SIAM J. CONTROL OPTIM.
Vol. 46, No. 5, pp. 1726–1753

CONSTRAINED DIRICHLET BOUNDARY CONTROL IN L^2 FOR A CLASS OF EVOLUTION EQUATIONS*

K. KUNISCH† AND B. VEXLER‡

Abstract. Optimal Dirichlet boundary control based on the very weak solution of a parabolic state equation is analyzed. This approach allows us to consider the boundary controls in L^2, which has advantages over approaches which consider control in Sobolev spaces involving (fractional) derivatives. Pointwise constraints on the boundary are incorporated by the primal-dual active set strategy. Its global and local superlinear convergences are shown. A discretization based on space-time finite elements is proposed and numerical examples are included.

Key words. Dirichlet boundary control, inequality constraints, parabolic equations, very weak solution

AMS subject classifications. 49J20, 35K05, 49K20, 49K05, 49M29

DOI. 10.1137/060670110

1. Introduction.

In this work we focus on the Dirichlet boundary optimal control problem with pointwise constraints on the boundary, formally given by

$$\left\{\begin{array}{ll} \min & J(y,u) \\ \text{subject to} & \partial_t y - \kappa\Delta y + b\cdot\nabla y = f \quad \text{in } Q, \\ & y = u, \quad u \le \psi \quad \text{on } \Sigma, \\ & y(0) = y_0 \quad \text{in } \Omega, \end{array}\right. \tag{1.1}$$

where $Q = (0,T] \times \Omega$, $\Sigma = (0,T] \times \partial\Omega$, and κ, b, f, y_0, ψ, and $T > 0$ are fixed. We propose and analyze a function space formulation which is amenable to efficient numerical realizations. To incorporate the constraints numerically the primal-dual active set (PDAS) strategy is used and its convergence is investigated. We also propose a space-time Galerkin approximation and provide numerical examples.

The specific difficulties involved in Dirichlet control problems result from the fact that they are not of variational type. In the literature several treatments of Dirichlet boundary control problems can be found, where the function space for the controls is H^s with $s \ge \frac{1}{2}$. As a consequence, the numerical realization by finite elements or finite differences is more involved than if the control space were L^2. Our approach will be based on the concept of very weak solutions to the state equation. This allows the use of L^2 as the control space.

Let us briefly describe possible approaches to treating Dirichlet boundary optimal control problems. While in our work we shall treat the time-dependent case, it will be convenient for the present purpose to restrict our attention to a tracking-type optimal

*Received by the editors September 18, 2006; accepted for publication (in revised form) March 20, 2007; published electronically November 9, 2007.
http://www.siam.org/journals/sicon/46-5/67011.html

†Institute for Mathematics and Scientific Computing, University of Graz, Heinrichstraße 36, A-8010 Graz, Austria (karl.kunisch@uni-graz.at).
‡Johann Radon Institute for Computational and Applied Mathematics (RICAM), Austrian Academy of Sciences, Altenberger Straße 69, 4040 Linz, Austria (boris.vexler@oeaw.ac.at).

control problem with the most simple stationary elliptic equation as constraint:

$$(1.2)\quad \begin{cases} \text{min} & \frac{1}{2}|y-z|^2_{L^2(\Omega)} + \frac{\beta}{2}\,|u|^2_{L^2(\partial\Omega)} \\ \text{over} & (y,u) \in L^2(\Omega)\times L^2(\partial\Omega) \\ \text{subject to} & -(y,\Delta v)_{L^2(\Omega)} = -(u,\frac{\partial}{\partial n}v)_{L^2(\partial\Omega)} \quad \text{for all } v\in H^2(\Omega)\cap H^1_0(\Omega) \\ \text{and} & u\le\psi \quad \text{on } \partial\Omega, \end{cases}$$

where $z \in L^2(\Omega)$ and $\partial\Omega$ denotes the boundary of the domain Ω. The variational equation in (1.2) is the very weak form of

$$\begin{cases} -\Delta y = 0 & \text{in } \Omega, \\ y = u & \text{on } \partial\Omega; \end{cases}$$

see [36]. In our work we shall use the analogue of (1.2). If the state variable y is considered in $H^1(\Omega)$, then a proper formulation is given by

$$(1.3)\quad \begin{cases} \text{min} & \frac{1}{2}|y-z|^2_{L^2(\Omega)} + \frac{\beta}{2}\,|u|^2_{H^{\frac{1}{2}}(\partial\Omega)} \\ \text{over} & (y,u) \in H^1(\Omega)\times H^{\frac{1}{2}}(\partial\Omega) \\ \text{subject to} & (\nabla y,\nabla v)_{L^2(\Omega)} = 0 \quad \text{for all } v\in H^1_0(\Omega) \text{ and } y=u \text{ on } \partial\Omega \\ \text{and} & u\le\psi \quad \text{on } \partial\Omega. \end{cases}$$

For both formulations (1.2) and (1.3) it is classical to argue existence of a unique solution; see, e.g., [36]. Numerically realizing the $H^{1/2}$-norm in (1.3) is more involved than realizing the L^2-norm in (1.2). To avoid difficulties with implementing the $H^{1/2}$-norm it was replaced in several publications by the H^1-norm. As a consequence the Laplace–Beltrami operator appears in the optimality condition. This formulation, properly modified for the specific application and without control constraints, was used in the context of optimal boundary control of the Navier–Stokes equations and the Boussinesq equations; see, e.g., [23, 24, 33]. For a numerical wavelet–based realization of H^s-norms in the context of Dirichlet control of elliptic equations, we refer to [30].

A third alternative is given by

$$(1.4)\quad \begin{cases} \text{min} & \frac{1}{2}|y-z|^2_{H^1(\Omega)} + \frac{\beta}{2}\,|u|^2_{L^2(\partial\Omega)} \\ \text{over} & (y,u) \in H^1(\Omega)\times H^{1/2}(\partial\Omega) \\ \text{subject to} & (\nabla y,\nabla v)_{(\Omega)} = 0 \quad \text{for all } v\in H^1_0(\Omega) \text{ and } y=u \quad \text{on } \partial\Omega \\ \text{and} & u\le\psi \quad \text{on } \partial\Omega. \end{cases}$$

Again existence can be argued by standard arguments, but for (1.4), differently from (1.2) and (1.3), the essential term for obtaining coercivity is the H^1-norm of the tracking functional. Just like (1.2) this formulation also avoids having to deal with fractional order Sobolev spaces. It was used in the context of boundary control of the stationary Navier–Stokes equations in [15], for example. In the adjoint equation, however, a Laplacian now appears in the source term acting on the defect $y - z$.

Besides the difficulties already mentioned with (1.3) and (1.4) there is yet another, possibly more essential, reason to favor the formulation in (1.2). For (1.2) the Lagrange multiplier associated to the constraint $u \le \psi$ is an L^2-function, whereas it is only a measure for the formulations in (1.3) and (1.4). As a consequence the complementarity conditions related to the inequality constraint can be expressed in a pointwise a.e. manner by the common pointwise complementarity functions like the max or the Fischer–Burmeister functions only for formulation (1.2). Such a pointwise formulation is a basis for efficient optimization algorithms such as the PDAS strategy or the semismooth Newton method.

Let us also recall the possibility of approximating Dirichlet boundary control problems by regularization based on Robin boundary controls of the form $\delta\frac{\partial y}{\partial n}+y=u$ for $\delta\to 0^+$. This results in the variational formulation

$$(1.5)\qquad \begin{cases} \text{min} & \frac{1}{2}|y-z|^2_{L^2(\Omega)}+\frac{\beta}{2}\,|u|^2_{L^2(\partial\Omega)}\\ \text{over} & (y,u)\in H^1(\Omega)\times L^2(\partial\Omega)\\ \text{subject to} & (\nabla y,\nabla v)_{L^2(\Omega)}=\frac{1}{\delta}(y-u,v)_{L^2(\partial\Omega)}\quad \text{for all } v\in H^1(\Omega)\\ \text{and} & u\le\psi\quad \text{on } \partial\Omega. \end{cases}$$

The choice of δ remains a delicate matter. This approach was used for stationary and nonstationary problems in [6] and [2], respectively. In [5] a numerical approach to Dirichlet boundary control based on a discretization using the Nitsche method was proposed.

We next point out some additional features of this paper. As already mentioned, the pointwise inequality constraint $u\le\psi$ will be treated by the PDAS algorithm. Its global, as well as local, superlinear convergence will be analyzed. Here it is essential that the Lagrange multiplier is an L^2-function and that the resulting complementarity condition involving the max operation is Newton differentiable. This is the case for (1.2), whereas this is not true for the other two formulations. Newton differentiability will be shown for (1.2) for time-dependent problems in the present paper. For stationary problems it easily follows as well.

Discretization of infinite-dimensional problems will be carried out by a space-time finite element method. This approach guarantees that the algorithm is invariant with respect to the ordering of discretization of the problem and gradient computations.

In spite of the fact that we use the very weak solution concept as our functional analytic setting for Dirichlet boundary control, the numerical discretization is based on standard space-time Galerkin finite-dimensional spaces. This will be justified by the fact that the solutions of the optimal control problems are more regular than those required by (1.2).

In our numerical implementation we use piecewise (bi)linear elements for spatial discretization of the primal and adjoint states as well as for the controls. This may appear to be incompatible at first, since the optimality condition involves $\frac{\partial p}{\partial n}$ and u in an additive manner, where p denotes the adjoint state. However, we replace $\frac{\partial p}{\partial n}$ by a variational expression in such a way that the resulting discretization is well balanced.

In section 2 we gather well-posedness results and a priori estimates for a class of evolution equations with Dirichlet boundary conditions in L^2. We include a convection term, due to future interest in considering similar problems for the Boussinesq systems, with specific nonconvex cost functionals, motivated by fluid mechanics considerations. In this case the convection coefficient is the velocity field of the fluid. Section 3 is devoted to the statements and analysis of the optimal control problems under consideration. In particular, we describe regularity properties of the optimal solutions. These are not only of interest in their own right, but are essential for superlinear convergence of the PDAS strategy, as explained in section 4. Section 5 contains a description of the finite element discretization, and section 6 is devoted to selected numerical examples.

2. On the state equation. In this section we provide the necessary existence and a priori estimates for very weak solutions to

$$\begin{cases} \partial_t y - \kappa\Delta y + b\cdot\nabla y = f \quad \text{in } Q,\\ y = u \quad \text{on } \Sigma,\\ y(0) = y_0 \quad \text{in } \Omega, \end{cases} \tag{2.1}$$

where $Q = (0,T]\times\Omega$, $\Sigma = (0,T]\times\partial\Omega$, and Ω is a bounded domain in $\mathbb{R}^n$, $n\geq 2$, with C^2 boundary $\partial\Omega$. This boundary regularity of Ω guarantees that the Laplacian with homogeneous Dirichlet boundary conditions, denoted by Δ_0, is an isomorphism form $H^2(\Omega)\cap H^1_0(\Omega)$ to $L^2(\Omega)$. We shall denote the adjoint of Δ_0, mapping from $L^2(\Omega)$ to $H^{-2}(\Omega) = (H^2(\Omega)\cap H^1_0(\Omega))^*$ by Δ_0 as well. For any Banach space Y, we use the abbreviations $L^2(Y) = L^2(0,T;Y)$, $H^s(Y) = H^s(0,T;Y)$, $s\in[0,\infty)$, and $C(Y) = C([0,T];Y)$.

Further $\kappa>0$, $y_0\in H^{-1}(\Omega)$, $f\in L^2(H^{-2}(\Omega))$, $u\in L^2(\Sigma)$ and $b\in\mathbb{L}^\infty(Q)$, div b $\in L^\infty(L^{\hat{n}}(\Omega))$, where $\hat{n}=\max(n,3)$, and $\mathbb{L}^\infty(Q) = \bigotimes_{i=1}^n L^\infty(Q)$. At times we shall simply write $L^p(Q)$ for $\mathbb{L}^p(Q)$.

The very weak form of (2.1) that we shall utilize is given by

$$\begin{cases} \langle\partial_t y(t), v\rangle - \kappa(y(t),\Delta v) - (y(t), \operatorname{div}(b(t))\,v) - (y(t), b(t)\nabla v)\\ \qquad = \langle f(t), v\rangle - \kappa(u(t), \frac{\partial}{\partial n}v)_{\partial\Omega} \quad \text{for all } v\in H^2(\Omega)\cap H^1_0(\Omega)\\ \qquad\qquad\qquad\qquad \text{and a.e. } t\in(0,T),\\ y(0) = y_0, \end{cases} \tag{2.2}$$

where $\langle\cdot,\cdot\rangle = \langle\cdot,\cdot\rangle_{H^{-2}(\Omega),H^2(\Omega)\cap H^1_0(\Omega)}$ and denotes the canonical duality pairing, and $(\cdot,\cdot)$ and $(\cdot,\cdot)_{\partial\Omega}$ stand for the inner products in $L^2(\Omega)$ and $L^2(\partial\Omega)$, respectively. The last equality in (2.2) is understood in $H^{-1}(\Omega)$. The existence and uniqueness of a very weak solution in the space $L^2(Q)\cap H^1(H^{-2}(\Omega))\cap C(H^{-1}(\Omega))$ is shown in the following theorem.

THEOREM 2.1. *For every $\kappa>0$, $b\in L^\infty(Q)$, with $\operatorname{div} b\in L^\infty(L^{\hat{n}}(\Omega))$, $y_0\in H^{-1}(\Omega)$, $f\in L^2(H^{-2}(\Omega))$, and $u\in L^2(\Sigma)$, there exists a unique very weak solution $y\in L^2(Q)\cap H^1(H^{-2}(\Omega))\cap C(H^{-1}(\Omega))$ satisfying*

$$|y|_{L^2(Q)\cap H^1(H^{-2}(\Omega))\cap C(H^{-1}(\Omega))} \leq C(|y_0|_{H^{-1}(\Omega)} + |f|_{L^2(H^{-2}(\Omega))} + |u|_{L^2(\Sigma)}), \tag{2.3}$$

where C depends continuously on $\kappa>0$, $|b|_{L^\infty(Q)}$, and $|\text{div } b|_{L^\infty(L^{\hat{n}}(\Omega))}$, and is independent of $f\in L^2(H^{-2}(\Omega))$, $u\in L^2(\Sigma)$, and $y_0\in H^{-1}(\Omega)$.

Proof. Let us first assume existence of y with the claimed regularity and verify the a priori estimate (2.3). Throughout, k will denote a generic embedding constant. Let us introduce the transformed state-variable $\hat{y}(t) = y(t)e^{-ct}$, $c\geq 0$, and note that if y is a very weak solution of (2.1), then $\hat{y}\in L^2(Q)\cap H^1(H^{-2}(\Omega))$ is a very weak solution of

$$\begin{cases} \partial_t\hat{y} + c\hat{y} - \kappa\Delta\hat{y} + b\cdot\nabla\hat{y} = \hat{f} \quad \text{in } Q,\\ \hat{y} = \hat{u} \quad \text{on } \Sigma,\\ \hat{y}(0) = y_0 \quad \text{in } \Omega, \end{cases}$$

where $\hat{f} = fe^{-ct}$, $\hat{u} = ue^{-ct}$. The constant c will be fixed below. We further introduce $\omega = (-\Delta_0)^{-1}\,\hat{y} \in L^2(H^2(\Omega) \cap H^1_0(\Omega)) \cap H^1(L^2(\Omega))$ and note that ω satisfies for all $v \in H^2(\Omega) \cap H^1_0(\Omega)$

$$\begin{aligned}\langle(-\Delta_0)\,\partial_t\omega(t),\, v\rangle &+ \kappa(\Delta_0\,\omega(t),\, \Delta v) + c(-\Delta_0\,\omega(t),\, v)\\ &+ (\Delta_0\,\omega(t),\, \operatorname{div} b(t)\, v) + (\Delta_0\,\omega(t),\, b(t)\nabla v) = \langle \hat{f}(t),\, v\rangle - \kappa\left(\hat{u}(t),\, \frac{\partial}{\partial n}\, v\right)_{\partial\Omega}\end{aligned}$$

for all $t \in (0,T)$. Setting $v = \omega(t)$ and integrating over $(0,t)$, we find

$$\begin{aligned}\frac{1}{2}|\nabla\omega(t)|^2 &- \frac{1}{2}\,|\nabla\omega(0)|^2 + \kappa\int_0^t |\Delta_0\;\omega(s)|^2\,ds + c\int_0^t |\nabla\omega(s)|^2\,ds\\ &+ \int_0^t (\Delta_0\,\omega(s),\, \operatorname{div} b(s)\,\omega(s))\,ds + \int_0^t (\Delta_0\,\omega(s),\, b(s)\nabla\omega(s))\,ds\\ &= \int_0^t \langle \hat{f}(s), \omega(s)\rangle\,ds - \kappa\int_0^t \left(\hat{u}(s), \frac{\partial}{\partial n} w(s)\right)_{\partial\Omega},\end{aligned}$$

and consequently using $\|\frac{\partial}{\partial n} w(s)\|_{L^2(\partial\Omega)} \le k\,\|\Delta_0 w(s)\|_{L^2(\Omega)}$ we obtain

$$\begin{aligned}&\frac{1}{2}|\nabla\omega(t)|^2\,ds + \kappa\int_0^t |\Delta_0\,\omega(s)|^2 ds + c\int_0^t |\nabla\omega(s)|^2\,ds\\ &\le \frac{1}{2}|\nabla\omega(0)|^2 + \frac{\kappa}{8}\int_0^t |\Delta_0\,\omega(s)|^2\,ds + \frac{2k}{\kappa}\,|\operatorname{div} b|^2_{L^\infty(L^{\hat{n}}(\Omega)}\int_0^t |\nabla\omega(s)|^2\,ds\\ &\quad + \frac{\kappa}{8}\int_0^t |\Delta_0\,\omega(s)|^2 + \frac{2|b|^2_{L^\infty(Q)}}{\kappa}\int_0^t |\nabla\omega(s)|^2\,ds + \frac{2k^2}{\kappa}\int_0^t |\hat{f}(s)|^2_{H^{-2}} + \frac{\kappa}{8}\int_0^t |\Delta_0\,\omega|^2\,ds\\ &\quad + 2\kappa^2\int_0^t |\hat{u}(s)|^2_{L^2(\partial\Omega)}\,ds + \frac{\kappa}{8}\int_0^t |\Delta_0\,\omega(s)|^2\,ds\\ &\le \frac{1}{2}|\nabla\omega(0)|^2 + \frac{4\kappa}{8}\int_0^t |\Delta_0\,\omega(s)|^2\,ds + \left(\frac{2k}{\kappa}\,|\operatorname{div} b|^2_{L^\infty(L^{\hat{n}}(\Omega))} + \frac{2|b|^2_{L^\infty(Q)}}{\kappa}\right)\int_0^t |\nabla\omega(s)|^2\,ds\\ &\quad + \frac{2k^2}{\kappa}\int_0^t |\hat{f}(s)|^2_{H^{-2}(\Omega)}\,ds + 2k^2\int_0^t |\hat{u}(s)|^2_{L^2(\partial\Omega)}\,ds\,.\end{aligned}$$

If we choose c such that

$$\frac{2k}{\kappa}\,|\operatorname{div} b|^2_{L^\infty(L^{\hat{n}}(\Omega))} + \frac{2\,|b|^2_{L^\infty(Q)}}{\kappa} \le \frac{c}{2}\,, \tag{2.4}$$

then

$$\begin{aligned}\frac{1}{2}\,|\nabla\omega(t)|^2 &+ \frac{k}{2}\int_0^t |\Delta_0\,\omega(s)|^2\,ds + \frac{c}{2}\int_0^t |\nabla\omega(s)|^2\,ds\\ &\le \frac{1}{2}\,|\nabla\omega(0)|^2 + \frac{2k^2}{\kappa}\int_0^t |\hat{f}(s)|^2_{H^{-2}(\Omega)}\,ds + 2k^2\int_0^t |\hat{u}(s)|^2_{L^2(\partial\Omega)}\,ds\,.\end{aligned} \tag{2.5}$$

From (2.5) we deduce the existence of a constant C with the specified properties such that for all $t \in [0,T]$

$$|\hat{y}(t)|_{H^{-1}(\Omega)} + \int_0^t |\hat{y}(s)|^2_{L^2(\Omega)}\,ds \le C(|y_0|_{H^{-1}(\Omega)} + |f|_{L^2(H^{-2}(\Omega))} + |u|_{L^2(\Sigma)})\,,$$

and, since $\hat{y}(t) = y(t)e^{-ct}$, we find for a possibly modified C:

$$|y(t)|_{H^{-1}(\Omega)} + \int_0^t |y(s)|^2_{L^2(\Omega)} ds \le C(|y_0|_{H^{-1}(\Omega)} + |f|_{L^2(H^{-2}(\Omega))} + |u|_{L^2(\Sigma)}) . \tag{2.6}$$

Finally using (2.2) we obtain

$$\begin{aligned}\int_0^T |\partial_t y(t)|^2_{H^{-2}(\Omega)}\, dt &= \int_0^T \sup_{\substack{v\in H^2(\Omega)\cap H_0^1(\Omega),\\ |\Delta_0 v|\le 1}} \langle \partial_t y(t), v\rangle^2\, dt \\ &\le \kappa^2 \int_0^T |y(t)|^2\, dt + \int_0^T (y(t), \operatorname{div} b\, v)^2_{L^2(\Omega)}\, dt \\ &\quad + |b|^2_{L^\infty(Q)} \int_0^T |y(t)|^2\, dt + |f|^2_{L^2(H^{-2}(\Omega))} + k\, |u|^2_{L^2(\Sigma)} .\end{aligned}$$

For the second term on the right-hand side we estimate for $n > 4$

$$\begin{aligned}\int_0^T (y(t), \operatorname{div} b\, v)^2_{L^2(\Omega)}\, dt &\le \int_0^T |y(t)|^2_{L^2(\Omega)}\, |\operatorname{div} b|^2_{L^{2p}(\Omega)}\, |v|^2_{L^{2q}(\Omega)}\, dt \\ &\le k \int_0^T |y(t)|^2_{L^2(\Omega)}\, |\operatorname{div} b|^2_{L^{\hat{n}}(\Omega)}\, dt ,\end{aligned}$$

where $q = \frac{n}{n-4}$, $p = \frac{n}{4}$, and we used that $H^2(\Omega) \hookrightarrow L^{\frac{2n}{n-4}}(\Omega)$ and $\hat{n} > 2p = \frac{n}{2}$. The same estimate for dimensions $n = 2, 3, 4$ follows quite easily.

We obtain

$$\begin{aligned}&\int_0^T |\partial_t y|^2_{H^{-2}(\Omega)}\, dt \\ &\quad\le (\kappa^2 + k\, |\operatorname{div} b|_{L^\infty(L^{\hat{n}}(\Omega))} + |b|_{L^\infty(Q)}) \int_0^T |y(t)|^2\, dt + |f|^2_{L^2(H^{-2}(\Omega))} + k\, |u|^2_{L^2(\Sigma)} .\end{aligned}$$

Together with (2.6) this gives the desired estimate (2.3), which in particular also implies the uniqueness of the very weak solution to (2.1). Existence follows, for example, by combining this a priori estimate with a Galerkin procedure; see, e.g., [14, Chapter 18]. Alternatively analytic semigroup theory as in [32] can be used, noting that $-\kappa\Delta - b\cdot\nabla + cI$ generates an analytic semigroup in $L^2(\Omega)$. $\square$

From the proof it follows that the solution y to (2.2) also satisfies the variational equation in Q given by

(2.7)
$$\begin{aligned}&\int_0^T \big(\langle \partial_t y(t), v(t)\rangle - \kappa(y(t), \Delta v(t)) - (y(t), \operatorname{div}(b(t))v(t)\,) - (y(t), b(t)\nabla v(t))\big)\, dt \\ &= \int_0^T \langle f(t), v(t)\rangle dt - \kappa \int_0^T \left(u(t), \frac{\partial}{\partial n} v(t)\right)_{L^2(\Omega)} dt \quad \text{for all } v \in L^2(H^2(\Omega)\cap H_0^1(\Omega)).\end{aligned}$$

The following result will allow us to consider cost functionals with pointwise-in-time evaluation of the trajectory.

Corollary 2.2. *If, in addition to the assumptions of Theorem* 2.1, $y_0 \in L^2(\Omega), f \in L^2(Q)$, *and* $u \in L^\infty(L^2(\partial\Omega))$, *then the very weak solution satisfies* $y \in$

$L^\infty(L^2(\Omega))$ and $y(\bar t)$ is a well-defined element in $L^2(\Omega)$ for every fixed $\bar t \in (0,T]$. Moreover, there exists a constant C independent of y_0, f, and u, such that for the corresponding solution $y = y(u)$ we have

$$|y(\bar t)|_{L^2(\Omega)} \le C(|y_0|_{L^2(\Omega)} + |f|_{L^2(Q)} + |u|_{L^\infty(L^2(\partial\Omega))}). \tag{2.8}$$

Proof. Fix $\kappa > 0$ and $b \in L^\infty(Q)$ with $\operatorname{div} b \in L^\infty(L^{\hat n}(\Omega))$. Without loss of generality we can assume that $A = -\kappa\Delta - b \cdot \nabla$ is uniformly elliptic. If not, we add a multiple c of the identity operator and accordingly multiply the constant C by the factor e^{cT}. Then A generates an analytic semigroup in $L^2(\Omega)$. For the equation with $u = 0$, estimate (2.8) follows by standard semigroup arguments. Using the superposition principle for (2.1) it therefore suffices to consider the case $y_0 = 0$, $f = 0$, and $u \in L^\infty(L^2(\partial\Omega))$. From [32] (see also [2]), we have the existence of $C > 0$ such that

$$|y|_{L^\infty(L^2(\Omega))} \le C|u|_{L^\infty(L^2(\partial\Omega))}\,. \tag{2.9}$$

From Theorem 2.1 we deduce $y \in C(H^{-1}(\Omega))$ and therefore

$$y(\bar t) = \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\int_{-\varepsilon}^{0} y(\bar t + \tau)\, d\tau, \tag{2.10}$$

where the integral and the equality are interpreted in $H^{-1}(\Omega)$. Denoting

$$g_\varepsilon = \frac{1}{\varepsilon}\int_{-\varepsilon}^{0} y(\bar t + \tau)\, d\tau\,,$$

we obtain using (2.9) that

$$|g_\varepsilon|_{L^2(\Omega)} \le C|u|_{L^\infty(L^2(\partial\Omega))}.$$

Therefore, there is a subsequence converging weakly in $L^2(\Omega)$ to $\tilde g$ with

$$|\tilde g|_{L^2(\Omega)} \le C|u|_{L^\infty(L^2(\partial\Omega))}.$$

Using (2.10) we obtain that $y(\bar t) = \tilde g$. The desired conclusion follows. ∎

3. The optimal control problems and regularity of optimal controls. We consider the following two optimal control problems:

$$\text{(P1)} \qquad \begin{cases} \text{min} & J(y,u) = G(y) + \frac{\beta}{2}\,|u|^2_{L^2(\Sigma)} \\ \text{over} & (y,u) \in L^2(Q) \times L^2(\Sigma) \\ \text{subject to (2.1) and } & u \le \psi \text{ on } \Sigma, \end{cases}$$

where $\beta > 0$, $\psi \in L^2(\Sigma)$, and $G : L^2(Q) \to \mathbb{R}$ is bounded below, C^1, and weakly lower semicontinuous. The second problem under consideration is

$$\text{(P2)} \qquad \begin{cases} \text{min} & J(y,u) = G\left(y(T)\right) + \frac{\beta}{2}\,|u|^2_{L^2(\Sigma)} \\ \text{over} & (y,u) \in L^2(Q) \times L^2_{T_1}(\Sigma) \\ \text{subject to (2.1),} & \varphi \le u \le \psi \text{ on } \Sigma, \end{cases}$$

where $\beta > 0$, $\varphi, \psi \in L^\infty(L^2(\partial\Omega))$, $\varphi(x) < \psi(x)$ a.e. on Σ, and $G : L^2(\Omega) \to \mathbb{R}$ is bounded below, weakly lower semicontinuous, and C^1. Here

$$L^2_{T_1}(\Sigma) = \{u \in L^2(\Sigma) : u(t,x) = 0 \quad \text{for } t \in (T_1, T)\},$$

with $T_1 \in [0,T]$. For (P2) we require that $\varphi \le 0 \le \psi$ a.e. on (T_1, T). In section 3.2 we shall require that $T_1 < T$. The practical interpretation of setting $u = 0$ in a neighborhood of T is that the controller and the observer are not acting simultaneously. This choice will be important for obtaining better regularity results for $y(T)$.

We refer to (y,u) as a solution of (2.1) if that equation is satisfied in the very weak sense (2.2). Throughout this section the regularity assumptions of Theorem 2.1 for b are supposed to hold, and

$$f \in L^2(Q), \quad y_0 \in L^2(\Omega).$$

Then we have the following result.

PROPOSITION 3.1. *There exist solutions* $(y^*, u^*) = (y(u^*), u^*)$ *to* (P1) *as well as* (P2)*, which are unique if* G *is convex.*

This follows from weak sequential limit arguments (see, e.g., [36]) utilizing Theorem 2.1, respectively, Corollary 2.2. For (P1) a lower bound $\varphi \le u$ can be added and treated as we do for (P2). In (P2) the simultaneous use of upper and lower bounds for the control is essential to guarantee the $L^\infty(L^2(\partial\Omega))$ bound for the controls, which is required by Corollary 2.2.

The above theorem is valid for all $T_1 \le T$. If one additionally assumes that $T_1 < T$, then the condition $\varphi, \psi \in L^\infty(L^2(\partial\Omega))$ can be weakened to $\varphi, \psi \in L^2(\Sigma)$, and the statement of the theorem follows from additional regularity of $y(T)$ in this case.

3.1. Problem (P1). To argue the existence of Lagrange multipliers for the inequality constraint in (P1), we introduce

$$\begin{aligned} e = (e_1, e_2) : (L^2(Q) \cap H^1(H^{-2})) \times L^2(\Sigma) &\to L^2(H^{-2}(\Omega)) \times H^{-1}(\Omega)\,, \\ g : L^2(\Sigma) &\to L^2(\Sigma) \end{aligned}$$

by

$$\begin{aligned} \langle e_1(y,u), v\rangle &= \int_0^T \left(\langle \partial_t y - f, v\rangle - (y \operatorname{div} b, v) - \kappa(y, \Delta v) - (y, b\cdot\nabla v) + \kappa \left(u, \frac{\partial}{\partial n} v\right)_{\partial\Omega} \right) dt\,, \\ e_2(y,u) &= y(0) - y_0\,, \\ g(u) &= u - \psi \end{aligned}$$

for arbitrary $v \in L^2(H^2(\Omega) \cap H^1_0(\Omega))$. Recall that $L^2(Q) \cap H^1(H^{-2}) \subset C(H^{-1}(\Omega))$, so that e_2 is well defined. The linearizations e' of e and g' of g are obtained from e and g by deleting the affine terms y_0, f, and ψ, respectively. We introduce the Lagrangian

$$\mathcal{L}(y,u,p,p_0,\lambda) = G(y) + \frac{\beta}{2}|u|^2_{L^2(\Sigma)} + \langle (p,p_0), e(y,u)\rangle + (\lambda, g(u)).$$

From Theorem 2.1 it follows that (e', g') is surjective, and hence there exists a Lagrange multiplier $(p, p_0, \lambda) \in L^2(H^2(\Omega) \cap H^1_0(\Omega)) \times H^1_0(\Omega) \times L^2(\Sigma)$ associated to the constraints (e,g); see, e.g., [37]. It follows that the optimality system satisfied by an optimal pair (y^*, u^*) is obtained by setting $\nabla_{y,u,p,p_0}\mathcal{L}(y,u,p,p_0,\lambda) = 0$, and $\lambda \geq 0$, $g(u) \leq 0$, $\lambda\, g(u) = 0$. Consequently the optimality system for (P1) is given by

$$
\begin{cases}
\partial_t y - \kappa\Delta y + b\cdot\nabla y = f \quad \text{in } Q,\\
y = u \text{ on } \ \Sigma,\ y(0) = y_0 \quad \text{in } \Omega,\\[1ex]
-\partial_t p - \kappa\Delta p - \operatorname{div} b\, p - b\cdot\nabla p = -G'(y) \quad \text{in } Q,\\
p = 0 \ \text{ on } \ \Sigma, \quad p(T) = 0 \quad \text{in } \Omega,\\[1ex]
\kappa \frac{\partial p}{\partial n} + \beta u + \lambda = 0 \quad \text{on } \Sigma,\\[1ex]
\lambda = \max(0, \lambda + c(u-\psi)) \quad \text{on } \Sigma
\end{cases}
\tag{3.1}
$$

for any $c > 0$. Moreover, $p(0) = p_0$. Note that the last equation in (3.1) is equivalent to $\lambda \geq 0$, $u \leq \psi$, and $\lambda(u - \psi) = 0$. The equations in the last two lines of (3.1) are equivalent to

$$u = \min\left(\psi, -\frac{\kappa}{\beta}\frac{\partial p}{\partial n}\right).$$

The equations in the first two lines of (3.1) are understood in the sense of very weak solutions. The time derivative in $\partial_t p$ must first be interpreted in variational form, but from the third equation in (3.1) it immediately follows that $p \in L^2(H^2(\Omega)\cap H_0^1(\Omega))\cap H^1(L^2(\Omega))$. This is consistent with the regularity results for parabolic equations, since $G'(y) \in L^2(Q)$; see, e.g., [31, p. 342]. If G is convex, then (3.1) is a necessary and sufficient optimal condition for (P1).

We now turn to regularity properties of the optimal solution on Σ. This result is essential for superlinear convergence of the PDAS method; see section 4. Henceforth let (y, u, p, λ) denote a solution to (3.1). The active and inactive sets at a solution are denoted by

$$\mathcal{A} = \{\,(t,x) \in \Sigma \,:\, u(t,x) = \psi\,\}, \quad \mathcal{I} = \{\,(t,x) \in \Sigma \,:\, u(t,x) < \psi\,\}.$$

THEOREM 3.2. *On the inactive set $\mathcal{I}$ we have for the optimal solution $u|\mathcal{I} \in L^{q_n}(\mathcal{I})$ with*

$$
q_n = \begin{cases} \frac{2(n+1)}{n} & \text{if } n \geq 3\,,\\ 3-\varepsilon & \text{if } n = 2\,. \end{cases}
\tag{3.2}
$$

On the active set the regularity of u is determined by ψ. Moreover,

$$\frac{\partial p}{\partial n} \in L^{q_n}(\Sigma) \quad \text{and} \quad \left\|\frac{\partial p}{\partial n}\right\|_{L^{q_n}(\Sigma)} \leq C\,\|p\|_{L^2(H^2(\Omega))\cap H^1(L^2(\Omega))}$$

with an embedding constant C.

Proof. As already noted, $p \in L^2(H^2(\Omega)) \cap H^1(L^2(\Omega))$. This implies that

$$\frac{\partial p}{\partial n} \in L^2(H^{\frac12}(\partial\Omega)) \cap H^{\frac14}(L^2(\partial\Omega));$$

see [21], or [31, Chapter II and p. 342]. Since $H^{\frac14}(L^2(\partial\Omega)) \hookrightarrow L^4(L^2(\partial\Omega))$ (see [1]), we find

$$\frac{\partial p}{\partial n} \in L^2(H^{\frac12}(\partial\Omega)) \cap L^4(L^2(\partial\Omega)), \tag{3.3}$$

and hence interpolation [42, Chapter 1] implies that

$$\frac{\partial p}{\partial n} \in L^{r_s}([H^{\frac{1}{2}}(\partial\Omega), L^2(\partial\Omega)]_s), \quad \text{where } \frac{1}{r_s} = \frac{1-s}{2} + \frac{s}{4}\,.$$

For $n \geq 3$ we use the fact that for $H^{\frac{1}{2}}(\partial\Omega) \hookrightarrow L^{\frac{2n-2}{n-2}}(\partial\Omega)$, and we obtain

$$[H^{\frac{1}{2}}(\partial\Omega), L^2(\partial\Omega)]_s \hookrightarrow L^{q_s}(\partial\Omega), \quad \text{where } \frac{1}{q_s} = \frac{(1-s)(n-2)}{2n-2} + \frac{s}{2}\,.$$

Next we choose s such that $r_s = q_s$, i.e.,

$$r_s = \frac{8}{4-2s} = \frac{2n-2}{n+s-2} = q_s.$$

This implies that $s = \frac{2}{n+1}$ and hence $q_s = \frac{2(n+1)}{n}$. Consequently for $n \geq 3$ we obtain $\frac{\partial p}{\partial n} \in L^{\frac{2(n+1)}{n}}(\Sigma)$.

For $n = 2$ we have that $H^{\frac{1}{2}}(\partial\Omega) \hookrightarrow L^r(\partial\Omega)$ for all $r < \infty$. Using similar arguments to those before, we deduce that $\frac{\partial p}{\partial n} \in L^{3-\frac{1}{r-1}}(\Sigma)$.

From (3.1) we have that $\frac{\partial p}{\partial n} = -\beta u$ on $\mathcal{I}$, and the asserted regularity of u follows. The desired estimate for $\|\frac{\partial p}{\partial n}\|_{L^{q_n}(\Sigma)}$ holds due to the continuity of all embeddings involved. $\square$

Our next objective is to show that for the optimal solution u the corresponding very weak solution y to the state equation is in fact a variational solution in the sense that $y \in L^2(H^1(\Omega)) \cap H^1(H^{-1}(\Omega))$, $y = u$ a.e. on Σ, and

$$\int_Q \partial_t y\, v\; dxdt = \int_Q (-\kappa \nabla y\, \nabla v - b \cdot \nabla y\, v + fv) dxdt$$

for all $v \in L^2(H^2(\Omega) \cap H_0^1(\Omega))$. This is important for numerical realizations which are conveniently based on this formulation. We shall require the following lemma, which uses the notion of uniform 1-smooth regularity property of the boundary, for which we refer to [1].

LEMMA 3.3. *Let D be a domain in $\mathbb{R}^n$, having the uniform 1-smooth regularity property and a bounded boundary, and let $s \in [0,1]$.*

(a) *If $v \in H^s(D)$, then $\max(0,v) \in H^s(D)$ and*

$$|\max(0,v)|_{H^s(D)} \leq |v|_{H^s(D)}.$$

(b) *If $v \in H^s(0,T;L^2(D))$, then $\max(0,v) \in H^s(0,T;L^2(D))$ and*

$$|\max(0,v)|_{H^s(0,T;L^2(D))} \leq |v|_{H^s(0,T;L^2(D))}\,.$$

Proof. (a) For $s = 0$ the claim is trivial and for $s = 1$ it is well known; see [42]. Thus let us consider the case $0 < s < 1$. Under the stated regularity properties for ∂D, the interpolation norm on $H^s(D)$ is equivalent to the intrinsic $H^s(D)$-norm on D given by

$$|v|^2_{L^2(D)} + \int_D \int_D \frac{|v(x)-v(y)|^2}{|x-y|^{n+2s}}\, dxdy; \tag{3.4}$$

see [1]. Let $S_i \subset D \times D$ be given by

$$S_1 = \{(x,y) : v(x) \geq 0,\, v(y) \geq 0\},\; S_2 = \{(x,y) : v(x) \geq 0,\, v(y) < 0\},$$
$$S_3 = \{(x,y) : v(x) < 0,\, v(y) \geq 0\},\; S_4 = \{(x,y) : v(x) < 0,\, v(y) < 0\}.$$

Then with $v^+ = \max(0, v)$

$$\int_D \int_D \frac{|v^+(x) - v^+(y)|^2}{|x-y|^{n+2s}}\, dxdy \le \int_{s_1 \cup s_2 \cup s_3} \int_{s_1 \cup s_2 \cup s_3} \frac{|v(x) - v(y)|^2}{|x-y|^{n+2s}}\, dxdy \le \int_D \int_D \frac{|v(x) - v(y)|^2}{|x-y|^{n+2s}}\, dxdy,$$

and (a) follows. Turning to (b), from [29, Theorem 1.7], it is known that for $s \in (0,1)$ up to equivalence of norms we have

$$|v|^2_{H^s(L^2(D))} = |v|^2_{L^2(L^2(D))} + 2\int_0^T \int_0^{T-t} t^{-1-2s}\, |v(\tau) - v(t+\tau)|^2_{L^2(D)} d\tau dt.$$

Setting $t + \tau = r$ the last term can equivalently be expressed as

$$\int_0^T \int_\tau^T |r - \tau|^{-1-2s}\, |v(\tau) - v(r)|^2\, dr d\tau,$$

and using the symmetry of this expression with respect to s and τ, we find

$$|v|^2_{H^s(L^2(D))} = |v|^2_{L^2(L^2(D))} + \int_0^T \int_0^T \frac{|v(\tau) - v(r)|^2_{L^2(D)}}{|\tau - r|^{1+2s}}\, dr d\tau,$$

which is analogous to (3.4). The integral term can be expressed as

$$\int_0^T \int_0^T \int_\Omega \frac{|v(\tau,x) - v(r,x)|^2}{|\tau - r|^{1+2s}}\, dx\, dr\, d\tau,$$

and hence the proof can be completed as in (a). ∎

THEOREM 3.4. *Let* (y, u) *denote a solution to* (P1) *and assume that* $\psi \in L^2(H^{\frac{1}{2}}(\partial\Omega)) \cap H^{\frac{1}{4}}(L^2(\partial\Omega))$. *Then* y *is a variational solution of the state equation with*

$$u \in L^2(H^{\frac{1}{2}}(\partial\Omega)) \cap H^{\frac{1}{4}}(L^2(\partial\Omega)) \quad \text{and} \quad y \in L^2(H^1(\Omega)) \cap H^{\frac{1}{2}}(L^2(\Omega)) \cap H^1(H^{-1}(\Omega)).$$

If, moreover, $G'(y) \in L^2(H^1(\Omega)) \cap H^{\frac{1}{2}}(L^2(\Omega))$, $y_0 \in H^{\frac{1}{2}-\epsilon}(\Omega)$, *and* $\psi \in L^2(H^1(\partial\Omega)) \cap H^{\frac{1}{2}}(L^2(\partial\Omega))$, *then*

$$u \in L^2(H^1(\partial\Omega)) \cap H^{\frac{1}{2}}(L^2(\partial\Omega)) \qquad \text{and} \qquad y \in L^2(H^{\frac{3}{2}-\epsilon}(\Omega)) \cap H^{\frac{3-2\epsilon}{4}}(L^2(\Omega))$$

for every $\epsilon \in (0, \frac{1}{2}]$. *In addition* $u = 0$ *on* $\mathcal{I} \cap (\{T\} \times \partial\Omega)$.

Proof. From the proof of Theorem 3.2 we have that

$$\frac{\partial p}{\partial n} \in L^2(H^{\frac{1}{2}}(\partial\Omega)) \cap H^{\frac{1}{4}}(L^2(\partial\Omega)).$$

From (3.1) with $\beta = c$ we deduce that $u = \min(0, -\frac{1}{\beta}\frac{\partial p}{\partial n} - \psi) + \psi$, and hence Lemma 3.3 implies that $u \in L^2(H^{\frac{1}{2}}(\partial\Omega)) \cap H^{\frac{1}{4}}(L^2(\partial\Omega))$. By regularity results for parabolic equations based on interpolation theory [35, p. 78] (with $s = -\frac{1}{2}$), we obtain that $y \in L^2(H^1(\Omega)) \cap H^{\frac{1}{2}}(L^2(\Omega))$. Therefore

$$\int_0^T \langle \partial_t y, v\rangle\, dt = \int_Q (-\kappa \nabla y\, \nabla v - b \cdot \nabla y\, v + fv) dxdt$$

for all $v \in L^2(H^2(\Omega) \cap H_0^1(\Omega))$. Since the right-hand side can uniquely be extended to a continuous functional on $L^2(H_0^1(\Omega))$, it follows that $\partial_t y \in L^2(H^{-1}(\Omega))$. From (2.7), moreover, $y = u$ in $L^2(H^{\frac{1}{2}}(\partial\Omega))$. We conclude that y is a variational solution to (2.2).

If $G'(y) \in L^2(H^1(\Omega)) \cap H^{\frac{1}{2}}(L^2(\Omega))$, then $p \in L^2(H^3(\Omega)) \cap H^{\frac{3}{2}}(L^2(\Omega))$ [35, p. 32] (with $k = 1$). It follows that $\frac{\partial p}{\partial n} \in L^2(H^{\frac{3}{2}}(\partial\Omega)) \cap H^{\frac{3}{4}}(L^2(\partial\Omega))$; see, e.g., [20, p. 9]. Due to the regularity assumption on ψ and Lemma 3.3, we find that $u \in L^2(H^1(\partial\Omega)) \cap H^{\frac{1}{2}}(L^2(\partial\Omega))$. This implies that $y \in L^2(H^{\frac{3}{2}-\epsilon}(\Omega)) \cap H^{\frac{3}{4}-\frac{\epsilon}{2}}(L^2(\Omega))$ for every $\epsilon > 0$ [35, p. 78] (with $s = -\frac{1}{4} - \frac{\epsilon}{2}$). Regularity of p implies that $p(T) \in H^{2-\epsilon}(\Omega)$ and hence $\frac{\partial p}{\partial n}(T) \in H^{\frac{1}{2}-\epsilon}(\partial\Omega)$. Since $p(T) = 0$ on Ω we find that $\frac{\partial p}{\partial n}(T) = 0$ on $\partial\Omega$. Hence from the fifth equation in (3.1) we deduce that $u = 0$ on $\mathcal{I} \cap (\{T\} \times \partial\Omega)$. $\square$

Remark 3.1. For $G(y) = \frac{1}{2}|y - y_d|^2$ the condition $G'(y) \in L^2(H^1(\Omega)) \cap H^{\frac{1}{2}}(L^2(\Omega))$ is satisfied if $y_d \in L^2(H^1(\Omega)) \cap H^{\frac{1}{2}}(L^2(\Omega))$ and $\psi \in L^2(H^{\frac{1}{2}}(\partial\Omega)) \cap H^{\frac{1}{4}}(L^2(\partial\Omega))$.

COROLLARY 3.5 (extra L^p regularity). *By interpolation one can show that if $u \in L^2(H^1(\partial\Omega)) \cap H^{\frac{1}{2}}(L^2(\partial\Omega))$, then $u \in L^{q_\epsilon}(\Sigma)$, where $q_\epsilon = \frac{2(n+1)}{n-1} - \epsilon$, for every $\epsilon > 0$.*

3.2. Problem (P2).

We first derive the optimality system for (P2). This requires more care than for (P1) since G in this case is not defined on the space of trajectories $L^2(Q)$.

Let (y, u) denote an optimal solution to (P2). We shall require that $G'(y(T)) \in H_0^1(\Omega)$. This will guarantee the required regularity of the adjoint state. In case $G(y(T)) = \frac{1}{2}\,|y(T) - z|^2$, this is the case if $y(T) - z \in H_0^1(\Omega)$, i.e., we require regularity of $y(T)$ (and z) beyond that which is guaranteed by Corollary 2.2, as well as boundary conditions for $y(T) - z$. The required regularity of y at T can be achieved by restricting u to be a function of t only in a neighborhood of T. To take into consideration the additional boundary condition, we require that $u = 0$ in a neighborhood of $T = 0$. Then by semigroup theory $y(T) \in H_0^1(\Omega) \cap H^2(\Omega)$ and, if $z \in H_0^1(\Omega)$, we have $y(T) - z \in H_0^1(\Omega)$. Thus for tracking-type functionals the requirement that $G'(y(T)) \in H_0^1(\Omega)$ holds if $u \in L^2_{T_1}(\Sigma)$ and $z \in H_0^1(\Omega)$. This motivates the use of $L^2_{T_1}(\Sigma)$ in (P2).

THEOREM 3.6. *Let (y, u) denote a solution to* (P2) *with $T_1 < T$ and assume that $G'(y(T)) \in H_0^1(\Omega)$. Then there exist $p \in L^2(H^2(\Omega) \cap H_0^1(\Omega)) \cap H^1(L^2(\Omega))$ and $\lambda \in L^2(\Sigma_{T_1})$ such that for all $c > 0$*

$$
(3.5) \qquad \begin{cases}
\partial_t y - \kappa\Delta y + b \cdot \nabla y = f \quad \text{in } Q, \\
y = u \quad \text{on } \Sigma, \quad y = y_0 \quad \text{in } \Omega, \\
-\partial_t p - \kappa\Delta p - \operatorname{div} b\, p - b \cdot \nabla p = 0 \quad \text{in } Q, \\
p = 0 \quad \text{on } \Sigma, \quad p(T) = -G'(y(T)) \quad \text{in } \Omega, \\
\kappa \frac{\partial p}{\partial n} + \beta u + \lambda = 0 \quad \text{on } \Sigma_{T_1}, \\
\lambda = \max(0, \lambda + c(u - \psi)) + \min(0, \lambda + c(u - \varphi)) \quad \text{on } \Sigma_{T_1}
\end{cases}
$$

holds, where $\Sigma_{T_1} = (0, T_1) \times \partial\Omega$.

Proof. From Theorem 2.1 the affine mapping $u \to y(u)$ is continuous from $L^2(\Sigma)$ to $L^2(Q) \cap H^1(H^{-2}(\Omega))$. The linearization $\dot{y}$ at u in direction h satisfies

$$
(3.6) \quad \langle \partial_t \dot{y}(t), v \rangle - \kappa(\dot{y}(t), \Delta v) - (\dot{y}(t), \operatorname{div}(b(t))v) - (\dot{y}(t), b(t)\nabla v) \\
= \kappa \left(h(t), \frac{\partial}{\partial n} v \right)_{\partial\Omega} \quad \text{for all } v \in H^2(\Omega) \cap H_0^1(\Omega) \text{ and a.e. } t \in (0, T).
$$

Moreover, by Corollary 2.2, the affine mapping $u \to y(T;u)$ is continuous from $L^\infty(\Sigma)$ to $L^2(\Omega)$, and hence it is differentiable at u in directions $h \in L^\infty(\Sigma)$. By assumption, $G'(y(T)) \in H_0^1(\Omega))$, and hence the solution to the adjoint equation satisfies $p \in L^2(H^2(\Omega) \cap H_0^1(\Omega)) \cap H^1(L^2(\Omega))$ [31]. Let $j(u) = J(y(u),u)$ denote the reduced cost functional corresponding to (P2). For the derivative at $u \in L^\infty(\Sigma)$ in direction $h \in L^2(\Sigma)$ we obtain by (3.6)

$$
\begin{aligned}
&(j'(u),h)_{L^2(\Sigma)} = (G'(y(T)),\, \dot y(T))_{L^2(\Omega)} + \beta(u,h)_{L^2(\Sigma)}\\
&= -(p(T),\, \dot y(T))_{L^2(\Omega)} + \beta(u,h)_{L^2(\Sigma)} = -\int_0^T \frac{d}{dt}(p(t),\dot y(t))_{L^2(\Omega)}dt + \beta(u,h)_{L^2(\Sigma)}\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad = \left(\kappa\frac{\partial p}{\partial n} + \beta u,\, h\right)_{L^2(\Sigma)}.
\end{aligned}
$$

At the solution we therefore have

$$
(j'(u),\, h-u) \ge 0 \quad \text{for all} \ \ h \in L^2_{T_1}(\Sigma), \ \text{with} \ \ \varphi \le h \le \psi. \tag{3.7}
$$

Note that the directions h in (3.7) are in $L^\infty_{T_1}(\Sigma)$ as well. Define

$$
\mathcal{A}_\varphi = \{(t,x) \in \Sigma_{T_1} : u = \varphi\}, \quad \mathcal{A}_\psi = \{(t,x) \in \Sigma_{T_1} : u = \psi\}, \quad \mathcal{I} = \Sigma_{T_1} \setminus (\mathcal{A}_\varphi \cup \mathcal{A}_\psi),
$$

where $\Sigma_1 = (0,T_1)\times\partial\Omega$. Set $\mathcal{S} = \{(t,x) \in \mathcal{I} : j'(u) \ge 0\}$ and define $\bar h = \varphi\,\chi_{\mathcal{S}} + u\,\chi_{\Sigma\setminus\,\mathcal{S}}$, which satisfies $\varphi \le \bar h \le \psi$ on Σ_{T_1}. By (3.7)

$$
0 \le (j'(u),\, \bar h - u)_{L^2(\Sigma_{T_1})} = (j'(u),\, \varphi - u)_{L^2(\mathcal{S})} \le 0,
$$

and hence $j'(u) = 0$ on $\mathcal{S}$, since $\varphi < u < \psi$ on $\mathcal{S}$. Analogously one shows that $j'(u) = 0$ on $\{(t,x) \in \mathcal{I} : j'(u) \le 0\}$ and hence $j'(u) = 0$ on $\mathcal{I}$. Next set $\mathcal{S}_\psi = \{(t,x) \in \mathcal{A}_\psi : j'(u) \ge 0\}$, and define $\bar h = \varphi\,\chi_{\mathcal{S}_\psi} + u\,\chi_{\Sigma\,\setminus\,\mathcal{S}_\psi}$. Then by (3.7)

$$
0 \le (j'(u),\, \bar h - u)_{L^2(\Sigma_{T_1})} = (j'(u),\, \varphi - \psi)_{L^2(\Sigma_{T_1})} \le 0.
$$

Since $\varphi < \psi$ a.e. on Σ_{T_1} this implies that $j'(u) = 0$ on $\mathcal{S}_\psi$ and hence $j'(u) \le 0$ on $\mathcal{A}_\psi$. Analogously one shows that $j'(u) \ge 0$ on $\mathcal{A}_\varphi$.

Setting

$$
\lambda = \begin{cases} -\kappa\frac{\partial p}{\partial n} - \beta u & \text{on } \ \Sigma_{T_1} \setminus \mathcal{I},\\ 0 & \text{on } \ \mathcal{I}, \end{cases}
$$

the last two equations of (3.5) follow and the optimality system is verified. ◻

COROLLARY 3.7. *Under the assumptions of Theorem* 3.4 *we have* $\frac{\partial p}{\partial n} \in L^{q_n}(\Sigma)$ *and* $u|\mathcal{I} \in L^{q_n}(\mathcal{I})$ *with* q_n *defined in* (3.2).

This is a direct consequence of Theorem 3.6, which asserts that $p \in L^2(H^2(\Omega)) \cap H^1(L^2(\Omega))$, and of the proof of Theorem 3.2.

COROLLARY 3.8. *Under the assumptions of Theorem* 3.6 *and if* $\varphi, \psi \in L^2(H^{\frac12}(\partial\Omega)) \cap H^{\frac14}(L^2(\partial\Omega))$, *then* y *is a variational solution of the state equation with*

$$
u \in L^2(H^{\frac12}(\partial\Omega)) \cap H^{\frac14}(L^2(\partial\Omega)) \ \ \textit{and} \ \ y \in L^2(H^1(\Omega)) \cap H^{\frac12}(L^2(\Omega)) \cap H^1(H^{-1}(\Omega)).
$$

If, moreover, $G'(y(T)) \in H^2(\Omega)\cap H_0^1(\Omega)$, $y_0 \in H^{\frac12-\epsilon}(\Omega)$, *and* $\varphi, \psi \in L^2(H^1(\partial\Omega)) \cap H^{\frac12}(L^2(\partial\Omega))$, *then*

$$
u \in L^2(H^1(\partial\Omega)) \cap H^{\frac{1-\varepsilon}{2}}(L^2(\partial\Omega)) \ \ \textit{and} \ \ y \in L^2(H^{\frac32-\epsilon}(\Omega)) \cap H^{\frac{3-2\epsilon}{4}}(L^2(\Omega))
$$

for every $\epsilon \in (0, \frac{1}{2}]$.

Proof. The proof of the first part is similar to that of Theorem 3.4. By the last two equations of (3.5) we find

$$u = \max\left(\varphi, \min\left(\psi, -\frac{\kappa}{\beta}\frac{\partial p}{\partial n}\right)\right) \qquad \text{a.e. on } \Sigma_{T_1}, \tag{3.8}$$

which is equivalent to $u = \max(0, \min(0, -\frac{\kappa}{\beta}\frac{\partial p}{\partial n} - \psi) + \psi - \varphi) + \varphi$. Since $\frac{\partial p}{\partial n} \in L^2(H^{\frac{1}{2}}(\partial\Omega)) \cap H^{\frac{1}{4}}(L^2(\partial\Omega))$ this implies by Lemma 3.3 that

$$u|(0, T_1) \in L^2(0, T_1; H^{\frac{1}{2}}(\partial\Omega)) \cap H^{\frac{1}{4}}(0, T_1; L^2(\partial\Omega)),$$

and by concatenation of functions in $H^{\frac{1}{4}}$ this implies that

$$u \in L^2(0, T; H^{\frac{1}{2}}(\partial\Omega)) \cap H^{\frac{1}{4}}(0, T; L^2(\partial\Omega))$$

(see [29, Proposition 1.13]), and hence $y \in L^2(H^1(\Omega)) \cap H^{\frac{1}{2}}(L^2(\Omega))$. Turning to the second part of the proof, we assume that $G'(y(T)) \in H^2(\Omega) \cap H_0^1(\Omega)$. Then $p \in L^2(H^3(\Omega)) \cap H^{\frac{3}{2}}(L^2(\Omega))$ [35, p. 32], and $\frac{\partial p}{\partial n} \in L^2(H^{\frac{3}{2}}(\partial\Omega)) \cap H^{\frac{3}{4}}(L^2(\partial\Omega))$. By (3.8) and concatenation of H^s-functions with $s \in [0, \frac{1}{2})$, we find that $u \in L^2(H^1(\partial\Omega)) \cap H^{\frac{1-\varepsilon}{2}}(L^2(\partial\Omega))$ for every $\varepsilon \in (0, 1)$. This implies that $y \in L^2(H^{\frac{3}{2}-\varepsilon}(\Omega)) \cap H^{\frac{3-2\varepsilon}{4}}(L^2(\Omega))$. ∎

4. The PDAS strategy and its convergence properties. The PDAS strategy has proved to be very efficient for solving constrained optimal control problems [8]. We describe it here for (P1) and defer the necessary modifications for (P2) to Remark 4.3.

In addition to the assumptions on $G : L^2(Q) \to \mathbb{R}$ made in section 3, we assume that G is convex so that all auxiliary optimal control problems that arise in this section have unique solutions.

The PDAS strategy is an iterative algorithm which, based on the current iterate (u_k, λ_k), defines the active set

$$\mathcal{A}_k = \{\, x \in \Omega \,:\, \lambda_k(x) + c(u_k - \psi)(x) > 0 \,\}$$

and the inactive set

$$\mathcal{I}_k = \Omega \setminus \mathcal{A}_k.$$

The subsequent step consists in solving the optimal control problem with equality constraints only:

$$(\mathrm{P_k}) \qquad \begin{cases} \min & J(y, u) = G(y) + \frac{\beta}{2}\,|u|^2_{L^2(\Sigma)} \\ \text{over} & (y, u) \in L^2(Q) \times L^2(\Sigma) \\ \text{subject to (2.1)} & \text{and } u = \psi \ \text{ on } \mathcal{A}_k. \end{cases}$$

This leads to the following iterative algorithm, in which step (iii) is the necessary and sufficient optimality condition for (P_k).

PDAS ALGORITHM.

(i) Choose $(u_1, \lambda_1) \in L^2(\Sigma) \times L^2(\Sigma)$, $c > 0$.

(ii) Define $\mathcal{A}_k = \{\, x \in \Omega \,:\, \lambda_k(x) + c(u_k - \psi)(x) > 0 \,\}$, $\mathcal{I}_k = \Omega \setminus \mathcal{A}_k$.

(iii) Solve for $(y_{k+1}, u_{k+1}, p_{k+1}) \in L^2(Q) \cap H^1(H^{-2}(\Omega)) \cap C(H^{-1}(\Omega)) \times L^2(\Sigma) \times L^2(H^2(\Omega) \cap H^1_0(\Omega))$:

$$
\begin{cases}
\partial_t y - \kappa\Delta y + b\cdot\nabla y = f \quad \text{in } Q,\\
y = u \quad \text{on } \Sigma, \quad y(0) = y_0 \quad \text{in } \Omega,\\
-\partial_t p - \kappa\Delta p - \operatorname{div} b\, p - b\cdot\nabla p = -G'(y) \quad \text{in } Q,\\
p = 0 \quad \text{on } \Sigma, \quad p(T) = 0 \quad \text{in } \Omega,\\
u = \psi \text{ on } \mathcal{A}_k, \quad \kappa\frac{\partial p}{\partial n} + \beta u = 0 \text{ on } \mathcal{I}_k.
\end{cases} \tag{4.1}
$$

(iv) Set

$$
\lambda_{k+1} = \begin{cases} 0 & \text{on } \mathcal{I}_k,\\ -\kappa\frac{\partial p_{k+1}}{\partial n} - \beta\psi & \text{on } \mathcal{A}_k. \end{cases}
$$

(v) Stop or return to (ii).

Remark 4.1. For practical features of this algorithm, we refer to [8] and [9], for example. Here, we mention only that

1. for $k \geq 2$ the iterates of the algorithm are independent of the choice of c, and
2. if the algorithm finds two successive active sets, for which $\mathcal{A}_k = \mathcal{A}_{k+1}$, then $(y(u_k), u_k)$ is the solution of the problem.

The latter will be used as a stopping criterion for numerical examples in section 6.

Remark 4.2. The equality-constrained optimization problem (P_k) is solved using the Newton method for the reduced cost functional $j(u) = G(y(u)) + \frac{\beta}{2}|u|^2_{L^2(\Sigma)}$. The required first and second derivatives of j are computed using solutions of the adjoint problems; see, e.g., [3]. In section 5 we describe the computation of these derivatives on the discrete level.

For the following result it will be convenient to choose a specific initialization for λ, given by

$$
\begin{cases}
\text{choose} & u_1 \in L^2(\Sigma),\\
\text{set} & \lambda_1 = -\kappa\frac{\partial p(u_1)}{\partial n} - \beta u_1,\\
\text{and set} & c = \beta \text{ for the first iteration.}
\end{cases} \tag{4.2}
$$

THEOREM 4.1. *If the PDAS algorithm is initialized by* (4.2), *and if further* $\psi \in L^{\frac{2(n+1)}{n}}(\Sigma)$, $G' : L^2(Q) \to L^2(Q)$ *is locally Lipschitz, and* $|u_1 - u^*|_{L^2(\Sigma)}$ *is sufficiently small, then the iterates* $(y_k, u_k, p_k, \lambda_k)$ *converge superlinearly in* $L^2(Q) \cap H^1(H^{-2}(\Omega)) \cap C(H^{-1}(\Omega)) \times L^2(\Sigma) \times L^2(H^2(\Omega) \cap H^1_0(\Omega)) \times L^2(\Sigma)$ *to* $(y^*, u^*, p^*, \lambda^*)$.

Proof. Let us consider λ in the last equation of (3.1) as a function of u. Then (3.1) can equivalently be expressed as

$$
F(u) = \lambda(u) - \max(0, \lambda(u) + \beta(u - \psi)) = 0, \text{ where } F : L^2(\Sigma) \to L^2(\Sigma). \tag{4.3}
$$

Note that (4.3) is equivalent to

$$
F(u) = \beta u - \beta\psi + \max\left(0, \kappa\frac{\partial p}{\partial n} + \beta\psi\right) = 0, \tag{4.4}
$$

due to the fifth equation in (3.1). By Theorem 3.2 and the assumption that $\psi \in L^{\frac{2(n+1)}{n}}(\Sigma)$ we have that $\kappa\frac{\partial p}{\partial n} + \beta\psi \in L^{q_n}(\Sigma)$ with q_n defined in (3.2). Due to the fact that $q_n > 2$ we obtain that

$$u \to F(u)$$

is Newton differentiable, as introduced in Definition 1 of [25] (see Proposition 4.1 of [25]), with the generalized derivate of F at u in direction $h \in L^2(\Sigma)$ given by

$$G_F(u)h = \beta h + G_{\max}\left(\kappa\frac{\partial p}{\partial n} + \beta\psi\right)\frac{\partial p(h)}{\partial n},$$

where

$$G_{\max}(u)(x) = \begin{cases} 1 & \text{if } \ u(x) > 0, \\ 0 & \text{if } \ u(x) \le 0. \end{cases}$$

It was proved in general terms in [25, Theorem 4.1] that $G_F(u)$ has a bounded inverse from $L^2(\Sigma)$ to itself for every $u \in L^2(\Sigma)$. Hence it follows that the semismooth Newton algorithm applied to $F(u) = 0$ is locally superlinearly convergent. The semismooth Newton iteration consists of the iteration

$$\begin{cases} G_F(u_\kappa)\delta u = -F(u_\kappa), \\ u_{k+1} = u_k + \delta u. \end{cases} \tag{4.5}$$

In the following arguments we show that the semismooth Newton iteration and the PDAS strategy coincide. In principle this argument can be extracted from [25], but we believe that it is instructive to carry it out for the present case. A short consideration shows that a semismooth Newton step (4.5) is equivalent to

$$\begin{cases} \partial_t y_{k+1} - \kappa\Delta y_{k+1} + b\cdot\nabla y_{k+1} = f \quad \text{in } Q, \\ y_{k+1} = u_{k+1} \quad \text{on } \Sigma,\ y(0) = y_0 \quad \text{in } \Omega, \\ -\partial_t p_{k+1} - \kappa\Delta p_{k+1} - \operatorname{div} b\, p_{k+1} - b\cdot\nabla p_{k+1} = -G'(y_{k+1}) \quad \text{in } Q, \\ p_{k+1} = 0 \quad \text{on } \Sigma, \quad p_{k+1}(T) = 0 \quad \text{in } \Omega, \\ u_{k+1} = \psi \quad \text{on } \mathcal{A}_k^{SN},\ \kappa\frac{\partial p_{k+1}}{\partial n} + \beta u_{k+1} = 0 \quad \text{on } \mathcal{I}_k^{SN}, \end{cases} \tag{4.6}$$

where

$$\mathcal{A}_k^{SN} = \left\{x : \left(-\kappa\frac{\partial p_k}{\partial n} - \beta\psi\right)(x) > 0\right\}, \qquad \mathcal{I}_k^{SN} = \Omega \setminus \mathcal{A}_k^{SN}.$$

We further set

$$\lambda_{k+1} = \begin{cases} 0 & \text{on } \mathcal{I}_k^{SN}, \\ -\kappa\frac{\partial p_{k+1}}{\partial n} - \beta\psi & \text{on } \mathcal{A}_k^{SN} \end{cases} \tag{4.7}$$

and observe that

$$\lambda_k + \beta(u_k - \psi) = -\kappa\frac{\partial p_{k+1}}{\partial n} - \beta\psi \quad \text{for } k = 2, 3, \ldots. \tag{4.8}$$

Note that

$$\lambda_k(u_k - \psi) = 0 \quad \text{for } k = 2, 3, \ldots . \tag{4.9}$$

Hence $\lambda_k + \beta(u_k - \psi) > 0$ if and only if $\lambda_k + c(u_k - \psi) > 0$ for any $c > 0$. From (4.8) we have that

$$\mathcal{A}_k = \mathcal{A}_k^{SN} \quad \text{and} \quad \mathcal{I}_k = \mathcal{I}_k^{SN} \quad \text{for } k = 2, 3, \ldots .$$

Therefore the PDAS strategy and the semismooth Newton iteration coincide, provided that their initialization phases coincide. For that it suffices to check that $\mathcal{A}_1 = \mathcal{A}_1^{SN}$. This is the case since for λ_1 as in (4.2) we have

$$\lambda_1 + \beta(u_1 - \psi) = -\kappa \frac{\partial p(u_1)}{\partial n} - \beta\psi_1.$$

Superlinear convergence of y_k to y^* in $L^2(Q) \cap H^1(H^{-2}(\Omega)) \cap C(H^{-1}(\Omega))$ follows from Theorem 2.1. Moreover, superlinear convergence of (p_k, λ_k) to (p^*, λ^*) in $L^2(H^2(\Omega) \cap H_0^1(\Omega)) \times L^2(\Sigma)$ is a consequence of (3.1) and (4.1),

$$\lambda^* - \lambda_k = -\beta(u^* - u_k) - \kappa \left(\frac{\partial p^*}{\partial n} - \frac{\partial p_k}{\partial n} \right),$$

and of Theorem 3.1. □

In Theorem 4.1 we addressed local convergence of the PDAS algorithm. We now turn to global convergence, i.e., to convergence from arbitrary initializations $(u_1, \lambda_1) \in L^2(\Sigma) \times L^2(\Sigma)$.

THEOREM 4.2. *If β is sufficiently large and $G(y) = \frac{1}{2}|y - z|^2_{L^2(Q)}$ for some $z \in L^2(Q)$, then the iterates $(y_k, u_k, p_k, \lambda_k) \to (y^*, u^*, p^*, \lambda^*)$ in $L^2(Q) \cap H^1(H^{-2}(\Omega)) \cap C(H^{-1}(\Omega)) \times L^2(\Sigma) \times L^2(H^2(\Omega) \cap H_0^1(\Omega)) \times L^2(\Sigma)$.*

Proof. Convergence of $(u_k, \lambda_k) \to (u^*, \lambda^*)$ in $L^2(\Sigma) \times L^2(\Sigma)$ follows from a general result in [27, Theorem 4.1]. It requires that $\beta > \|T\|_{\mathcal{L}(L^2(\Sigma), L^2(Q))}$, where $Tu = y(u)$. Convergence of (y_k, u_k) in the specified norms is a consequence of the governing equations for y_k and p_k. □

Remark 4.3. For (P2), under the assumptions of Theorem 3.6, identical assertions to Theorems 4.1 and 4.2 hold. (P2) differs from (P1) in that it involves a terminal observation and bilateral constraints. We again have, provided by Corollary 3.7, the necessary additional regularity $\frac{\partial p}{\partial n} \in L^{q_n}(\Sigma)$. Global convergence and local superlinear convergence for bilaterally constrained problems were treated in [27, Theorems 4.1 and 6.1].

5. Finite element discretization. In this section we discuss the space-time finite element discretization of the optimization problem under consideration. The space discretization of the state equation is based on the usual H^1-conforming finite elements, whereas the time discretization is done by a discontinuous Galerkin method as proposed in [16, 17]. We refer to [3, 38] for a detailed description of the space-time finite element methods for parabolic optimization problems including adaptivity. We emphasize that space-time Galerkin discretizations of optimal control problems allow a natural translation of the optimality system and the optimization algorithms from the continuous to the discrete level: in fact, the approaches "discretize-then-optimize" and "optimize-then-discretize" coincide. We return to this aspect in Remark 6.2 below.

Since the state equation (2.2) is considered in the very weak sense, it may appear at first that its approximation by finite elements based on the standard variational formulation may not be appropriate. However, such an approach is justified since the optimal state and control which need to be approximated possess the common regularity of a variational solution; see Theorem 3.4. For an interesting discussion of finite element discretizations of equations with rough boundary data, we refer to [7] in the elliptic case and to [19] in the parabolic case. Finite element approximation of Dirichlet optimal control problems governed by elliptic equations are discussed in [11, 43].

For this section it is convenient to introduce the following notation: $V = H^1(\Omega)$, $V_0 = H_0^1(\Omega)$, $H = L^2(\Omega)$, and $X = L^2(0,T;V) \cap H^1(0,T;V^*)$. We introduce a bilinear form $a\colon X \times X \to \mathbb{R}$ corresponding to the standard variational formulation of the state equation:

$$a(y,v) = \int_0^T \{(\partial_t y, v) + \kappa(\nabla y, \nabla v) + (b \cdot \nabla y, v)\}\, dt\,.$$

To define the discretization in time, let us partition the time interval $\bar{I} = [0,T]$ as

$$\bar{I} = \{0\} \cup I_1 \cup I_2 \cup \cdots \cup I_M$$

with subintervals $I_m = (t_{m-1}, t_m]$ of size k_m and time points

$$0 = t_0 < t_1 < \cdots < t_{M-1} < t_M = T.$$

We define the discretization parameter k as a piecewise constant function by setting $k|_{I_m} = k_m$ for $m = 1, \ldots, M$.

By means of the subintervals I_m, we define for $r \in \mathbb{N}_0$ a semidiscrete space X_k^r consisting of discontinuous-in-time piecewise polynomial functions:

$$X_k^r = \{\, v_k \in L^2(I, V_0) : v_k\big|_{I_m} \in \mathcal{P}^r(I_m, V_0) \text{ and } v_k(0) \in H \,\}\,.$$

Here, $\mathcal{P}^r(I_m, V_0)$ denotes the space of polynomials up to order r defined on I_m with values in V_0. For the definition of the discontinuous Galerkin method we introduce the following notation for a function $v_k \in X_k^r$:

$$v_{k,m}^+ := \lim_{t\to 0^+} v_k(t_m + t), \quad v_{k,m}^- := \lim_{t\to 0^+} v_k(t_m - t) = v_k(t_m), \quad [v_k]_m := v_{k,m}^+ - v_{k,m}^-\,.$$

Using this notation we define a discretized version of the bilinear form a:

$$\begin{aligned} a_k(y_k, v_k) = \sum_{m=1}^{M} \int_{I_m} &\{(\partial_t y_k, v_k) + \kappa(\nabla y_k, \nabla v_k) + (b \cdot \nabla y_k, v_k)\}\, dt \\ &+ \sum_{m=0}^{M-1} ([y_k]_{m-1}, v_{k,m-1}^+) + (y_{k,0}^-, v_{k,0}^-)\,. \end{aligned}$$

For the space discretization, we consider two- or three-dimensional shape-regular meshes; see, e.g., [12]. A mesh consists of quadrilateral or hexahedral cells K, which constitute a nonoverlapping cover of the computational domain Ω. The corresponding mesh is denoted by $\mathcal{T}_h = \{K\}$, where we define the discretization parameter h as a cellwise constant function by setting $h\big|_K = h_K$ with the diameter h_K of the cell K.

On the mesh $\mathcal{T}_h$ we construct a conforming finite element space $V_h \subset V$ in a standard way:

$$V_h^s = \{\, v \in V : v\big|_K \in \mathcal{Q}^s(K) \text{ for } K \in \mathcal{T}_h \,\}.$$

Here, $\mathcal{Q}^s(K)$ consists of shape functions obtained via bi- or trilinear transformations of polynomials in $\widehat{\mathcal{Q}^s}(\widehat{K})$ defined on the reference cell $\widehat{K} = (0,1)^n$, where

$$\widehat{\mathcal{Q}^s}(\widehat{K}) = \operatorname{span} \left\{ \prod_{j=1}^{n} x_j^{k_j} \, : \; k_j \in \mathbb{N}_0, \; k_j \leq s \right\}.$$

Remark 5.1. The definition of V_h^s can be extended to the case of triangular meshes in the obvious way.

The discrete space with homogeneous Dirichlet boundary conditions is denoted by $V_{h,0}^s = V_h^s \cap H_0^1(\Omega)$. Moreover, we introduce the space of traces of function in V_h^s:

$$W_h^s = \{\, w_h \in H^{1/2}(\partial\Omega) : w_h = \gamma(v_h), v_h \in V_h^s \,\},$$

where $\gamma\colon H^1(\Omega) \to H^{1/2}(\partial\Omega)$ is the trace operator.

With these preliminaries, we define the discrete spaces for the control and state variables:

$$\begin{aligned} X_{k,h}^{r,s} &= \{\, v_{kh} \in L^2(I, V_{h,0}^s) : v_{kh}\big|_{I_m} \in \mathcal{P}^r(I_m, V_{h,0}^s) \text{ and } v_{kh}(0) \in V_h^s \,\} \subset X_k^r\,, \\ U_{k,h}^{r,s} &= \{\, u_{kh} \in L^2(I, W_h^s) : u_{kh}\big|_{I_m} \in \mathcal{P}^r(I_m, W_h^s) \,\}. \end{aligned}$$

Remark 5.2. In the above definition of the discrete spaces $X_{k,h}^{r,s}$ and $U_{k,h}^{r,s}$, we assumed that the spatial discretization is fixed for all time intervals. However, in many situations the use of different meshes $\mathcal{T}_h^m$ for each of the subintervals I_m is required for efficient adaptive discretizations. The consideration of such dynamically changing meshes can be included in the above definitions in a natural way [41].

For a function $u_{kh} \in U_{k,h}^{r,s}$ we define an extension $\widehat{u}_{kh} \in X_{k,h}^{r,s}$ such that

$$\gamma(\widehat{u}_{kh}(t,\cdot)) = u_{kh}(t,\cdot) \text{ and } \widehat{u}_{kh}(t,x_i) = 0 \text{ on all interior nodes } x_i \text{ of } \mathcal{T}_h. \tag{5.1}$$

The optimization problem on the discrete level is then formulated as follows:

$$\min J(y_{kh}, u_{kh}), \quad u_{kh} \in U_{k,h}^{r,s} \cap U_{ad} \tag{5.2}$$

subject to

(5.3)

$$y_{kh} \in \widehat{u}_{kh} + X_{k,h}^{r,s}, \quad a_k(y_{kh}, v_{kh}) = \int_0^T (f, v_{kh})\,dt + (y_0, v_{kh,0}^-) \quad \text{for all } v_{kh} \in X_{k,h}^{r,s}.$$

The discrete state equation (5.3) defines a discrete solution operator S_{kh} which maps a given discrete control u_{kh} to the (unique) solution of (5.3). As on the continuous level we introduce a discrete reduced cost functional

$$j_{kh}(u_{kh}) = J(S_{kh}(u_{kh}), u_{kh}). \tag{5.4}$$

The discrete optimization problem (5.2)–(5.3) is solved by the PDAS strategy described in the previous section. In each step an equality-constrained optimization problem is solved by the Newton method for the discrete reduced cost functional j_{kh}; see Remark 4.2. For the realization of the Newton method, we need representations of the first and second directional derivatives of j_{kh}.

PROPOSITION 5.1. *Let the discrete reduced cost functional j_{kh} be defined as in* (5.4)*. Then the following hold:*

(a) *The first directional derivative in direction $\delta u_{kh} \in U^{r,s}_{k,h}$ can be expressed as*

$$j'_{kh}(u_{kh})(\delta u_{kh}) = J'_y(y_{kh}, u_{kh})(\widehat{\delta u}_{kh}) - a_k(\widehat{\delta u}_{kh}, p_{kh}) + J'_u(y_{kh}, u_{kh})(\delta u_{kh}), \tag{5.5}$$

where $y_{kh} = S_{kh}(u_{kh})$, the extension $\widehat{\delta u}_{kh}$ is defined in (5.1)*, and $p_{kh} \in X^{r,s}_{k,h}$ is the solution of the discrete adjoint equation*

$$a_k(v_{kh}, p_{kh}) = J'_y(y_{kh}, u_{kh})(v_{kh}) \quad \text{for all } v_{kh} \in X^{r,s}_{k,h}. \tag{5.6}$$

(b) *The second derivatives of j_{kh} in directions $\delta u_{kh}, \tau u_{kh} \in U^{r,s}_{k,h}$ can be expressed as*

$$\begin{aligned} j''_{kh}(u_{kh})(\delta u_{kh}, \tau u_{kh}) = J''_{yy}(y_{kh}, u_{kh})(\delta y_{kh}, \widehat{\tau u}_{kh}) - a_k(\widehat{\tau u}_{kh}, \delta p_{kh}) \\ + J''_{uu}(y_{kh}, u_{kh})(\delta u_{kh}, \tau u_{kh}), \end{aligned} \tag{5.7}$$

where δy_{kh} is the solution of the discrete tangent equation

$$\delta y_{kh} \in \widehat{\delta u}_{kh} + X^{r,s}_{k,h} \ : \ a_k(\delta y_{kh}, v_{kh}) = 0 \quad \text{for all } v_{kh} \in X^{r,s}_{k,h}\,, \tag{5.8}$$

$\delta p_{kh} \in X^{r,s}_{k,h}$ is given by

$$a_k(v_{kh}, \delta p_{kh}) = J''_{yy}(y_{kh}, u_{kh})(\delta y_{kh}, v_{kh}) \quad \text{for all } v_{kh} \in X^{r,s}_{k,h}\,, \tag{5.9}$$

and $\widehat{\delta u}_{kh}, \widehat{\tau u}_{kh}$ are the extensions of $\delta u_{kh}, \tau u_{kh}$ defined as in (5.1)*.*

Proof. Using the solution $\delta y_{kh} = S'_{kh}(u_{kh})(\delta u_{kh})$ of the discretized tangent equation (5.8), we obtain

$$j'_{kh}(u_{kh})(\delta u_{kh}) = J'_y(y_{kh}, u_{kh})(\delta y_{kh}) + J'_u(y_{kh}, u_{kh})(\delta u_{kh}).$$

We rewrite the first term using (5.8) and (5.6):

$$\begin{aligned} &J'_y(y_{kh}, u_{kh})(\delta y_{kh}) = J'_y(y_{kh}, u_{kh})(\delta y_{kh} - \widehat{\delta u}_{kh}) + J'_y(y_{kh}, u_{kh})(\widehat{\delta u}_{kh}) \\ &= a_k(\delta y_{kh} - \widehat{\delta u}_{kh}, p_{kh}) + J'_y(y_{kh}, u_{kh})(\widehat{\delta u}_{kh}) = -a_k(\widehat{\delta u}_{kh}, p_{kh}) + J'_y(y_{kh}, u_{kh})(\widehat{\delta u}_{kh}). \end{aligned}$$

This gives the desired representation (5.5). The representation of the second derivatives is obtained in a similar way. ∎

Remark 5.3. On the continuous level, similar representations of the derivatives hold. They can be equivalently expressed using the normal derivatives of the adjoint state on the boundary; see (3.1). A direct discretization of the representation involving normal fluxes is in general not equivalent to (5.5) and would not lead to the exact representation of the derivatives of j_{kh} due to the lack of appropriate formulas for integration by parts of the discretized solutions.

Remark 5.4. In the convection dominated case, i.e., if $|b| \gg \kappa$, the finite element discretization may lead to strongly oscillatory solutions. Several stabilization methods are known to improve the approximation properties of the pure Galerkin discretization and to reduce the oscillatory behavior; see, e.g., [10, 22, 28, 39, 40]. For the stabilized finite elements in the context of stationary optimal control problems, we refer the reader to [13, 4].

6. Numerical examples. In this section we discuss numerical examples illustrating our results and give some details on the numerical realization.

Due to the fact that the trial and the test spaces in the formulation of the discrete state equation (5.3) are discontinuous in time, this formulation results in a time stepping scheme. In our numerical realization we use bilinear finite elements for the space discretization and piecewise constant functions in time resulting in spaces $X^{0,1}_{k,h}$ and $U^{0,1}_{k,h}$. In the following we describe the state equation (5.3), the adjoint equation (5.6), equations (5.8) and (5.9), and the evaluation of the derivatives of the discrete reduced cost functional for this choice of discretization. We define

$$U_m = u_{kh}\big|_{I_m},\ Y_m = y_{kh}\big|_{I_m},\ P_m = p_{kh}\big|_{I_m}, \quad i = 1, \ldots, M,$$

$$Y_0 = y^-_{kh,0},\ P_0 = p^-_{kh,0}\,.$$

The discrete state equation reads as follows for $Y_0 \in V_h$ and $Y_m \in U_m + V_{h,0}$:

$$(Y_0, \phi) = (y_0, \phi) \quad \text{for all } \phi \in V_h\,,$$

$$(Y_m, \phi) + k_m\,(\nabla Y_m, \nabla\phi) + k_m \left(\int_{I_m} b(s)\, ds \cdot \nabla Y_m, \phi \right) = (Y_{m-1}, \phi) + k_m \left(\int_{I_m} f(s)\, ds, \phi \right) \quad \text{for all } \phi \in V_{h,0}\,,\ m = 1, \ldots, M\,.$$

Remark 6.1. If the time integrals are approximated by the box rule, then the resulting scheme is equivalent to the implicit Euler method. However, a better approximation of these time integrals leads to a scheme which allows for better error estimates with respect to the required smoothness of the solution and to long time integration ($T \gg 1$); see, e.g., [18]. For the numerical examples which follow, the trapezoidal rule is used, which guarantees this improved convergence behavior.

In order to cover both problem (P1) with a time-distributed cost functional and the problem (P2) with a terminal time functional, we write the cost functional in the form

$$J(y,u) = \int_0^T I(y(s))\, ds + K(y(T)) + \frac{\beta}{2}|u|^2_{L^2(\Sigma)}\,.$$

The discrete adjoint equation reads as follows for $P_0 \in V_h$ and $P_m \in V_{h,0}$:

$$(\phi, P_M) + k_M\,(\nabla\phi, \nabla P_M) + k_M \left(\int_{I_M} b(s)\, ds \cdot \nabla\phi, P_M \right) = K'(Y_M)(\phi) + k_M\, I'(Y_M)(\phi) \quad \text{for all } \phi \in V_{h,0}\,,$$

$$(\phi, P_m) + k_m\,(\nabla\phi, \nabla P_m) + k_m \left(\int_{I_m} b(s)\, ds \cdot \nabla\phi, P_m \right) = (\phi, P_{m+1}) + k_m\, I'(Y_m)(\phi) \quad \text{for all } \phi \in V_{h,0}\,,\ m = M-1, \ldots, 1\,,$$

$$(\phi, P_0) = (\phi, P_1) \quad \text{for all } \phi \in V_h\,.$$

Remark 6.2. There are two possible ways to obtain the above equations for P_m, $m = 0, \ldots, M$:

- discretization of the continuous adjoint equation with dG(0) in time and with H^1-conforming finite elements in space (optimize-then-discretize approach);
- application of the Lagrange formalism on the discrete level for the optimization problem with the state equation discretized by dG(0) in time and H^1-conforming finite elements in space (discretize-then-optimize approach).

The resulting schemes for P_m coincide independent of the temporal grid. This fact relies on the space-time Galerkin discretization.

For a standard formulation of the implicit Euler scheme, i.e.,

$$\frac{1}{k_m}(Y_m - Y_{m-1}, \phi) + (\nabla Y_m, \nabla \phi) + (b(t_m)\nabla Y_m, \phi) = (f(t_m), \phi) \quad \text{for all } \phi \in V_{h,0},$$

the optimize-then-discretize approach leads to the discrete adjoint

$$\frac{1}{k_{m+1}}(\phi, P_m - P_{m+1}) + (\nabla \phi, \nabla P_m) + (b(t_m)\nabla \phi, P_m) = (I'(Y_m), \phi) \quad \text{for all } \phi \in V_{h,0},$$

whereas the discretize-then-optimize approach produces

$$\frac{1}{k_m}(\phi, P_m) - \frac{1}{k_{m+1}}(\phi, P_{m+1}) + (\nabla \phi, \nabla P_m) + (b(t_m)\nabla \phi, P_m) = (I'(Y_m), \phi) \quad \text{for all } \phi \in V_{h,0}.$$

These schemes are different for nonconstant time steps k_m.

For the optimization algorithm we need the evaluation of the derivatives of j_{kh} for basis functions in $U^{0,1}_{k,h}$. We consider the following basis of $U^{0,1}_{k,h}$:

$$w_{i,m}(t,x) = \begin{cases} \phi_i(x), & t \in I_m, \\ 0 & \text{otherwise}, \end{cases} \tag{6.1}$$

where $\phi_i = \gamma(\widehat{\phi}_i)$ and $\widehat{\phi}_i \in V_h$ is a finite element nodal basis function for a boundary node i. We obtain the following corollary from Proposition 5.1.

Corollary 6.1. *The following representation holds:*

$$\begin{aligned} j'_{kh}(u_{kh})(w_{i,M}) &= \beta(U_M, \phi_i)_{\partial\Omega} + K'(Y_M)(\widehat{\phi}_i) + k_M\, I'(Y_M)(\widehat{\phi}_i) \\ &\quad - (\widehat{\phi}_i, P_M) - k_M\,(\nabla\widehat{\phi}_i, \nabla P_M) - k_M \left(\int_{I_M} b(s)\, ds \cdot \nabla \widehat{\phi}_i, P_M \right), \end{aligned}$$

$$\begin{aligned} j'_{kh}(u_{kh})(w_{i,m}) &= \beta(U_m, \phi_i)_{\partial\Omega} + k_m\, I'(Y_m)(\widehat{\phi}_i) + (\widehat{\phi}_i, P_{m+1}) \\ &\quad - (\widehat{\phi}_i, P_m) - k_m\,(\nabla\widehat{\phi}_i, \nabla P_m) - k_m \left(\int_{I_m} b(s)\, ds \cdot \nabla \widehat{\phi}_i, P_m \right), \\ &\qquad m = M-1, \dots, 1\,. \end{aligned}$$

Remark 6.3. Due to the fact that $\widehat{\phi}_i$ has local support, the spatial integration in the representations above is done only over cells adjacent to the boundary.

Next, we describe (5.8) and (5.9) and the evaluation of the second derivatives. We define

$$\delta U_m = \delta u_{kh}\big|_{I_m},\ \delta Y_m = \delta y_{kh}\big|_{I_m},\ \delta P_m = \delta p_{kh}\big|_{I_m}, \quad i = 1, \dots, M,$$

$$\delta Y_0 = \delta y^-_{kh,0},\ \delta P_0 = \delta p^-_{kh,0}\,.$$

The discrete tangent equation reads as follows for $\delta Y_0 \in V_h$ and $\delta Y_m \in \delta U_m + V_{h,0}$:

$$\delta Y_0 = 0\,,$$

$$(\delta Y_m, \phi) + k_m\,(\nabla \delta Y_m, \nabla \phi) + k_m \left(\int_{I_m} b(s)\,ds \cdot \nabla \delta Y_m, \phi \right) = (\delta Y_{m-1}, \phi)$$
$$\text{for all } \phi \in V_h\,,\ m = 1, \ldots, M\,.$$

The discrete equation (5.9) reads as follows for $\delta P_0 \in V_h$ and $\delta P_m \in V_{h,0}$:

$$(\phi, \delta P_M) + k_M\,(\nabla \phi, \nabla \delta P_M) + k_M \left(\int_{I_M} b(s)\,ds \cdot \nabla \phi, \delta P_M \right) = K''(Y_M)(\delta Y_M, \phi)$$
$$+\, k_M\, I''(Y_M)(\delta Y_M, \phi) \quad \text{for all } \phi \in V_h\,,$$

$$(\phi, \delta P_m) + k_m\,(\nabla \phi, \nabla \delta P_m) + k_m \left(\int_{I_m} b(s)\,ds \cdot \nabla \phi, \delta P_m \right) = (\phi, \delta P_{m+1})$$
$$+\, k_m\, I''(Y_m)(\delta Y_m, \phi) \quad \text{for all } \phi \in V_h\,,\ m = M-1, \ldots, 1\,,$$

$$(\phi, \delta P_0) = (\phi, \delta P_1) \quad \text{for all } \phi \in V_h\,.$$

Using the basis (6.1) we obtain the following representation of $j''_{kh}(u_{kh})(\delta u_{kh}, w_{i,m})$ as a corollary to Proposition 5.1.

COROLLARY 6.2. *The following representation holds:*

$$j''_{kh}(u_{kh})(\delta u_{kh}, w_{i,M}) = \beta(\delta U_M, \phi_i)_{\partial\Omega} + K''(Y_M)(\delta Y_M, \widehat{\phi}_i) + k_M\, I''(Y_M)(\delta Y_M, \widehat{\phi}_i)$$
$$-\, (\widehat{\phi}_i, \delta P_M) - k_M\,(\nabla \widehat{\phi}_i, \nabla \delta P_M) - k_M \left(\int_{I_M} b(s)\,ds \cdot \nabla \widehat{\phi}_i, \delta P_M \right)$$

$$j''_{kh}(u_{kh})(\delta u_{kh}, w_{i,m}) = \beta(\delta U_m, \phi_i)_{\partial\Omega} + k_m\, I''(Y_m)(\delta Y_M, \widehat{\phi}_i) + (\widehat{\phi}_i, \delta P_{m+1})$$
$$-\, (\widehat{\phi}_i, \delta P_m) - k_m\,(\nabla \widehat{\phi}_i, \nabla \delta P_m) - k_m \left(\int_{I_m} b(s)\,ds \cdot \nabla \widehat{\phi}_i, \delta P_m \right),$$
$$m = M-1, \ldots, 1\,.$$

We close the paper with three numerical examples. The first two examples correspond to problems (P1) and (P2), respectively, and examine the behavior of the PDAS method if the dimension of the discrete problem increases due to the refinement of spatial and time meshes. The third example is devoted to the superlinear convergence of the PDAS method.

Our special interest in considering the behavior of the algorithm as the mesh is refined results from previous experience with constrained optimal control problems with distributed controls. Pointwise control, respectively, state constraints, result in a very different behavior of the algorithm in the sense that it is mesh-independent for the former but strongly mesh-dependent for the latter; see [8] and [26]. Analytically this is reflected in the fact that for the former the Lagrange multipliers are L^2-functions, whereas they are only measures in the case of state constraints. Regularization or nested iteration can be used in the latter case to nearly restore mesh-independence.

For the case of Dirichlet boundary control with pointwise constraints on the controls the practical performance of the algorithm and specifically its behavior with respect to mesh refinement cannot easily be predicted from previous experience. On the one hand, as in the case of distributed control, the associated Lagrange multipliers are L^2-regular and we can prove superlinear convergence. However, at least formally, inequality constraints on the control along the boundary are equivalent to inequality constraints on the state on the boundary, and, second, the extra regularity on the adjoint states and the optimal controls obtained in section 3 is rather less than that in the case of distributed controls.

In section 4, we have shown the superlinear convergence of the PDAS method on the continuous level for Dirichlet optimal control problems. On the discrete level, we typically have finite step convergence (cf. the stopping criterion discussed in Remark 4.1), which is, of course, better than superlinear convergence. In our last numerical example, presented in section 6.3, we observe the behavior of the PDAS method corresponding to superlinear convergence also before the method stops finding the optimal discrete solution.

6.1. Example 1: Time-distributed functional. We consider the following Dirichlet optimal control problem on $\Omega \times (0,T)$ with $\Omega = (0,1)^2 \subset \mathbb{R}^2$ and $T = 1$:

$$\begin{array}{rll} \min & J(u,y) = \dfrac{1}{2}\|y - y_d\|^2_{L^2(Q)} + \dfrac{\beta}{2}\|u\|^2_{L^2(\Sigma)}\,, & \\ \text{subject to} & y_t - \kappa\Delta y + b\cdot\nabla u = f & \text{in } \Omega\times(0,T), \\ & y = u & \text{on } \partial\Omega\times(0,T), \\ & y(0) = y_0 & \text{in } \Omega \end{array}$$

and control constraints

$$u \geq \phi\,.$$

The data are given as follows:

$$f = 0, \quad \kappa = 1, \quad b(t,x) = 15\,(\sin(2\pi t), \cos(2\pi t)), \quad y_0 = 0, \quad \beta = 10^{-4}\,,$$

$$y_d(t,x) = x_1 x_2(\cos(\pi t) - x_1)(\sin(\pi t) - x_2), \quad \phi = -0.25\,.$$

This optimal control problem is discretized by space-time finite elements as described above. The resulting finite-dimensional problem is solved by the PDAS method. In Table 6.1 the number of iterations of the method is shown for a sequence of uniformly refined discretizations. Here, M denotes the number of time-steps and N is the number of nodes in the space discretization. In all cases the algorithm terminated with two consecutive active sets coinciding, so that the exact solution of the discretized problem is found.

We present the results for two choices of initial guesses for the control variable: the same choice for all discretization levels ($u_0 = 1$), and an interpolated solution from the previous discretization level (nested iteration). The goal consists in obtaining practical experience as to which degree the weak additional regularity established in Theorem 3.2 and Corollary 3.7 is sufficient for near mesh-independent behavior. The results indicate that the additional regularity is sufficient for nearly mesh-independent behavior and that nested iterations provide only a relatively moderate improvement. The algorithm was also tested with other initial guesses and led to very similar results.

Table 6.1
PDAS method on the sequence of uniformly refined discretizations.

N	M	$\dim X_h = M \cdot N$	$\dim U_h$	PDAS iterations	PDAS nested iterations
25	2	50	32	2	2
81	4	324	128	3	3
289	8	2312	512	4	3
1089	16	17424	2048	4	3
4225	32	135200	8192	5	4
16641	64	1065024	32768	6	4

6.2. Example 2: Terminal functional. In this example we consider a Dirichlet optimal control problem with a terminal cost functional:

$$\begin{array}{lll} \min & J(u,y) = \dfrac{1}{2}\|y(T) - y_d^T\|^2_{L^2(\Omega)} + \dfrac{\beta}{2}\|u\|^2_{L^2(\Sigma)}\,, & \\ \text{subject to} & y_t - \kappa\Delta y + b\cdot\nabla u = f & \text{in } \Omega\times(0,T), \\ & y = u & \text{on } \partial\Omega\times(0,T), \\ & y(0) = y_0 & \text{in } \Omega\,, \end{array}$$

and control constraints

$$\phi \le u \le \psi, \quad u = 0 \text{ on } \partial\Omega \times (T_1, T).$$

The data are given as follows:

$$f = 0, \quad \kappa = 1, \quad b(t,x) = 15\,(\sin(2\pi t), \cos(2\pi t)), \quad y_0 = 0, \quad \beta = 10^{-4}, \quad T_1 = 0.75\,,$$

$$y_d^T(x) = 3\,\left(x_1x_2 + \sin(12\pi x_1^2(1-x_1)^2)\,\sin(12\pi x_2^2(1-x_2)^2)\right), \quad \phi = -0.1, \quad \psi = 2.5\,.$$

In Table 6.2 we present the corresponding results.

Table 6.2
PDAS method on the sequence of uniformly refined discretizations.

N	M	$\dim X_h = M \cdot N$	$\dim U_h$	PDAS iterations	PDAS nested iterations
25	2	50	32	3	3
81	4	324	128	3	3
289	8	2312	512	4	4
1089	16	17424	2048	5	4
4225	32	135200	8192	5	5
16641	64	1065024	32768	6	5

6.3. Example 3. In this example, we consider the following Dirichlet optimal control problem on $\Omega \times (0,T)$ with $\Omega = (0,1)^2 \subset \mathbb{R}^2$ and $T = 1$:

$$\begin{array}{lll} \min & J(u,y) = \dfrac{1}{2}\|y - y_d\|^2_{L^2(Q)} + \dfrac{\beta}{2}\|u\|^2_{L^2(\Sigma)}\,, & \\ \text{subject to} & y_t - \kappa\Delta y + b\cdot\nabla u = f & \text{in } \Omega\times(0,T), \\ & y = u & \text{on } \partial\Omega\times(0,T), \\ & y(0) = y_0 & \text{in } \Omega \end{array}$$

and control constraints

$$\phi \le u \le \psi\,.$$

The data are given as follows:

$$f = \begin{cases} 2, & x_1 \le 0.25, \\ -35 & \text{else}, \end{cases} \quad \kappa = 1, \;\; b(t,x) = 10\,(\sin(2\pi t), \cos(2\pi t)), \;\; y_0 = 0, \;\; \beta = 10^{-5}\,,$$

$$y_d(t,x) = \begin{cases} 2 - 2x_1, & x_1 \le 0.5, \\ 2 - 2x_2 & \text{else}, \end{cases} \qquad \phi = -1, \qquad \psi = 2\,.$$

For a fixed discretization with $M = 64$ time-steps and $N = 1089$ nodes in the spacial mesh, we consider the iteration error

$$e_i = \|u_{kh}^{(i)} - u_{kh}\|_{L^2(\Sigma)},$$

where $u_{kh}^{(i)}$ is the ith iterate, and u_{kh} is the optimal discrete solution. The results presented in Table 6.3 demonstrate superlinear convergence of the algorithm.

TABLE 6.3
Iteration error for the PDAS method.

i	0	1	2	3	4	5	6	7
e_i	2.3e-1	1.7e-1	4.1e-2	1.9e-2	6.5e-3	3.8e-4	4.5e-6	0
e_{i+1}/e_i	7.4e-1	2.4e-1	4.6e-1	3.4e-1	5.8e-2	1.2e-2	0	–

REFERENCES

[1] R. A. ADAMS, *Sobolev Spaces*, Academic Press, Amsterdam, 2005.
[2] N. ARADA AND J.-P. RAYMOND, *Dirichlet boundary control of semilinear parabolic equations, Part* 1: *Problems with no state constraints*, Appl. Math. Optim., 45 (2002), pp. 125–143.
[3] R. BECKER, D. MEIDNER, AND B. VEXLER, *Efficient numerical solution of parabolic optimization problems by finite element methods*, Optim. Methods Softw., to appear.
[4] R. BECKER AND B. VEXLER, *Optimal control of the convection-diffusion equation using stabilized finite element methods*, Numer. Math., 106 (2007), pp. 349–367.
[5] R. BECKER, *Mesh adaptation for Dirichlet flow control via Nitsche's method*, Comm. Numer. Methods Engrg., 18 (2002), pp. 669–680.
[6] F. B. BELGACEM, H. E. FEKIH, AND H. METOUI, *Singular perturbation for the Dirichlet boundary control of elliptic problems*, M2AN Math. Model. Numer. Anal., 37 (2003), pp. 883–850.
[7] M. BERGGREN, *Approximations of very weak solutions to boundary-value problems*, SIAM J. Numer. Anal., 42 (2004), pp. 860–877.
[8] M. BERGOUNIOUX, K. ITO, AND K. KUNISCH, *Primal-dual strategy for constrained optimal control problems*, SIAM J. Control Optim., 37 (1999), pp. 1176–1194.
[9] M. BERGOUNIOUX, M. HADDOU, M. HINTERMÜLLER, AND K. KUNISCH, *A comparison of a Moreau–Yosida-based active set strategy and interior point methods for constrained optimal control problems*, SIAM J. Optim., 11 (2000), pp. 495–521.
[10] E. BURMAN AND P. HANSBO, *Edge stabilization for Galerkin approximations of convection–diffusion–reaction problems*, Comput. Methods Appl. Mech. Engrg., 193 (2004), pp. 1437–1453.
[11] E. CASAS AND J.-P. RAYMOND, *Error estimates for the numerical approximation of Dirichlet boundary control for semilinear elliptic equations*, SIAM J. Control Optim., 45 (2006), pp. 1586–1611.

[12] P. G. Ciarlet, *The Finite Element Method for Elliptic Problems*, Classics Appl. Math. 40, SIAM, Philadelphia, 2002.

[13] S. S. Collis and M. Heinkenschloss, *Analysis of the Streamline Upwind/Petrov Galerkin Method Applied to the Solution of Optimal Control Problems*, CAAM TR02-01, Rice University, Houston, TX, 2002.

[14] R. Dautray and J.-L. Lions, *Mathematical Analysis and Numerical Methods for Science and Technology: Evolution Problems* I, Vol. 5, Springer-Verlag, Berlin, 1992.

[15] J. De Los Reyes and K. Kunisch, *A semi-smooth Newton method for control constrained boundary optimal control of the Navier-Stokes equations*, Nonlinear Anal., 62 (2005), pp. 1289–1316.

[16] K. Eriksson, D. Estep, P. Hansbo, and C. Johnson, *Computational Differential Equations*, Cambridge University Press, Cambridge, UK, 1996.

[17] K. Eriksson, C. Johnson, and V. Thomée, *Time discretization of parabolic problems by the discontinuous Galerkin method*, RAIRO Modél. Math. Anal. Numér., 19 (1985), pp. 611–643.

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

[19] D. French and J. King, *Analysis of a robust finite element approximation for a parabolic equation with rough boundary data*, Math. Comp., 60 (1993), pp. 79–104.

[20] A. Fursikov, *Optimal Control of Distributed Systems: Theory and Applications*, Transl. Math. Monogr. 187, AMS, Providence, RI, 1999.

[21] P. Grisvard, *Commutativité de deux foncteurs d'interpolation et applications*, J. Math. Pures Appl., 45 (1966), pp. 143–206.

[22] J.-L. Guermond, *Stabilization of Galerkin approximations of transport equations by subgrid modeling*, Modél. Math. Anal. Numér., 36 (1999), pp. 1293–1316.

[23] M. Gunzburger, L. Hou, and T. Svobodny, *Analysis and finite element approximation of optimal control problems for the stationary Navier-Stokes equations with Dirichlet controls*, RAIRO Modél. Math. Anal. Numér., 25 (1991), pp. 711–748.

[24] M. D. Gunzburger and S. Manservisi, *The velocity tracking problem for Navier–Stokes flows with boundary control*, SIAM J. Control Optim., 39 (2000), pp. 594–634.

[25] M. Hintermüller, K. Ito, and K. Kunisch, *The primal-dual active set strategy as a semi-smooth Newton method*, SIAM J. Optim., 13 (2003), pp. 865–888.

[26] M. Hintermüller and K. Kunisch, *Feasible and noninterior path-following in constrained minimization with low multiplier regularity*, SIAM J. Control Optim., 45 (2006), pp. 1198–1221.

[27] K. Ito and K. Kunisch, *The primal-dual active set method for nonlinear optimal control problems with bilateral constraints*, SIAM J. Control Optim., 43 (2004), pp. 357–376.

[28] C. Johnson, *Numerical Solution of Partial Differential Equations by Finite Element Method*, Cambridge University Press, Cambridge, UK, 1987.

[29] F. Kappel and K. Kunisch, *Invariance results for delay and Volterra equations in fractional order Sobolev spaces*, Trans. Amer. Math. Soc., 304 (1987), pp. 1–51.

[30] A. Kunoth, *Adaptive wavelet schemes for an elliptic control problem with Dirichlet boundary control*, Numer. Algorithms, 39 (2005), pp. 199–200.

[31] O. A. Ladyženskaja, V. A. Solonnikov, and N. N. Uralćeva, *Linear and Quasilinear Equations of Parabolic Type*, AMS, Providence, RI, 1968.

[32] I. Lasiecka, *Galerkin approximation of abstract parabolic boundary value problems with rough boundary data-l^p theory*, Math. Comp., 175 (1986), pp. 55–75.

[33] H.-C. Lee, *Analysis and computational methods of Dirichlet boundary optimal control problems for 2d Boussinesq equations*, Adv. Comput. Math., 19 (2003), pp. 255–275.

[34] J. L. Lions and E. Magenes, *Non-homogeneous Boundary Value Problems and Applications*, Vol. I, Springer-Verlag, Berlin, 1972.

[35] J. L. Lions and E. Magenes, *Non-homogeneous Boundary Value Problems and Applications*, Vol. II, Springer-Verlag, Berlin, 1972.

[36] J.-L. Lions, *Optimal Control of Systems Governed by Partial Differential Equations*, Grundlehren Math. Wiss. 170, Springer-Verlag, Berlin, 1971.

[37] H. Maurer and J. Zowe, *First and second-order necessary and sufficient optimality conditions for infinite-dimensional programming problems*, Math. Program., 16 (1979), pp. 98–110.

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

[39] A. Quarteroni and A. Valli, *Numerical Approximation of Partial Differential Equations*, Springer Ser. Comput. Math. 23, Springer-Verlag, Berlin, 1994.

[40] H.-G. Roos, M. Stynes, and L. Tobiska, *Numerical Methods for Singularly Perturbed Dif-*

ferential Equations, Springer Ser. Comput. Math. 24, Springer-Verlag, Berlin, 1996.

[41] M. Schmich and B. Vexler, *Adaptivity with dynamic meshes for space-time finite element discretizations of parabolic equations*, SIAM J. Sci. Comput., to appear.

[42] H. Triebel, *Interpolation Theory, Function Spaces, Differential Operators*, J. A. Barth Verlag, Heidelberg, 1995.

[43] B. Vexler, *Finite element approximation of elliptic Dirichlet optimal control problems*, Numer. Funct. Anal. Optim., to appear.