Department of Mathematics
and Computer Science

College of Science

University of the Philippines Baguio

Mixed and Hybrid Petrov–Galerkin Finite Element Discretization for Optimal Control of the Wave Equation

MIXED AND HYBRID PETROV–GALERKIN FINITE ELEMENT DISCRETIZATION FOR OPTIMAL CONTROL OF THE WAVE EQUATION

GILBERT PERALTA AND KARL KUNISCH*

Abstract.
A mixed finite element discretization of an optimal control problem for the linear wave equation with homogeneous Dirichlet boundary condition is considered. For the temporal discretization, a Petrov–Galerkin scheme is utilized and the Raviart–Thomas finite elements for spatial discretization is used. A priori error analysis is proved for this numerical scheme. A hybridized formulation is proposed and if the Arnold–Brezzi post-processing method is applied, better convergence rates with respect to space are observed. The interchangeability of discretization and optimization holds both for mixed and hybrid formulations. Numerical experiments illustrating the theoretical results are presented using the lowest-order Raviart–Thomas elements.

2020 Mathematics Subject Classification.
49J20, 35L05, 65M60

Keywords.
Mixed and hybrid finite elements, Petrov–Galerkin method, wave equation, optimality system, error estimates, post-processing.

Citation.
G. Peralta and K. Kunisch, *Mixed and hybrid Petrov–Galerkin finite element discretization for optimal control of the wave equation*, Numerische Mathematik 150 (2), pp. 591-627, 2022.
DOI: https://doi.org/10.1007/s00211-021-01258-9

*Institute for Mathematics and Scientific Computing, University of Graz, Heinrichstrasse 36, 8010 Graz, Austria, and RICAM Institute, Austrian Academy of Sciences, Altenbergerstrasse 69, 4040 Linz Austria

Department of Mathematics and Computer Science, University of the Philippines Baguio, Governor Pack Road, Baguio, 2600 Philippines. Email: grperalta@up.edu.ph.

Table of Contents

1. Introduction

In this paper, we propose mixed and hybrid space–time finite element discretizations for the optimal control of the wave equation written as a first-order hyperbolic system in terms of the pressure and velocity. We consider the following linear-quadratic optimal control problem:

$$\text{Minimize } J(u,p,\boldsymbol{v}) = \frac{1}{2}\int_0^T\int_\Omega \alpha(p(t,x)-p_d(t,x))^2 + \beta|\boldsymbol{v}(t,x)-\boldsymbol{v}_d(t,x)|^2\,\mathrm{d}x\,\mathrm{d}t + \frac{\gamma}{2}\int_0^T\int_\Omega u(t,x)^2\,\mathrm{d}x\,\mathrm{d}t \tag{1.1}$$

subject to the state equation

$$\begin{cases} \partial_t p - \operatorname{div}\boldsymbol{v} = u & \text{in } (0,T)\times\Omega,\\ \partial_t \boldsymbol{v} - \nabla p = 0 & \text{in } (0,T)\times\Omega,\\ p = 0 & \text{on } (0,T)\times\partial\Omega,\\ p(0)=p_0,\ \boldsymbol{v}(0)=\boldsymbol{v}_0 & \text{in } \Omega, \end{cases} \tag{1.2}$$

over all distributed controls

$$u \in L^2(0,T;L^2(\Omega)). \tag{1.3}$$

Equation (1.2) describes the evolution of small amplitude pressure waves in a Newtonian fluid or elastic solid, where $p = p(t,x) \in \mathbb{R}$ and $\boldsymbol{v} = \boldsymbol{v}(t,x) \in \mathbb{R}^2$ denote the pressure and velocity field at time $t \in [0,T]$ and position $x \in \Omega$, respectively. These

equations can be obtained by linearizing Euler's continuity and momentum equations. Throughout this paper, we assume that Ω is a bounded convex polygonal domain in $\mathbb{R}^2$. Here, we focus on the two-dimensional case, nevertheless, three-dimensional bounded domains can be treated in analogous manner.

The state equation (1.2) with an additional linear term in the momentum equation also arises in modeling heat dynamics with finite speed of propagation. Indeed, starting from the energy balance law for the rate of change of temperature θ we have

$$c\rho\partial_t\theta = -\operatorname{div}\boldsymbol{q} + u \tag{1.4}$$

where c is the specific heat, ρ is the density of the material, u is an external source or control, and $\boldsymbol{q}$ is the heat flux. Suppose that there is a time delay between the heat flux and the temperature gradient. This assumption results in the following Cattaneo heat flux law

$$\boldsymbol{q}(t+\tau, x) = -\kappa\nabla\theta(t, x) \tag{1.5}$$

where $\kappa > 0$ is the heat conductivity and $\tau > 0$ is a constant representing the relaxation time. If $\tau = 0$ then Cattaneo's law reduces to the well-known Fourier's law of heat conduction. Applying a first-order Taylor approximation to (1.5) yields

$$\tau\partial_t\boldsymbol{q} + \boldsymbol{q} = -\kappa\nabla\theta. \tag{1.6}$$

After time-reversal and normalization, we observe that (1.4) and (1.6) take the form of (1.2) with an additional linear term in the momentum equation. For simplicity of exposition, we only consider the case of (1.2), nevertheless, the analysis presented here can be adapted to the system (1.4) and (1.6). We refer the reader to [34] for more details and related models.

The use of mixed finite elements for the optimal control of elliptic and parabolic partial differential equations has been of interest for researchers in the past decade, see for instance [13, 37, 38]. The choice of mixed finite elements is advatangeous when one needs to keep track of the flux instead of the displacement. However, for hyperbolic problems there is little work in this direction. The authors in [23] address a priori error analysis for the semidiscretization of the optimal control problem under the positivity constraint on the mean of the distributed control, that is,

$$\frac{1}{T|\Omega|}\int_0^T\int_\Omega u(t, x)\,\mathrm{d}t\,\mathrm{d}x \geq 0.$$

There are, however, some works that deal with mixed finite element approximations of hyperbolic partial differential equations, see [6, 9, 15, 19, 29] for example and the references therein.

There are two main approaches in the discretization of the wave equation

$$\partial_{tt}w - \Delta w = u \tag{1.7}$$

with homogeneous Dirichlet boundary condition via mixed method, namely the velocity–pressure and displacement–stress formulations. In the velocity–pressure formulation, the wave equation (1.7) is rewritten as a first-order system in the form of (1.2) with $p = w$ and $\boldsymbol{v} = \nabla w$. On the other hand, in the displacement–stress formulation, one introduces the stress field $\boldsymbol{\sigma} = \nabla w$ and rewrites the above wave

equation as the following system

$$\begin{cases} \partial_{tt} w - \operatorname{div} \boldsymbol{\sigma} = u, \\ \boldsymbol{\sigma} - \nabla w = 0. \end{cases} \tag{1.8}$$

One can then discretize this system with respect to space as in the elliptic case, and proceed with a centered difference time-stepping scheme for the approximation of the acceleration to obtain a fully discrete system. This strategy has been also considered for acoustic wave equations with Neumann boundary conditions in [5, 16, 24]. Regardless of the formulation (1.2) or (1.8) for mixed finite elements, the Dirichlet boundary condition is a natural one, while the Neumann boundary condition is an essential boundary condition.

For the spatial discretization of the state equation (1.2), we shall use mixed finite elements, specifically the Raviart–Thomas finite elements [33]. Implementing mixed finite elements for this system is well-suited when the problem is written as a Cauchy problem on its usual state-space setting. For smooth initial data and control, the resulting finite elements will be conformal. With respect to time-discretization, we shall use a Petrov–Galerkin scheme consisting of continuous piecewise-linear ansatz functions and discontinuous piecewise constant test functions in time. The same strategy has been employed in [25] for the optimal control of wave equations with either distributed, Dirichlet, or Neumann control in displacement–velocity formulations.

Although the proposed Petrov–Galerkin scheme is formulated globally in time, it results in a time-stepping scheme by approximating the integrals through the trapezoidal rule. As an outcome, the tracking part of the cost functional will be discretized by the trapezoidal rule as well. The key by-product of the described numerical scheme is that the two approaches discretize-then-optimize and optimize-then-discretize coincide. In other words, the adjoint system of the discretized optimal control problem is a discretization of the adjoint system for the continuous optimal control problem.

It is well-known that the above Petrov–Galerkin scheme is a variant of the Crank–Nicolson scheme and hence has the capability to be second-order accurate with respect to time. However, when this method is applied to the optimal control problem (1.1)-(1.3), we are only able to obtain a linear order of convergence due to the fact that the time-discretization of the adjoint equations will consist of discontinuous piecewise constant ansatz functions. For optimal control governed by parabolic problems, second-order accuracy can be obtained using an appropriate post-processing strategy that utilizes the midpoints of the subintervals induced by the temporal partition, see [2, 28].

Recent advances in space–time finite element methods for wave equations and related hyperbolic problems consider higher order methods in time, for which better convergence rates were shown in comparison to linear approximations. For instance, discontinuous Petrov–Galerkin methods in the space–time cylinder were studied in [18, 21] and post-processing techniques for continuous Petrov–Galerkin methods with respect to time were analyzed in [1, 8]. Space–time-discretizations are advantageous for large-scale problems where it is possible to solve the solutions in parallel

both in space and time. It would be interesting to look at the performance of these schemes in the context of the optimal control problem (1.1)-(1.3).

The rest of the paper is organized as follows: In Section 2, the well-posedness of (1.1)-(1.3) based on semigroup theory is briefly presented. Space–time mixed finite element discretization, and the hybridization of this problem will be discussed in Section 3, and a priori error estimates will be proved in Section 4. In the hybrid formulation, the continuity of the normal components of the discretized stress along the inter-element boundaries is relaxed by introducing a Lagrange multiplier. As in the elliptic case, simple post-processing of this Lagrange multiplier yields better convergence rates with respect to space for the optimal pressure and control, at the expense of an additional computing time. In Section 5, specific details for the implementation will be discussed. Finally, numerical examples based on the lowest order Raviart–Thomas element will be presented in Section 6.

2. Well-Posedness of the Optimal Control Problem

In this section, we briefly discuss the theoretical framework of our optimization problem, specifically the well-posedness of the state equation (1.2). In order to set up the weak formulation of (1.2), we introduce several notations: Let $X_0 = L^2(\Omega)$, $X_1 = H_0^1(\Omega)$, $\boldsymbol{V}_0 = L^2(\Omega)^2$, and $\boldsymbol{V}_1 = \boldsymbol{H}(\operatorname{div}, \Omega) = \{\boldsymbol{v} \in \boldsymbol{V}_0 : \operatorname{div} \boldsymbol{v} \in X_0\}$. We shall use the notations $(\cdot, \cdot)$ and $\|\cdot\|$ for the inner products and norms on X_0 and $\boldsymbol{V}_0$, and likewise $(v, w)_I$ and $\|v\|_I$ for the inner product and norm on $L^2(I, X)$, where $I = (0, T)$, and X is a given Banach space.

Define the product space $H = X_0 \times \boldsymbol{V}_0$ and the linear operator $A : D(A) \subset H \to H$ by

$$A \begin{pmatrix} p \\ \boldsymbol{v} \end{pmatrix} = - \begin{pmatrix} \operatorname{div} \boldsymbol{v} \\ \nabla p \end{pmatrix}$$

with domain $D(A) = X_1 \times \boldsymbol{V}_1$. Applying the divergence theorem, we can see that A is a skew-adjoint operator, that is, $A^* = -A$, and that it has a dense domain. As a consequence of Stone's theorem, A generates a unitary group on H. System (1.2) can be recasted as an abstract Cauchy problem on H as

$$\begin{cases} \dfrac{d}{dt} \begin{pmatrix} p(t) \\ \boldsymbol{v}(t) \end{pmatrix} + A \begin{pmatrix} p(t) \\ \boldsymbol{v}(t) \end{pmatrix} = \begin{pmatrix} u(t) \\ \mathbf{0} \end{pmatrix} \quad t \in I, \\ p(0) = p_0, \ \ \boldsymbol{v}(0) = \boldsymbol{v}_0. \end{cases} \tag{2.1}$$

Applying classical results for semigroups of bounded linear operators, see [30] for instance, given an initial data $(p_0, \boldsymbol{v}_0)^\top \in H$ and control $u \in U := L^2(I, X_0)$, (2.1) has a unique mild solution $(p, v)^\top \in C(\overline{I}, H)$ and there exists a constant $C > 0$ independent of the data, solution, and control such that

$$\sup_{t \in \overline{I}} \{\|p(t)\| + \|\boldsymbol{v}(t)\|\} \leq C(\|u\|_I + \|p_0\| + \|\boldsymbol{v}_0\|). \tag{2.2}$$

A variational formulation of (2.1) is given by the following equation

$$(p, -\varphi_t + \operatorname{div} \boldsymbol{\psi})_I + (\boldsymbol{v}, -\boldsymbol{\psi}_t + \nabla \varphi)_I = (u, \varphi)_I + (p_0, \varphi(0)) + (\boldsymbol{v}_0, \boldsymbol{\psi}(0)) \tag{2.3}$$

for every $(\varphi, \boldsymbol{\psi})^\top \in H^1(I, X_1) \times H^1(I, \boldsymbol{V}_1)$ such that $\varphi(T) = 0$ and $\boldsymbol{\psi}(T) = \mathbf{0}$. By a density argument a pair $(p, v)^\top \in C(\overline{I}, H)$ is a solution to (2.1) if and only if it satisfies (2.3). If in addition, the data satisfies

$$(p_0, \boldsymbol{v}_0)^\top \in D(A) \text{ and } u \in W^{1,1}(I, X_0) \tag{2.4}$$

then the mild solution of (2.1) satisfies

$$(p, \boldsymbol{v})^\top \in C(\overline{I}, D(A)) \times C^1(\overline{I}, H) \tag{2.5}$$

where $D(A)$ is equipped with the graph norm. Moreover, there is a constant $C > 0$ such that

$$\begin{aligned} &\sup_{t\in\overline{I}}\{\|\partial_t p(t)\| + \|\partial_t \boldsymbol{v}(t)\| + \|\nabla p(t)\| + \|\operatorname{div} \boldsymbol{v}(t)\|\} \\ &\leq C(\|u\|_{W^{1,1}(I,X_0)} + \|\nabla p_0\| + \|\boldsymbol{v}_0\| + \|\operatorname{div} \boldsymbol{v}_0\|). \end{aligned}$$

Under the regularity assumptions (2.4) on the data and control, we can see that the solution $(p, \boldsymbol{v})^\top$ satisfies the variational equation

$$(\partial_t p, \varphi)_I - (\operatorname{div} \boldsymbol{v}, \varphi)_I + (\partial_t \boldsymbol{v}, \boldsymbol{\psi})_I + (p, \operatorname{div} \boldsymbol{\psi})_I = (u, \varphi)_I$$

for all $(\varphi, \boldsymbol{\psi})^\top \in L^2(I, X_0) \times L^2(I, \boldsymbol{V}_1)$. This variational formulation will be the basis for the finite element discretization of our optimal control problem.

All throughout this work, we assume that $\alpha, \beta \geq 0$ and $\gamma > 0$, and at the very least the desired states satisfy $p_d \in L^2(I, X_0)$ and $\boldsymbol{v}_d \in L^2(I, \boldsymbol{V}_0)$.

Theorem 2.1. *Given $p_0 \in H_0$ and $\boldsymbol{v}_0 \in \boldsymbol{V}_0$, the optimal control problem (1.1)-(1.3) admits a unique solution $(\bar{u}, \bar{p}, \bar{\boldsymbol{v}})^\top \in U \times L^2(I, H)$.*

Proof. The proof of this theorem follows from standard weak sequential arguments as in [35] for linear-quadratic optimal control problems. □ □

By means of the control-to-state mapping $u \mapsto (p, \boldsymbol{v})^\top = (p(u), \boldsymbol{v}(u))^\top$ from U into $L^2(I, H)$ defined through (1.2), we introduce the reduced cost functional $j : U \to \mathbb{R}$ given by

$$j(u) = J(u, p(u), \boldsymbol{v}(u)).$$

Then the optimal control problem (1.1)-(1.3) can be equivalently expressed as

$$\min_{u\in U} j(u). \tag{2.6}$$

The first-order necessary optimality condition for the control problem (2.6) is

$$j'(\bar{u})\delta u = 0 \quad \forall \delta u \in U. \tag{2.7}$$

This condition is also sufficient due to the linear-quadratic structure of the optimal control problem under consideration. Using the mild solution $(w, \boldsymbol{y})^\top = (w(u), \boldsymbol{y}(u))^\top$ of the adjoint system

$$\begin{cases} -w_t + \operatorname{div} \boldsymbol{y} = \alpha(p(u) - p_d) & \text{in } I \times \Omega, \\ -\boldsymbol{y}_t + \nabla w = \beta(\boldsymbol{v}(u) - \boldsymbol{v}_d) & \text{in } I \times \Omega, \\ w = 0 & \text{on } I \times \partial\Omega, \\ w(T) = 0,\ \boldsymbol{y}(T) = \mathbf{0} & \text{in } \Omega, \end{cases} \tag{2.8}$$

the first derivative of the reduced cost functional in the direction of $\delta u \in U$ is given by

$$j'(u)\delta u = (\gamma u + w(u), \delta u)_I. \tag{2.9}$$

This can be proved using an approximation argument applied to the initial data, desired state, and control, and then applying the regularity (2.5) and the continuity of the solutions (2.2) with respect to time. Consequently, from (2.9) the optimality condition (2.7) reduces to

$$\bar{u} = -\gamma^{-1} w(\bar{u}). \tag{2.10}$$

The existence and uniqueness of mild solutions in $C(\overline{I}, H)$ to the adjoint system (2.8) follows from the fact that the generator A is skew-adjoint. Indeed, we can express (2.8) as the backward-in-time Cauchy problem

$$\begin{cases} -\dfrac{d}{dt}\begin{pmatrix} w(t) \\ \boldsymbol{y}(t) \end{pmatrix} + A^* \begin{pmatrix} w(t) \\ \boldsymbol{y}(t) \end{pmatrix} = \begin{pmatrix} \alpha(p(t) - p_d(t)) \\ \beta(\boldsymbol{v}(t) - \boldsymbol{v}_d(t)) \end{pmatrix} & t \in I, \\ w(T) = 0, \ \boldsymbol{y}(T) = \boldsymbol{0}. \end{cases}$$

In terms of the variational formulation, the mild solution to (2.8) satisfies the equation

$$(\partial_t \xi - \operatorname{div} \boldsymbol{\eta}, w)_I + (\partial_t \boldsymbol{\eta} - \nabla \xi, y)_I = \alpha(w - w_d, \xi)_I + \beta(\boldsymbol{y} - \boldsymbol{y}_d, \eta)_I \tag{2.11}$$

for every $(\xi, \boldsymbol{\eta})^\top \in H^1(I, X_1) \times H^1(I, \boldsymbol{V}_1)$ such that $\xi(0) = 0$ and $\boldsymbol{\eta}(0) = \boldsymbol{0}$.

In the case that $u \in W^{1,1}(I, X_0)$, $p_d \in W^{1,1}(I, X_0)$ if $\alpha > 0$, and $\boldsymbol{v}_d \in W^{1,1}(I, \boldsymbol{V}_0)$ if $\beta > 0$, then from (2.5) we have $(p, \boldsymbol{v})^\top \in W^{1,1}(I, H)$, and as a consequence $(w, \boldsymbol{y})^\top \in C(\overline{I}, D(A)) \cap C^1(\overline{I}, H)$.

3. Mixed and Hybrid Finite Element Discretizations

In this section, we describe the Petrov–Galerkin scheme for the discretization of the state equation using a mixed finite element method for the spatial variable. Suppose that $\{\mathcal{T}_h\}$ is a family of quasi-regular triangulations of Ω parametrized by the meshsize $h = \max_{K \in \mathcal{T}_h} \operatorname{diam}(K)$. For a nonnegative integer r, define the space of piecewise polynomial and Raviart–Thomas finite element spaces

$$\begin{aligned} X_h^r &= \{p_h \in X_0 : p_h|_K \in \mathbb{P}_r \ \forall K \in \mathcal{T}_h\}, \\ \boldsymbol{V}_h^r &= \{\boldsymbol{v}_h \in \boldsymbol{V}_1 : \boldsymbol{v}_h|_K \in \mathbb{P}_r^2 \oplus \mathbf{x}\mathbb{P}_r \ \forall K \in \mathcal{T}_h\}, \end{aligned}$$

where $\mathbf{x} = (x, y)^\top$ represents the spatial variable and $\mathbb{P}_r$ is the space of polynomials in K of degree at most r. It is known that

$$\operatorname{div} \boldsymbol{V}_h^r = X_h^r \tag{3.1}$$

and there exists a projection operator $\boldsymbol{\rho}_h : H^1(\Omega)^2 \to \boldsymbol{V}_h^r$ such that

$$\operatorname{div}(\boldsymbol{\rho}_h \boldsymbol{v}) = \pi_h \operatorname{div} \boldsymbol{v} \tag{3.2}$$

where π_h is the orthogonal projection from X_0 onto X_h^r. Moreover, we have

$$\|p - \pi_h p\| \le C h^{r+1} \|p\|_{H^{r+1}} \qquad \forall p \in H^{r+1}(\Omega), \tag{3.3}$$

$$\|\boldsymbol{v} - \boldsymbol{\rho}_h \boldsymbol{v}\| \le C h^{r+1} \|\boldsymbol{v}\|_{H^{r+1}} \qquad \forall \boldsymbol{v} \in H^{r+1}(\Omega)^2. \tag{3.4}$$

Partition the interval $\overline{I}$ into $0 = t_0 < t_1 < \cdots < t_M = T$. Let $I_0 = \{0\}$ and $I_j = (t_{j-1}, t_j]$ for $j = 1, \ldots, M$. We consider the case of uniform time stepsizes, that

is, $\tau = t_j - t_{j-1}$ for every j. To formulate the space–time-discretization of (1.2), we introduce the following space consisting of continuous piecewise-linear functions in time with values in X_h^r and $\boldsymbol{V}_h^r$

$$X_{hk}^r = \{p_{hk} \in C(\overline{I}, X_h^r) : p_{hk}|_{I_j} \in \mathcal{P}^1(I_j, X_h^r)\},$$
$$\boldsymbol{V}_{hk}^r = \{\boldsymbol{v}_{hk} \in C(\overline{I}, \boldsymbol{V}_h^r) : \boldsymbol{v}_{hk}|_{I_j} \in \mathcal{P}^1(I_j, \boldsymbol{V}_h^r)\},$$

and the space of discontinuous piecewise constant functions in time with values in X_h^r and $\boldsymbol{V}_h^r$

$$\tilde{X}_{hk}^r = \{p_{hk} \in L^2(I, X_h^r) : p_{hk}|_{I_j} \in \mathcal{P}^0(I_j, X_h^r),\ p_{hk}(0) \in X_h^r\},$$
$$\tilde{\boldsymbol{V}}_{hk}^r = \{\boldsymbol{v}_{hk} \in L^2(I, \boldsymbol{V}_h^r) : \boldsymbol{v}_{hk}|_{I_j} \in \mathcal{P}^0(I_j, \boldsymbol{V}_h^r),\ \boldsymbol{v}_{hk}(0) \in \boldsymbol{V}_h^r\}.$$

Here, $\mathcal{P}^r(I_j, X)$ is the space of polynomial functions in I_j of degree at most r with values in a Banach space X.

3.1. Discretization. Define the multilinear form $a_{hk} : X_{hk}^r \times \boldsymbol{V}_{hk}^r \times \tilde{X}_{hk}^r \times \tilde{\boldsymbol{V}}_{hk}^r \to \mathbb{R}$ by

$$a_{hk}(p_{hk}, \boldsymbol{v}_{hk}, \varphi_{hk}, \boldsymbol{\psi}_{hk}) = (\partial_t p_{hk}, \varphi_{hk})_I - (\operatorname{div} \boldsymbol{v}_{hk}, \varphi_{hk})_I + (\partial_t \boldsymbol{v}_{hk}, \boldsymbol{\psi}_{hk})_I + (p_{hk}, \operatorname{div} \boldsymbol{\psi}_{hk})_I + (p_{hk}(0), \varphi_{hk}(0)) + (\boldsymbol{v}_{hk}(0), \boldsymbol{\psi}_{hk}(0)).$$

For the discretization of the state equation (1.2) with a given control $u_{hk} \in U_{hk}$, we consider the following: Find $(p_{hk}, \boldsymbol{v}_{hk})^\top = (p_{hk}(u_{hk}), \boldsymbol{v}_{hk}(u_{hk}))^\top \in X_{hk}^r \times \boldsymbol{V}_{hk}^r$ such that

$$a_{hk}(p_{hk}, \boldsymbol{v}_{hk}, \varphi_{hk}, \boldsymbol{\psi}_{hk}) = (u_{hk}, \varphi_{hk})_I + (p_0, \varphi_{hk}(0)) + (\boldsymbol{v}_0, \boldsymbol{\psi}_{hk}(0)) \tag{3.5}$$

for all $(\varphi_{hk}, \boldsymbol{\psi}_{hk})^\top \in \tilde{X}_{hk}^r \times \tilde{\boldsymbol{V}}_{hk}^r$. Here, $U_{hk} \subset U$ and different possible choices will be mentioned in Remark 3.1 below. The corresponding discrete optimal control problem reads as follows:

$$\min_{u_{hk} \in U_{hk}} J(u_{hk}, p_{hk}, \boldsymbol{v}_{hk}) \text{ subject to (3.5).} \tag{3.6}$$

As in the continuous case, by means of the discrete control-to-state operator $u_{hk} \mapsto (p_{hk}(u_{hk}), \boldsymbol{v}_{hk}(u_{hk}))^\top$ from U_{hk} into $X_{hk}^r \times \boldsymbol{V}_{hk}^r$ and the discrete reduced cost functional $j_{hk} : U_{hk} \to \mathbb{R}$ defined by

$$j_{hk}(u_{hk}) = J(u_{hk}, p_{hk}(u_{hk}), \boldsymbol{v}_{hk}(u_{hk})),$$

(3.6) can be reformulated as

$$\min_{u_{hk} \in U_{hk}} j_{kh}(u_{hk}). \tag{3.7}$$

The first-order necessary and sufficient condition for $\bar{u}_{hk} \in U_{hk}$ to be the optimal control is $j'_{hk}(\bar{u}_{hk})\delta u_{hk} = 0$ for every $\delta u_{hk} \in U_{hk}$. In terms of the solution $(w_{hk}, \boldsymbol{y}_{hk})^\top = (w_{hk}(u_{hk}), \boldsymbol{y}_{hk}(u_{hk}))^\top \in \tilde{X}_{hk}^r \times \tilde{\boldsymbol{V}}_{hk}^r$ of the discrete adjoint equation

$$a_{hk}(\xi_{hk}, \boldsymbol{\eta}_{hk}, w_{hk}, \boldsymbol{y}_{hk}) = \alpha(\xi_{hk}, p_{hk}(u_{hk}) - p_d)_I + \beta(\boldsymbol{\eta}_{hk}, \boldsymbol{v}_{hk}(u_{hk}) - \boldsymbol{v}_d)_I \tag{3.8}$$

for every $(\xi_{hk}, \boldsymbol{\eta}_{hk})^\top \in X_{hk}^r \times \boldsymbol{V}_{hk}^r$, the first-order directional derivative of the discrete cost is given by

$$j'_{hk}(u_{hk})\delta u_{hk} = (\gamma u_{hk} + w_{hk}(u_{hk})_I, \delta u_{hk})_I,$$

for every $u_{hk}, \delta u_{hk} \in U_{hk}$.

Remark 3.1. Let us mention possible discretizations of the control space. Let $\bar{u}_{hk}$ be the solution of (3.7). With the choice $U_{hk} = \tilde{X}^r_{hk}$, we have $\bar{u}_{hk} = -\gamma^{-1} w_{hk}(\bar{u}_{hk})$. On the other hand, if we take $U_{hk} = X^r_{hk}$, then $\bar{u}_{hk} = -\gamma^{-1}\Pi_{hk} w_{hk}(\bar{u}_{hk})$, where $\Pi_{hk} : \tilde{X}^r_h \to X^r_{hk}$ is the projection operator such that

$$(\Pi_{hk} w_{hk}, \varphi_{hk})_I = (w_{hk}, \varphi_{hk})_I, \qquad \forall (w_{hk}, \varphi_{hk})^\top \in \tilde{X}^r_{hk} \times X^r_{hk}.$$

Let us now express the discrete problem (3.5) more explicitly. Though (3.5) is formulated globally in time, it results in a time-stepping scheme. By approximating the time integrals through the trapezoidal rule, we obtain a variant of the Crank–Nicholson method. Here, for the discretization of the control space, we choose the same discretization as for the pressure, that is, $U_{hk} = X^r_{hk}$.

We follow the usual strategy of discretizing the optimal control problem (2.6), that is, by discretizing explicitly both the state and control spaces. Another quite popular method is the so-called *variational discretization* introduced by Hinze [22]. In this discretization concept, the state-space is discretized while the space of controls is not. Although the control lies in a continuous control space, it is a discrete object since it is implicitly discretized in terms of the discrete adjoint variable and through projection.

Define the bilinear form $b : \boldsymbol{V}^r_h \times X^r_h \to \mathbb{R}$ by

$$b(\boldsymbol{v}_h, p_h) = (\operatorname{div} \boldsymbol{v}_h, p_h).$$

In the following, a superscript will refer to the value at the time nodes, for example, $p^m = p_{hk}(t_m)$ and $\boldsymbol{v}^m = \boldsymbol{v}_{hk}(t_m)$. Also, we denote by χ_O to be the indicator function of a set O.

Let $\varphi_h \in X^r_h$ and $\boldsymbol{\psi}_h \in \boldsymbol{V}^r_h$. Taking $\varphi_{hk} = \chi_{I_m}\varphi_h$ and $\boldsymbol{\psi}_{hk} = \chi_{I_m}\boldsymbol{\psi}_h$ in (3.5), we obtain the following: For $m = 0$ we find

$$(p^0, \varphi_h) = (p_0, \varphi_h), \quad (\boldsymbol{v}^0, \boldsymbol{\psi}_h) = (\boldsymbol{v}_0, \boldsymbol{\psi}_h) \qquad \forall \varphi_h \in X^r_h, \boldsymbol{\psi}_h \in \boldsymbol{V}^r_h, \tag{3.9}$$

and for $m = 1, \dots, M$ we have

$$\begin{cases} \dfrac{1}{\tau}(p^m - p^{m-1}, \varphi_h) - \dfrac{1}{2} b(\boldsymbol{v}^m + \boldsymbol{v}^{m-1}, \varphi_h) \\ \qquad\qquad = \dfrac{1}{2}(u^m + u^{m-1}, \varphi_h) \quad \forall \varphi_h \in X^r_h, \\ \dfrac{1}{\tau}(\boldsymbol{v}^m - \boldsymbol{v}^{m-1}, \boldsymbol{\psi}_h) + \dfrac{1}{2} b(\boldsymbol{\psi}_h, p^m + p^{m-1}) = 0 \quad \forall \boldsymbol{\psi}_h \in \boldsymbol{V}^r_h. \end{cases} \tag{3.10}$$

If $u_{hk} = 0$ then this scheme preserves the energy at every time step, that is, $\|p^m\|^2 + \|\boldsymbol{v}^m\|^2 = \|p^0\|^2 + \|\boldsymbol{v}^0\|^2$ for every m.

For the realization of the discrete adjoint equation (3.8) as a time-stepping scheme, we take $\xi_{hk} = \phi_m \xi_h$ and $\boldsymbol{\eta}_{hk} = \phi_m \boldsymbol{\eta}_h$, where $\xi_h \in X^r_h$, $\boldsymbol{\eta}_h \in \boldsymbol{V}^r_h$. Further, $\{\phi_m : 0 \le m \le M\}$ are the linear Lagrange elements (hat functions) with respect to the temporal partition. Thus, we obtain the following: For $m = M$ we have

$$\begin{cases} \dfrac{1}{\tau}(\xi_h, w^M) + \dfrac{1}{2} b(\boldsymbol{y}^M, \xi_h) = \dfrac{\alpha}{2}(p^M - p^M_d, \xi_h) \quad \forall \xi_h \in X^r_h, \\ \dfrac{1}{\tau}(\boldsymbol{\eta}_h, \boldsymbol{y}^M) - \dfrac{1}{2} b(\boldsymbol{\eta}_h, w^M) = \dfrac{\beta}{2}(\boldsymbol{v}^M - \boldsymbol{v}^M_d, \boldsymbol{\eta}_h) \quad \forall \boldsymbol{\eta}_h \in \boldsymbol{V}^r_h. \end{cases} \tag{3.11}$$

For $m = M-1, \dots, 1$ we have

$$\begin{cases} \frac{1}{\tau}(\xi_h, w^m - w^{m+1}) + \frac{1}{2}b(\boldsymbol{y}^m + \boldsymbol{y}^{m+1}, \xi_h) \\ \qquad\qquad = \alpha(p^m - p_d^m, \xi_h) \quad \forall \xi_h \in X_h^r, \\ \frac{1}{\tau}(\boldsymbol{\eta}_h, \boldsymbol{y}^m - \boldsymbol{y}^{m+1}) - \frac{1}{2}b(\boldsymbol{\eta}_h, w^m + w^{m+1}) \\ \qquad\qquad = \beta(\boldsymbol{v}^m - \boldsymbol{v}_d^m, \boldsymbol{\eta}_h) \quad \forall \boldsymbol{\eta}_h \in \boldsymbol{V}_h^r, \end{cases} \tag{3.12}$$

and for $m = 0$ we have

$$\begin{cases} \frac{1}{\tau}(\xi_h, w^0 - w^1) + \frac{1}{2}b(\boldsymbol{y}^1, \xi_h) = \frac{\alpha}{2}(p^0 - p_d^0, \xi_h) \quad \forall \xi_h \in X_h^r, \\ \frac{1}{\tau}(\boldsymbol{\eta}_h, \boldsymbol{y}^0 - \boldsymbol{y}^1) - \frac{1}{2}b(\boldsymbol{\eta}_h, w^1) = \frac{\beta}{2}(\boldsymbol{v}^0 - \boldsymbol{v}_d^0, \boldsymbol{\eta}_h) \quad \forall \boldsymbol{\eta}_h \in \boldsymbol{V}_h^r. \end{cases} \tag{3.13}$$

Note that the fully discrete adjoint system (3.11)-(3.13) is an approximation of (3.8) and under appropriate regularity assumptions the error between the solutions at the time nodes is second order in time. In contrast to the usual Crank–Nicolson scheme, we have half-steps at the first and last time-steps, and the right-hand sides are approximated by the box-rule.

The existence of solutions for the fully discrete state and adjoint equations can be established immediately by induction over m and the fact that for each m the system matrix is injective, hence bijective since the associated linear system is finite-dimensional and square.

Due to the approximation of the integrals, the solutions of the above discrete state and adjoint equations are different from the original Petrov–Galerkin formulations (3.5) and (3.8). However, by using an appropriate approximation of the cost functional, we will see below that the associated adjoint equations corresponds to the system (3.11)-(3.13). Consider the discretized cost functional

$$\begin{aligned} &J_{hk}(u_{hk}, p_{hk}, \boldsymbol{v}_{hk}) \\ &= \frac{\alpha\tau}{2}\left(\frac{1}{2}\|p^0 - p_d^0\|^2 + \sum_{m=1}^{M-1}\|p^m - p_d^m\|^2 + \frac{1}{2}\|p^M - p_d^M\|^2\right) \\ &+ \frac{\beta\tau}{2}\left(\frac{1}{2}\|\boldsymbol{v}^0 - \boldsymbol{v}_d^0\|^2 + \sum_{m=1}^{M-1}\|\boldsymbol{v}^m - \boldsymbol{v}_d^m\|^2 + \frac{1}{2}\|\boldsymbol{v}^M - \boldsymbol{v}_d^M\|^2\right) + \frac{\gamma}{2}\|u_{hk}\|_I^2. \end{aligned} \tag{3.14}$$

Here the tracking part of the functional is approximated by the trapezoidal rule, while the control cost is computed exactly. Set

$$j_{hk}(u_{hk}) = J_{hk}(u_{hk}, p_{hk}(u_{hk}), \boldsymbol{v}_{hk}(u_{hk})).$$

Given $u_{hk} \in U_{hk}$, the first-order derivative of the reduced cost j_{hk} in the direction of $\delta u_{hk} \in U_{hk}$ is given by

$$\begin{aligned} &j'_{hk}(u_{hk})\delta u_{hk} = \gamma(u_{hk}, \delta u_{hk})_I \\ &+ \alpha\tau\left(\frac{1}{2}(p^0 - p_d^0, \delta p^0) + \sum_{m=1}^{M-1}(p^m - p_d^m, \delta p^m) + \frac{1}{2}(p^M - p_d^M, \delta p^M)\right) \end{aligned} \tag{3.15}$$

$$+\beta\tau\left(\frac{1}{2}(\boldsymbol{v}^0-\boldsymbol{v}_d^0,\delta\boldsymbol{v}^0)+\sum_{m=1}^{M-1}(\boldsymbol{v}^m-\boldsymbol{v}_d^m,\delta\boldsymbol{v}^m)+\frac{1}{2}(\boldsymbol{v}^M-\boldsymbol{v}_d^M,\delta\boldsymbol{v}^M)\right)$$

where $(\delta p_{hk},\delta\boldsymbol{v}_{hk})^\top=(p_{hk}(\delta u_{hk}),\boldsymbol{v}_{hk}(\delta u_{hk}))^\top$. We claim that

$$j'_{hk}(u_{hk})\delta u_{hk}=(\gamma u_{hk}+w_{hk},\delta u_{hk})_I, \tag{3.16}$$

for every $u_{hk},\delta u_{hk}\in U_{hk}$, where $(w_{hk},\boldsymbol{y}_{hk})^\top=(w_{hk}(u_{hk}),\boldsymbol{y}_{hk}(u_{hk}))^\top\in\tilde{X}_{hk}^r\times\tilde{\boldsymbol{V}}_{hk}^r$ is the solution of (3.11)-(3.13). Because we are in the unconstrained case, it is enough to prove the claim in the case where $\delta p^0=0$ and $\delta\boldsymbol{v}^0=0$. First, let us define the sum

$$S:=\sum_{m=1}^{M}\{(\delta p^m,w^m)-\frac{\tau}{2}b(\delta\boldsymbol{v}^m,w^m)+(\delta\boldsymbol{v}^m,\boldsymbol{y}^m)+\frac{\tau}{2}b(\boldsymbol{y}^m,\delta p^m)\}.$$

This sum is related to the discrete adjoint equation as we shall see in (3.18) below. Using $(\varphi_h,\boldsymbol{\psi}_h)^\top=(w^m,\boldsymbol{y}^m)^\top$ as a test function in the discrete state equation (3.10) with u^m replaced by δu^m, we obtain

$$\begin{aligned}S=&\sum_{m=2}^{M}\{(\delta p^{m-1},w^m)+\frac{\tau}{2}b(\delta\boldsymbol{v}^{m-1},w^m)+(\delta\boldsymbol{v}^{m-1},\boldsymbol{y}^m)-\frac{\tau}{2}b(\boldsymbol{y}^m,\delta p^{m-1})\}\\&+\sum_{m=1}^{M}\frac{\tau}{2}(\delta u^m+\delta u^{m-1},w^m).\end{aligned}\tag{3.17}$$

Because δu_{hk} is piecewise-linear in time and w_{hk} is piecewise constant in time, the trapezoidal rule applied to $(\delta u_{hk},w_{hk})_{I_m}$ is exact on each subinterval I_m, and thus

$$\sum_{m=1}^{M}\frac{\tau}{2}(\delta u^m+\delta u^{m-1},w^m)=(\delta u_{hk},w_{hk})_I.$$

On the other hand, taking $(\xi_h,\boldsymbol{\eta}_h)^\top=(\delta p^m,\delta\boldsymbol{v}^m)^\top$ as a test function in the discrete adjoint equation (3.11)-(3.13), we have

$$\begin{aligned}S=&\sum_{m=1}^{M-1}\{(\delta p^m,w^{m+1})-\frac{\tau}{2}b(\boldsymbol{y}^{m+1},\delta p^m)+(\delta\boldsymbol{v}^m,\boldsymbol{y}^{m+1})+\frac{\tau}{2}b(\delta\boldsymbol{v}^m,w^{m+1})\}\\&+\sum_{m=1}^{M-1}\{\alpha\tau(p^m-p_d^m,\delta p^m)+\beta\tau(\boldsymbol{v}^m-\boldsymbol{v}_d^m,\delta\boldsymbol{v}^m)\}\\&+\frac{\alpha\tau}{2}(p^M-p_d^M,\delta p^M)+\frac{\beta\tau}{2}(\boldsymbol{v}^M-\boldsymbol{v}_d^M,\delta\boldsymbol{v}^M).\end{aligned}\tag{3.18}$$

Reindexing the first sum in (3.18), comparing to (3.17), and using (3.15) proves our claim (3.16). As a consequence, the process of discretization and optimization commute for the above Petrov–Galerkin approximation.

3.2. Hybridization. We discuss a hybridized formulation of the mixed finite element method presented in the previous subsection. Following the idea in the elliptic case [3], we relax the continuity of the normal component of the velocity along the inter-element boundaries by introducing Langrange multipliers on each of the interior edges. We note that hybridization offers an easier construction of a basis

for the finite element spaces. The advantages and disadvantages of hybridization will be discussed in the numerical section. We recall the *discontinuous* Raviart–Thomas finite element space

$$\boldsymbol{Y}_h^r = \{\boldsymbol{v}_h \in \boldsymbol{V}_0 : \boldsymbol{v}_h|_K \in \mathbb{P}_r^2 \oplus \mathbf{x}\mathbb{P}_r \ \forall K \in \mathcal{T}_h\}$$

and the space of multipliers

$$M_h^r = \{\lambda_h \in L^2(\mathcal{E}_h^i) : \lambda_h|_e \in \mathbb{P}_r \ \forall e \in \mathcal{E}_h\}$$

where $\mathcal{E}_h^i$ is the set of all interior edges of $\mathcal{T}_h$. The spaces $\boldsymbol{Y}_{hk}^r$, $\tilde{\boldsymbol{Y}}_{hk}^r$, M_{hk}^r, and $\tilde{M}_{hk}^r$, are defined analogously as before. Here, r is a nonnegative even integer. Let $\Pi_h^r \in \mathcal{L}(X_h^{r+1}, M_h^r)$ and $P_h^{r-2} \in \mathcal{L}(X_h^{r+1}, X_h^{r-2})$ be the L^2-projections onto M_h^r and X_h^{r-2}, respectively. Also, define $R_h^{r+1} : M_h^r \times X_h^r \to X_h^{r+1}$ according to

$$\begin{aligned}
&\Pi_h^r R_h^{r+1}(\lambda_h, p_h) = \lambda_h, \\
&P_h^{r-2}(R_h^{r+1}(\lambda_h, p_h) - p_h) = 0, \quad \text{for } r \geq 2,
\end{aligned}$$

for each $(\lambda_h, p_h)^\top \in M_h^r \times X_h^r$. In the case $r = 0$, we simply write $R_h^1\lambda_h$ instead of $R_h^1(\lambda_h, p_h)$.

In [3] the restriction that r is even was imposed to obtain a unified proof for the stability of the post-processing operator R_h^{r+1}. *Ad hoc* nonconforming approximations are needed to prove stability in the case of an odd r, and constructions to the cases $k = 1$ and $k = 3$ were given in the said paper.

Define the bilinear forms $b_h : \boldsymbol{Y}_h^r \times X_h^r \to \mathbb{R}$ and $d_h : \boldsymbol{Y}_h^r \times M_h^r \to \mathbb{R}$ by

$$\begin{aligned}
b_h(\boldsymbol{v}_h, p_h) &= \sum_{K \in \mathcal{T}_h} \int_K p_h \operatorname{div} \boldsymbol{v}_h \, \mathrm{d}x, \\
d_h(\boldsymbol{v}_h, \lambda_h) &= \sum_{e \in \mathcal{E}_h^i} \int_e [\boldsymbol{v}_h] \cdot \boldsymbol{\nu} \lambda_h \, \mathrm{d}x,
\end{aligned}$$

where $[\boldsymbol{v}_h]$ is the jump across the edge e, and the corresponding bilinear forms $b_{h,I} : \boldsymbol{Y}_{hk}^r \times X_{hk}^r \to \mathbb{R}$ and $d_{h,I} : \boldsymbol{Y}_{hk}^r \times M_{hk}^r \to \mathbb{R}$ for time-dependent variables

$$\begin{aligned}
b_{h,I}(\boldsymbol{v}_{hk}, p_{hk}) &= \int_I b_h(\boldsymbol{v}_{hk}(t), p_{hk}(t)) \, \mathrm{d}t \\
d_{h,I}(\boldsymbol{v}_{hk}, \lambda_{hk}) &= \int_I d_h(\boldsymbol{v}_{hk}(t), \lambda_{hk}(t)) \, \mathrm{d}t.
\end{aligned}$$

Also, consider the multilinear form $\tilde{a}_{hk} : X_{hk}^r \times \boldsymbol{Y}_{hk}^r \times M_{hk}^r \times \tilde{X}_{hk}^r \times \tilde{\boldsymbol{Y}}_{hk}^r \times \tilde{M}_{hk}^r \to \mathbb{R}$ defined by

$$\begin{aligned}
&\tilde{a}_{hk}(p_{hk}, \boldsymbol{v}_{hk}, \lambda_{hk}, \varphi_{hk}, \boldsymbol{\psi}_{hk}, \mu_{hk}) \\
&\qquad = (\partial_t p_{hk}, \varphi_{hk})_I - b_{h,I}(\boldsymbol{v}_{hk}, \varphi_{hk}) + (\partial_t \boldsymbol{v}_{hk}, \boldsymbol{\psi})_I \\
&\qquad + b_{h,I}(\boldsymbol{\psi}_{hk}, p_{hk}) - d_{h,I}(\boldsymbol{v}_{hk}, \mu_{hk}) + d_{h,I}(\boldsymbol{\psi}_{hk}, \lambda_{hk}) \\
&\qquad + (p_{hk}(0), \varphi_{hk}(0)) + (\boldsymbol{v}_{hk}(0), \boldsymbol{\psi}_{hk}(0)) + (\lambda_{hk}(0), \mu_{hk}(0)).
\end{aligned}$$

The hybrid formulation of (1.2) that we consider is the following: Find a triple $(p_{hk}, \boldsymbol{v}_{hk}, \lambda_{hk})^\top \in X_{hk}^r \times \boldsymbol{Y}_{hk}^r \times M_{hk}^r$ such that

$$\tilde{a}_{hk}(p_{hk}, \boldsymbol{v}_{hk}, \lambda_{hk}, \varphi_{hk}, \boldsymbol{\psi}_{hk}, \mu_{hk})$$

$$= (u_{hk}, \varphi_{hk})_I + (p_0, \varphi_{hk}(0)) + (\boldsymbol{v}_0, \boldsymbol{\psi}_{hk}(0)) + (\lambda_0, \mu_{hk}(0)) \tag{3.19}$$

for every $(\varphi_{hk}, \boldsymbol{\psi}_{hk}, \mu_{hk})^\top \in \tilde{X}^r_{hk} \times \tilde{\boldsymbol{Y}}^r_{hk} \times \tilde{M}^r_{hk}$. We take $\lambda_0 \in L^2(\mathcal{E}^i_h)$ such that $\lambda_0|_e = p_0|_e$ for every $e \in \mathcal{E}^i_h$. This is a natural choice since physically, the Lagrange multiplier approximates the trace of the pressure on the inter-element boundaries.

For the hybrid formulation, the discrete reduced cost functional is defined by

$$j_H(u_{hk}) = J(u_{hk}, p_{hk}(u_{hk}), \boldsymbol{v}_{hk}(u_{hk}))$$

where $(p_{hk}, \boldsymbol{v}_{hk}, \lambda_{hk})^\top \in \tilde{X}^r_{hk} \times \tilde{\boldsymbol{Y}}^r_{hk} \times \tilde{M}^r_{hk}$ is the solution of (3.19), and the corresponding optimal control problem is given as

$$\min_{u_{hk} \in X^{r+1}_{hk}} j_H(u_{hk}). \tag{3.20}$$

Notice that the space X^{r+1}_{hk} is used as a discretization of the control space. The motivation of this choice originates from the case of elliptic PDEs, where the post-processed Lagrange multiplier posseses better convergence properties than the computed scalar state u_{hk}, see [3] for instance. More precisely, instead of using the dual pressure $w_{hk} \in \tilde{X}^r_{hk}$ to update the control in the gradient algorithm, one utilizes the post-processed dual Lagrange multiplier $R^{r+1}_h \mu_{hk} \in X^{r+1}_{hk}$, see the algorithm in Section 5.3 for the details. In this way, the control is an element of X^{r+1}_{hk}.

To compensate for the additional degrees of freedom arising from the inter-element Lagrange multipliers, we will solve the discrete state equations with an additional penalization term, at the expense of an additional error and larger condition number. As a result, the corresponding linear systems at each time step can be reduced. For this purpose, we replace the form a_{hk} by the form

$$\begin{aligned}&\tilde{a}^\varepsilon_{hk}(p_{hk}, \boldsymbol{v}_{hk}, \lambda_{hk}, \varphi_{hk}, \boldsymbol{\psi}_{hk}, \mu_{hk})\\ &= \tilde{a}_{hk}(p_{hk}, \boldsymbol{v}_{hk}, \lambda_{hk}, \varphi_{hk}, \boldsymbol{\psi}_{hk}, \mu_{hk}) + \varepsilon s_{h,I}(\lambda_{hk}, \mu_{hk}).\end{aligned}$$

where $s_{h,I} : M^r_{hk} \times M^r_{hk} \to \mathbb{R}$ is given by

$$s_{h,I}(\lambda_{hk}, \mu_{hk}) := \int_I s_h(\lambda_{hk}(t), \mu_{hk}(t))\,\mathrm{d}t = \sum_{e \in \mathcal{E}^i_h} \frac{1}{|e|} \int_I \int_e \lambda_{hk}(t)\mu_{hk}(t)\,\mathrm{d}s\,\mathrm{d}t.$$

This penalization is adapted from the case of elliptic equations, see [31] for instance. The effect of the penalty parameter ε on the system matrix will be discussed in the numerical section.

With the above penalized discrete state equation, the first-order directional derivative of the discrete cost is given by

$$j'_H(u_{hk})\delta u_{hk} = (\gamma u_{hk} + w_{hk}(u_{hk}), \delta u_{hk})_I,$$

for every $u_{hk}, \delta u_{hk} \in X^{r+1}_{hk}$, where

$$(w_{hk}, \boldsymbol{y}_{hk}, \mu_{hk})^\top = (w_{hk}(u_{hk}), \boldsymbol{y}_{hk}(u_{hk}), \mu_{hk}(u_{hk}))^\top \in \tilde{X}^r_{hk} \times \tilde{\boldsymbol{Y}}^r_{hk} \times \tilde{M}^r_{hk}$$

is the solution of the penalized discrete adjoint equation satisfying the variational equation

$$\begin{aligned}&\tilde{a}^\varepsilon_{hk}(\xi_{hk}, \boldsymbol{\eta}_{hk}, \zeta_{hk}, w_{hk}, \boldsymbol{y}_{hk}, \mu_{hk})\\ &\quad = \alpha(\xi_{hk}, p_{hk}(u_{hk}) - p_d)_I + \beta(\boldsymbol{\eta}_{hk}, \boldsymbol{v}_{hk}(u_{hk}) - \boldsymbol{v}_d)_I\end{aligned} \tag{3.21}$$

for every $(\xi_{hk}, \boldsymbol{\eta}_{hk}, \zeta_{hk})^\top \in X^r_{hk} \times \boldsymbol{Y}^r_{hk} \times M^r_{hk}$. However, instead of solving the above adjoint equation, we replace the discrete pressure $p_{hk}(u_{hk})$ by the post-processed Lagrange multiplier $R^{k+1}_h(\lambda_{hk}(u_{hk}), p_{hk}(u_{hk}))$.

Using a similar process as in the mixed formulation, we have the following time-stepping formulation of (3.19) if we take $\varphi_{hk} = \phi_m \varphi_h$, $\boldsymbol{\psi}_{hk} = \phi_m \boldsymbol{\psi}_h$, and $\mu_{hk} = \phi_m \mu_h$, where $\varphi_h \in X^r_h$, $\boldsymbol{\psi}_h \in \boldsymbol{Y}^r_h$, and $\mu_h \in M^r_h$: For $m = 0$ we have

$$(p^0, \varphi_h) = (p_0, \varphi_h), \quad (\boldsymbol{v}^0, \boldsymbol{\psi}_h) = (\boldsymbol{v}_0, \boldsymbol{\psi}_h), \quad (\lambda^0, \mu_h) = (\lambda_0, \mu_h)$$

for all $\varphi_h \in X^r_h, \boldsymbol{\psi}_h \in \boldsymbol{V}^r_h, \mu_h \in M^r_h$, and for $m = 1, \ldots, M$ we have

$$\begin{cases} \dfrac{1}{\tau}(p^m - p^{m-1}, \varphi_h) - \dfrac{1}{2} b_h(\boldsymbol{v}^m + \boldsymbol{v}^{m-1}, \varphi_h) \\ \qquad\qquad = \dfrac{1}{2}(u^m + u^{m-1}, \varphi_h) \quad \forall \varphi_h \in X^r_h, \\ \dfrac{1}{\tau}(\boldsymbol{v}^m - \boldsymbol{v}^{m-1}, \boldsymbol{\psi}_h) + \dfrac{1}{2} b_h(\boldsymbol{\psi}_h, p^m + p^{m-1}) \\ \qquad\qquad + \dfrac{1}{2} d_h(\boldsymbol{\psi}_h, \lambda^m + \lambda^{m-1}) = 0 \quad \forall \boldsymbol{\psi}_h \in \boldsymbol{Y}^r_h, \\ d_h(\boldsymbol{v}^m + \boldsymbol{v}^{m-1}, \mu_h) - \varepsilon s_h(\lambda^m + \lambda^{m-1}, \mu_h) = 0 \quad \forall \mu_h \in M^r_h. \end{cases} \tag{3.22}$$

Taking $\xi_{hk} = \phi_m \xi_h$, $\boldsymbol{\eta}_{hk} = \phi_m \boldsymbol{\eta}_h$, and $\zeta_{hk} = \phi_m \zeta_h$, where $\xi_h \in X^r_h$, $\boldsymbol{\eta}_h \in \boldsymbol{Y}^r_h$, and $\zeta_h \in M^r_h$ in (3.21), the discrete adjoint equation of the penalized problem is the following: For $m = M$ we have

$$\begin{cases} \dfrac{1}{\tau}(\xi_h, w^M) + \dfrac{1}{2} b_h(\boldsymbol{y}^M, \xi_h) \\ \qquad\qquad = \dfrac{\alpha}{2}(R^{r+1}_h(\lambda^M, p^M) - p^M_d, \xi_h) \quad \forall \xi_h \in X^r_h, \\ \dfrac{1}{\tau}(\boldsymbol{\eta}_h, \boldsymbol{y}^M) - \dfrac{1}{2} b_h(\boldsymbol{\eta}_h, w^M) - \dfrac{1}{2} d_h(\boldsymbol{\eta}_h, \mu^M) \\ \qquad\qquad = \dfrac{\beta}{2}(\boldsymbol{v}^M - \boldsymbol{v}^M_d, \boldsymbol{\eta}_h) \quad \forall \boldsymbol{\eta}_h \in \boldsymbol{Y}^r_h, \\ d_h(\boldsymbol{y}^M, \zeta_h) + \varepsilon s_h(\zeta_h, \mu^M) = 0 \quad \forall \zeta_h \in M^r_h, \end{cases} \tag{3.23}$$

for $m = M - 1, \ldots, 1$ we have

$$\begin{cases} \dfrac{1}{\tau}(\xi_h, w^m - w^{m+1}) + \dfrac{1}{2} b_h(\boldsymbol{y}^m + \boldsymbol{y}^{m+1}, \xi_h) \\ \qquad\qquad = \alpha(R^{r+1}_h(\lambda^m, p^m) - p^m_d, \xi_h) \quad \forall \xi_h \in X^r_h, \\ \dfrac{1}{\tau}(\boldsymbol{\eta}_h, \boldsymbol{y}^m - \boldsymbol{y}^{m+1}) - \dfrac{1}{2} b_h(\boldsymbol{\eta}_h, w^m + w^{m+1}) - \dfrac{1}{2} d_h(\boldsymbol{\eta}_h, \mu^m + \mu^{m+1}) \\ \qquad\qquad = \beta(\boldsymbol{v}^m - \boldsymbol{v}^m_d, \boldsymbol{\eta}_h) \quad \forall \boldsymbol{\eta}_h \in \boldsymbol{Y}^r_h, \\ d_h(\boldsymbol{y}^m + \boldsymbol{y}^{m+1}, \zeta_h) + \varepsilon s_h(\zeta_h, \mu^m + \mu^{m+1}) = 0 \quad \forall \zeta_h \in M^r_h, \end{cases} \tag{3.24}$$

and for $m = 0$ we have

$$\begin{cases} \frac{1}{\tau}(\xi_h, w^0 - w^1) + \frac{1}{2}b_h(\boldsymbol{y}^1, \xi_h) = \frac{\alpha}{2}(R_h^{r+1}(\lambda^0, p^0) - p_d^0, \xi_h) \quad \forall \xi_h \in X_h^r, \\ \frac{1}{\tau}(\boldsymbol{\eta}_h, \boldsymbol{y}^0 - \boldsymbol{y}^1) - \frac{1}{2}b_h(\boldsymbol{\eta}_h, w^1) - \frac{1}{2}d_h(\boldsymbol{\eta}_h, \mu^1) \\ \qquad\qquad = \frac{\beta}{2}(\boldsymbol{v}^0 - \boldsymbol{v}_d^0, \boldsymbol{\eta}_h) \quad \forall \boldsymbol{\eta}_h \in \boldsymbol{Y}_h^r, \\ \frac{\tau}{2}d_h(\boldsymbol{y}^1, \zeta_h) + \frac{\tau}{2}\varepsilon s_h(\zeta_h, \mu^1) + \varepsilon s_h(\zeta_h, \mu^0) = 0 \quad \forall \zeta_h \in M_h^r. \end{cases} \tag{3.25}$$

If $\varepsilon > 0$, then one can see that the systems (3.23)-(3.25) are coercive, hence, existence of solutions follows from the Lax–Milgram Lemma. If $\varepsilon = 0$, then one can prove the existence of a triple $(p^m, \boldsymbol{v}^m, \lambda^m)$ by induction using a similar method as in the elliptic case, see for instance [33]. Furthermore, if the initial discrete velocity is chosen in such a way that $\boldsymbol{v}^0 \in \boldsymbol{V}_h^r$, then it follows that $\boldsymbol{v}^m \in \boldsymbol{V}_h^r$ for every m. In particular, this implies that the pressure and velocity components of the solution to (3.22) with $\varepsilon = 0$ are equal to those of (3.10). Indeed, this follows immediately from the fact that if $\boldsymbol{v}_h \in \boldsymbol{Y}_h^r$, then $d_h(\boldsymbol{v}_h, \mu_h) = 0$ for every $\mu_h \in M_h^r$ if and only if $\boldsymbol{v}_h \in \boldsymbol{V}_h^r$, see [3] for the proof. The said remark also applies to (3.23)-(3.25).

Observe that the above discrete state and adjoint equations, without the penalization term, are the hybridization of the discrete state and adjoint equations in the mixed formulation. This means that discretization and optimization commute even in the hybrid formulation.

4. Error Analysis for the Mixed Formulation

In this section, we prove a priori error estimates for the solution of the continuous optimal control problem (2.6) and its discretization

$$\min_{u_{hk} \in X_{hk}^r} j_{hk}(u_{hk}) := J_{kh}(u_{hk}, p_{hk}(u_{hk}), \boldsymbol{v}_{hk}(u_{hk})) \tag{4.1}$$

where $(p_{hk}(u_{hk}), \boldsymbol{v}_{hk}(u_{hk}))^\top \in X_{hk}^r \times \boldsymbol{V}_{hk}^r$ is the solution of the fully discrete state equation (3.9)-(3.10). To do this, let us also introduce the following semidiscrete optimal control problem

$$\min_{u \in U} j_{hk}(u) := J_{hk}(u, p_{hk}(u), \boldsymbol{v}_{hk}(u)). \tag{4.2}$$

Denote by $\bar{u}_{hk}$ and u_{hk}^* the solutions to (4.1) and (4.2), respectively.

Lemma 4.1. *Let X be a Hilbert space. Then for each $u \in C(\overline{I}, X)$ such that $u|_{I_m} \in \mathcal{P}^1(I_m, X)$ for all $m = 1, \dots, M$, we have*

$$\|u\|_I^2 = \frac{\tau}{3}\|u(0)\|^2 + \frac{\tau}{3}\|u(T)\|^2 + \frac{2\tau}{3}\sum_{m=1}^{M-1}\{\|u(t_m)\|^2 + (u(t_m), u(t_{m-1}))\}. \tag{4.3}$$

In particular, it holds that

$$\frac{\tau}{6}\sum_{m=1}^{M}\|u(t_m) + u(t_{m-1})\|^2 \le \|u\|_I^2. \tag{4.4}$$

Proof. Let $\{\phi_m\}_{m=0}^M$ be the linear Lagrange basis functions on $\overline{I}$ with respect to the above partition so that $\phi_j(t_k) = \delta_{jk}$. Then we can write $u = \sum_{m=0}^M \phi_m u(t_m)$. Using this representation of u along with the identities $\|\phi_m\|^2 = \frac{2\tau}{3}$, $(\phi_m, \phi_{m-1}) = \frac{\tau}{6}$, and $(\phi_m, \phi_\mu) = 0$ for $|m-\mu| > 1$, we obtain (4.3). Expanding the norm in the sum yields

$$\begin{aligned}&\tau \sum_{m=1}^M \|u(t_m) + u(t_{m-1})\|^2 \\ &= \tau\|u(0)\|^2 + 2\tau \sum_{m=1}^M \{\|u(t_m)\|^2 + (u(t_m), u(t_{m-1}))\} - \tau\|u(T)\|^2,\end{aligned}$$

from which the estimate (4.4) follows. □ □

Recall that the solution $(p_{hk}, \boldsymbol{v}_{hk})^\top$ of (3.9)-(3.10) can be written as

$$p_{hk} = \sum_{m=0}^M \phi_m p^m, \qquad \boldsymbol{v}_{hk} = \sum_{m=0}^M \phi_m \boldsymbol{v}^m$$

where $p^m = p_{hk}(t^m)$ and $\boldsymbol{v}^m = \boldsymbol{v}_{hk}(t_m)$. Therefore, we have the estimate

$$\|p_{hk}\|_I + \|\boldsymbol{v}_{hk}\|_I \le C_T \max_{0\le m\le M} \{\|p^m\| + \|\boldsymbol{v}^m\|\}. \tag{4.5}$$

To establish the stability of the semidiscrete state equation, we shall use the discrete analogue of the Gronwall lemma: For nonnegative numbers u_n, v_n, and w_n, if $u_n \le v_n + \sum_{k=0}^{n-1} w_k u_k$, then $u_n \le v_n \exp(\sum_{k=0}^{n-1} w_k)$. For a given control $u \in C(\overline{I}, L^2(\Omega))$, we denote its piecewise-linear Lagrange interpolation by

$$P_I u = \sum_{m=0}^M \phi_m u(t_m).$$

Obviously, $P_I u \in C(\overline{I}, L^2(\Omega))$ and $P_I u|_{I_m} \in \mathcal{P}^1(I_m, L^2(\Omega))$ for each $m = 1, \ldots, M$.

Lemma 4.2. *Let $(p_{hk}(u), \boldsymbol{v}_{hk}(u))^\top$ be the solution of the fully discrete state equation (3.9)-(3.10) with control $u \in W^{1,1}(I, L^2(\Omega))$. Then there exists a constant $C_T > 0$ independent of u such that for every $0 < \tau < T$ we have*

$$\max_{0\le m\le M} \{\|p^m\| + \|\boldsymbol{v}^m\|\} \le C_T(\|p^0\| + \|\boldsymbol{v}^0\| + \|P_I u\|_I). \tag{4.6}$$

In particular, it holds that

$$\|p_{hk}(u)\|_I + \|\boldsymbol{v}_{hk}(u)\|_I \le C_T(\|p^0\| + \|\boldsymbol{v}^0\| + \|P_I u\|_I). \tag{4.7}$$

Proof. Using the test functions $\varphi_h = p^\ell + p^{\ell-1}$ and $\boldsymbol{\psi}_h = \boldsymbol{v}^\ell + \boldsymbol{v}^{\ell-1}$ in the discrete state equation (3.9)-(3.10) and the Cauchy-Schwarz inequality, we obtain

$$\|p^\ell\|^2 + \|\boldsymbol{v}^\ell\|^2 - \|p^{\ell-1}\|^2 - \|\boldsymbol{v}^{\ell-1}\|^2 \le \tau\|u^\ell + u^{\ell-1}\|^2 + 2\tau\{\|p^\ell\|^2 + \|p^{\ell-1}\|^2\}$$

for every $1 \le \ell \le M$. Taking the sum over all $1 \le \ell \le m$ for a given $1 \le m \le M$, one can deduce that

$$\|p^m\|^2 + \|\boldsymbol{v}^m\|^2 \le \|p^0\|^2 + \|\boldsymbol{v}^0\|^2 + \tau \sum_{\ell=1}^m \|u^\ell + u^{\ell-1}\|^2 + \sum_{\ell=0}^m 4\tau\|p^\ell\|^2.$$

Applying the discrete Gronwall lemma, the inequality $\tau(m+1) \leq 2T$ and the fact that $P_I u$ and u coincide at the temporal nodes, we obtain

$$\|p^m\|^2 + \|\boldsymbol{v}^m\|^2 \leq e^{8T}\left(\|p^0\|^2 + \|\boldsymbol{v}^0\|^2 + \tau\sum_{\ell=1}^{m}\|P_I u^\ell + P_I u^{\ell-1}\|^2\right).$$

By Lemma 4.1, we obtain the estimate (4.6) after taking square roots. Finally, (4.7) is a direct consequence of (4.5) and (4.6) □ □

Now let us prove a priori error estimates for the discrete state equations under additional regularity assumptions on the state equations and the control.

Theorem 4.3. *Let $(p_{hk}(u), \boldsymbol{v}_{hk}(u))^\top$ be the solution of the fully discrete state equation (3.9)-(3.10) for a given control $u \in W^{1,1}(I, L^2(\Omega))$, and suppose that $p \in H^3(I, X_0) \cap H^1(I, H^{r+1}(\Omega))$ and $\boldsymbol{v} \in H^3(I, \boldsymbol{V}_0) \cap H^1(I, H^{r+1}(\Omega)^2)$. Then there is a constant $C > 0$ independent of τ and h such that*

$$\|p_{hk}(u) - p(u)\|_I + \|\boldsymbol{v}_{hk}(u) - \boldsymbol{v}(u)\|_I \leq C(\tau^2 + h^{r+1}).$$

Proof. It is enough to prove the following estimate

$$\|p^m - p(t_m)\| + \|\boldsymbol{v}^m - \boldsymbol{v}(t_m)\| \leq C(\tau^2 + h^{r+1})$$

for each $0 \leq m \leq M$. First, let us separate the error into discrete and projection errors according to

$$\begin{aligned} p^m - p(t_m) &= \hat{e}_h^m + \tilde{e}_h^m := (p^m - \pi_h p(t_m)) + (\pi_h p(t_m) - p(t_m)) \\ \boldsymbol{v}^m - \boldsymbol{v}(t_m) &= \hat{\boldsymbol{r}}_h^m + \tilde{\boldsymbol{r}}_h^m := (\boldsymbol{v}^m - \boldsymbol{\rho}_h \boldsymbol{v}(t_m)) + (\boldsymbol{\rho}_h \boldsymbol{v}(t_m) - \boldsymbol{v}(t_m)). \end{aligned}$$

The projection errors can be estimated from above thanks to (3.3) and (3.4) as follows:

$$\|\tilde{e}_h^m\| + \|\tilde{\boldsymbol{r}}_h^m\| \leq Ch^{r+1}(\|p\|_{L^\infty(I,H^{r+1}(\Omega))} + \|\boldsymbol{v}\|_{L^\infty(I,H^{r+1}(\Omega)^2)}). \tag{4.8}$$

On the other hand, for each $1 \leq m \leq M$, the errors $\hat{e}_h^m$ and $\hat{\boldsymbol{r}}_h^m$ satisfy the equations

$$\begin{aligned} &\frac{1}{\tau}(\hat{e}_h^m - \hat{e}_h^{m-1}, \varphi_h) + \frac{1}{\tau}(\hat{\boldsymbol{r}}_h^m - \hat{\boldsymbol{r}}_h^{m-1}, \boldsymbol{\psi}_h) - \frac{1}{2}b(\hat{\boldsymbol{r}}_h^m + \hat{\boldsymbol{r}}_h^{m-1}, \varphi_h) \\ &\qquad + \frac{1}{2}b(\boldsymbol{\psi}_h, \hat{e}_h^m + \hat{e}_h^{m-1}) = -(\varepsilon_{1h}^m, \varphi_h) + (\boldsymbol{\varepsilon}_{2h}^m, \boldsymbol{\psi}_h) - (\boldsymbol{\varepsilon}_{3h}^m, \boldsymbol{\psi}_h) \end{aligned}$$

where the terms on the right hand side are given by

$$\begin{aligned} \varepsilon_{1h}^m &= \frac{1}{\tau}(p(t_m) - p(t_{m-1})) - \frac{1}{2}\partial_t p(t_m) - \frac{1}{2}\partial_t p(t_{m-1}) \\ \boldsymbol{\varepsilon}_{2h}^m &= \frac{1}{\tau}(\boldsymbol{\rho}_h \boldsymbol{v}(t_m) - \boldsymbol{v}(t_m) - \boldsymbol{\rho}_h \boldsymbol{v}(t_{m-1}) + \boldsymbol{v}(t_{m-1})) \\ \boldsymbol{\varepsilon}_{3h}^m &= \frac{1}{\tau}(\boldsymbol{v}(t_m) - \boldsymbol{v}(t_{m-1})) - \frac{1}{2}\partial_t \boldsymbol{v}(t_m) - \frac{1}{2}\partial_t \boldsymbol{v}(t_{m-1}). \end{aligned}$$

By rewriting the term ε_{1h}^m as the integral

$$\varepsilon_{1h}^m = \frac{1}{2\tau}\int_{I_m}(t_m - s)(t_{m-1} - s)\partial_t^3 p(s)\,\mathrm{d}s$$

and similarly for $\boldsymbol{\varepsilon}_{3h}^m$, we have the estimate

$$\|\varepsilon_{1h}^m\|_{I_m}^2 + \|\boldsymbol{\varepsilon}_{3h}^m\|_{I_m}^2 \leq C\tau^3(\|\partial_t^3 p(s)\|_{I_m}^2 + \|\partial_t^3 \boldsymbol{v}(s)\|_{I_m}^2).$$

On the other hand, from interpolation theory we likewise have the estimate

$$\begin{aligned}\|\boldsymbol{\varepsilon}_{2h}^m\|_{I_m}^2 &\leq C\tau^{-2}h^{2(r+1)}\|\boldsymbol{v}(t_m)-\boldsymbol{v}(t_{m-1})\|_{L^2(I_m,H^{r+1}(\Omega)^2)}^2\\ &\leq C\tau^{-1}h^{2(r+1)}\|\partial_t\boldsymbol{v}\|_{L^2(I_m,H^{r+1}(\Omega)^2)}^2.\end{aligned}$$

Taking the test functions $\varphi_h = \hat{e}_h^m + \hat{e}_h^{m-1}$ and $\boldsymbol{\psi}_h = \hat{\boldsymbol{r}}_h^m + \hat{\boldsymbol{\eta}}_h^{m-1}$, and applying the same strategy as in the previous lemma, we obtain

$$\|\hat{e}_h^m\|^2 + \|\hat{\boldsymbol{r}}_h^m\|^2 \leq C_T(\tau^2+h^{r+1})^2 + \sum_{\ell=0}^{m-1} 4\tau\|\hat{e}_h^\ell\|^2,$$

where C_T is a constant depending on $\|\partial_t^3 p(s)\|_I$, $\|\partial_t^3\boldsymbol{v}(s)\|_I$, and $\|\partial_t\boldsymbol{v}\|_{L^2(I,H^{r+1}(\Omega)^2)}$. Therefore, by the discrete Gronwall Lemma, we have

$$\|\hat{e}_h^m\| + \|\hat{\boldsymbol{r}}_h^m\| \leq C(\tau^2+h^{r+1}).$$

Combining this inequality with (4.8) proves the desired estimate. □ □

Recall that $(w_{hk}(u), \boldsymbol{y}_{hk}(u))^\top \in \tilde{X}_{hk}^r \times \tilde{\boldsymbol{V}}_{hk}^r$ is the solution of the corresponding discrete adjoint equation with the pair $(p_{hk}, \boldsymbol{v}_{hk})^\top = (p_{hk}(u), \boldsymbol{v}_{hk}(u))^\top$. One can write

$$w_{hk} = \sum_{m=0}^{M} \chi_{I_m} w^m, \qquad \boldsymbol{y}_{hk} = \sum_{m=0}^{M} \chi_{I_m}\boldsymbol{y}^m,$$

where $w^m = w_{hk}(t^m)$, $\boldsymbol{y}^m = \boldsymbol{y}_{hk}(t_m)$, and χ_{I_m} is the indicator function on I_m, so that

$$\|w_{hk}\|_I + \|\boldsymbol{y}_{hk}\|_I \leq C_T \max_{1\leq m\leq M}\{\|w^m\| + \|\boldsymbol{y}^m\|\}.$$

Lemma 4.4. *Let $(w_{hk}(u), \boldsymbol{y}_{hk}(u))^\top$ be the solution of the fully discrete adjoint equation (3.11)-(3.13). Then there exists a constant $C>0$ independent on u such that*

$$\max_{1\leq m\leq M}\{\|w^m\| + \|\boldsymbol{y}^m\|\} \leq C(\|p_{hk}-p_d\|_I + \|\boldsymbol{v}_{hk}-\boldsymbol{v}_d\|_I),$$

and thus we have

$$\|w_{hk}(u)\|_I + \|\boldsymbol{y}_{hk}(u)\|_I \leq C(\|p_{hk}(u)-p_d\|_I + \|\boldsymbol{v}_{hk}(u)-\boldsymbol{v}_d\|_I)$$

Proof. The proof is similar to the one given for the discrete state equation where we take $(\xi_h, \boldsymbol{\eta}_h) = (w^M, \boldsymbol{y}^M)$ and $(\xi_h, \boldsymbol{\eta}_h) = (w^m + w^{m+1}, \boldsymbol{y}^m + \boldsymbol{y}^{m+1})$ for $1 \leq m \leq M-1$ as the test functions. □ □

Theorem 4.5. *Let $(w_{hk}(u), \boldsymbol{y}_{hk}(u))^\top$ be the solution of the fully discrete adjoint equation (3.11)-(3.13) for a given control $u \in W^{1,1}(I, L^2(\Omega))$, and assume that $w \in H^3(I, X_0)\cap H^1(I, H^{r+1}(\Omega))$ and $\boldsymbol{y} \in H^3(I, \boldsymbol{V}_0)\cap H^1(I, H^{r+1}(\Omega)^2)$. Then there is a constant $C>0$ independent of τ and h such that*

$$\|w_{hk}(u)-w(u)\|_I + \|\boldsymbol{y}_{hk}(u)-\boldsymbol{y}(u)\|_I \leq C(\tau+h^{r+1}).$$

Proof. As in the proof of the previous theorem, it suffices to establish the following a priori estimate at the time nodes

$$\|w^m - w(t_m)\| + \|\boldsymbol{y}^m - \boldsymbol{y}(t_m)\| \leq C(\tau+h^{r+1})$$

for each $1 \leq m \leq M$. Using a similar decomposition as in the case of state equations, we only need to estimate the error terms $\hat{e}_h^m = \pi_h w(t_m) - w^m$ and $\hat{\boldsymbol{r}}_h^m = \boldsymbol{\rho}_h \boldsymbol{y}(t_m) - \boldsymbol{y}^m$.

First, we consider the case where $m = M$. Recall from (2.11) that for each $\xi \in H^1(I, X_0)$ and $\boldsymbol{\eta} \in H^1(I, \boldsymbol{V}_1)$ such that $\xi(0) = 0$ and $\boldsymbol{\eta}(0) = \boldsymbol{0}$, we have

$$(\partial_t \xi, w) + b(\boldsymbol{y}, \xi) + (\partial_t \boldsymbol{\eta}, \boldsymbol{y}) - b(\boldsymbol{\eta}, w) = \alpha(p - p_d, \xi) + \beta(\boldsymbol{v} - \boldsymbol{v}_d, \boldsymbol{\eta}).$$

Taking $\xi = \phi_M \xi_h$, $\boldsymbol{\eta} = \phi_M \boldsymbol{\eta}_h$, evaluating at $t = T$, and using the fact that $w(T) = 0$ and $\boldsymbol{y}(T) = 0$ yields

$$\alpha(p(T) - p_d(T), \xi_h) + \beta(\boldsymbol{v}(T) - \boldsymbol{v}_d(T), \boldsymbol{\eta}_h) = 0,$$

for every $\xi_h \in X_h^r$ and $\boldsymbol{\eta}_h \in \boldsymbol{V}_h^r$. Using this equation together with (3.11) and the fact that $\hat{e}_h^M = -w^M$ and $\hat{\boldsymbol{r}}_h^M = -\boldsymbol{y}^M$, we obtain

$$\|\hat{e}_h^M\|^2 + \|\hat{\boldsymbol{r}}_h^M\|^2 = \frac{\alpha\tau}{2}(p(T) - p_{hk}(T), \hat{e}_h^M) + \frac{\beta\tau}{2}(\boldsymbol{v}(T) - \boldsymbol{v}_{hk}(T), \hat{\boldsymbol{r}}_h^M).$$

Applying the Cauchy-Schwarz inequality and Theorem 4.3, we deduce that

$$\|\hat{e}_h^M\| + \|\hat{\boldsymbol{r}}_h^M\| \leq C\tau^2.$$

In the case where $1 \leq m \leq M - 1$, we follow the same method as in the fully discrete state equation to get

$$\begin{aligned}
&\frac{1}{\tau}(\xi_h, \hat{e}_h^m - \hat{e}_h^{m+1}) + \frac{1}{\tau}(\boldsymbol{\eta}_h, \hat{\boldsymbol{r}}_h^m - \hat{\boldsymbol{r}}_h^{m+1}) + \frac{1}{2} b(\hat{\boldsymbol{r}}_h^m + \hat{\boldsymbol{r}}_h^{m+1}, \xi_h) \\
&\quad - \frac{1}{2} b(\boldsymbol{\eta}_h, \hat{e}_h^m + \hat{e}_h^{m+1}) = (\xi_h, \varepsilon_{1h}^m) + (\boldsymbol{\eta}_h, \boldsymbol{\varepsilon}_{2h}^m) + (\boldsymbol{\eta}_h, \boldsymbol{\varepsilon}_{3h}^m) \\
&\quad + \alpha(p(t_m) - p^m, \xi_h) + \beta(\boldsymbol{v}(t_m) - \boldsymbol{v}^m, \boldsymbol{\eta}_h) \\
&\quad + \frac{\alpha}{2}(\partial_t p(t_m) - \partial_t p_d(t_m), \xi_h) + \frac{\beta}{2}(\partial_t \boldsymbol{v}(t_m) - \partial_t \boldsymbol{v}_d(t_m), \boldsymbol{\eta}_h)
\end{aligned}$$

where ε_{1h}^m, $\boldsymbol{\varepsilon}_{2h}^m$, and $\boldsymbol{\varepsilon}_{3h}^m$ are defined by

$$\begin{aligned}
\varepsilon_{1h}^m &= \frac{1}{\tau}(w(t_m) - w(t_{m+1})) - \frac{1}{2}\partial_t w(t_m) - \frac{1}{2}\partial_t w(t_{m+1}) \\
\boldsymbol{\varepsilon}_{2h}^m &= \frac{1}{\tau}(\boldsymbol{\rho}_h \boldsymbol{y}(t_m) - \boldsymbol{y}(t_m) - \boldsymbol{\rho}_h \boldsymbol{y}(t_{m+1}) + \boldsymbol{y}(t_{m+1})) \\
\boldsymbol{\varepsilon}_{3h}^m &= \frac{1}{\tau}(\boldsymbol{y}(t_m) - \boldsymbol{y}(t_{m+1})) - \frac{1}{2}\partial_t \boldsymbol{y}(t_m) - \frac{1}{2}\partial_t \boldsymbol{y}(t_{m+1}).
\end{aligned}$$

With these quantities, one can now proceed as before to establish the desired a priori estimate by taking the test functions $\xi_h = \hat{e}_h^m + \hat{e}_h^{m+1}$ and $\boldsymbol{\eta}_h = \hat{\boldsymbol{r}}_h^m + \hat{\boldsymbol{r}}_h^{m+1}$. The main difference is that the last two terms in the above equation are estimated from above as

$$\begin{aligned}
&\tau(\|\partial_t p(t_m) - \partial_t p_d(t_m)\| + \|\partial_t \boldsymbol{v}(t_m) - \partial_t \boldsymbol{v}_d(t_m)\|) \\
&\qquad \leq C\tau(\|p - p_d\|_{H^1(I_{m+1}, X_0)} + \|\boldsymbol{v} - \boldsymbol{v}_d\|_{H^1(I_{m+1}, X_0)^2}).
\end{aligned}$$

This leads to a linear order with respect to τ. □ □

Theorem 4.6. *Let $(\bar{u}, \bar{p}, \bar{\boldsymbol{v}})^\top$ and $(\bar{u}_{hk}, \bar{p}_{hk}, \bar{\boldsymbol{v}}_{hk})^\top$ be the solutions of the continuous and fully discrete optimal control problems (2.6) and (4.1), respectively. Suppose that*

$\bar{u} \in H^1(I, H^{r+1}(\Omega))$, $\bar{p}, \bar{w} \in H^3(I, X_0) \cap H^1(I, H^{r+1}(\Omega))$, *and* $\bar{\boldsymbol{v}}, \bar{\boldsymbol{y}} \in H^3(I, \boldsymbol{V}_0) \cap H^1(I, H^{r+1}(\Omega)^2)$. *Then*

$$\|\bar{u} - \bar{u}_{hk}\|_I + \|\bar{p} - \bar{p}_{hk}\|_I + \|\bar{\boldsymbol{v}} - \bar{\boldsymbol{v}}_{hk}\|_I \leq C(\tau + h^{r+1}).$$

Moreover, if $(\bar{w}, \bar{\boldsymbol{y}})^\top$ *and* $(\bar{w}_{hk}, \bar{\boldsymbol{y}}_{hk})^\top$ *are the corresponding optimal adjoint states, then*

$$\|\bar{w} - \bar{w}_{hk}\|_I + \|\bar{\boldsymbol{y}} - \bar{\boldsymbol{y}}_{hk}\|_I \leq C(\tau + h^{r+1}).$$

Proof. The proof follows from the stability estimates given Lemmas 4.2 and 4.4, the a priori error estimates in Theorems 4.3 and 4.5, and by adapting the methodologies presented in [28] in the case of parabolic equations.

Let P_{hk}^r be the orthogonal projection from U onto X_{hk}^r and let $\hat{u}_{hk} = P_{hk}^r \bar{u}$. Applying the error estimates (3.3) and (3.4), we have

$$\|\hat{u}_{hk} - \bar{u}\|_I \leq C(\tau^2 + h^{r+1}). \tag{4.9}$$

By optimality, we deduce that

$$j'(\bar{u})(\hat{u}_{hk} - \bar{u}_{hk}) = j'_{hk}(u^*_{hk})(\hat{u}_{hk} - \bar{u}_{hk}) = j'_{hk}(\bar{u}_{hk})(\hat{u}_{hk} - \bar{u}_{hk}) = 0. \tag{4.10}$$

According to the linear-quadratic structure of the optimal control problem and (4.10), we obtain

$$\begin{aligned}
\gamma\|\hat{u}_{hk} - \bar{u}_{hk}\|_I^2 &\leq j''_{kh}(u^*_{hk})(\hat{u}_{hk} - \bar{u}_{hk}, \hat{u}_{hk} - \bar{u}_{hk}) \\
&= j'_{kh}(\hat{u}_{hk})(\hat{u}_{hk} - \bar{u}_{hk}) - j'_{kh}(\bar{u}_{hk})(\hat{u}_{hk} - \bar{u}_{hk}) \\
&= j'_{kh}(\hat{u}_{hk})(\hat{u}_{hk} - \bar{u}_{hk}) - j'_{kh}(\bar{u})(\hat{u}_{hk} - \bar{u}_{hk}) + j'_{kh}(\bar{u})(\hat{u}_{hk} - \bar{u}_{hk}) \\
&\quad - j'(\bar{u})(\hat{u}_{hk} - \bar{u}_{hk}).
\end{aligned}$$

Utilizing the representations of discrete reduced cost in terms of the adjoint variable w_{hk}, we have the following estimate

$$\begin{aligned}
&j'_{kh}(\hat{u}_{hk})(\hat{u}_{hk} - \bar{u}_{hk}) - j'_{kh}(\bar{u})(\hat{u}_{hk} - \bar{u}_{hk}) \\
&\quad = (\gamma u_{hk} + w_{hk}(u_{hk}), \hat{u}_{hk} - \bar{u}_{hk})_I - (\gamma\bar{u} + w_{hk}(\bar{u}), \hat{u}_{hk} - \bar{u}_{hk})_I \\
&\quad \leq \{\gamma\|\hat{u}_{hk} - \bar{u}\|_I + \|w_{hk}(\hat{u}_{hk}) - w_{hk}(\bar{u})\|_I\}\|\hat{u}_{hk} - \bar{u}_{hk}\|_I.
\end{aligned}$$

Using the equations satisfied by $w_{hk}(\hat{u}_{hk})$ and $w_{hk}(\bar{u})$, one has

$$\begin{aligned}
\|w_{hk}(\hat{u}_{hk}) - w_{hk}(\bar{u})\|_I &\leq C(\|p_{hk}(\hat{u}_{hk}) - p_{hk}(\bar{u})\|_I + \|\boldsymbol{v}_{hk}(\hat{u}_{hk}) - \boldsymbol{v}_{hk}(\bar{u})\|_I) \\
&\leq C\|P_I(\hat{u}_{hk} - \bar{u})\|_I \\
&= C(\|\hat{u}_{hk} - \bar{u}\|_I + \|\bar{u} - P_I\bar{u}\|_I) \\
&\leq C(\|\hat{u}_{hk} - \bar{u}\|_I + \tau\|\bar{u}\|_{H^1(I,L^2(\Omega))}),
\end{aligned}$$

where C is a constant independent of h and k. Similarly, it holds that

$$\begin{aligned}
&j'_{kh}(\bar{u})(\hat{u}_{hk} - \bar{u}_{hk}) - j'(\bar{u})(\hat{u}_{hk} - \bar{u}_{hk}) \\
&\quad = (\gamma\bar{u} + w_{hk}(\bar{u}), \hat{u}_{hk} - \bar{u}_{hk}) - (\gamma\bar{u} + w(\bar{u}), \hat{u}_{hk} - \bar{u}_{hk}) \\
&\quad \leq C\|w(\bar{u}) - w_{hk}(\bar{u})\|_I\|\hat{u}_{hk} - \bar{u}_{hk}\|_I.
\end{aligned}$$

Therefore, by the triangle inequality,

$$\|\hat{u}_{hk} - \bar{u}_{hk}\|_I \leq C_\gamma(\tau + \|\hat{u}_{hk} - \bar{u}\|_I + \|w(\bar{u}) - w_{hk}(\bar{u})\|_I).$$

According to Theorem 4.5 and the estimate (4.9), it holds that

$$\|\bar{u}_{hk} - \bar{u}\|_I \leq C_\gamma(\tau + h^{r+1}). \tag{4.11}$$

The error estimate for the optimal states can be established by writing $(\bar{p} - \bar{p}_{hk}, \bar{\boldsymbol{v}} - \bar{\boldsymbol{v}}_{hk})^\top = (p(\bar{u}) - p_{hk}(\bar{u}), \boldsymbol{v}(\bar{u}) - \boldsymbol{v}_{hk}(\bar{u}))^\top + (p_{hk}(\bar{u}) - p_{hk}(\bar{u}_{hk}), \boldsymbol{v}_{hk}(\bar{u}) - \boldsymbol{v}_{hk}(\bar{u}_{hk}))^\top$, the stability estimates for the discrete state equations given in Lemma 4.2, and the error estimate (4.11) for the optimal controls. Similar decompositions for the optimal adjoint states can be done to prove the corresponding a priori error estimates. □ □

5. Implementation with the Lowest-Order Raviart–Thomas Elements

In this section, we present the corresponding linear systems for the proposed Petrov–Galerkin mixed and hybrid finite element discretization of the state and adjoint equations. A gradient algorithm approximating the solutions of the fully discretized reduced problem will also be given.

5.1. Linear Systems. Consider a fixed triangulation $\mathcal{T}_h$. Let $\mathcal{T}_h = \{K_{jh} : 1 \leq j \leq N_h\}$ be the list of triangles in $\mathcal{T}_h$ and let $\mathcal{E}_h = \{e_{\ell h} : 1 \leq \ell \leq M_h\}$ be the list of edges with a given fixed global orientation. Given an edge $e_{\ell h}$, there exist two triangles say $K_{j(\ell)h}$ and $K_{i(\ell)h}$ sharing the common edge $e_{\ell h}$ or there is one element $K_{j(\ell)h}$ containing it. Let $\mathbf{x}_{j(\ell)h}$ and $\mathbf{x}_{i(\ell)h}$ be the nodes in $K_{j(\ell)h}$ and $K_{i(\ell)h}$ opposite to $e_{\ell h}$, respectively. Let $K_{j(\ell)h}$ be the element that contains $e_{\ell h}$ having the same orientation with $e_{\ell h}$ and let $\boldsymbol{\nu}_{\ell h}$ be the unit normal inward to $K_{j(\ell)h}$, hence outward to $K_{i(\ell)h}$. On a boundary edge we set $\boldsymbol{\nu}_{\ell h} = \boldsymbol{\nu}$, where $\boldsymbol{\nu}$ is the unit outward normal to $\partial\Omega$.

First, let us introduce a basis for the lowest-order Raviart–Thomas finite element space, see [4, 20] for instance. Define

$$\boldsymbol{\psi}_{\ell h}(\mathbf{x}) = \frac{1}{2}|e_{\ell h}||K_{j(\ell)h}|^{-1}(\mathbf{x} - \mathbf{x}_{j(\ell)h})\chi_{K_{j(\ell)h}} - \frac{1}{2}|e_{\ell h}||K_{i(\ell)h}|^{-1}(\mathbf{x} - \mathbf{x}_{i(\ell)h})\chi_{K_{i(\ell)h}},$$

where χ_A denotes the characteristic function of a set A. In the case where there is only one element containing the edge $e_{\ell h}$, we ignore the second term in the above definition. Then $\{\boldsymbol{\psi}_{\ell h} : 1 \leq \ell \leq M_h\}$ forms a basis for $\boldsymbol{V}_h^0$. On the other hand, letting $\varphi_{jh} = \chi_{K_{jh}}$, the set $\{\varphi_{jh} : 1 \leq j \leq N_h\}$ forms a basis for X_h^0. Let $C_h \in \mathbb{R}^{N_h \times N_h}$, $B_h \in \mathbb{R}^{M_h \times N_h}$, and $A_h \in \mathbb{R}^{M_h \times M_h}$, be the matrices with the following components

$$(C_h)_{ij} = (\varphi_{ih}, \varphi_{jh}), \quad (B_h)_{i\ell} = (\varphi_{ih}, \operatorname{div} \boldsymbol{\psi}_{\ell h}), \quad (A_h)_{\ell k} = (\boldsymbol{\psi}_{\ell h}, \boldsymbol{\psi}_{kh}).$$

Now we can write the resulting linear system for the fully discrete state equation. For $m = 1, \ldots, M$, equation (3.10) can be expressed as

$$\begin{pmatrix} C_h & -\frac{\tau}{2}B_h^\top \\ \frac{\tau}{2}B_h & A_h \end{pmatrix} \begin{pmatrix} p^m \\ \boldsymbol{v}^m \end{pmatrix} = \begin{pmatrix} C_h & \frac{\tau}{2}B_h^\top \\ -\frac{\tau}{2}B_h & A_h \end{pmatrix} \begin{pmatrix} p^{m-1} \\ \boldsymbol{v}^{m-1} \end{pmatrix} + \frac{\tau}{2} \begin{pmatrix} \tau C_h(u^m + u^{m-1}) \\ \mathbf{0} \end{pmatrix}.$$

This system may be solved by various methods such as LU factorization. Alternatively, a reduction can be done by eliminating p^m and then substituting it in the second equation. Performing this process, we obtain the following equation for $\boldsymbol{v}^m$

$$R_h^+ \boldsymbol{v}^m = R_h^- \boldsymbol{v}^{m-1} - \frac{\tau^2}{4} B_h (u^m + u^{m-1}) - \tau B_h p^{m-1}$$

and after solving for $\boldsymbol{v}^m$, we can determine p^m from

$$p^m = p^{m-1} + \frac{\tau}{2} C_h^{-1} B^\top (\boldsymbol{v}^m + \boldsymbol{v}^{m-1}) + \frac{\tau}{2} (u^m + u^{m-1})$$

where

$$R_h^\pm = A_h \pm \frac{\tau^2}{4} B_h C_h^{-1} B_h^\top.$$

Note that C_h is a diagonal matrix, hence it can be easily inverted. In fact, the entries of C_h are given by $(C_h)_{ij} = \delta_{ij} |K_{ih}|$. Thus, in the case of uniform triangulations, $C_h = |K| I_{N_h}$ where $|K|$ denotes the common area of triangles in the mesh. Also, since A_h is symmetric and positive-definite, so is the matrix R_h^+. Therefore, we can solve for $\boldsymbol{v}^m$ via the conjugate gradient (CG) method.

For the approximations of the desired states, we take p_{dh} and $\boldsymbol{v}_{dh}$ to be the projections of p_d and $\boldsymbol{v}_d$ in X_{hk}^0 and $\boldsymbol{V}_{hk}^0$, respectively. By doing a similar procedure for the fully discrete adjoint equations (3.11)-(3.13), we obtain the following: For $m = M$ we have

$$\begin{cases} R_h^+ \boldsymbol{y}^M = \dfrac{\beta}{2} \tau A_h (\boldsymbol{v}^M - \boldsymbol{v}_d^M) + \dfrac{\alpha}{4} \tau^2 B_h (p^M - p_d^M) \\ w^M = \dfrac{\alpha}{2} \tau (p^M - p_d^M) - \dfrac{\tau}{2} C_h^{-1} B_h^\top \boldsymbol{y}^M, \end{cases}$$

for $m = M-1, \ldots, 1$ we have

$$\begin{cases} R_h^+ \boldsymbol{y}^m = R_h^- \boldsymbol{y}^{m+1} + \tau B_h w^m + \beta \tau A_h (\boldsymbol{v}^m - \boldsymbol{v}_d^m) + \dfrac{\alpha}{2} \tau^2 B_h (p^m - p_d^m) \\ w^m = w^{m+1} - \dfrac{\tau}{2} C_h^{-1} B_h^\top (\boldsymbol{y}^m + \boldsymbol{y}^{m+1}) + \alpha \tau (p^m - p_d^m), \end{cases}$$

and for $m = 0$ we have

$$\begin{cases} \boldsymbol{y}^0 = \boldsymbol{y}^1 + \dfrac{\tau}{2} A_h^{-1} B_h w^1 + \dfrac{\beta}{2} \tau (\boldsymbol{v}^0 - \boldsymbol{v}_d^0) \\ w^0 = w^1 - \dfrac{\tau}{2} C_h^{-1} B_h^\top \boldsymbol{y}^1 + \dfrac{\alpha}{2} \tau (p^0 - p_d^0). \end{cases}$$

The discretized cost functional can be computed through, see Lemma 4.1,

$$\begin{aligned} j_{kh}(u_{kh}, p_{kh}, \boldsymbol{v}_{kh}) &= \frac{\tau}{6} \{\alpha (C_h p^0, p^0) + \beta (A_h \boldsymbol{v}^0, \boldsymbol{v}^0) + (C_h u^0, u^0)\} \\ &+ \frac{\tau}{6} \sum_{m=1}^{M} \{\alpha (C_h (p^m + p^{m-1}), p^m) + \beta (A_h (\boldsymbol{v}^m + \boldsymbol{v}^{m-1}), \boldsymbol{v}^m) \\ &\qquad + (C_h (u^m + u^{m-1}), u^m)\}. \end{aligned}$$

5.2. Hybridization. Now we discuss the corresponding linear systems for the (regularized) hybrid formulation. For this approach, we use the following basis functions (see [4])

$$\begin{aligned}\tilde{\boldsymbol{\psi}}_{3j+1,h}(x,y) &= (x,0)^\top \chi_{K_j},\\ \tilde{\boldsymbol{\psi}}_{3j+2,h}(x,y) &= (0,y)^\top \chi_{K_j},\\ \tilde{\boldsymbol{\psi}}_{3j+3,h}(x,y) &= (x-\bar{x}_j, y-\bar{y}_j)^\top \chi_{K_j}\end{aligned}$$

for $j = 1,\dots,N_h$, where $(\bar{x}_j,\bar{y}_j)^\top$ is the centroid of the triangle K_j in the mesh $\mathcal{T}_h$. The set $\{\tilde{\boldsymbol{\psi}}_{jh} : 1 \le j \le 3N_h\}$ forms a basis for $\boldsymbol{Y}_h^0$. Let $\tilde{M}_h$ be the number of interior edges and for simplicity, we suppose that they are the first $\tilde{M}_h$ in the list of edges $\mathcal{E}_h$. This is of course not necessary in the implementation. Define $\tilde{A}_h \in \mathbb{R}^{3N_h\times 3N_h}$, $\tilde{B}_h \in \mathbb{R}^{3N_h\times N_h}$, and $D_h \in \mathbb{R}^{3N_h\times \tilde{M}_h}$ by

$$(\tilde{A}_h)_{k\ell} = (\tilde{\boldsymbol{\psi}}_{hk}, \tilde{\boldsymbol{\psi}}_{\ell h}), \quad (\tilde{B}_h)_{j\ell} = (\varphi_{\ell h}, \operatorname{div}\tilde{\boldsymbol{\psi}}_{jh}), \quad (D_h)_{i\ell} = \int_{\partial K(\tilde{\boldsymbol{\psi}}_\ell)} \tilde{\boldsymbol{\psi}}_\ell \cdot \boldsymbol{\nu}\chi_{e_i}\,\mathrm{d}x$$

for $1 \le k,\ell \le 3N_h$, $1 \le j \le N_h$, $1 \le i,r \le \tilde{M}_h$, and $K(\tilde{\boldsymbol{\psi}}_\ell)$ is the element containing the support of $\tilde{\boldsymbol{\psi}}_\ell$. Furthermore, from the basis $\{\tilde{\varphi}_{hj}\}_{j=1}^{3N_h}$ of X_h^1 consisting of discontinuous linear Langrange elements on $\mathcal{T}_h$, we define the matrix $F_h \in \mathbb{R}^{N_h\times 3N_h}$ by

$$(F_h)_{i\ell} = (\varphi_{hi}, \tilde{\varphi}_{h\ell}).$$

Before we proceed, let us discuss an efficient implementation of post-processing the Langrange multiplier. For each element K_j, denote by $e_{hj(k)}$ for $k = 1,2,3$ the sides of K_j following the given orientation of the element. Given $\lambda_h \in M_h^0$, let $\tilde{\lambda}_h \in \mathbb{R}^{3N_h}$ satisfy

$$\tilde{\lambda}_{h,3j+k} = \lambda_{hj(k)}$$

for each $j = 1,\dots,N_h$ and $k = 1,2,3$. By introducing the $N_h \times 3N_h$ matrix

$$L_h = I_{N_h} \otimes \begin{pmatrix} -1 & 1 & 1\\ 1 & -1 & 1\\ 1 & 1 & -1\end{pmatrix},$$

where I_{N_h} is the identity matrix of size N_h and $\otimes$ is the Kronecker product, one can easily check that

$$R_h^1\lambda_h = L_h\tilde{\lambda}_h.$$

Following the same procedure as in the mixed formulation, the linear system for the regularized hybridized discrete state equations (3.22) is

$$\begin{cases} \tilde{R}_h^+ \boldsymbol{v}^m = \tilde{R}_h^- \boldsymbol{v}^{m-1} - \dfrac{\tau^2}{4}\tilde{B}_h F_h(u^m + u^{m-1}) - \tau\tilde{B}_h p^{m-1}\\ p^m = p^{m-1} + \dfrac{\tau}{2}C_h^{-1}\tilde{B}^\top(\boldsymbol{v}^m + \boldsymbol{v}^{m-1}) + \dfrac{\tau}{2}F_h(u^m + u^{m-1})\\ \lambda^m = -\lambda^{m-1} + \dfrac{1}{\varepsilon}D_h^\top(\boldsymbol{v}^m + \boldsymbol{v}^{m-1})\end{cases}$$

for $m = 1,\dots,M$, where

$$\tilde{R}_h^\pm = \tilde{A}_h \pm \frac{\tau^2}{4}\tilde{B}_h C_h^{-1}\tilde{B}_h^\top \pm \frac{\tau}{2\varepsilon}D_h D_h^\top.$$

The approximations of the desired states that we take are the projections p_{dh} and $\boldsymbol{v}_{dh}$ of p_d and $\boldsymbol{v}_d$ in X^1_{hk} and $\boldsymbol{V}^0_{hk}$, respectively. Similar representations can be obtained from the hybridized adjoint equations: For $m = M$ we have

$$\begin{cases} \tilde{R}_h^+ \boldsymbol{y}^M = \dfrac{\beta}{2}\tau \tilde{A}_h(\boldsymbol{v}^M - \boldsymbol{v}_d^M) + \dfrac{\alpha}{4}\tau^2 \tilde{B}_h F_h(L_h\tilde{\lambda}^M - p_d^M) \\ w^M = \dfrac{\alpha}{2}\tau F_h(L_h\tilde{\lambda}^M - p_d^M) - \dfrac{\tau}{2}C_h^{-1}B_h^\top \boldsymbol{y}^M \\ \mu^M = -\dfrac{1}{\varepsilon}D_h^\top \boldsymbol{y}^M \end{cases}$$

for $m = M-1, \ldots, 1$ we have

$$\begin{cases} \tilde{R}_h^+ \boldsymbol{y}^m = \tilde{R}_h^- \boldsymbol{y}^{m+1} + \tau \tilde{B}_h w^m + \beta\tau \tilde{A}_h(\boldsymbol{v}^m - \boldsymbol{v}_d^m) + \dfrac{\alpha}{2}\tau^2 \tilde{B}_h F_h(L_h\tilde{\lambda}^m - p_d^m) \\ w^m = w^{m+1} - \dfrac{\tau}{2}C_h^{-1}\tilde{B}_h^\top(\boldsymbol{y}^m + \boldsymbol{y}^{m+1}) + \alpha\tau F_h(L_h\tilde{\lambda}^m - p_d^m) \\ \mu^m = -\mu^{m+1} - \dfrac{1}{\varepsilon}D_h^\top(\boldsymbol{y}^{m+1} + \boldsymbol{y}^m) \end{cases}$$

and for $m = 0$ we have

$$\begin{cases} \boldsymbol{y}^0 = \boldsymbol{y}^1 + \dfrac{\tau}{2}\tilde{A}_h^{-1}(\tilde{B}_h w^1 + D_h\mu^1) + \dfrac{\beta}{2}\tau(\boldsymbol{v}^0 - \boldsymbol{v}_d^0) \\ w^0 = w^1 + \dfrac{\alpha}{2}\tau F_h(L_h\tilde{\lambda}^0 - p_d^0) - \dfrac{\tau}{2}C_h^{-1}\tilde{B}_h^\top \boldsymbol{y}^1 \\ \mu^0 = -\dfrac{\tau}{2}\mu^1 - \dfrac{\tau}{2\varepsilon}D_h^\top \boldsymbol{y}^1. \end{cases}$$

The corresponding reduced cost can be also calculated as in the mixed case, however, we replace the discrete pressure p_{hk} by the post-processed Lagrange multiplier $R_h^1\lambda_{hk}$.

It is important to point out that the efficient solution for $\boldsymbol{v}^m$ by the CG method requires the use of a preconditioner. In our implementation, we use $\tilde{R}_h^+$ without its last term as preconditioner, that is,

$$P_h = \tilde{A}_h + \frac{\tau^2}{4}\tilde{B}_h C_h^{-1}\tilde{B}_h^\top. \tag{5.1}$$

The choice of our basis functions for $\boldsymbol{Y}_h^0$ and X_h^0 imply that the matrices $\tilde{A}_h$ and $\tilde{B}_h$ are diagonal, hence, the preconditioner P_h is also diagonal and therefore P_h^{-1} can be easily computed. This leads to cheaper computations in the use of the preconditioner for the CG method.

5.3. Optimization. We present the algorithm for numerically solving the discrete optimization problem using the Barzilai-Borwein (BB) version of the gradient method in [7]. We only present the hybridized formulation, the case of mixed formulation being similar. In the following, we denote by $(p_{hk}^i, \boldsymbol{v}_{hk}^i, \lambda_{hk}^i)^\top$ and $(w_{hk}^i, \boldsymbol{y}_{hk}^i, \mu_{hk}^i)^\top$ the discrete state and adjoint variables corresponding to the control u_{hk}^i in the ith iteration.

The BB gradient method can be viewed as the secant version of Newton's method and is known to be superlinearly convergent in the quadratic case for two-dimensions [7]. In step 5, the second iteration for the gradient method is taken from the steepest descent method. Alternatively, one may consider an inexact line search with a suitable steplength selection criterion, for example, Armijo's rule. The steplengths

Algorithm: BB Gradient Method for (3.20) with Post-processing

1 Initialize the control u_{hk}^0, tolerance $0 < \epsilon \ll 1$, iteration number $i = 0$, and maximum number of iterations $i_{\max} \gg 1$.
2 Compute the discrete state $(p_{hk}^i, \boldsymbol{v}_{hk}^i, \lambda_{hk}^i)^\top$.
3 Compute the discrete adjoint $(w_{hk}^i, \boldsymbol{y}_{hk}^i, \mu_{hk}^i)^\top$.
4 Set $g^i = -(\gamma u_{hk}^i + \Pi_{hk} R_h^1 \mu_{hk}^i)$.
5 If $i = 0$, then take $s_i = 1$. Otherwise, take

$$s_i = \begin{cases} (u_{hk}^i - u_{hk}^{i-1}, g^i - g^{i-1})/\|g^i - g^{i-1}\|^2 & \text{if } i \text{ is even,} \\ \|u_{hk}^i - u_{hk}^{i-1}\|^2/(u_{hk}^i - u_{hk}^{i-1}, g^i - g^{i-1}) & \text{if } i \text{ is odd.} \end{cases}$$

6 Set $u_{hk}^{i+1} = u_{hk}^i + s_i g^i$ and solve $(p_{hk}^{i+1}, \boldsymbol{v}_{hk}^{i+1}, \lambda_{hk}^{i+1})^\top$.
7 Compute the cost $j^i = J_{hk}(u_{hk}^{i+1}, R_h^1 \lambda_{hk}^{i+1}, \boldsymbol{v}_{hk}^{i+1})^\top$.
8 If $i > 0$ and $|j^i - j^{i-1}|/j^i < \epsilon$, then stop. Otherwise, increment i and return to step 3. If $i = i_{\max}$, then terminate algorithm.

in step 5 are alternately computed, however, one may choose either of the given formulas for all iterates. Numerical experiments in [17] shows better performance when using alternate BB stepsizes instead of a single one. The steplengths are calculated with respect to the Euclidean inner product and norm. In the computation of the gradients, we used the post-processed Lagrange multiplier of the discrete adjoint equation instead of the dual pressure.

We use the linear interpolant determined by the values of $R_h^1 \mu_{hk}^i$ at the time nodes, that is, one may replace Π_{hk} in step 4 by the operator $\tilde{\Pi}_{hk}$ defined by

$$\tilde{\Pi}_{hk} R_h^1 \mu_{hk}^i = \sum_{m=0}^{M} \phi_j R_h^1 \mu_{hk}^i(t_m).$$

Finally, an alternative stopping criterion is $\|\gamma u_{hk}^i + \tilde{\Pi}_{hk} R_h^1 \mu_{hk}^i\|_I < \epsilon$, that is, when the optimality residual is less than the prescribed tolerance.

6. Numerical Examples

In this section, we present numerical examples illustrating the performance of the above schemes. In all examples, we utilized a uniform triangulation of the unit square $\Omega = (0,1)^2$ and a final time $T = 1$. The algorithm presented in the previous section was implemented in Python 3.7.6 (Python Software Foundation, https://www.python.org/) on a 2.3 GHz Intel Core i5 with 8 GB RAM. The repository for the source codes and the iteration histories is available at https://github.com/grperalta/pgrtwave.

Example 1. We partition the time domain $[0,T] = [0,1]$ into a uniform grid with stepsize $\tau = 0.01$, and the spatial domain Ω with mesh size $h = \sqrt{2}/20$, which corresponds to a triangulation with 441 nodes, 800 elements, and 1240 edge elements. The parameters in the cost functional are $\alpha = 10$, $\beta = 1$, and $\gamma = 10^{-5}$. Each linear system is solved by the CG method, with the preconditioner (5.1) in the

hybrid case, and stop the loop if the relative error between successive iterates is less than 10^{-12}. The BB gradient algorithm is terminated if the relative error between consecutive cost function values is less than 10^{-6}.

For the target states, we take the following functions

$$\begin{aligned} p_d(t,x) &= \cos(\pi t)\sin(2\pi x)\sin(2\pi y) \\ v_{d1}(t,x) &= 2(1+\sin(\pi t))\cos(2\pi x)\sin(2\pi y) \\ v_{d2}(t,x) &= 2(1+\sin(\pi t))\sin(2\pi x)\cos(2\pi y). \end{aligned}$$

In the mixed method, the gradient algorithm converges after 317 iterations with optimality residual $\|\gamma u_{kh} + w_{kh}\|_I \approx 1.953614 \cdot 10^{-4}$ and a relative error $6.842802 \cdot 10^{-7}$. On the other hand, for the hybrid method with penalization parameter $\varepsilon = 10^{-10}$, the algorithm terminated after 179 iterations with a relative error $9.718834 \cdot 10^{-7}$ and optimality residual $\|\gamma u_{kh} + \tilde{\Pi}_{hk} R_h^1 \mu_{kh}\|_I \approx 2.074387 \cdot 10^{-4}$.

The corresponding optimal costs are given by $j^* \approx 5.567564 \cdot 10^{-2}$ in the mixed formulation and $j^* \approx 1.425830 \cdot 10^{-3}$ in the hybrid formulation. Therefore, in this example, the hybrid method performs significantly better than the mixed method with respect to the cost value and number of gradient iterations, however, at the cost of additional computing time. Even though the hybrid method requires less BB iterations, each iteration takes a longer time compared to the mixed method. This is due to the fact that the linear systems have larger dimensions, hence more iterations are needed in the preconditioned CG method at each time step in the solution of the primal and dual state variables. On the average, the computing time of the hybrid method is around twice slower than that of the mixed method.

The performance of the BB gradient method is presented in Figure 1. Here, we notice from part (b) the non-monotone property of method with respect to the gradient norms. The first few iterations lead to a fast decrease in the cost values and stagnates afterwards (a), which is typical for certain gradient-based algorithms. Note the oscillatory behavior of the norm of the computed optimal control as a function of time in the hybrid formulation and accordingly the oscillations in the residual of the post-processed optimal and target states. Based on this observation, temporal averaging on the computed pressure and Lagrange multiplier in the state equation and the associated dual variables in the adjoint equation were utilized (see plots at the bottom of Figure 1). With these additional processes, the hybrid formulation converges after 201 iterations, with $j^* \approx 1.484320 \cdot 10^{-3}$ and $\|\gamma u_{kh} + \tilde{\Pi}_{hk} R_h^1 \mu_{kh}\|_I \approx 1.115701 \cdot 10^{-4}$.

Example 2. In this example, we verify the a priori error estimates presented in the previous section. We take $\alpha = \beta = \gamma = 1$. With a fixed time step $\tau = 10^{-3}$, we consider uniform triangulations with mesh sizes $h_k = \sqrt{2}/2^k$ for $k = 3, 4, 5, 6, 7$. On the other hand, with a fixed mesh size $h = \sqrt{2}/2^7$, we consider uniform time steps $\tau_k = 2^{-k}$ for $k = 2, 3, 4, 5, 6$.

Let us construct an exact solution of the optimal control problem (1.1)–(1.3). For this purpose, we consider the following functions as our exact state variables

$$\begin{aligned} \bar{p}(t,x) &= \cos(\pi t)\sin(2\pi x)\sin(2\pi y) \\ \bar{v}_1(t,x) &= 2(1+\sin(\pi t))\cos(2\pi x)\sin(2\pi y) \end{aligned}$$

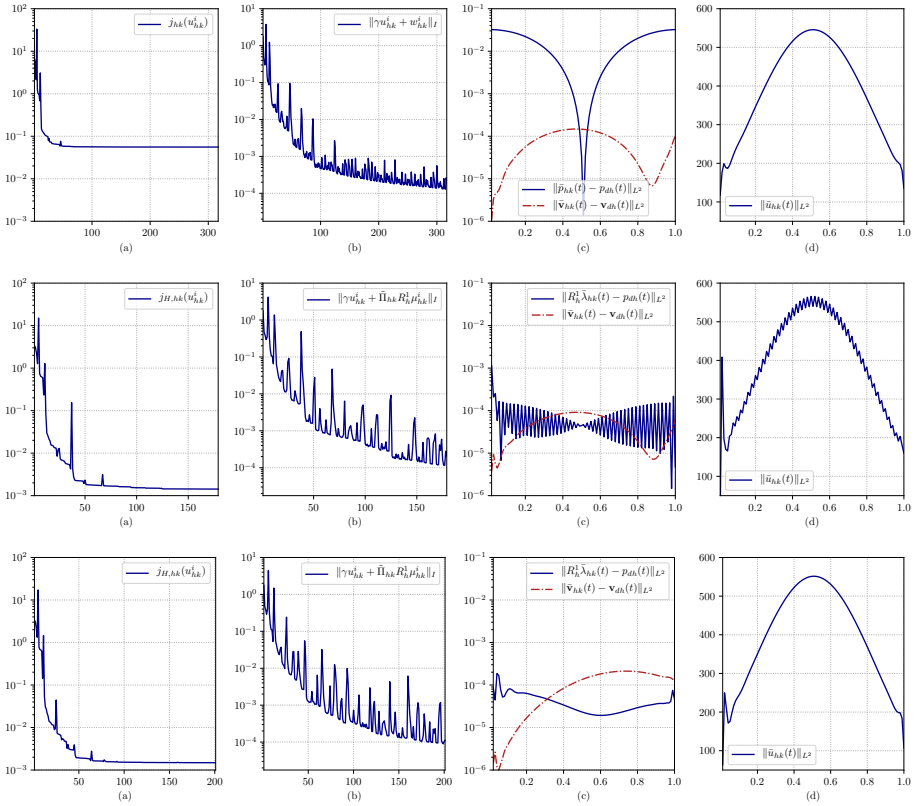

FIGURE 1. Results and performance of the Barzilai-Borwein gradient algorithm with post-processing and alternating stepsize selection in the mixed formulation (top), the hybrid formulation (middle), and the hybrid formulation with temporal averaging on the pressure and Lagrange multiplier (bottom): cost values (a) and gradient norms (b) as functions of the gradient iterations, residuals of the pressure and velocity with respect to the desired states as functions of time (c), and L^2-norm of controls as a function of time (d).

$$\bar{v}_2(t,x) = 2(1+\sin(\pi t))\sin(2\pi x)\cos(2\pi y)$$

and the following functions as our exact adjoint state variables

$$\begin{aligned}\bar{w}(t,x) &= -\sin(\pi t)\sin(2\pi x)\sin(2\pi y)\\ \bar{y}_1(t,x) &= 2(1+\cos(\pi t))\cos(2\pi x)\sin(2\pi y)\\ \bar{y}_2(t,x) &= 2(1+\cos(\pi t))\sin(2\pi x)\cos(2\pi y).\end{aligned}$$

For the exact control we take $\bar{u} = -\gamma^{-1}\bar{w}$. In order for these to be the solution of the optimal control problem, we add the source term $f = \bar{p}_t - \operatorname{div}\bar{\boldsymbol{v}} - \bar{u}$ on the right hand side of the pressure equation. On the other hand, we take $\boldsymbol{v}_d = \bar{\boldsymbol{v}}$ and $p_d = \bar{p} + \alpha^{-1}(\bar{w}_t - \operatorname{div}\bar{\boldsymbol{y}})$ as our desired states.

In order to compare the exact solutions with the numerical solutions of the discretized optimal control problems, we used the projections of $\bar{p}$, $\bar{w}$, $\bar{u}$ in X^1_{hk}, and the Fortin projections of $\bar{\boldsymbol{v}}$ and $\bar{\boldsymbol{y}}$ in $\boldsymbol{V}^1_{hk}$ as our reference exact solutions. Components of the Fortin projection are approximated using one-dimensional Gaussian quadrature

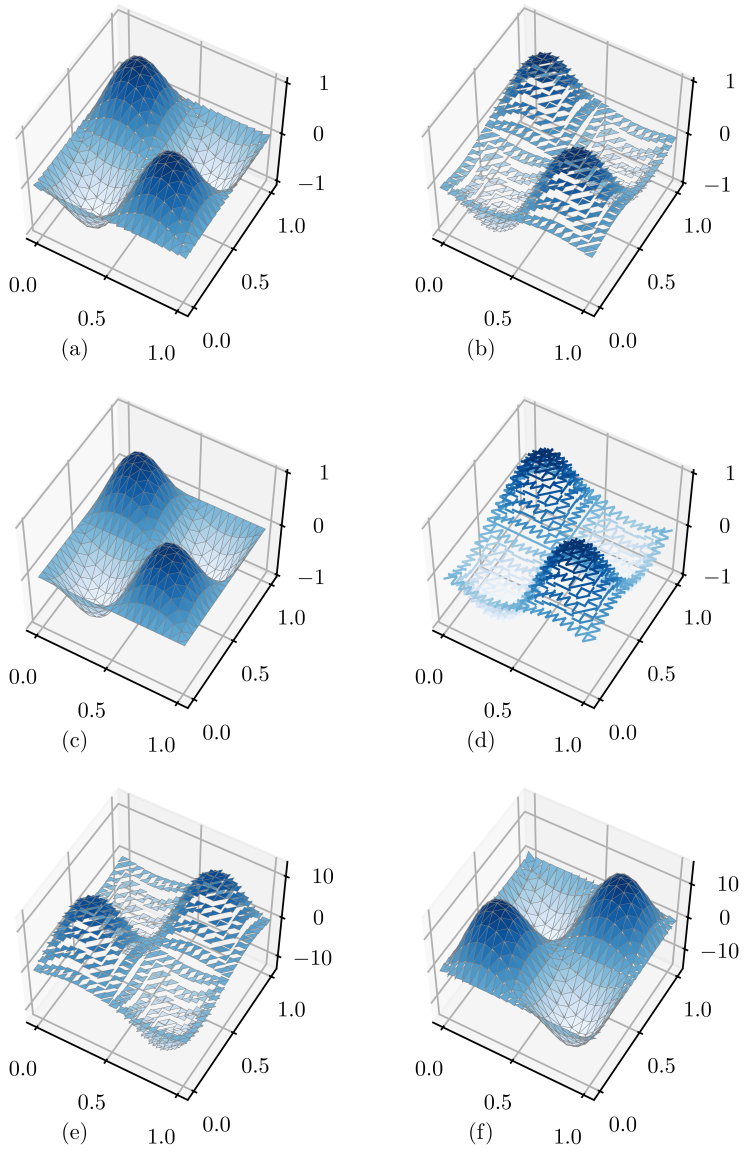

FIGURE 2. The desired pressure (c), optimal pressure (b), Lagrange multiplier (d), post-processed Langrange multiplier (a) obtained from the hybrid formulation, and the computed optimal controls in the mixed formulation (e) and the hybrid formulation (f) at the final time $T = 1$ in Example 1.

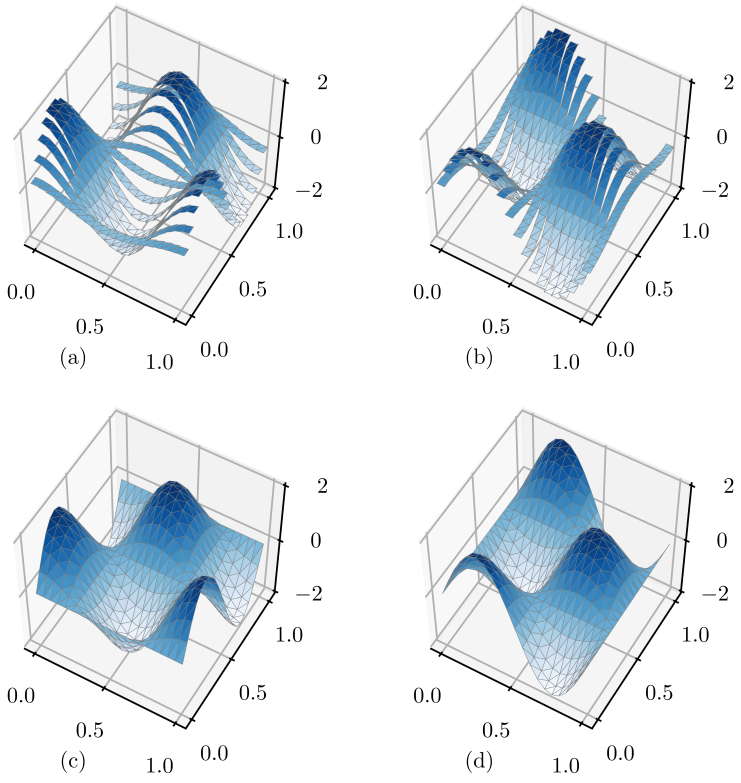

Figure 3. The components of desired velocity (c and d) and computed optimal velocity (a and b) in the hybrid formulation at the final time $T = 1$ in Example 1.

of order 5. The temporal and spatial discretization errors and the corresponding experimental orders of convergence are given in Tables 1 and 2. We observe the approximate orders of convergence $\mathcal{O}(\tau)$ and $\mathcal{O}(h)$, respectively.

Notice that we have an approximate quadratic order of convergence for spatial discretizations errors for the velocity and dual velocity. This superconvergence property of the discrete velocity to the Fortin projection of the exact velocity in $\boldsymbol{V}_{hk}^1$ has been observed in the elliptic case for uniform triangulations, see [11, 31]. In fact, Brandts established the rate $\mathcal{O}(h^{3/2})$ for elliptic problems with Dirichlet data in [11]. The quadratic convergence rates for the velocity and dual velocity have been also observed for the above example, when quasi-uniform triangulations are utilized, with mesh refinement based on bisection. These numerical results suggest that we might have the rate $\mathcal{O}(h^2)$ for sufficiently smooth solutions. Further analytical and numerical investigations are needed to confirm these observations.

Example 3. We repeat the previous example, but now using the hybrid formulation. In Tables 3 and 4 are the spatial and temporal discretization errors using

k	$\|\bar{\boldsymbol{v}}_{hk} - \bar{\boldsymbol{v}}\|_I$	eoc	$\|\bar{p}_{hk} - \bar{p}\|_I$	eoc
2	1.974567e-2	—	2.426421e-2	—
3	9.881673e-3	0.998709	1.057259e-2	1.198501
4	4.351001e-3	1.183408	5.198032e-3	1.024291
5	2.090567e-3	1.057453	3.799925e-3	0.451995
6	1.183192e-3	0.821209	3.479410e-3	0.127128

k	$\|\bar{u}_{hk} - \bar{u}\|_I$	eoc	$\|\bar{\boldsymbol{y}}_{hk} - \bar{\boldsymbol{y}}\|_I$	eoc	$\|\bar{w}_{hk} - \bar{w}\|_I$	eoc
2	4.903817e-2	—	2.706410e-1	—	4.871087e-2	—
3	3.591443e-2	0.449341	1.133596e-1	1.255474	3.591500e-2	0.439638
4	1.853659e-2	0.954188	5.593515e-2	1.019079	1.869639e-2	0.941827
5	9.821253e-3	0.916397	2.779501e-2	1.008929	9.911550e-3	0.915577
6	5.732314e-3	0.776789	1.375234e-2	1.015149	5.783482e-3	0.777172

Table 1. Experimental orders of convergence (eoc) for the temporal discretization errors of the primal, dual, and control variables in the mixed formulation with a fixed mesh size $h = \sqrt{2}/2^7$ and decreasing temporal stepsizes 2^{-k}.

k	$\|\bar{\boldsymbol{v}}_{hk} - \bar{\boldsymbol{v}}\|_I$	eoc	$\|\bar{p}_{hk} - \bar{p}\|_I$	eoc
3	1.473548e-1	—	7.442448e-2	—
4	3.890624e-2	1.921221	3.011163e-2	1.305457
5	9.868665e-3	1.979075	1.383103e-2	1.122412
6	2.485708e-3	1.989198	6.746917e-3	1.035608
7	6.342321e-4	1.970574	3.352089e-3	1.009168

k	$\|\bar{u}_{hk} - \bar{u}\|_I$	eoc	$\|\bar{\boldsymbol{y}}_{hk} - \bar{\boldsymbol{y}}\|_I$	eoc	$\|\bar{w}_{hk} - \bar{w}\|_I$	eoc
3	5.916516e-2	—	1.051876e-1	—	5.919348e-2	—
4	2.751500e-2	1.104529	2.702965e-2	1.960350	2.749597e-2	1.106218
5	1.345100e-2	1.032505	6.289935e-3	2.103426	1.343856e-2	1.032842
6	6.690677e-3	1.007489	1.146246e-3	2.456129	6.690060e-3	1.006287
7	3.351606e-3	0.997300	6.541195e-4	0.809290	3.363973e-3	0.991853

Table 2. Experimental orders of convergence (eoc) for the spatial discretization errors of the primal, dual, and control variables in the mixed formulation with a fixed time stepsize $\tau = 10^{-3}$ and decreasing mesh sizes $\sqrt{2}/2^k$.

the penalization parameter $\varepsilon = 10^{-10}$. A discussion on the choice of the tuning parameter ε will be discussed below. Approximate orders of convergence $\mathcal{O}(\tau)$ and $\mathcal{O}(h)$ are depicted. Also, aside from the quadratic convergence for the velocity and dual velocity as in the mixed case, we also have the quadratic convergence of the post-processed Lagrange multiplier to the optimal pressure. In general, the hybrid method produces smaller discretization errors than the mixed method.

Let us discuss the conditioning of the linear systems for the mixed and hybrid formulations provided in Examples 2 and 3. The approximate extreme eigenvalues

k	$\|\bar{\boldsymbol{v}}_{hk} - \bar{\boldsymbol{v}}\|_I$	eoc	$\|R_h^1\bar{\lambda}_{hk} - \bar{p}\|_I$	eoc
2	1.912979e-2	—	2.333909e-2	—
3	9.278768e-3	1.043816	9.355121e-3	1.318920
4	4.315141e-3	1.104525	3.866334e-3	1.274790
5	2.089870e-3	1.045995	1.718777e-3	1.169584
6	1.185135e-3	0.818362	8.911245e-4	0.947684

k	$\|\bar{u}_{hk} - \bar{u}\|_I$	eoc	$\|\bar{\boldsymbol{y}}_{hk} - \bar{\boldsymbol{y}}\|_I$	eoc	$\|R_h^1\bar{\mu}_{hk} - \bar{w}\|_I$	eoc
2	4.683819e-2	—	2.614458e-1	—	4.683861e-2	—
3	3.380741e-2	0.470346	1.133171e-1	1.206146	3.380658e-2	0.470394
4	1.811959e-2	0.899789	5.592169e-2	1.018886	1.824368e-2	0.889907
5	9.228530e-3	0.973378	2.779203e-2	1.008737	9.309845e-3	0.970568
6	4.659173e-3	0.986027	1.375150e-2	1.015082	4.715472e-3	0.981355

Table 3. Experimental orders of convergence (eoc) for the temporal discretization errors of the primal, dual, and control variables in the hybrid formulation with a fixed mesh size $h = \sqrt{2}/2^7$ and decreasing temporal stepsizes 2^{-k}.

k	$\|\bar{\boldsymbol{v}}_{hk} - \bar{\boldsymbol{v}}\|_I$	eoc	$\|R_h^1\bar{\lambda}_{hk} - \bar{p}\|_I$	eoc
3	1.481898e-1	—	5.876289e-2	—
4	3.906684e-2	1.923430	1.548843e-2	1.923715
5	9.915554e-3	1.978179	3.975572e-3	1.961956
6	2.497349e-3	1.989296	1.026295e-3	1.953717
7	6.371507e-4	1.970691	2.870094e-4	1.838276

k	$\|\bar{u}_{hk} - \bar{u}\|_I$	eoc	$\|\bar{\boldsymbol{y}}_{hk} - \bar{\boldsymbol{y}}\|_I$	eoc	$\|R_h^1\bar{\mu}_{hk} - \bar{w}\|_I$	eoc
3	3.205965e-2	—	1.047692e-1	—	3.199172e-2	—
4	8.824014e-3	1.861252	2.751281e-2	1.929040	8.872961e-3	1.850211
5	2.037239e-3	2.114820	6.242397e-3	2.139931	1.951703e-3	2.184682
6	4.991417e-4	2.029094	1.132843e-3	2.462152	4.878823e-4	2.000128
7	2.898751e-4	0.784018	6.535607e-4	0.793555	4.068231e-4	0.262131

Table 4. Experimental orders of convergence (eoc) for the spatial discretization errors of the primal, dual, and control variables in the hybrid formulation with a fixed time stepsize $\tau = 10^{-3}$ and decreasing mesh sizes $\sqrt{2}/2^k$.

and condition numbers of the system matrices in the CG method are presented in Table 5. The eigenvalues were calculated using the function eigs in the SciPy package. This function is a wrapper to some ARPACK functions that utilizes the Implicitly Restarted Arnoldi Method [27]. We used the tolerance 10^{-10} as the relative accuracy in the computation of the smallest eigenvalue, while the machine epsilon was used for the largest eigenvalue. In general, ARPACK is better at locating large eigenvalues than small ones.

We see that the system matrix for the penalized hybrid method have large condition numbers in comparison to the mixed method. This is attributed to the small penalty parameter ε. Decreasing the time stepsize for a fixed meshsize leads to a significant change on the condition number, while decreasing the meshsize for a fixed stepsize yields a small difference. Also, we have observed that a decrease of the parameter ε leads to an increase of the condition number by the same factor. Finally, for the pair of stepsizes $(\tau, h) = (10^{-3}, \sqrt{2}/2^6)$, we observed that the discretization error decreases for $\varepsilon = 10^{-k}$ with $k = 6, 7, 8, 9, 10$, before a small increase of the error in the optimal control is observed at $\varepsilon = 10^{-11}$. Based on this, the parameter $\varepsilon = 10^{-10}$ was used in the numerical experiments.

	Mixed			Hybrid		
k	$\Lambda_{\max}$	$\Lambda_{\min}$	κ	$\Lambda_{\max}$	$\Lambda_{\min}$	κ
2	3.07214e+3	2.92892e-1	1.04890e+4	1.49995e+10	1.0e+0	1.49995e+10
3	7.68285e+2	2.92890e-1	2.62312e+3	7.49975e+9	1.0e+0	7.49975e+9
4	1.92321e+2	2.92880e-1	6.56655e+2	3.74988e+9	1.0e+0	3.74988e+9
5	4.83304e+1	2.92842e-1	1.65039e+2	1.87494e+9	1.0e+0	1.87494e+9
6	1.23326e+1	2.92689e-1	4.21356e+1	9.37469e+8	1.0e+0	9.37469e+8

	Mixed			Hybrid		
k	$\Lambda_{\max}$	$\Lambda_{\min}$	κ	$\Lambda_{\max}$	$\Lambda_{\min}$	κ
3	9.93321e-1	1.66731e-1	5.95764e+0	5.99890e+7	1.0e+0	5.99890e+7
4	9.98362e-1	1.66923e-1	5.98099e+0	5.99767e+7	1.0e+0	5.99767e+7
5	9.99596e-1	1.67688e-1	5.96106e+0	5.99900e+7	1.0e+0	5.99900e+7
6	9.99900e-1	1.70712e-1	5.85722e+0	5.99970e+7	1.0e+0	5.99970e+7
7	9.99976e-1	1.82246e-1	5.48697e+0	5.99990e+7	1.0e+0	5.99990e+7

TABLE 5. Approximations of the largest eigenvalues ($\Lambda_{\max}$), smallest eigenvalues ($\Lambda_{\min}$), and condition numbers ($\kappa = \Lambda_{\max}/\Lambda_{\min}$) rounded up to 6 decimal digits of the system matrices R_h^+ and $P_h^{-1}\tilde{R}_h^+$ in the mixed and hybrid formulations, respectively, under fixed spatial mesh size $h = \sqrt{2}/2^6$ and decreasing temporal stepsizes $\tau_k = 2^{-k}$ (top table), and fixed temporal stepsize $\tau = 10^{-3}$ and decreasing spatial mesh sizes $h_k = \sqrt{2}/2^k$ (bottom table).

References

[1] M. Anselmann, M. Bause, S, Becher and G. Matthies, ESAIM Math. Model. Numer. Anal. 54, pp. 2099-2123, (2020)

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

[3] D. Arnold and F. Brezzi, Mixed and nonconforming finite element methods: Implementation, post-processing and error estimates, RAIRO Model. Math. Anal. Numer. 19, pp. 7-32, (1985)

[4] C. Bahriawati and C. Carstensen, Three Matlab implementations of the lowest-order Raviart–Thomas MFEM with a posteriori error control, Comput. Methods Appl. Math. 5, pp. 333-361, (2005)

[5] W. Bangerth and R. Rannacher, Adaptive finite element techniques for the acoustic wave equation, J. Comput. Acoustics 1, pp. 1-17, (1999)

[6] G. A. Baker, Error estimates for finite element methods for second order hyperbolic equations, SIAM J. Numer. Anal. 13, pp. 564-576, (1976)

[7] J. Barzilai and J. Borwein, Two-point step size gradient method. IMA J. Numerical Analysis 8, 141-148, (1988)

[8] M. Bause, U. Köcher, F. A. Radu and F. Schieweck, Post-processed Galerkin approximation of improved order for wave equations, Math. Comp. 89, pp. 595-627, (2020)

[9] E. Bècache, P. Joly and C. Tsogka, An analysis of new mixed finite elements for the approximation of wave propagation problems, SIAM J. Numer. Anal. 37, pp. 1053-1084, (2000)

[10] R. Becker, D. Meidner, and B. Vexler, Efficient numerical solution of parabolic optimization problems by finite element methods, Optim. Methods Softw. 22, pp. 813-833, (2007)

[11] J. H. Brandts, Superconvergence and a posteriori error estimation for triangular mixed finite elements, Numer. Math. 68, pp. 311-324, (1994)

[12] F. Brezzi and M. Fortin, Mixed and Hybrid Finite Element Methods, Springer-Verlag, New York, (1991)

[13] Y. Chen and Z. Lu, Error estimates for parabolic optimal control problem by fully discrete mixed finite element methods, Finite Elem. Anal. Des. 46, pp. 957-965, (2010)

[14] B. Cockburn and J. Gopalakhrishnan, A characterication of hybridized mixed methods for second order elliptic problems, SIAM J. Numer. Anal. 42, pp. 283-301, (2004)

[15] L. C. Cowsar, T. F. Dupont and M. F. Wheeler, A priori estimates for mixed finite element methods for the wave equations, Comput. Methods Appl. Mech. Engrg. 8, pp.205-222, (1990)

[16] L. C. Cowsar, T. F. Dupont and M. F. Wheeler, A priori estimates for mixed finite element approximations of second-order hyperbolic equations with absorbing boundary conditions, SIAM J. Numer. Anal. 33, pp. 492-504, (1996)

[17] Y.H. Dai and R. Fletcher, Projected Barzilai-Borwein methods for large-scale box constrained quadratic programming, Numer. Math. 100, pp. 21-47 (2005)

[18] W. Dö, S. Findeisen, C. Wieners and D. Ziegler, Parallel adaptive discontinuous Galerkin discretizations in space and time for linear elastic and acoustic waves, In U. Langer and O. Steinbach (Eds.), Space–Time Methods: Applications to Partial Differential Equations, De Gruyter, Berlin, (2019)

[19] H. Egger and B. Radu, Super-convergence and post-processing for mixed finite element approximations of the wave equation, B. Numer. Math. 140, pp. 427–447, (2018)

[20] A. Ern and J.-L. Guermond, Theory and Practice of Finite Elements, Springer-Verlag, New York, (2004)

[21] J. Ernesti and C. Wieners, Space–time discontinuous Petrov–Galerkin methods for linear wave equations in heterogeneous media, Comput. Methods Appl. Math. 19, pp. 465Ð481, (2019)

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

[23] T. Hou, Error estimates and superconvergence of semidiscrete mixed methods for optimal control problems governed by hyperbolic equations, Numerical Analysis and Applications 5, pp. 348-362, (2012)

[24] E. W. Jenkins, B. Rivière and M. F. Wheeler, A priori error estimates for mixed finite element approximations fo the acoustic wave equation, SIAM J. Numer. Anal. 40, pp. 1698-1715, (2002)

[25] A. Kröner, K. Kunisch and B. Vexler, Semismooth Newton methods for optimal control of the wave equation with control constraints, SIAM J. Control Optim. 49, pp. 830-858, (2011)

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

[27] R. B. Lehoucq, D. C. Sorensen, and C. Yang, ARPACK USERS GUIDE: Solution of Large Scale Eigenvalue Problems by Implicitly Restarted Arnoldi Methods, SIAM, Philadelphia, PA, (1998)

[28] D. Meidner and B. Vexler, A priori error analysis of the Petrov–Galerkin Crank–Nicholson Scheme for parabolic optimal control problems, SIAM J. Control. Optim. 49, pp. 2183-2211, (2011)

[29] Ch. G. Makridakis, On mixed finite element methods for linear elastodynamics, Numer. Math. 61, pp. 235-260, (1992)

[30] A. Pazy, Semigroups of Linear Operators and Applications to Partial Differential Equations, Springer-Verlag, New York, (1983)

[31] G. Peralta, Error estimates for mixed and hybrid FEM for elliptic optimal control problems with penalizations, submitted.

[32] A. Quarteroni and A. Valli, Numerical Approximations of Partial Differential Equations, Springer, Heidelberg, (2008)

[33] P. A. Raviart and J. M. Thomas, A mixed finite element method for second order elliptic problems, Lect. Notes Math 606, pp. 292-315, (1977)

[34] A. Sellitto, V. A. Cimmelli and D. Jou, Mesoscopic Theories of Heat Transport in Nanosystems, Spinger, New York, (2016)

[35] F. Tröltzsch, Optimal Control of Partial Differential Equations: Theory, Methods and Applications, American Mathematical Society, Providence, Rhode Island, (2010)

[36] B. I. Wohlmuth and R. H.W. Hoppe, A comparison of a priori error estimators for mixed finite element discretizations by Raviart–Thomas elements, Math. Comp. 68, pp. 1347-1378, (1999)

[37] X. Xing and Y. Chen, Error estimates of mixed methods for optimal control problems governed by parabolic equations, Int. J. Numer. Method. Engng 75, pp. 735-754, (2008)

[38] X. Xing, Y. Chen, N. Yi, Error estimates of mixed finite element methods for quadratic control problems, J Comput. Appl. Math. 233, pp. 1812-1820, (2010)