} b(r_{1h}, \bar{\phi} - q_h)$ we obtain,

$$-(1/2)\|r^n_{1h+}\|^2_{\mathbf{L}^2(\Omega)} + (1/2)\|[r^n_{1h}]\|^2_{\mathbf{L}^2(\Omega)} + (1/2)\|r^{n-1}_{1h+}\|^2_{\mathbf{L}^2(\Omega)} + (\nu/4)\int_{t^{n-1}}^{t^n} \|r_{1h}\|^2_{\mathbf{H}^1(\Omega)}dt$$
$$\leq C\int_{t^{n-1}}^{t^n} \Big((1/\nu)\|\bar{e}\|^2_{\mathbf{L}^2(\Omega)} + (1/\nu)\|r_p\|^2_{\mathbf{H}^1(\Omega)} + \|\bar{\phi} - q_h\|^2_{L^2(\Omega)}\Big)dt. \tag{4.7}$$

Summing inequalities (4.7), we obtain the estimate in $L^2[0, T; \mathbf{H}^1(\Omega)]$, and at partition points using triangle inequality. Once the estimate for $\|\bar{r}\|_{L^2[0,T;\mathbf{H}^1(\Omega)]}$ is obtained, the estimate in $L^\infty[0, T; \mathbf{L}^2(\Omega)]$ follows using the arguments of Theorem [10, Theorem 4.7], modified to handle the backwards in time Stokes equation.

Estimates (3) and (4): We turn our attention to the last two estimates. In order to obtain the improved rate for the $L^2[0, T; \mathbf{L}^2(\Omega)]$ norm we employ a duality argument to derive a better bound for the quantity $\|e_{1h}\|^2_{L^2[0,T;\mathbf{L}^2(\Omega)]}$. For this purpose, we employ the duality argument of [5, Section 3] or [9, Lemma 4.3] in order to handle arbitrary order schemes, and the discrete incompressibility constraint. We define a backwards in time evolutionary problem with right hand side $e_{1h} \in L^2[0, T; \mathbf{L}^2(\Omega)]$, and zero terminal data, i.e., for $n = 1, ..., N$ and for all $v \in L^2[0, T; \mathbf{H}^1(\Omega)] \cap H^1[0, T; \mathbf{H}^{-1}(\Omega)]$, we seek $(z, \psi) \in W(0, T) \times L^2[0, T; L^2_0(\Omega)]$ such that

$$\begin{cases} \int_0^T \big(\langle z, v_t\rangle + a(v, z) + b(v, \psi)\big)dt + (z(0), v(0)) = \int_0^T (e_{1h}, v)dt, \\ \int_0^T b(z, q)dt = 0 \quad \forall\, q \in L^2[0, T; L^2_0(\Omega)]. \end{cases} \tag{4.8}$$

Note that since $e_{1h} \in L^\infty[0, T; \mathbf{W}(\Omega)]$, then Remark 2.1 implies that the following estimate hold:

$$\|z\|_{L^2[0,T;\mathbf{H}^2(\Omega)]} + \|z_t\|_{L^2[0,T;\mathbf{L}^2(\Omega)]} + \|\psi\|_{L^2[0,T;H^1(\Omega)]} \leq C\|e_{1h}\|_{L^2[0,T;\mathbf{L}^2(\Omega)]}. \tag{4.9}$$

The lack of regularity of the right hand side of (4.8) due to the presence of discontinuities, implies that we can not improve regularity of z in $[0, T]$. The associated discontinuous time-stepping scheme can be defined as follows: Given, terminal data $z^N_{h+} = 0$, we seek $(z_h, \psi_h) \in \mathcal{Y}_h \times \mathcal{Q}_h$ such that for all $v_h \in \mathcal{P}_k[t^{n-1}, t^n; \mathbf{Y}_h]$,

$q_h \in \mathcal{P}_k[t^{n-1}, t^n; Q_h]$,

$$\begin{cases} \quad -(z^n_{h+}, v^n_h) + \displaystyle\int_{t^{n-1}}^{t^n} \big((z_h, v_{ht}) + a(z_h, v_h) + b(\psi_h, v_h) \big) dt \\ \qquad +(z^{n-1}_{h+}, v^{n-1}_{h+}) = \displaystyle\int_{t^{n-1}}^{t^n} (e_{1h}, v_h) dt, \\ \quad \displaystyle\int_{t^{n-1}}^{t^n} b(z_h, q_h) dt = 0. \end{cases} \tag{4.10}$$

Hence using Lemma 3.6, we obtain $\|z_h\|_{L^\infty[0,T;\mathbf{H}^1(\Omega)]} \leq C_k \|e_{1h}\|_{L^2[0,T;\mathbf{L}^2(\Omega)]}$. It is now clear that we have the following estimate for $z - z_h$, which is a straightforward application of the previous estimates in $L^2[0, T; \mathbf{H}^1(\Omega)]$, the approximation properties of Lemma 4.3, of projections P^{loc}_h, Ξ^{loc}_h, (see for instance [10, Theorem 4.6]),

$$\begin{aligned} &\nu \|z - z_h\|_{L^2[0,T;\mathbf{H}^1(\Omega)]} \\ &\leq C \left(h + \tau^{1/2} \right) \big(\|z\|_{L^2[0,T;\mathbf{H}^2(\Omega)]} + \|z_t\|_{L^2[0,T;\mathbf{L}^2(\Omega)]} + \|\psi\|_{L^2[0,T;H^1(\Omega)]} \big) \\ &\leq C \big(h + \tau^{1/2} \big) \|e_{1h}\|_{L^2[0,T;\mathbf{L}^2(\Omega)]}. \end{aligned} \tag{4.11}$$

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 $l \geq 1$, $k = 0$, even if high order schemes (in time) are chosen. Setting $v_h = e_{1h}$, into (4.10), and using the fact that $\int_{t^{n-1}}^{t^n} b(e_{1h}, \psi_h) dt = 0$ we obtain,

$$-(z^n_{h+}, e^n_{1h}) + \int_{t^{n-1}}^{t^n} (z_h, e_{1ht}) + a(e_{1h}, z_h) dt + (z^{n-1}_{h+}, e^{n-1}_{1h+}) = \int_{t^{n-1}}^{t^n} \|e_{1h}\|^2_{\mathbf{L}^2(\Omega)} dt.$$

Integrating by parts in time, we deduce,

$$\begin{aligned} &-(z^n_{h+}, e^n_{1h}) + (z^n_h, e^n_{1h}) + \int_{t^{n-1}}^{t^n} \big(-(z_{ht}, e_{1h}) + a(z_h, e_{1h}) \big) dt \\ &\qquad = \int_{t^{n-1}}^{t^n} \|e_{1h}\|^2_{\mathbf{L}^2(\Omega)} dt. \end{aligned} \tag{4.12}$$

Setting $v_h = z_h$ into (4.4) and using $\bar{e} = e_p + e_{1h}$, the definition of projection Ξ^{loc}_h of Definition 4.2, and the fact that $\int_{t^{n-1}}^{t^n} b(z_h, \bar{p} - p_{1h}) dt = \int_{t^{n-1}}^{t^n} b(z_h, \bar{p} - q_h) dt$ we obtain,

$$\begin{aligned} &(e^n_{1h}, z^n_h) + \int_{t^{n-1}}^{t^n} \big(-(e_{1h}, z_{ht}) + a(e_{1h}, z_h) \big) dt - (e^{n-1}_{1h}, z^{n-1}_{h+}) \\ &= -\int_{t^{n-1}}^{t^n} \big(a(e_p, z_h) + b(z_h, p - q_h) \big) dt. \end{aligned} \tag{4.13}$$

Here, we have also used the fact that the definition of projection Ξ^{loc}_h of Definition 4.2, implies that $(e^n_p, z^n_h) = 0$, $\int_{t^{n-1}}^{t^n} (e_p, v_{ht}) dt = 0$ and $(e^{n-1}_p, z^{n-1}_{h+}) = 0$. Using (4.12) to replace the first three terms of (4.13) we arrive to

$$\begin{aligned}
&(z_{h+}^n, e_{1h}^n) - (e_{1h}^{n-1}, z_{h+}^{n-1}) + \int_{t^{n-1}}^{t^n} \|e_{1h}\|^2_{\mathbf{L}^2(\Omega)} dt \\
&= -\int_{t^{n-1}}^{t^n} \big(a(e_p, z_h) + b(z_h, \bar{p} - q_h)\big) dt \\
&= -\int_{t^{n-1}}^{t^n} \big(a(e_p, z_h - z) + a(e_p, z) + b(z_h - z, \bar{p} - q_h)\big) dt \\
&= -\int_{t^{n-1}}^{t^n} \big(a(e_p, z_h - z) - \nu(e_p, \Delta z) + b(z_h - z, \bar{p} - q_h)\big) dt,
\end{aligned}$$

where at the last two equalities we have used integration by parts (in space), and the incompressibility constraint which implies that $\int_{t^{n-1}}^{t^n} b(z, p - q_h) dt = 0$. Therefore,

$$\begin{aligned}
\int_{t^{n-1}}^{t^n} \|e_{1h}\|^2_{\mathbf{L}^2(\Omega)} dt + (z_{h+}^n, e_{1h}^n) - (e_{1h}^{n-1}, z_{h+}^{n-1}) &\leq \int_{t^{n-1}}^{t^n} \nu \|z_h - z\|_{\mathbf{H}^1(\Omega)} \|e_p\|_{\mathbf{H}^1(\Omega)} dt \\
&+ \int_{t^{n-1}}^{t^n} \big(\nu \|e_p\|_{\mathbf{L}^2(\Omega)} \|\Delta z\|_{\mathbf{L}^2(\Omega)} + \|z - z_h\|_{\mathbf{H}^1(\Omega)} \|\bar{p} - q_h\|_{L^2(\Omega)}\big) dt.
\end{aligned}$$

Then summing the above inequalities and using the fact that $z_+^N \equiv 0$ and $e_{1h}^0 = 0$ (by definition) and rearranging terms, we obtain

$$\begin{aligned}
(1/2)\|e_{1h}\|^2_{L^2[0,T;\mathbf{L}^2(\Omega)]} &\leq C\Big(\nu \|e_p\|_{L^2[0,T;\mathbf{L}^2(\Omega)]} \|z\|_{L^2[0,T;\mathbf{H}^2(\Omega)]} \\
&+ \nu \|z_h - z\|_{L^2[0,T;\mathbf{H}^1(\Omega)]} \big(\|e_p\|_{L^2[0,T;\mathbf{H}^1(\Omega)]} + (1/\nu)\|p - q_h\|_{L^2[0,T;L^2(\Omega)]}\big)\Big) \\
&\leq C\Big(\nu \|e_p\|_{L^2[0,T;\mathbf{L}^2(\Omega)]} \|e_{1h}\|_{L^2[0,T;\mathbf{L}^2(\Omega)]} \\
&+ (h + \tau^{1/2}) \|e_{1h}\|_{L^2[0,T;\mathbf{L}^2(\Omega)]} \big(\|e_p\|_{L^2[0,T;\mathbf{H}^1(\Omega)]} + (1/\nu)\|p - q_h\|_{L^2[0,T;L^2(\Omega)]}\big)\Big).
\end{aligned}$$

Here, we have used the Cauchy-Schwarz inequality, the stability bounds of dual equation (4.9), i.e., and the error estimates (4.11) on $z_h - z$. Finally, the estimate on $\|\bar{r}\|_{L^2[0,T;\mathbf{L}^2(\Omega)]}$ follows by using a similar duality argument. ∎

Remark 4.6. *The combination of the last two Theorems implies the "symmetric, regularity free" structure of our estimate. In particular, suppose that the initial data $y_0 \in \mathbf{W}(\Omega)$, and the forcing term $f \in L^2[0, T; \mathbf{H}^{-1}(\Omega)]$, and we define the natural energy norm $|||(v_1, v_2)|||_{W(0,T)} \equiv \|v_1\|_{W(0,T)} + \|v_2\|_{W(0,T)}$ endowed by the weak formulation. Then, the estimate under minimal regularity assumptions can be written as follows:*

$$|||(\bar{e}, \bar{r})|||_{W(0,T)} \leq C(|||(e_p, r_p)|||_{W(0,T)} + \|\bar{p} - q_h\|_{L^2[0,T;L^2(\Omega)]} + \|\bar{\phi} - q_h\|_{L^2[0,T;L^2(\Omega)]}).$$

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. [11]). Hence, the rates of convergence for $\bar{e}, \bar{r}$ depend only on the approximation and regularity results, via the projection error e_p as indicated in Lemma 4.3 and Remark 4.4. For example, if the Taylor-Hood element is being used, and $\bar{y} \in L^2[0, T; \mathbf{V}(\Omega)] \cap H^1[0, T; \mathbf{H}^{-1}(\Omega)]$, $\bar{p} \in L^2[0, T; L^2_0(\Omega)]$, then for for $\tau \leq Ch^2$ we obtain that

1. $\|e_p\|_{L^2[0,T;\mathbf{H}^1(\Omega)]} \leq C, \qquad \|\bar{p} - q_h\|_{L^2[0,T;L^2(\Omega)]} \leq C,$
2. $\|e_p\|_{L^2[0,T;\mathbf{L}^2(\Omega)]} \leq Ch\|y\|_{L^2[0,T;\mathbf{H}^1(\Omega)]} + \tau^{1/2}\|y_t\|_{L^2[0,T;\mathbf{H}^{-1}(\Omega)]}.$

Therefore, the above estimates, and Theorem 4.5, imply $\|\bar{e}\|_{L^2[0,T;\mathbf{L}^2(\Omega)]} \approx \mathcal{O}(h)$, *for* $\tau \leq Ch^2$. *Obviously the estimate of Theorem 4.5 is applicable even in case more regular solutions. For example, for smooth solutions, the Taylor-Hood element combined with the dG time-stepping scheme of order k will allow the following rates,*

1. $\|e_p\|_{L^2[0,T;\mathbf{H}^1(\Omega)]} \leq C(h^2 + \tau^{k+1})$
2. $\|e_p\|_{L^2[0,T;\mathbf{L}^2(\Omega)]} \leq C(h^3 + \tau^{k+1})$

Thus, Theorem 4.5, implies that $\|\bar{e}\|_{L^2[0,T;\mathbf{H}^1(\Omega)]} \approx \mathcal{O}(h^2 + \tau^{k+1})$, $\|\bar{r}\|_{L^2[0,T;\mathbf{H}^1(\Omega)]} \approx \mathcal{O}(h^2 + \tau^{k+1})$, $\|\bar{e}\|_{L^2[0,T;\mathbf{L}^2(\Omega)]} \approx \mathcal{O}(h^3 + \tau^{k+1})$ *and* $\|\bar{r}\|_{L^2[0,T;\mathbf{L}^2(\Omega)]} \approx \mathcal{O}(h^3 + \tau^{k+1})$.

4.2. Unconstrained Controls: Symmetric estimates for the optimality system. It remains to compare the discrete optimality system (3.6)-(3.7)-(3.8) to the auxiliary system (4.1)-(4.2).

LEMMA 4.7. *Let* $(\bar{y}_h, \bar{p}_h),(\bar{\mu}_h, \bar{\phi}_h),(w_h, p_{1h}),(z_h, \phi_{1h}) \in \mathcal{Y}_h \times \mathcal{Q}_h$ *be the solutions the discrete optimality system (3.6)-(3.7)-(3.8) and of the auxiliary system (4.1)-(4.2) respectively. Denote by* $\bar{e} \equiv \bar{y} - w_h$, $\bar{r} \equiv \bar{\mu} - z_h$, *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:*

$$\|e_{2h}\|_{L^2[0,T;\mathbf{L}^2(\Omega)]} + (1/\alpha^{1/2})\|r_{2h}\|_{L^2[0,T;\mathbf{L}^2(\Omega)]} \leq C(1/\alpha^{1/2})\|\bar{r}\|_{L^2[0,T;\mathbf{L}^2(\Omega)]}.$$

In addition, the following estimates hold:

$$\|e^N_{2h}\|^2_{\mathbf{L}^2(\Omega)} + \sum_{i=0}^{N-1} \|[e^i_{2h}]\|^2_{\mathbf{L}^2(\Omega)} + \nu \int_0^T \|e_{2h}\|^2_{\mathbf{H}^1(\Omega)} dt \leq (C/\alpha^{3/2}) \int_{t^{n-1}}^{t^n} \|\bar{r}\|^2_{\mathbf{L}^2(\Omega)}) dt,$$

$$\|r^0_{2h+}\|^2_{\mathbf{L}^2(\Omega)} + \sum_{i=1}^{N} \|[r^i_{2h}]\|^2_{\mathbf{L}^2(\Omega)} + \nu \int_0^T \|r_{2h}\|^2_{\mathbf{H}^1(\Omega)} dt \leq (C/\alpha^{1/2}) \int_0^T \|\bar{r}\|^2_{\mathbf{L}^2(\Omega)} dt,$$

where C is constant depending only upon Ω.

Proof. Subtracting (3.7) from (4.2) we obtain the equation: For $n = 1, ..., N$, $v_h \in \mathcal{P}_k[t^{n-1}, t^n, \mathbf{Y}_h]$, $q_h \in \mathcal{P}_k[t^{n-1}, t^n; Q_h]$

$$\left\{ \begin{aligned} & -(r^n_{2h+}, v^n) + \int_{t^{n-1}}^{t^n} \Big(\langle r_{2h}, v_{ht} \rangle + a(r_{2h}, v_h) + b(v_h, \phi_{1h} - \bar{\phi}_h) \Big) dt \\ & \qquad = -(r^{n-1}_{2h+}, v^{n-1}_+) + \int_{t^{n-1}}^{t^n} (e_{2h}, v_h) dt \\ & \int_{t^{n-1}}^{t^n} b(r_{2h}, q_h) dt = 0. \end{aligned} \right. \tag{4.14}$$

Subtracting (3.6) from (4.1) and using the optimality conditions (2.6), and (3.8), to replace $\bar{g}$ and $\bar{g}_h$ respectively, we obtain: For $n = 1, ..., N$, for all $v_h \in \mathcal{P}_k[t^{n-1}, t^n, \mathbf{Y}_h]$, $q_h \in \mathcal{P}_k[t^{n-1}, t^n; Q_h]$,

$$\left\{ \begin{aligned} & (e^n_{2h}, v^n) + \int_{t^{n-1}}^{t^n} \Big(- \langle e_{2h}, v_{ht} \rangle + a(e_{2h}, v_h) + b(v_h, p_{1h} - \bar{p}_h) \Big) dt \\ & \qquad = (e^{n-1}_{2h}, v^{n-1}_+) + \int_{t^{n-1}}^{t^n} -(1/\alpha)(\bar{\mu} - \bar{\mu}_h, v_h) dt, \\ & \int_{t^{n-1}}^{t^n} b(e_{2h}, q_h) dt = 0. \end{aligned} \right. \tag{4.15}$$

We set $v_h = e_{2h}$ into (4.14) and note that $\int_{t^{n-1}}^{t^n} b(e_{2h}, \phi_{1h} - \bar{\phi}_h)dt = 0$, to obtain

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

Setting $v_h = r_{2h}$ into (4.15), and noting $\int_{t^{n-1}}^{t^n} b(r_{2h}, p_{1h} - \bar{p}_h)dt = 0$ we deduce,

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

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

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

Using Young's inequality to bound the right hand side, adding the resulting inequalities 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 first estimate. For the second estimate, we simply set $v_h = e_{2h}$ into (4.15) and use the previous estimate on r_{2h}. Finally, the third estimate easily follows by setting $v_h = r_{2h}$ into (4.14), the estimate on $\|e_{2h}\|_{L^2[0,T;\mathbf{L}^2(\Omega)]}$ and standard algebra. □

Various estimates can be derived, using results of Theorem 4.5 and Lemma 4.7 and standard approximation theory results. We begin by stating an almost symmetric error estimates which can be viewed as the analogue of the classical Céa's Lemma.

Theorem 4.8. *Let $(\bar{y}_h, \bar{p}_h), (\bar{\mu}_h, \bar{\phi}_h) \in \mathcal{Y}_h \times \mathcal{Q}_h$ and $(\bar{y}, \bar{p})$, $(\bar{\mu}, \bar{\phi}) \in W(0,T) \times L^2[0,T;L^2_0(\Omega)]$ denote the solutions of the discrete and continuous optimality systems (3.6)-(3.7)-(3.8) and (2.8)-(2.9)-(2.6) respectively. Let $e_p = \bar{y} - \Xi^{loc}_h \bar{y}$, $r_p = \bar{\mu} - P^{loc}_h \bar{\mu}$ denote the projection error, where P^{loc}_h, Ξ^{loc}_h 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 $r = \bar{\mu} - \bar{\mu}_h$:*

$$|||(e,r)|||_{W(0,T)} \le \tilde{\mathbf{C}}(1/\alpha^{3/2})(|||(e_p, r_p)|||_{W(0,T)} + \|\bar{p} - q_h\|_{L^2[0,T;L^2(\Omega)]}$$
$$+\|\bar{\phi} - q_h\|_{L^2[0,T;L^2(\Omega)]})$$

where $\tilde{\mathbf{C}}$ depends upon constants of Theorem 4.5, and Lemma 4.7, $1/\nu^2$, and is independent of τ, h, α, and $q_h \in \mathcal{Q}_h$ arbitrary.

Proof. First, we observe that an estimate for $\|e_{2h}\|_{L^\infty[0,T;\mathbf{L}^2(\Omega)]}$ and $\|r_{2h}\|_{L^\infty[0,T;\mathbf{L}^2(\Omega)]}$ can be derived identical to [10, Theorem 4.6] since the (4.14)-(4.15) are uncoupled due to the estimate of Lemma 4.7. Therefore, the estimate follows by using triangle inequality and previous estimates of Theorem 4.5 and Lemma 4.7. □

An improved estimate for the $L^2[0,T;L^2(\Omega)]$ norm for the state, and adjoint follow by combining the estimates of Theorem 4.5, and the first estimate of Lemma 4.7.

THEOREM 4.9. *Suppose that $y_0 \in \mathbf{W}(\Omega)$, $f \in L^2[0,T;\mathbf{H}^{-1}(\Omega)]$, and the assumptions of Theorem 4.5 and Lemma 4.7 hold. Let $e_p = \bar{y} - \Xi_h^{loc}\bar{y}$, $r_p = \bar{\mu} - P_h^{loc}\bar{\mu}$ denote the projection error, where P_h^{loc}, Ξ_h^{loc} defined in Definition of 4.1, and 4.2 respectively. Then, there exists a constant C depending upon Ω, $1/\nu$ such that,*

$$\begin{aligned}
\|e\|_{L^2[0,T;\mathbf{L}^2(\Omega)]} \leq C(1/\alpha^{1/2})\Big(&\|e_p\|_{L^2[0,T;\mathbf{L}^2(\Omega)]} + \|r_p\|_{L^2[0,T;\mathbf{L}^2(\Omega)]}\\
&+\tau^{1/2}(\|e_p\|_{L^2[0,T;\mathbf{H}^1(\Omega)]} + \|\bar{p} - q_h\|_{L^2[0,T;L^2(\Omega)]})\\
&+\tau^{1/2}(\|r_p\|_{L^2[0,T;\mathbf{H}^1(\Omega)]} + \|\bar{\phi} - q_h\|_{L^2[0,T;L^2(\Omega)]})\Big)\\
\|r\|_{L^2[0,T;\mathbf{L}^2(\Omega)]} \leq C\Big(&\|e_p\|_{L^2[0,T;\mathbf{L}^2(\Omega)]} + \|r_p\|_{L^2[0,T;\mathbf{L}^2(\Omega)]}\\
&+\tau^{1/2}(\|e_p\|_{L^2[0,T;\mathbf{H}^1(\Omega)]} + \|\bar{p} - q_h\|_{L^2[0,T;L^2(\Omega)]})\\
&+\tau^{1/2}(\|r_p\|_{L^2[0,T;\mathbf{H}^1(\Omega)]} + \|\bar{\phi} - q_h\|_{L^2[0,T;L^2(\Omega)]})\Big).
\end{aligned}$$

Proof. Both estimates follow by using triangle inequality and previous estimates of Theorem 4.5, Lemma 4.7. ∎

We close this subsection by stating convergence rates in two cases for the Taylor-Hood element, depending on the available regularity. Obviously a variety of other estimates can be derived, depending on the chosen elements.

PROPOSITION 4.10. *Suppose that the assumptions of Theorem 4.5 and Lemma 4.7 hold.*
1) Let $y_0 \in \mathbf{W}(\Omega)$, $f \in L^2[0,T;\mathbf{H}^{-1}(\Omega)]$, and there exists $\bar{p} \in L^2[0,T;L^2_0(\Omega)]$, such that the weak formulation (2.8) is valid. Assume that the Taylor-Hood element are being used to construct the subspaces and piecewise constants polynomials $k = 0$ for the temporal discretization. Then, for $\tau \leq Ch^2$ we obtain,

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

2) Let $\bar{y}, \bar{\mu} \in L^2[0,T;\mathbf{H}^3(\Omega) \cap \mathbf{V}(\Omega)] \cap H^{k+1}[0,T;\mathbf{H}^1(\Omega)]$, $\bar{p}, \bar{\phi} \in L^2[0,T;H^2(\Omega) \cap L^2_0(\Omega)$. Suppose that the Taylor-Hood element combined with piecewise polynomials of degree k for the temporal discretization are being used, then the following rates hold:

$$\begin{aligned}
\|(e,r)\|_{W(0,T)} &\leq \tilde{\mathbf{C}}(1/\alpha^{3/2})\big(h^2 + \tau^{k+1}\big),\\
\|e\|_{L^2[0,T;\mathbf{L}^2(\Omega)]} &\leq C(1/\alpha^{1/2})\big(h^3 + \tau^{k+1} + \tau^{1/2}(h^2 + \tau^{k+1})\big),\\
\|r\|_{L^2[0,T;\mathbf{L}^2(\Omega)]} &\leq C\big(h^3 + \tau^{k+1} + \tau^{1/2}(h^2 + \tau^{k+1})\big).
\end{aligned}$$

Proof. The rates directly follow from Theorem 4.5, Theorem 4.9, Lemma 4.3 and Remark 4.6. ∎

4.3. Control Constraints: The variational discretization approach. We demonstrate that the variational discretization approach of Hinze ([26]) can be used within our framework. In the variational discretization approach the control is not discretized explicitly, and in particular we define $\mathcal{A}^d_{ad} \equiv \mathcal{A}_{ad}$. Thus, our discrete optimal control problem now coincides to: Minimize functional

$$J_h(y_h(g), g) = \frac{1}{2}\int_0^T \|y_h(g) - y_d\|^2_{\mathbf{L}^2(\Omega)}dt + \frac{\alpha}{2}\int_0^T \|g\|^2_{\mathbf{L}^2(\Omega)}dt$$

subject to (3.5), where $y_h(g) \in \mathcal{Y}_h$ denotes the solution of (3.5) with right hand side given control $g \in L^2[0,T;\mathbf{L}^2(\Omega)]$. The optimal control (abusing the notation, denoted again by $\bar{g}_h$) satisfies the following first order optimality condition,

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

where $\bar{g}_h$ takes the form $\bar{g}_h = Proj_{[g_a,g_b]}(-\frac{1}{\alpha}\bar{\mu}_h(\bar{g}_h))$, 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. Thus its algorithmic construction requires extra care (see e.g. [26]). However, in most cases the quantity of interest is the state variable, and not the control. For the second derivative we easily obtain an estimate independent of $\bar{g}, \bar{g}_h$, and in particular,

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

THEOREM 4.11. *Let $y_0 \in \mathbf{W}(\Omega)$, $f \in L^2[0,T;\mathbf{H}^{-1}(\Omega)]$, and $y_d \in L^2[0,T;\mathbf{L}^2(\Omega)]$, and the there exists an associated pressure $\bar{p} \in L^2[0,T;L^2_0(\Omega)]$. 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:*

$$\begin{aligned}
&\|\bar{g}-\bar{g}_h\|_{L^2[0,T;\mathbf{L}^2(\Omega)]} \leq C(1/\alpha)\|\mu(\bar{g})-\mu_h(\bar{g})\|_{L^2[0,T;\mathbf{L}^2(\Omega)]}\\
&\leq C(\|e_p\|_{L^2[0,T;\mathbf{L}^2(\Omega)]} + \|r_p\|_{L^2[0,T;\mathbf{L}^2(\Omega)]}\\
&\qquad +\tau^{1/2}(\|e_p\|_{L^2[0,T;\mathbf{H}^1(\Omega)]} + \|\bar{p}-q_h\|_{L^2[0,T;L^2(\Omega)]})\\
&\qquad +\tau^{1/2}(\|r_p\|_{L^2[0,T;\mathbf{H}^1(\Omega)]} + \|\bar{\phi}-q_h\|_{L^2[0,T;L^2(\Omega)]}),
\end{aligned}$$

where $(\mu_h(\bar{g}), \phi_h(\bar{g}))$ and $(\mu(\bar{g}), \bar{\phi})$ denote the solutions of (3.7) and (2.9) respectively, and $e_p \equiv y(\bar{g}) - \Xi_h^{loc}y(\bar{g})$, $r_p = \mu(\bar{g}) - P_h^{loc}\mu(\bar{g})$ the corresponding projection errors. Furthermore, if $\tau \leq Ch^2$,

$$\|\bar{g}-\bar{g}_h\|_{L^2[0,T;\mathbf{L}^2(\Omega)]} \leq Ch.$$

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

$$J_h^{'}(\bar{g}_h)(\bar{g}-\bar{g}_h) \geq 0 \qquad \text{and} \qquad J^{'}(\bar{g})(\bar{g}-\bar{g}_h) \leq 0. \tag{4.19}$$

Therefore, using the second order condition and the mean value theorem, we obtain for any $u \in L^2[0,T;\mathbf{L}^2(\Omega)]$, (and hence for the one resulting from the mean value theorem) and inequalities (4.19),

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

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 (3.7) from (2.9) and setting $\bar{r} = \mu_h(\bar{g}) - \mu(\bar{g})$, and $\bar{e} = y_h(\bar{g}) - y(\bar{g})$. Using the estimates of Theorem 4.5, and the rates of Proposition 4.10, we obtain the desired estimate, after noting the reduced regularity of $\bar{e}$. □

4.4. Numerical experiments. The following examples are based on the one presented in [21]. The pressure and the velocity must be discretized in compatible finite element spaces, satisfying appropriate inf-sup condition. Here, we employ the Taylor Hood $P2/P1$ element for the spatial approximation of the velocity/pressure. For the time approximation we use dG time stepping schemes of order $k = 0$, $k = 1$, i.e., piecewise constants and piecewise linears respectively. Our example, focus on the unconstrained control case, where a classical boot-strap argument implies smooth solutions for the state and adjoint variables, for smooth data. We consider a numerical test in the case $k = 0$, and three examples in the case $k = 1$ for the model problem in $\Omega \times [0, T] = [0, 2]^2 \times [0, 0.1]$, choosing $\bar{y}|_\Gamma = \mathbf{0}$ with known analytical exact solution:

$$\begin{aligned}
&\bar{y} = (\bar{y}_1, \bar{y}_2) = ((\cos(2jx) - 1)\sin(2my), \sin(2mx)(1 - \cos(2jy)))e^{-\nu t\lambda^2/2},\\
&\bar{p} = e^{-\nu t\lambda^2}((\sin(jx)^2 \sin(my)^2\lambda^2)/k^2 + (\cos(2jx) - 1)^2 \sin(2my)^2\\
&\quad + \sin(2mx)^2(1 - \cos(2jy))^2)/2,\\
&\bar{g} = (\bar{g}_1, \bar{g}_2),
\end{aligned}$$

where

$$\begin{aligned}
\bar{g}_1 = &((((j\nu\sin(jx)^2 - j\nu\cos(jx)^2 + j\nu)\cos(my)\sin(my)\lambda^2 + ((-8km^2 - 8j^3)\sin(jx)^2\\
&+(8jm^2 + 8j^3)\cos(jx)^2 - 8jm^2)\cos(my)\sin(my)))/j,\\
\bar{g}_2 = &(((j^2\nu\sin(2mx)\cos(2jy) - j^2\nu\sin(2mx))\lambda^2 + (-8j^2m^2 - 8j^4)\sin(2mx)\cos(2jy)\\
&+8j^2m^2\sin(2mx)))/(2j^2))e^{-\nu t\lambda^2/2},
\end{aligned}$$

initial velocity $\bar{y}_0 = ((\cos(2jx)-1)\sin(2my), \sin(2mx)(1-\cos(2jy)))$ and target $\mathbf{U}_d = (U_{d_1}, U_{d_2}) = (0.5, 0.5)$.

The forcing term $f = (f_1, f_2)$ can be easily computed according to the Stokes equation. We expect for the velocity $\mathcal{O}(h^3 + \tau^{k+1})$ and $\mathcal{O}(h^2 + \tau^{k+1})$ rates of convergence for the $L^2[0, T; \mathbf{L}^2(\Omega)]$ and $L^2[0, T; \mathbf{H}^1(\Omega)]$ norms respectively. In all examples, we fix the regularization parameter in the functional chosen as $\alpha = 10^{-4}$, and the free parameters (adapted from [10]) $\nu = 1$, $j = \pi$, $m = \pi$, and $\lambda = 1$. The optimal control problem is solved by the finite element toolkit FreeFem++ (see [24]) using a gradient algorithm method.

Numerical Test 1 ($k = 0$). In the first example, we will use $\tau = h^2/8$. We expect $\|e\|_{L^2[0,T;\mathbf{L}^2(\Omega)]} = \mathcal{O}(h^2)$ and $\|e\|_{L^2[0,T;\mathbf{H}^1(\Omega)]} = \mathcal{O}(h^2)$. For this choice of mesh the corresponding errors are shown in the Table 4.1, and the expected average rate is also validated.

Table 4.1
Experiment 1-Rates of convergence for $k = 0$ and $\tau = h^2/8$.

Discretization	Velocity - Control Error		
$\tau = h^2/8$	$\|e\|_{L^2[0,T;\mathbf{L}^2(\Omega)]}$	$\|e\|_{L^2[0,T;\mathbf{H}^1(\Omega)]}$	$\|\bar{g} - \bar{g}_h\|_{L^2[0,T;\mathbf{L}^2(\Omega)]}$
$h = 0.4714050$	0.110215	1.81853	5.33150
$h = 0.2357022$	0.011512	0.43118	0.63211
$h = 0.1178511$	0.002031	0.11109	0.11369
$h = 0.0589255$	0.001255	0.02922	0.07081
Conv. rate	2.152143	1.98600	2.07596

Numerical Test 2 ($k = 1$). Now, we turn our attention to the case of piecewise linear (in time) discretization. Recall, that our analysis allows to time-stepping approaches of the form $\tau \approx h$, hence in the following example we set $\tau = h/16$. We expect $\|e\|_{L^2[0,T;\mathbf{L}^2(\Omega)]} = \mathcal{O}(h^2)$, $\|e\|_{L^2[0,T;\mathbf{H}^1(\Omega)]} = \mathcal{O}(h^2)$. For this choice of mesh the corresponding errors are shown in the Table 4.2. We also emphasize that the almost "coarse" time stepping choice $\tau \approx h$ still gives the expected theoretical rates, which highlights the "implicit" nature of dG time stepping schemes. Here, we also note that the penalty parameter satisfies $\alpha << h$, in all mesh-size choices.

Table 4.2
Experiment 2-Rates of convergence for k=1 with $\tau = h/16$.

Discretization	Velocity - Control Error		
$\tau = h/16$	$\|e\|_{L^2[0,T;\mathbf{L}^2(\Omega)]}$	$\|e\|_{L^2[0,T;\mathbf{H}^1(\Omega)]}$	$\|\bar{g} - \bar{g}_h\|_{L^2[0,T;\mathbf{L}^2(\Omega)]}$
$h = 0.4714050$	0.108866	2.315120	5.470750
$h = 0.2357022$	0.010535	0.453111	0.607322
$h = 0.1178511$	0.001838	0.113375	0.083115
$h = 0.0589255$	0.000832	0.028927	0.020270
Conv. rate	2.343953	2.107000	2.686666

Numerical Test 3 ($k = 1$). In the third test, our focus is to validate an estimate of order $\mathcal{O}(h^3)$, when $\tau = h^{3/2}/10$. Our estimates leads to the predicted rates $\|e\|_{L^2[0,T;\mathbf{L}^2(\Omega)]} = \mathcal{O}(h^3)$, $\|e\|_{L^2[0,T;\mathbf{H}^1(\Omega)]} = \mathcal{O}(h^2)$ respectively. For this choice of mesh the corresponding errors are shown in the Table 4.3. Here, we recover the rate for the $L^2[0,T;\mathbf{L}^2(\Omega)]$ norm, with an almost "coarse" choice of time-stepping.

Table 4.3
Experiment 3-Rates of convergence for $k = 1$ with $\tau = h^{3/2}/10$.

Discretization	Velocity - Control Error		
$\tau = h^{3/2}/10$	$\|e\|_{L^2[0,T;\mathbf{L}^2(\Omega)]}$	$\|e\|_{L^2[0,T;\mathbf{H}^1(\Omega)]}$	$\|\bar{g} - \bar{g}_h\|_{L^2[0,T;\mathbf{L}^2(\Omega)]}$
$h = 0.4714050$	0.1138780	2.420150	5.718610
$h = 0.2357022$	0.0104282	0.455479	0.610602
$h = 0.1178511$	0.0014891	0.112681	0.082763
$h = 0.0589255$	0.0004965	0.028212	0.020051
Conv. rate	2.6137833	2.140366	2.718333

Finally, we close this section by presenting a computational example with rough (discontinuous) data y_0, U, and unknown true solution. Once again, the model problem is posed in $\Omega \times [0,T] = [0,2]^2 \times [0,0.1]$. Here, the obvious choice for the discretization in time is piecewise constants (in time) $k = 0$ combined with the standard Taylor-Hood element in space. We consider as solution the solution computed in the most advanced partitioned grid of the square, comparing it with our computations in each one of the previous meshes using interpolation.

Numerical Test 4 ($k = 0$ and rough initial data y_0). Here, the obvious choice for the discretization in time is piecewise constants (in time) $k = 0$ combined with the standard Taylor-Hood element in space. We consider as "known analytical" solution the solution computed in the most advanced partitioned grid of the square, comparing it with our computations in each one of the previous meshes using interpolation.

In this setting, we choose $\tau = h^2/8$, which implies rates (at least) $\|e\|_{L^2[0,T;\mathbf{L}^2(\Omega)]} = \mathcal{O}(h), \|r\|_{L^2[0,T;\mathbf{L}^2(\Omega)]} = \mathcal{O}(h)$. Since, we consider as true solution the one computed by the most advanced grid, we actually expect to compute better rates in both norms. To this end, we also point out that y_0 possesses slightly better regularity than $\mathbf{W}(\Omega)$.

In particular, we choose "rough" initial data to our problem, i.e., $y_0 = (y_{0,1}, y_{0,2})$ where

$$y_{0,1} = \begin{cases} 6 + (\cos(2jx_1) - 1)\sin(2mx_2) & \text{for } x_1, x_2 \geq 0.5 \\ (\cos(2jx_1) - 1)\sin(2mx_2) & \text{for } x_1, x_2 < 0.5 \end{cases},$$

and

$$y_{0,2} = \begin{cases} 6 + (\cos(2jx_2) - 1)\sin(2mx_1) & \text{for } x_1, x_2 \geq 0.5 \\ (\cos(2jx_2) - 1)\sin(2mx_1) & \text{for } x_1, x_2 < 0.5 \end{cases}.$$

The target is given by:

$$U_d = (U_{d1}, U_{d2}) \text{ where } U_{d1} = U_{d2} = \begin{cases} 6.5, & \text{for } x_1, x_2 \geq 0.5 \\ 0.5, & \text{for } x_1, x_2 < 0.5 \end{cases},$$

Table 4.4

Experiment 4-Rates of convergence for $k = 0$ with $\tau = h^2/8$ and discontinuous initial data and target function.

Discretization	Velocity Error	
$\tau = h^2/8$	$\|e\|_{L^2[0,T;\mathbf{L}^2(\Omega)]}$	$J(\bar{y}_\sigma, \bar{g}_\sigma)$
$h = 0.4714050$	0.1268288547	14.80282714
$h = 0.2357022$	0.0362554882	9.742095817
$h = 0.1178511$	0.0140523956	9.608375932
$h = 0.0589255$	0.0044720938	9.619787446
$h = 0.0294627$	-	9.612306775
Conv. rate	1.6085962400	-

Remark 4.12. *We note that when computing the errors (especially in $L^2[0, T; \mathbf{L}^2(\Omega)]$ norms) the change from $h = 0.1178511$ to $h = 0.0589255$ seems to show lower rates of convergence than the predicted. This is because the temporal and spatial integration procedures accumulate errors, and it is expected due to high resolution of the time-space mesh (21219 number of degrees of freedom for each variable y, g, μ, and 9409 for p in space with computations in each of the 231 time points). This rate reduction does not appear in the numerical experiment of the non smooth data because of the "exact" solution is constructed at the smallest grid). More specifically, integration errors don't effect the final error computations in $L^2[0, T; \mathbf{L}^2(\Omega)]$ norm, and asymptotic rate of convergence is visible for small h too. We note that for the smooth case we used a common double dual-core processor. For the non smooth case, the computation of the exact solution required 37249 number of degrees of freedom for each variable y, g, μ in space and 922 time points, and a 4 six-core processor was used.*

REFERENCES

[1] F. Abergel and R.Temam, On some control problems in fluid mechanics, *Theor. Comput. Fluid Dyn.*, **1**(1990), pp. 303-326.

[2] T. Apel and T. Flaig, Crank-Nicolson schemes for optimal control problems with evolution equations, *SIAM J. Numer. Anal.*, **50** (2012), pp. 1484-1512.

[3] E. Casas, An optimal control problem governed by the evolution Navier-Stokes equations. In S.S. Sritharan, editor, *Optimal Control of Viscous Flows*, Philadelphia, 1998. Frontiers in Applied Mathematics, SIAM.

[4] 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.
[5] E. Casas, and K. Chrysafinos, Error estimates for the discretization of the velocity tracking problem, *submitted, available at http://www.math.ntua.gr/∼chrysafinos.*
[6] 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.
[7] 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.
[8] 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.
[9] K. Chrysafinos and E. karatzas, Error estimates for discontinuous Galerkin time-stepping schemes for Robin boundary control problems constrained to parabolic pdes', *submitted, available at http://www.math.ntua.gr/∼chrysafinos.*
[10] K. Chrysafinos and N. J. Walkington, Discontinuous Galerkin approximations of the Stokes and Navier-Stokes equations, Math. Comp., **79** (2010), pp. 2135-2167.
[11] P. Ciarlet, The finite element method for elliptic problems, SIAM Classics, Philadelphia, 2002.
[12] P. Constantin, Navier-Stokes equations, The University of Chicago press, Chicago, 1988.
[13] R. Dautray, and J.-L. Lions, Mathematical analysis and Numerical Methods for Science and Technology, Vol 6, Spinger-Verlag, Berlin, 1993.
[14] K. Deckelnick, and M. Hinze, Error estimates in space and time for tracking type control of the instationary Stokes system, *Inter. Ser. Numer. Math.*, **143** (2002), pp. 87-103.
[15] 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.
[16] J.-C. de los Reyes, and K. Kunisch, A semi-smooth Newton method for control constrained boundary control of the Navie-Stokes equtions, *Nonl. Anal.*, **62** (2005), pp. 1289-1316.
[17] L. Evans, Partial Differential Equations, AMS, Providence RI, 1998.
[18] V. Girault and P.-A. Raviart, Finite Element Methods for Navier-Stokes, Springer-Verlag, New York, 1986.
[19] P. Grisvard, Elliptic problems in nonsmooth domains, Pitman, Boston, 1985.
[20] M.D. Gunzburger Perspectives in flow control and optimization, SIAM, Advances in Design and Control, Philadelphia, 2003.
[21] 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.
[22] M.D. Gunzburger and S. Manservisi, The velocity tracking problem for Navier-Stokes flows with bounded distributed control, *SIAM J.Control and Optim.* **37** (2000), pp. 1913-1945.
[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. 594-634.
[24] F. Hecht, FreeFem++, Third edition, Version 3.13, 2011. Available from: http://www.freefem.org/ff++.
[25] J. Heywood and R. Rannacher, Finite element approximation of the nonstationary Navier-Stokes problem. I. Regularity of solutions and second order estimates for spatial discretization, *SIAM J. Numer. Anal.*, **19** (1982), pp. 275-311.
[26] M. Hinze, A variational discretization concept in control constrained optimization: The linear-quadratic case, *Comput. Optim. Appl.*, **30** (2005), pp. 45-61.
[27] 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.
[28] 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.
[29] L.S. Hou, Error estimates for semidiscrete finite element approximation of the evolutionary Stokes equations under minimal regularity assumptions, *J. Scient. Comput.*, **16** (2001), pp. 287-317.
[30] 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.
[31] I. Lasiecka and R. Triggiani, Control theory for partial differential equations, Cambridge University press, 2000.
[32] J.-L. Lions, Some aspects of the control of distributed parameter systems, SIAM publications, 1972.

[33] 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.

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

[35] 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.

[36] 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.

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

[38] 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.

[39] T. Richter, A. Springer, and B. Vexler, Efficient numerical realization of Discontinuous Galerkin methods for temporal discretization of parabolic problems, *Numer. Math.*, **124** (2013), pp. 151-182.

[40] A. Springer, and B. Vexler, Third order convergent time discretization for parabolic optimal control problems with control constraints, *Comput. Optim. and Appl., published online* (2013).

[41] S.S. Sritharan. *Optimal control of viscous flow*. SIAM, Philadelphia, 1998.

[42] R. Temam, Navier-Stokes equations, Amer. Math. Soc., Chelsea, Providence, RI, 2001.

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

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

[45] F. Tröltzsch and D. Wachsmuth, Second-order suficcient optimality conditions for the optimal control of Navier-Stokes equations. *ESAIM: COCV*, **12** (2006), pp. 93-119.