SEMISMOOTH NEWTON METHODS FOR OPTIMAL CONTROL OF THE WAVE EQUATION WITH CONTROL CONSTRAINTS[*]

AXEL KRÖNER[†], KARL KUNISCH[‡], AND BORIS VEXLER[†]

Abstract. In this paper optimal control problems governed by the wave equation with control constraints are analyzed. Three types of control action are considered: distributed control, Neumann boundary control, and Dirichlet control, and proper functional analytic settings for them are discussed. For treating inequality constraints, semismooth Newton methods are discussed and their convergence properties are investigated. In the case of distributed and Neumann control, superlinear convergence is shown. For Dirichlet boundary control, superlinear convergence is proved for a strongly damped wave equation. For numerical realization, a space-time finite element discretization is discussed. Numerical examples illustrate the results.

Key words. semismooth Newton methods, wave equation, optimal control, control constraints, superlinear convergence, space-time finite elements

AMS subject classifications. 49J20, 35L05, 45M37, 65N30

DOI. 10.1137/090766541

1. Introduction. In this paper we consider optimal control problems governed by the wave equation and subject to pointwise inequality control constraints. We discuss three different control actions: distributed control, Neumann boundary control, and Dirichlet boundary control.

The optimal control problem under consideration is formulated as follows:

$$(1.1) \quad \text{Minimize} \quad J(y,u) = \mathcal{G}(y) + \frac{\alpha}{2}\|u\|_U^2 \quad \text{subject to} \quad y = S(u), \quad y \in Y, \ u \in U_{\mathrm{ad}},$$

for the control u and the state y from appropriate functional spaces U and Y to be specified later. The set of admissible controls U_{ad} is given by bilateral box constraints

$$(1.2) \quad U_{\mathrm{ad}} = \{u \in U \,|\, u_a \leq u \leq u_b\} \quad \text{with} \quad u_a, u_b \in U.$$

The control-to-state operator $S\colon U \to Y$ is the solution operator of the wave equation with control u entering either the right-hand side, or the Neumann boundary conditions, or the Dirichlet boundary conditions. The functional $\mathcal{G}\colon Y \to \mathbb{R}$ will be defined in more detail in the next section.

Throughout, we assume that $\Omega \subset \mathbb{R}^n$, $n \geq 1$, is a bounded domain with a C^2-boundary. For $T > 0$ we denote $I = (0,T)$, $Q = I \times \Omega$, and $\Sigma = I \times \partial\Omega$. In the case of distributed control, the state equation defining the operator S is given as

$$(1.3) \quad \begin{aligned} y_{tt} - \Delta y &= u \text{ in } Q, \\ y(0) = y_0, \quad y_t(0) &= y_1 \text{ in } \Omega, \quad y = 0 \text{ on } \Sigma; \end{aligned}$$

in the case of the Neumann boundary control we have

$$
\begin{aligned}
y_{tt} - \Delta y &= f \text{ in } Q, \\
y(0) &= y_0, \quad y_t(0) = y_1 \text{ in } \Omega, \quad \partial_n y = u \text{ on } \Sigma;
\end{aligned}
$$
(1.4)

and in the case of the Dirichlet boundary control we have

$$
\begin{aligned}
y_{tt} - \Delta y &= f \text{ in } Q, \\
y(0) &= y_0, \quad y_t(0) = y_1 \text{ in } \Omega, \quad y = u \text{ on } \Sigma.
\end{aligned}
$$
(1.5)

For this class of optimal control problems we will discuss a proper functional analytic setting, which is suitable for application of the semismooth Newton methods. These methods have proven their efficiency for a large class of optimization problems with partial differential equations; see, e.g., [17, 19, 41, 42, 22], where superlinear convergence was shown in several situations. It is well known (see [17]) that semismooth Newton methods are equivalent to primal-dual active set (PDAS) strategies, which exploit pointwise information from Lagrange multipliers for updating active sets. Here it is essential that the Lagrange multipliers are L^2-functions rather than measures, which can be achieved by setting $U = L^2(Q)$ for distributed control and $U = L^2(\Sigma)$ for both Neumann and Dirichlet boundary control problems; cf. the discussion in [22].

Our goal here is to analyze semismooth Newton methods for optimal control problems governed by the wave equation with respect to superlinear convergence. An important ingredient in proving superlinear convergence is a smoothing property of the operator mapping the control variable u to the adjoint state p or to a trace of p. For distributed and Neumann boundary control we will establish this smoothing property and prove superlinear convergence. For the case of Dirichlet boundary control we will provide an example illustrating the fact that such a property cannot hold in general. In addition we will consider a Dirichlet boundary control problem governed by the strongly damped wave equation given as

$$
\begin{aligned}
y_{tt} - \Delta y - \rho \Delta y_t &= f \text{ in } Q, \\
y(0) &= y_0, \quad y_t(0) = y_1 \text{ in } \Omega, \quad y = u \text{ on } \Sigma,
\end{aligned}
$$
(1.6)

with a positive damping parameter $\rho > 0$. This equation appears often in models with loss of energy; e.g., it arises in the modeling of longitudinal vibrations in a homogeneous bar, in which there are viscous effects; cf. [31]. The corresponding optimal control problem (with small ρ) can also be regarded as a regularization of the Dirichlet boundary control problem for the wave equation. For the resulting optimal control problem, we will establish the required smoothing property and prove superlinear convergence of the semismooth Newton method.

For numerical realization the infinite-dimensional problems have to be discretized. Following [5, 33, 22] we use space-time finite element methods for discretization. This approach guarantees that the algorithm is invariant with respect to the ordering of discretization of the problem and derivative computations. This means that the approaches "optimize-then-discretize" and "discretize-then-optimize" coincide.

There is a rich literature on the controllability and the stabilization of the wave equation. Optimal control problems governed by wave equations are considered in various publications; see the monographs [27, 28, 26]. Domain decomposition in the context of optimal control of the wave equation is considered in [23, 24]. In [14, 7] boundary control for the wave equation, and the strongly damped wave equation, respectively, is discussed. For optimal control of nonlinear wave equations see [9, 11].

For state constrained optimal control problems with the wave equation, see [16, 35, 36]. Finite difference approximations in the context of control of the wave equation are discussed in [44]. To the best of our knowledge this is the first paper analyzing convergence of semismooth Newton methods in this context and providing details on the corresponding numerical realization.

The paper is organized as follows. In the next section we discuss the semismooth Newton method for an abstract optimal control problem with control constraints and formulate a set of assumptions for its superlinear convergence. Section 3 is devoted to relevant existence, uniqueness, and regularity results for the wave equation and for the strongly damped wave equation. In section 4 we discuss functional analytic settings for distributed and boundary (Dirichlet and Neumann) control problems for these equations and check the assumptions for superlinear convergence of the semismooth Newton method. In section 5 we describe the space-time finite element discretization for optimal control problem under consideration, and in section 6 we present numerical examples illustrating our results.

2. Semismooth Newton methods and the primal-dual active set strategy. In this section we summarize known results for semismooth Newton methods, which are relevant for the analysis in this paper. Moreover, we provide a set of assumptions for superlinear convergence of an abstract optimal control problem with control constraints. Later on, in section 4 we will check these assumptions for the considered situation of distributed, Neumann boundary, and Dirichlet boundary control of the wave equation, as well as for Dirichlet boundary control of the strongly damped wave equation.

Let X and Z be Banach spaces, and let $F: D \subset X \to Z$ be a nonlinear mapping with open domain D. Moreover, let $\mathcal{L}(X, Z)$ be the set of continuous, linear mappings from X to Z.

DEFINITION 2.1. *The mapping $F: D \subset X \to Z$ is called Newton differentiable in the open subset $U \subset D$ if there exists a family of generalized derivatives $G: U \to \mathcal{L}(X, Z)$ such that*

$$\lim_{h \to 0} \frac{1}{\|h\|_X} \|F(x+h) - F(x) - G(x+h)h\|_Z = 0$$

for every $x \in U$.

The following theorem provides a generic result on superlinear convergence for semismooth Newton methods; see [17].

THEOREM 2.2. *Suppose that $x^* \in D$ is a solution to $F(x) = 0$ and that F is Newton differentiable with Newton derivative G in an open neighborhood U containing x^* and that*

$$\{\|G(x)^{-1}\|_{\mathcal{L}(Z,X)} | x \in U\}$$

is bounded. Then for $x_0 \in D$ the Newton iteration

$$x_{k+1} = x_k - G(x_k)^{-1}F(x_k), \quad k = 0, 1, 2, \ldots,$$

converges superlinearly to x^* provided that $\|x_0 - x^*\|_X$ is sufficiently small.

We shall require Newton differentiability of the max-operator. For this purpose let X denote a function space of real-valued functions on $\omega \subset \mathbb{R}^n$, and let $\max(0, v)$ denote the pointwise max-operation for $v \in X$. We introduce candidates for the

generalized derivative in the form

$$(2.1) \qquad G_{m,\delta}(v)(x) = \begin{cases} 1 & \text{if } v(x) > 0, \\ 0 & \text{if } v(x) < 0, \\ \delta & \text{if } v(x) = 0, \end{cases}$$

where $v \in X$, and $\delta \in \mathbb{R}$ is arbitrary.

PROPOSITION 2.3. *The mapping* $\max(0, \cdot) \colon L^q(\omega) \to L^p(\omega)$ *with* $1 \le p < q < \infty$ *is Newton differentiable on* $L^q(\omega)$, *and* $G_{m,\delta}$ *is a generalized derivative.*

For the proof we refer to [17]. We also have the following chain rule [19].

LEMMA 2.4 (chain rule). *Let* $H \colon D \subset L^p(\omega) \to L^q(\omega)$, $1 \le p < q < \infty$, *be continuously Fréchet differentiable at* $y^* \in D$, *and let* $\phi \colon L^q(\omega) \to L^p(\omega)$ *be Newton differentiable at* $H(y^*)$ *with a generalized derivative* G. *Then* $F = \phi(H) \colon D \subset L^p(\omega) \to L^p(\omega)$ *is Newton differentiable at* y^* *with a generalized derivative given by* $G(H)H' \in \mathcal{L}(L^p(\omega), L^p(\omega))$.

We consider a general (linear-quadratic) optimal control problem (1.1), (1.2) with $\alpha > 0$, with the control space $U = L^2(\omega)$ and ω being a subset of $\mathbb{R}^n$, which later will be either $\omega = Q$ or $\omega = \Sigma$. The solution operator $S \colon U \to Y$ is assumed to be affine-linear with

$$S(u) = Tu + \bar{y},$$

where $T \in \mathcal{L}(U, Y)$ and $\bar{y} \in Y$. For the state space we set $Y = L^2(Q)$, and the functional $\mathcal{G} \colon L^2(Q) \to \mathbb{R}$ is assumed to be quadratic with $\mathcal{G}'$ being an affine operator from $L^2(Q)$ to itself, and $\mathcal{G}''$ is assumed to be nonnegative. From standard subsequential limit arguments (see, e.g., [27]), we have the following.

PROPOSITION 2.5. *Under the above assumptions there exists a unique global solution of the optimal control problem* (1.1), (1.2).

We define the reduced cost functional

$$j \colon U \to \mathbb{R}, \quad j(u) = \mathcal{G}(S(u)) + \frac{\alpha}{2}\|u\|_U^2$$

and reformulate the optimal control problem under consideration as follows:

$$\text{Minimize } j(u), \quad u \in U_{\text{ad}}.$$

The first (directional) derivative of j is given as

$$j'(u)(\delta u) = (\alpha u - q(u), \delta u)_\omega,$$

where the operator $q \colon U \to U$ is given by

$$(2.2) \qquad q(u) = -T^* \mathcal{G}'(S(u))$$

and $(\cdot, \cdot)_\omega$ denotes the inner product in $U = L^2(\omega)$.

PROPOSITION 2.6. *Let the above assumptions be fulfilled. Then the necessary and sufficient optimality conditions for* (1.1), (1.2) *can be expressed as the variational inequality*

$$(2.3) \qquad (\alpha u - q(u), \delta u - u)_\omega \ge 0 \quad \text{for all } \delta u \in U_{ad}.$$

This can alternatively be expressed as an optimality system for the control $u \in U$ and the Lagrange multiplier $\lambda \in U$ as

(2.4)
$$
\begin{cases}
\alpha u + \lambda = q(u), \\
\lambda = \max(0, \lambda + c(u - u_b)) + \min(0, \lambda + c(u - u_a))
\end{cases}
$$

with an arbitrary $c > 0$.

To set up a semismooth Newton method for the optimal control problem under consideration, we set $c = \alpha$, eliminate the Lagrange multiplier λ from the optimality system (2.4), and obtain the equivalent formulation

(2.5)
$$
\mathcal{F}(u) = 0,
$$

with the operator $\mathcal{F}\colon L^2(\omega) \to L^2(\omega)$ defined by

$$
\mathcal{F}(u) = \alpha(u - u_b) + \max(0, \alpha u_b - q(u)) + \min(0, q(u) - \alpha u_a).
$$

We will use the generalized derivatives of max- and min-operators (see (2.1)) chosen as

$$
(G_{\max}(v)\phi)(x) = \begin{cases} \phi(x) & \text{if } v(x) \geq 0, \\ 0 & \text{if } v(x) < 0, \end{cases} \quad \text{and} \quad (G_{\min}(v)\phi)(x) = \begin{cases} \phi(x) & \text{if } v(x) \leq 0, \\ 0 & \text{if } v(x) > 0 \end{cases}
$$

for $v, \phi \in L^2(\omega)$. The following assumption will ensure the superlinear convergence of the semismooth Newton method applied to (2.5).

ASSUMPTION 2.7. *We assume that the operator q defined in (2.2) is a continuous affine-linear operator $q\colon L^2(\omega) \to L^r(\omega)$ for some $r > 2$.*

In the following sections we will check this assumption for the optimal control problems under consideration.

LEMMA 2.8. *Let Assumption 2.7 be fulfilled, and let $u_a, u_b \in L^r(\omega)$ for some $r > 2$. Then the operator $\mathcal{F}\colon L^2(\omega) \to L^2(\omega)$ is Newton differentiable, and a generalized derivative $G_{\mathcal{F}}(u) \in \mathcal{L}(L^2(\omega), L^2(\omega))$ is given as*

$$
G_{\mathcal{F}}(u)h = \alpha h + G_{\max}(\alpha u_b - q(u)) \, T^* \mathcal{G}''(S(u)) T h - G_{\min}(q(u) - \alpha u_a) \, T^* \mathcal{G}''(S(u)) T h.
$$

The statement of this lemma follows from the chain rule in Lemma 2.4, the Newton differentiability of max- and min-operators, and Assumption 2.7.

For the operators $G_{\mathcal{F}}(u)$ we have the following lemma.

LEMMA 2.9. *There exists a constant C_G such that*

(2.6)
$$
\|G_{\mathcal{F}}(u)^{-1}(w)\|_{L^2(\omega)} \leq C_G \, \|w\|_{L^2(\omega)} \quad \textit{for all } w \in L^2(\omega)
$$

and for each $u \in L^2(\omega)$.

Proof. Let χ_I denote the characteristic function of the set

$$
I = \{x \in \omega : \alpha u_a(x) \leq q(u)(x) \leq \alpha u_b(x)\},
$$

and analogously let χ_A be the characteristic function of $A = \omega \setminus I$. Let $h \in L^2(\omega)$ and set

(2.7)
$$
w = G_{\mathcal{F}}(u)(h).
$$

On A there holds

$$G_{\mathcal{F}}(u)(h) = \alpha h,$$

and on I,

$$G_{\mathcal{F}}(u)(h) = \alpha h + T^* \mathcal{G}''(S(u))Th.$$

Hence, we deduce

(2.8) $$\|h\chi_A\|_{L^2(\omega)} \leq \frac{1}{\alpha} \|w\chi_A\|_{L^2(\omega)},$$

and taking the inner product of (2.7) with $h\chi_I$ we find

$$\alpha \|h\chi_I\|_{L^2(\omega)}^2 + (\mathcal{G}''(S(u))Th, Th\chi_I) = (w, h\chi_I).$$

This implies that

$$\alpha \|h\chi_I\|_{L^2(\omega)}^2 + (\mathcal{G}''(S(u))Th\chi_I, Th\chi_I) = (w, h\chi_I) - (\mathcal{G}''(S(u))Th\chi_A, Th\chi_I).$$

Thus, since $\mathcal{G}''$ is nonnegative and $\mathcal{G}$ quadratic, we deduce further that

$$\alpha \|h\chi_I\|_{L^2(\omega)}^2 \leq \|w\chi_I\|_{L^2(\omega)} \|h\chi_I\|_{L^2(\omega)} + K\|h\chi_A\|_{L^2(\omega)} \|h\chi_I\|_{L^2(\omega)}$$

for a constant K independent of h and u. Consequently,

(2.9) $$\alpha \|h\chi_I\|_{L^2(\omega)} \leq \|w\chi_I\|_{L^2(\omega)} + K\|h\chi_A\|_{L^2(\omega)} \leq \|w\chi_I\|_{L^2(\omega)} + \frac{K}{\alpha} \|w\chi_A\|_{L^2(\omega)}.$$

Combining (2.8) and (2.9), the desired result follows.   □

After these considerations we can formulate the following theorem.

THEOREM 2.10. *Let Assumption 2.7 be fulfilled and suppose that $u^* \in L^2(\omega)$ is a solution to the optimal control problem under consideration. Then, for $u_0 \in L^2(\omega)$ with $\|u_0 - u^*\|_{L^2(\omega)}$ sufficiently small, the semismooth Newton method*

(2.10) $$G_{\mathcal{F}}(u_k)(u_{k+1} - u_k) + \mathcal{F}(u_k) = 0, \quad k = 0, 1, 2, \ldots,$$

converges superlinearly.

Proof. This follows from Theorem 2.2 and Lemmas 2.8 and 2.9.   □

This semismooth Newton method is known to be equivalent [17, 19] to a PDAS method, which is defined as follows.

ALGORITHM 2.11 (PDAS method).
1. *Choose u_0 and set $\lambda_0 = q(u_0) - \alpha u_0$.*
2. *Given (u_k, λ_k) determine*

$$\mathcal{A}_{k+1}^b = \{x \in \omega | \lambda_k(x) + \alpha(u_k - u_b)(x) > 0\},$$
$$\mathcal{A}_{k+1}^a = \{x \in \omega | \lambda_k(x) + \alpha(u_k - u_a)(x) < 0\},$$
$$\mathcal{I}_{k+1} = \omega \backslash (\mathcal{A}_{k+1}^b \cup \mathcal{A}_{k+1}^a).$$

3. *Determine u_{k+1} as the solution to the following:*

$$\begin{cases} \text{Minimize} \quad j(u_{k+1}), \quad u_{k+1} \in U, \\ \text{subject to} \quad u_{k+1} = u_b \text{ on } \mathcal{A}_{k+1}^b, \quad u_{k+1} = u_a \text{ on } \mathcal{A}_{k+1}^a. \end{cases}$$

4. *Update λ_{k+1} according to*

$$\lambda_{k+1} = q(u_{k+1}) - \alpha u_{k+1}.$$

5. *Update $k = k + 1$.*

The equivalence of the two methods follows immediately (see, e.g., [19]) since the Newton iteration (2.10) and the PDAS method can be equivalently expressed as

$$\alpha(u_{k+1} - u_b) - G_{\max}(\alpha u_b - q(u_k))(q(u_{k+1}) - q(u_k))$$
$$+ G_{\min}(q(u) - \alpha u_a)(q(u_{k+1}) - q(u_k))$$
$$+ \max(0, \, \alpha u_b - q(u_k)) + \min(0, \, q(u_k) - \alpha u_a) = 0.$$

Remark 2.12. If the algorithm finds two successive active sets, for which $\mathcal{A}_k = \mathcal{A}_{k+1}$, then u_k is the solution of the problem. We apply this condition as a stopping criterion.

3. On the state equation. In this section we formulate existence and regularity results for the wave equation and for the strongly damped wave equation to be used in what follows. We denote V to be either $H^1(\Omega)$ or $H_0^1(\Omega)$ depending on the prescribed type of boundary conditions (homogeneous Neumann or homogeneous Dirichlet). Together with $H = L^2(\Omega)$, the Hilbert space V and its dual V^* build a Gelfand triple $V \hookrightarrow H \hookrightarrow V^*$. Here and in what follows, we employ the usual notion of Lebesgue and Sobolev spaces. 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 $C(Y) = C([0, T]; Y)$, $C^1(Y) = C^1([0, T]; Y)$. Moreover we use $(\cdot, \cdot)$ for the $L^2(\Omega)$-inner product, $\|\cdot\|$ for the corresponding norm, $\langle \cdot, \cdot \rangle$ for the $L^2(\partial\Omega)$-inner product, $(\cdot, \cdot)_I$ for the inner product in $L^2(L^2(\Omega))$, and $\langle \cdot, \cdot \rangle_I$ for the inner product in $L^2(L^2(\Sigma))$.

THEOREM 3.1. *Suppose that $f \in L^2(H)$, $y_0 \in V$, $y_1 \in H$. Then*

(3.1)
$$\begin{cases} y_{tt} - \Delta y = f & in \ Q, \\ y(0) = y_0, \quad y_t(0) = y_1 & on \ \Omega \end{cases}$$

with either homogeneous Neumann or homogeneous Dirichlet boundary conditions admits a unique solution $y \in C(V)$ with $y_t \in C(H)$ such that $(f, y_0, y_1) \mapsto (y, y_t)$ is continuous from $L^2(H) \times V \times H$ to $C(V) \times C(H)$.

For the proof we refer to [29, p. 275].

It is well known that (3.1) can be equivalently formulated as a first-order system as follows:

(3.2)
$$\begin{cases} y_t^2 - \Delta y^1 = f & in \ Q, \\ \quad\quad y_t^1 = y^2 & in \ Q, \\ \quad\quad y^1(0) = y_0 & in \ \Omega, \\ \quad\quad y^2(0) = y_1 & in \ \Omega, \\ \quad\quad y^1|_\Sigma = 0 & on \ \Sigma. \end{cases}$$

This formulation is the basis for our discretizations of the control problems under consideration; see section 5.

For the inhomogeneous Neumann problem

(3.3)
$$\begin{cases} y_{tt} - \Delta y = f & in \ Q, \\ y(0) = y_0, \quad y_t(0) = y_1 & in \ \Omega, \\ \partial_n y = u & on \ \Sigma \end{cases}$$

we have the following result with $V = H^1(\Omega)$ and $H = L^2(\Omega)$.

THEOREM 3.2. *For every $(f, y_0, y_1, u) \in L^1((H^1(\Omega))^*) \times L^2(\Omega) \times (H^1(\Omega))^* \times L^2(\Sigma)$ there exists a unique very weak solution $y \in L^2(Q)$ to (3.3) depending continuously on the data. It satisfies*

$$(3.4) \qquad (y, \zeta_{tt} - \Delta\zeta)_I = (f, \zeta)_I - (y_0, \zeta_t(0)) + \langle y_1, \zeta(0) \rangle_{(H^1(\Omega))^*, H^1(\Omega)} + \langle u, \zeta \rangle_I,$$

where ζ is the solution to

$$(3.5) \qquad \begin{cases} \zeta_{tt} - \Delta\zeta = g & in\ Q, \\ \zeta(T) = \zeta_t(T) = 0 & in\ \Omega, \\ \partial_n \zeta = 0 & on\ \Sigma \end{cases}$$

for any $g \in L^2(L^2(\Omega))$.

If we assume that $(f, y_0, y_1, u) \in L^1(L^2(\Omega)) \times H^1(\Omega) \times L^2(\Omega) \times L^2(\Sigma)$, then $(y, y_t) \in C(H^{\frac{1}{2}}(\Omega)) \times C((H^{\frac{1}{2}}(\Omega))^*)$ holds.

Proof. The proof of the first assertion follows classical arguments: From Theorem 3.1 we deduce that $(\zeta, \zeta_t) \in C(H^1(\Omega)) \times C(L^2(\Omega))$, and hence the mapping

$$g \to \mathcal{F} = (f, \zeta)_I - (y_0, \zeta_t(0)) + \langle y_1, \zeta(0) \rangle_{(H^1(\Omega))^*, H^1(\Omega)} + \langle u, \zeta \rangle_I$$

defines a continuous linear form on $L^2(L^2(\Omega))$; i.e., there exists $y \in L^2(L^2(\Omega))$ such that

$$(y, g)_I = \mathcal{F}(g).$$

This implies (3.4). From (3.4)–(3.5) we deduce further that there exists a constant C independent of $(f, y_0, y_1, u) \in L^1((H^1(\Omega))^*) \times L^2(\Omega) \times (H^1(\Omega))^* \times L^2(\Sigma)$ such that

$$\|y\|_{L^2(L^2(\Omega))} \le C \|(f, y_0, y_1, u)\|_{L^1((H^1(\Omega))^*) \times L^2(\Omega) \times (H^1(\Omega))^* \times L^2(\Sigma)}.$$

Uniqueness of the weak solution and continuous dependence on the data follows from this estimate.

The additional regularity has been proved in [36]. ☐

For the inhomogeneous Dirichlet problem we consider

$$(3.6) \qquad \begin{cases} y_{tt} - \Delta y = f & in\ Q, \\ y(0) = y_0, \quad y_t(0) = y_1 & in\ \Omega, \\ y = u & on\ \Sigma. \end{cases}$$

THEOREM 3.3. *For every $(f, y_0, y_1, u) \in L^1((H_0^1(\Omega))^*) \times L^2(\Omega) \times (H_0^1(\Omega))^* \times L^2(\Sigma)$ there exists a unique very weak solution $(y, y_t) \in C(L^2(\Omega)) \times C(H^{-1}(\Omega))$ depending continuously on the data. It satisfies*

$$(y, \zeta_{tt} - \Delta\zeta)_I = (f, \zeta)_I - (y_0, \zeta_t(0)) + \langle y_1, \zeta(0) \rangle_{(H^1(\Omega))^*, H^1(\Omega)} - \langle u, \partial_n\zeta \rangle_I,$$

where ζ is the solution to

$$\begin{cases} \zeta_{tt} - \Delta\zeta = g & in\ Q, \\ \zeta(T) = \zeta_t(T) = 0 & in\ \Omega, \\ \zeta = 0 & on\ \Sigma \end{cases}$$

for any $g \in L^1(L^2(\Omega))$.

For the proof we refer to [28, p. 240] and [35]. It depends on the hidden regularity result, which states that if $(f, y_0, y_1) \in L^1(L^2(\Omega)) \times H_0^1(\Omega) \times L^2(\Omega)$ and $u = 0$, then the solution to (3.6) satisfies $\partial_n y \in L^2(\Sigma)$; see, e.g., [28, p. 233].

As announced in the introduction we also consider the strongly damped wave equation with a damping parameter ρ, $0 < \rho < \rho_0$, $\rho_0 \in \mathbb{R}^+$, i.e.,

$$
(3.7) \quad
\begin{cases}
y_{tt} - \Delta y - \rho \Delta y_t = f & \text{in } Q, \\
y(0) = y_0 & \text{in } \Omega, \\
y_t(0) = y_1 & \text{in } \Omega, \\
y = u & \text{on } \Sigma
\end{cases}
$$

for $u \in L^2(\Sigma)$.

To prove a regularity result we first consider the damped wave equation with homogeneous Dirichlet data:

$$
(3.8) \quad
\begin{cases}
y_{tt} - \Delta y - \rho \Delta y_t = f & \text{in } Q, \\
y(0) = y_0 & \text{in } \Omega, \\
y_t(0) = y_1 & \text{in } \Omega, \\
y = 0 & \text{on } \Sigma.
\end{cases}
$$

The following theorem can be obtained.

THEOREM 3.4. *For $f \in L^2(L^2(\Omega))$, $y_0 \in H_0^1(\Omega) \cap H^2(\Omega)$, and $y_1 \in H_0^1(\Omega)$, there exists a unique weak solution of* (3.8),

$$
(3.9) \qquad y \in H^2(L^2(\Omega)) \cap C^1(H_0^1(\Omega)) \cap H^1(H^2(\Omega)),
$$

defined by the conditions $y(0) = y_0$, $y_t(0) = y_1$, and
(3.10)
$$
(y_{tt}(s), \phi) + (\nabla y(s), \nabla \phi) + \rho(\nabla y_t(s), \nabla \phi) = (f(s), \phi) \quad \text{for all } \phi \in H_0^1(\Omega) \text{ a.e. in } (0, T).
$$

Moreover, the a priori estimate
(3.11)
$$
\|y\|_{H^2(L^2(\Omega)) \cap C^1(H_0^1(\Omega)) \cap H^1(H^2(\Omega))} \leq C \left(\|f\|_{L^2(L^2(\Omega))} + \|\nabla y_0\| + \|\Delta y_0\| + \|\nabla y_1\| \right)
$$

holds, where the constant $C = C(\rho)$ tends to infinity as ρ tends to zero.

Here, we present a direct proof. Similar results can also be extracted from [8]. Further regularity results for some structurally damped problems can be found in [40].

To prove Theorem 3.4 we use a Galerkin procedure, i.e., we build a solution of (3.10) by first constructing solutions y^m, $m \in \mathbb{N}$, of certain finite-dimensional approximations to (3.10) and then passing to the limits $m \to \infty$; for details of such a Galerkin procedure we refer, e.g., to [10, Chap. 7]. Thus, we have to prove the above estimate (3.11) for these functions y^m, where $y^m(s)$ are linear combinations of eigenfunctions of the Laplacian in $H_0^1(\Omega)$. Then the existence of a solution in $H^2(L^2(\Omega)) \cap W^{1,\infty}(H_0^1(\Omega)) \cap H^1(H^2(\Omega))$ can be ensured by standard arguments; cf. [10]. Furthermore, with usual techniques we derive that if $v \in H^2(L^2(\Omega)) \hookrightarrow C^1(L^2(\Omega))$ and $v \in W^{1,\infty}(H_0^1(\Omega))$, then there holds $v \in C^1(H_0^1(\Omega))$. Thus we obtain the asserted regularity in (3.9).

The estimate (3.11) is shown in four steps in the following four lemmas. To improve clarity we omit the index m.

LEMMA 3.5. *Let the conditions of Theorem* 3.4 *be fulfilled. Then the following estimate holds for almost every* $t \in (0,T)$:

$$\|y_t(t)\|^2 + \|\nabla y(t)\|^2 + \rho \int_0^t \|\nabla y_t(s)\|^2 ds \le C \left(\|\nabla y_0\|^2 + \|y_1\|^2 + \|f\|_{L^2(L^2(\Omega))}^2 \right).$$

Proof. Note that within a Galerkin procedure, $y = y_m$ is a linear combination of eigenfunctions of Δ satisfying $y_t(t) \in H_0^1(\Omega)$. We set $\phi = y_t$ in (3.10) and obtain

$$(y_{tt}(s), y_t(s)) + (\nabla y(s), \nabla y_t(s)) + \rho\|\nabla y_t(s)\|^2 = (f(s), y_t(s)).$$

Hence,

$$\frac{1}{2}\frac{d}{dt}\|y_t\|^2 + \frac{1}{2}\frac{d}{dt}\|\nabla y\|^2 + \rho\|\nabla y_t(s)\|^2 = (f(s), y_t(s)).$$

Integrating in time from 0 to t we find

$$\|y_t(t)\|^2 + \|\nabla y(t)\|^2 + 2\rho \int_0^t \|\nabla y_t(s)\|^2\, ds$$
$$\le \|f\|_{L^2(L^2(\Omega))}^2 + \|y_1\|^2 + \|\nabla y_0\|^2 + \int_0^t \|y_t(s)\|^2\, ds.$$

Using Gronwall's lemma we obtain

$$\|y_t(t)\|^2 \le C \left(\|\nabla y_0\|^2 + \|y_1\|^2 + \|f\|_{L^2(L^2(\Omega))}^2 \right).$$

This gives the desired result. □

LEMMA 3.6. *Let the conditions of Theorem* 3.4 *be fulfilled. Then the following estimate holds for almost every* $t \in (0,T)$:

$$\int_0^t \|\Delta y(s)\|^2 ds + \rho\|\Delta y(t)\|^2 \le \frac{C}{\rho} \left(\|\nabla y_0\|^2 + \|\Delta y_0\|^2 + \|y_1\|^2 + \|f\|_{L^2(L^2(\Omega))}^2 \right).$$

Proof. Recalling that within a Galerkin procedure $y(s)$ is a linear combination of eigenfunctions of the Laplacian in $H_0^1(\Omega)$, we know that Δy vanishes on Σ. We use $\phi = -\Delta y$ as a test function in (3.10) and obtain

$$-(y_{tt}(s), \Delta y(s)) + \|\Delta y(s)\|^2 + \rho(\Delta y_t(s), \Delta y(s)) = -(f(s), \Delta y(s))$$

or, equivalently,

$$-(y_{tt}(s), \Delta y(s)) + \|\Delta y(s)\|^2 + \frac{\rho}{2}\frac{d}{dt}\|\Delta y(s)\|^2 = -(f(s), \Delta y(s)).$$

Integrating in time from 0 to t implies that

$$-\int_0^t (y_{tt}(s), \Delta y(s))\, ds + \int_0^t \|\Delta y(s)\|^2\, ds + \frac{\rho}{2}\|\Delta y(t)\|^2$$
$$\le \frac{1}{2}\|f\|_{L^2(L^2(\Omega))}^2 + \frac{1}{2}\int_0^t \|\Delta y(s)\|^2\, ds + \frac{\rho}{2}\|\Delta y_0\|^2.$$

For the first term on the left-hand side we get for almost every $t \in (0, T)$

$$-\int_0^t (y_{tt}(s), \Delta y(s))\, ds = \int_0^t (y_t(s), \Delta y_t(s))\, ds - (y_t(t), \Delta y(t)) + (y_t(0), \Delta y(0))$$

$$= -\int_0^t \|\nabla y_t(s)\|^2\, ds + (\nabla y_t(t), \nabla y(t)) - (\nabla y_t(0), \nabla y(0))$$

$$= -\int_0^t \|\nabla y_t(s)\|^2\, ds - (y_t(t), \Delta y(t)) + (y_1, \Delta y_0).$$

Here, we have used the facts that $y_{tt} = y_t = 0$ on Σ and $y_1 = 0$ on $\partial\Omega$. This yields

$$\int_0^t \|\Delta y(s)\|^2\, ds + \frac{\rho}{2}\|\Delta y(t)\|^2$$

$$\leq \frac{1}{2}\|f\|_{L^2(L^2(\Omega))}^2 + \frac{1}{2}\int_0^t \|\Delta y(s)\|^2\, ds + \frac{\rho}{2}\|\Delta y_0\|^2$$

$$+ \int_0^t \|\nabla y_t(s)\|^2\, ds + \frac{1}{\rho}\|y_t(t)\|^2 + \frac{\rho}{4}\|\Delta y(t)\|^2 + \frac{1}{2}\|y_1\|^2 + \frac{1}{2}\|\Delta y_0\|^2.$$

Absorbing terms, we obtain

$$\frac{1}{2}\int_0^t \|\Delta y(s)\|^2\, ds + \frac{\rho}{4}\|\Delta y(t)\|^2$$

$$\leq \frac{1}{2}\|f\|_{L^2(L^2(\Omega))}^2 + \frac{\rho+1}{2}\|\Delta y_0\|^2 + \int_0^t \|\nabla y_t(s)\|^2\, ds + \frac{1}{\rho}\|y_t(t)\|^2 + \frac{1}{2}\|y_1\|^2.$$

Using the result from the previous lemma, we obtain the desired estimate. □

LEMMA 3.7. *Let the conditions of Theorem 3.4 be fulfilled. Then the following estimate holds for almost every $t \in (0, T)$:*

$$\|\nabla y_t(t)\|^2 + \|\Delta y(t)\|^2 + \rho \int_0^t \|\Delta y_t(s)\|^2\, ds \leq \frac{1}{\rho}\|f\|_{L^2(L^2(\Omega))}^2 + \|\nabla y_1\|^2 + \|\Delta y_0\|^2.$$

Proof. We proceed as in the proofs of the previous lemmas. Since within the Galerkin procedure, Δy_t vanishes on Σ, we can choose $\phi = -\Delta y_t$ and obtain

$$-(y_{tt}(s), \Delta y_t(s)) + (\Delta y(s), \Delta y_t(s)) + \rho\|\Delta y_t(s)\|^2 = -(f(s), \Delta y_t(s)).$$

We integrate by parts in the first term and obtain for almost every s

$$\frac{1}{2}\frac{d}{dt}\|\nabla y_t(s)\|^2 + \frac{1}{2}\frac{d}{dt}\|\Delta y(s)\|^2 + \rho\|\Delta y_t(s)\|^2 = -(f(s), \Delta y_t(s)).$$

Integrating in time from 0 to t we obtain

$$\frac{1}{2}\|\nabla y_t(t)\|^2 + \frac{1}{2}\|\Delta y(t)\|^2 + \rho \int_0^t \|\Delta y_t(s)\|^2\, ds$$

$$\leq \frac{1}{2\rho}\|f\|_{L^2(L^2(\Omega))}^2 + \frac{\rho}{2}\int_0^t \|\Delta y_t(s)\|^2\, ds + \frac{1}{2}\|\nabla y_1\|^2 + \frac{1}{2}\|\Delta y_0\|^2.$$

This implies the desired estimate. □

LEMMA 3.8. *Let the conditions of Theorem* 3.4 *be fulfilled. Then the following estimate holds:*

$$\int_0^t \|y_{tt}(s)\|^2 ds \leq \frac{C}{\rho} \left(\|f\|_{L^2(L^2(\Omega))}^2 + \|\nabla y_0\|^2 + \|\Delta y_0\|^2 + \|\nabla y_1\|^2 \right).$$

Proof. We proceed as in the proof of Lemma 3.6 and choose $\phi = y_{tt}$. This yields

$$\|y_{tt}(s)\|^2 - (\Delta y(s), y_{tt}(s)) - \rho(\Delta y_t, y_{tt}) = (f(s), y_{tt}(s)).$$

Hence,

$$\int_0^t \|y_{tt}(s)\|^2 ds + \int_0^t (\Delta y_t(s), y_t(s)) ds - (\Delta y(t), y_t(t)) + (\Delta y(0), y_t(0))$$
$$= \int_0^t (f, y_{tt}) ds + \rho \int_0^t (\Delta y_t(s), y_{tt}(s)),$$

and thus, we obtain

$$\int_0^t \|y_{tt}(s)\|^2 ds \leq \|f\|_{L^2(L^2(\Omega))}^2 + \frac{1}{4} \int_0^t \|y_{tt}(s)\|^2 ds + \frac{\rho^2}{2} \int_0^t \|\Delta y_t(s)\|^2 ds$$
$$+ \frac{1}{2} \int_0^t \|y_{tt}(s)\|^2 ds + \int_0^t \|\nabla y_t(s)\|^2 ds + \frac{1}{2} \|\nabla y(t)\|^2$$
$$+ \frac{1}{2} \|\nabla y_t\|^2 + \frac{1}{2} \|\Delta y_0\|^2 + \frac{1}{2} \|y_1\|^2.$$

Absorbing terms and using Lemmas 3.5 and 3.7, we obtain the desired estimate. $\quad\square$

Now, we consider the strongly damped wave equation with inhomogeneous Dirichlet boundary conditions (3.7). In order to define a suitable weak formulation we proceed as follows. For given $v \in L^2(L^2(\Omega))$ let ζ be the solution of the adjoint equation

(3.12)
$$\begin{cases} \zeta_{tt} - \Delta\zeta + \rho\Delta\zeta_t = v & \text{in } Q, \\ \zeta(T) = 0 & \text{in } \Omega, \\ \zeta_t(T) = 0 & \text{in } \Omega, \\ \zeta = 0 & \text{on } \Sigma. \end{cases}$$

Using the transformation $t \mapsto T - t$, this equation can be written in the same form as (3.8). Therefore, Theorem 3.4 can be applied, leading to $\zeta \in H^2(L^2(\Omega)) \cap W^{1,\infty}(H_0^1(\Omega)) \cap H^1(H^2(\Omega))$. If a solution of (3.7) exists, then there holds (by testing with ζ and integrating in time)

$$(\zeta_{tt} - \Delta\zeta + \rho\Delta\zeta_t, y)_I + (y_0, \zeta_t(0)) - (y_1, \zeta(0)) + \langle y, \partial_n\zeta \rangle_I$$
$$- \rho\langle y, \partial_n\zeta_t \rangle_I + \rho(y_0, \Delta\zeta(0)) - \rho\langle y_0, \partial_n\zeta(0)\rangle = (f, \zeta)_I.$$

This suggests the following definition. A function $y \in L^2(L^2(\Omega))$ is called a very weak solution of (3.7) if the following variational equation holds for all $v \in L^2(L^2(\Omega))$:

(3.13) $(v, y)_I = -(y_0, \zeta_t(0)) + (y_1, \zeta(0)) - \langle u, \partial_n\zeta \rangle_I$
$$+ \rho\langle u, \partial_n\zeta_t \rangle_I - \rho(y_0, \Delta\zeta(0)) + \rho\langle y_0, \partial_n\zeta(0)\rangle + (f, \zeta)_I,$$

where ζ is the solution to (3.12). This leads to the following theorem.

THEOREM 3.9. *For $u \in L^2(\Sigma)$, $f \in L^2(L^2(\Omega))$, $y_0 \in H^1(\Omega)$, and $y_1 \in L^2(\Omega)$, (3.7) possess a unique very weak solution defined by (3.13), and there the estimate*

$$\|y\|_{L^2(L^2(\Omega))} \le C \left(\|u\|_{L^2(\Sigma)} + \|f\|_{L^2(L^2(\Omega))} + \|y_0\|_{H^1(\Omega)} + \|y_1\| \right)$$

holds, where the constant $C = C(\rho)$ tends to infinity as ρ tends to zero.

Proof. The right-hand side of (3.13) defines a linear functional $G(v)$ on $L^2(L^2(\Omega))$. This functional is bounded. In fact as a consequence of Theorem 3.4 we have

$$\|\zeta_t(0)\| + \|\zeta(0)\| + \|\Delta\zeta(0)\| + \|\partial_n\zeta(0)\|_{L^2(\partial\Omega)}$$
$$+ \|\partial_n\zeta\|_{L^2(\Sigma)} + \|\partial_n\zeta_t\|_{L^2(\Sigma)} + \|\zeta\|_{L^2(L^2(\Omega))} \le C\|v\|_{L^2(L^2(\Omega))}.$$

The representative of this functional in $L^2(L^2(\Omega))$ is y. This implies the desired result. $\square$

4. Optimal control problems. In this section we discuss the functional analytic settings for distributed and Neumann boundary control of the wave equation and for Dirichlet boundary control of the wave equation as well as the strongly damped wave equation. We provide the corresponding optimality systems and verify the assumptions for superlinear convergence of the semismooth Newton method formulated in section 2. In particular we will check Assumption 2.7. Furthermore, we formulate some regularity results for the optimal control and the optimal state.

4.1. Distributed control. In this section we analyze the optimal control problem with distributed control, i.e.,

$$(4.1) \quad \begin{cases} \min & J(y,u) = \mathcal{G}(y) + \frac{\alpha}{2}\|u\|_{L^2(Q)}^2, \quad y \in L^2(Q), \ u \in L^2(Q), \\ \text{subject to} & y_{tt} - \Delta y = u \quad \text{in } Q, \\ & y(0) = y_0, \quad y_t(0) = y_1 \quad \text{in } \Omega, \\ & y = 0 \quad \text{on } \Sigma, \\ & u_a \le u \le u_b \quad \text{a.e. in } Q, \end{cases}$$

where $y_0 \in H_0^1(\Omega)$, $y_1 \in L^2(\Omega)$, and the state equation is understood in the sense of Theorem 3.1. Further, we assume that u_a, u_b are in $L^r(Q)$ for some $r > 2$. The optimality system can be derived by standard techniques (see [27, p. 296] and [17]) and is found to be

$$(4.2) \quad \begin{cases} y_{tt} - \Delta y = u, \\ y(0) = y_0, \quad y_t(0) = y_1, \quad y|_\Sigma = 0, \\ \\ p_{tt} - \Delta p = -\mathcal{G}'(y), \\ p(T) = 0, \quad p_t(T) = 0, \quad p|_\Sigma = 0, \\ \\ \alpha u + \lambda = p, \\ \lambda = \max(0, \lambda + c(u - u_b)) + \min(0, \lambda + c(u - u_a)) \end{cases}$$

for any $c > 0$, $\lambda \in L^2(Q)$, and $p \in C(H^1(\Omega)) \cap C^1(L^2(\Omega))$.

We verify Assumption 2.7 in the next theorem.

THEOREM 4.1. *In the case of distributed control there holds, for the operator q defined in (2.2),*

$$q \colon L^2(Q) \to L^r(Q)$$

with some $r > 2$.

Proof. A direct comparison between the general optimality systems (2.4) and (4.2) shows that in this case for a given control $u \in L^2(Q)$ we have $q(u) = p$, where p is the solution of the corresponding dual equation. From Theorem 3.1 we deduce that $p \in C(H^1(\Omega))$, and hence for $n = 2$ we have $p \in L^r(Q)$ for all $1 \leq r < \infty$, and for $n \geq 3$ we have $p \in L^{\frac{2n}{n-2}}(Q)$. $\quad\square$

From this result superlinear convergence of the semismooth Newton method follows by Theorem 2.10 for the distributed control case.

As a further consequence of Theorem 4.1 we obtain the following regularity results for the optimal control and the optimal state.

COROLLARY 4.2. *Let $u_a, u_b \in H^1(L^2(\Omega)) \cap L^2(H^1(\Omega))$. Then there holds for the optimal control u*

$$u \in H^1(L^2(\Omega)) \cap L^2(H^1(\Omega)).$$

Proof. In section 2 the optimality condition is equivalent to

$$\alpha(u - u_b) + \max(0, \alpha u_b - q(u)) + \min(0, q(u) - \alpha u_a) = 0$$

with $q(u) = p$. From [22, Lem. 3.3] we deduce that the regularity of $q(u)$ is transferred to $\max(0, \alpha u_b - q(u))$ and $\min(0, q(u) - \alpha u_a)$ and therefore also to u. $\quad\square$

As a consequence, we can formulate some improved regularity result for the optimal state.

COROLLARY 4.3. *Assume that $y_0 \in H^2(\Omega) \cap H^1_0(\Omega)$, $y_1 \in H^1_0(\Omega)$, and $u_a, u_b \in H^1(L^2(\Omega))$. Then, for the optimal state there holds*

$$y \in L^\infty(H^2(\Omega)), \quad y_t \in L^\infty(H^1_0(\Omega)),$$
$$y_{tt} \in L^\infty(L^2(\Omega)), \quad y_{ttt} \in L^2((H^1_0(\Omega))^*).$$

Proof. Using an argument similar to that in Corollary 4.2 we obtain $u \in H^1(L^2(\Omega))$, and thus the assertion follows by [10, p. 389]. $\quad\square$

Thus, under the assumptions of Corollary 4.3 the optimal state y satisfies the state equation in the following weak sense:

$$(4.3) \qquad (y_{tt}, \zeta)_I + (\nabla y, \nabla \zeta)_I - (y(0) - y_0, \zeta_t(0)) + (y_t(0) - y_1, \zeta(0)) = (u, \zeta)_I$$

for all $\zeta \in H^2(L^2(\Omega)) \cap L^2(H^1_0(\Omega))$.

This variational formulation—rewritten as a first-order system—is the basis of our numerical realizations; see section 5.

4.2. Neumann boundary control. We consider the optimal control problem with Neumann boundary control, i.e.,

$$(4.4) \quad \begin{cases} \min & J(y, u) = \mathcal{G}(y) + \frac{\alpha}{2} \|u\|^2_{L^2(\Sigma)}, \quad y \in L^2(Q), \ u \in L^2(\Sigma), \\ \text{subject to} & y_{tt} - \Delta y = f \quad \text{in } Q, \\ & y(0) = y_0, \quad y_t(0) = y_1 \quad \text{in } \Omega, \\ & \partial_n y = u \quad \text{on } \Sigma, \\ & u_a \leq u \leq u_b \quad \text{a.e. in } \Sigma, \end{cases}$$

where $y_0 \in L^2(\Omega)$, $y_1 \in (H^1(\Omega))^*$, $f \in L^1((H^1(\Omega))^*)$, $u_a, u_b \in L^r(\Sigma)$ with some $r > 2$, and the state equation is understood in the sense of Theorem 3.2. The optimality

system can be derived by standard techniques (see, e.g., [30]) and is found to be

$$
(4.5) \quad
\begin{cases}
y_{tt} - \Delta y = f, \\
y(0) = y_0, \quad y_t(0) = y_1, \quad \partial_n y|_\Sigma = u, \\[4pt]
p_{tt} - \Delta p = -\mathcal{G}'(y), \\
p(T) = 0, \quad p_t(T) = 0, \quad \partial_n p|_\Sigma = 0, \\[4pt]
\alpha u + \lambda = p|_\Sigma, \\
\lambda = \max(0, \lambda + c(u - u_b)) + \min(0, \lambda + c(u - u_a))
\end{cases}
$$

for any $c > 0$, $\lambda \in L^2(\Sigma)$, and $p \in C(H^1(\Omega)) \cap C^1(L^2(\Omega))$.

In the next theorem we verify Assumption 2.7 for the Neumann boundary control problem.

THEOREM 4.4. *In the case of Neumann boundary control there holds, for the operator q defined in (2.2),*

$$
q \colon L^2(\Sigma) \to L^r(\Sigma)
$$

with some $r > 2$.

Proof. A direct comparison between the general optimality systems (2.4) and (4.5) shows that in this case, for a given control $u \in L^2(\Sigma)$ we have $q(u) = p|_\Sigma$, where p is the solution of the corresponding dual equation. From Theorem 3.1 we deduce that $p \in L^2(H^1(\Omega)) \cap H^1(L^2(\Omega))$, and hence by [30, p. 9],

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

By [1, p. 218] we have $H^{\frac{1}{2}}(L^2(\partial\Omega)) \hookrightarrow W^{\frac{1}{r}, r}(L^2(\partial\Omega)) \hookrightarrow L^r(L^2(\partial\Omega))$ for all $2 \leq r < \infty$. Consequently, we deduce

$$
p \in L^2(H^{\frac{1}{2}}(\partial\Omega)) \cap L^r(L^2(\partial\Omega)) \quad \text{for all } 2 \leq r < \infty,
$$

and hence interpolation (see, e.g., [39, Chap. 1]) implies that

$$
p \in L^{r_s}([H^{\frac{1}{2}}(\partial\Omega), L^2(\partial\Omega)]_s), \quad \text{where} \quad \frac{1}{r_s} = \frac{(1-s)}{2} + \frac{s}{r}, \quad s \in [0, 1].
$$

For $n \geq 3$ we use $H^{\frac{1}{2}}(\partial\Omega) \hookrightarrow L^{\frac{2n-2}{n-2}}(\partial\Omega)$ and get

$$
[H^{\frac{1}{2}}(\partial\Omega), L^2(\partial\Omega)]_s \hookrightarrow L^{q_s}(\partial\Omega), \quad \text{where} \quad \frac{1}{q_s} = \frac{(1-s)(n-2)}{2n-2} + \frac{s}{2}, \quad s \in [0, 1].
$$

We choose s in such a way that $r_s = q_s$. This implies

$$
s = \frac{r}{2 + nr - 2n}, \quad r \geq 2,
$$

and hence

$$
q_s = \frac{8n - 4n^2 - 4 + 2n^2 r - 2nr}{6n - 4 - 2n^2 + n^2 r - 2nr + r}.
$$

q_s is monotonic increasing in r, and hence we deduce $p \in L^{\frac{2n}{n-1} - \varepsilon}(\Sigma)$ for all $\varepsilon > 0$.

For $n = 2$ we have $H^{\frac{1}{2}}(\partial\Omega) \hookrightarrow L^q(\partial\Omega)$ for all $q < \infty$. Using similar arguments as before we obtain $p \in L^{4-\varepsilon}(\Sigma)$ for all $\epsilon > 0$. $\quad\square$

As for the distributed case we obtain additional regularity results for the optimal control and the optimal state.

COROLLARY 4.5. *Let* $u_a, u_b \in L^2(H^{\frac{1}{2}}(\partial\Omega)) \cap H^{\frac{1}{2}}(L^2(\partial\Omega))$. *Then, the optimal control satisfies* $u \in L^2(H^{\frac{1}{2}}(\partial\Omega)) \cap H^{\frac{1}{2}}(L^2(\partial\Omega))$.

COROLLARY 4.6. *For* $f \in L^2(L^2(\Omega))$, $y_0 \in H^1(\Omega)$, $y_1 \in L^2(\Omega)$, *and under the assumptions of Corollary 4.5 the optimal state satisfies* $y \in C(H^1(\Omega)) \cap C^1(L^2(\Omega))$.

Proof. We consider the equation

$$y_{tt} - \Delta y = 0, \quad y(0) = 0, \quad y_t(0) = 0, \quad \partial_n y|_\Sigma = g$$

with $g \in L^2(H^{\frac{1}{2}}(\partial\Omega))$. This equation admits a solution $y \in C(H^1(\Omega)) \cap C^1(L^2(\Omega))$ (see [26, Vol. II, Thm. 8A.5]; cf. also [34]). From Corollary 4.5 with $g = u \in L^2(H^{\frac{1}{2}}(\partial\Omega))$ and by Theorem 3.1 we obtain for the optimal state y of (4.4)

$$y \in C(H^1(\Omega)) \cap C^1(L^2(\Omega)). \qquad \square$$

As a direct consequence we deduce that under the assumptions of Corollary 4.6 the very weak solution y of the state equation, which corresponds to the optimal control u, is in fact a variational solution in the sense that $y \in C(H^1(\Omega)) \cap C^1(L^2(\Omega))$ and

$$-(y_t, \zeta_t)_I + (\nabla y, \nabla \zeta)_I - \langle u, \zeta \rangle_I - (y(0) - y_0, \zeta_t(0)) - (y_1, \zeta(0)) = (f, \zeta)_I$$

for all $\zeta \in C(H^1(\Omega)) \cap C^1(L^2(\Omega))$. This is important for numerical realizations; see the corresponding discussion in [22].

4.3. Dirichlet control. Here we consider the optimal control problem with Dirichlet boundary control:

$$(4.6) \quad \begin{cases} \min & J(y,u) = \mathcal{G}(y) + \frac{\alpha}{2} \|u\|_{L^2(\Sigma)}^2, \quad y \in L^2(Q), \quad u \in L^2(\Sigma), \\ \text{subject to} & y_{tt} - \Delta y = f \quad \text{in } Q, \\ & y(0) = y_0, \, y_t(0) = y_1 \quad \text{in } \Omega, \\ & y = u \quad \text{on } \Sigma, \\ & u_a \leq u \leq u_b \quad \text{a.e. on } \Sigma, \end{cases}$$

where $y_0 \in L^2(\Omega)$, $y_1 \in (H_0^1(\Omega))^*$, $f \in L^1((H_0^1(\Omega))^*)$, and the state equation is understood in the sense of Theorem 3.3. We have the following optimality system:

$$(4.7) \quad \begin{cases} y_{tt} - \Delta y = f, \\ y(0) = y_0, \quad y_t(0) = y_1, \quad y|_\Sigma = u, \\ p_{tt} - \Delta p = -\mathcal{G}'(y), \\ p(T) = 0, \quad p_t(T) = 0, \quad p|_\Sigma = 0, \\ \alpha u + \lambda = -\partial_n p|_\Sigma, \\ \lambda = \max(0, \lambda + c(u - u_b)) + \min(0, \lambda + c(u - u_a)) \end{cases}$$

for $c > 0$, $\lambda \in L^2(\Sigma)$, and $p \in C(H^1(\Omega)) \cap C^1(L^2(\Omega))$.

In the case of Dirichlet boundary control the operator q defined in (2.2) turns out to be given by $q(u) = -\partial_n p$, where p is the solution of the corresponding adjoint equation in (4.7). From the hidden regularity result (see, e.g., [28, p. 233]), we obtain that $\partial_n p \in L^2(\Sigma)$ and the operator q is a continuous affine-linear operator $q\colon L^2(\Sigma) \to L^2(\Sigma)$. In the following we provide a one-dimensional example showing that in general

the operator q does not map any control $u \in L^2(\Sigma)$ to $L^r(\Sigma)$ with $r > 2$. Therefore, Assumption 2.7 is not fulfilled in the case of Dirichlet boundary control.

We consider the one-dimensional wave equation with Dirichlet boundary control:

$$
\begin{aligned}
y_{tt} - y_{xx} &= 0 && \text{in } (0,1) \times (0,1), \\
y(t,0) &= u(t), && y(t,1) = 0, \\
y(0,x) &= 0, && y_t(0,x) = 0
\end{aligned}
$$

with $u \in L^2(0,1)$. We denote

$$\xi = t + x, \ \xi \in [0,2], \qquad \eta = t - x, \ \eta \in [-1,1]$$

and obtain

$$
y(\xi, \eta) = \begin{cases} 0, & \eta < 0, \\ u(\eta), & \eta \geq 0. \end{cases}
$$

Considering the adjoint equation

$$
\begin{aligned}
p_{tt} - p_{xx} &= y && \text{in } (0,1) \times (0,1), \\
p(t,0) &= 0, && p(t,1) = 0, \\
p(1,x) &= 0, && p_t(1,x) = 0,
\end{aligned}
$$

we obtain for the solution

$$
p(\xi,\eta) = \frac{1}{4} \begin{cases} U(\eta)\xi - (2-\eta)U(\eta) - \hat{U}(\eta) + (2-2\xi)U(\xi) + \hat{U}(\xi), & \eta \geq 0, \ \xi < 1, \\ U(\eta)\xi - (2-\eta)U(\eta) - \hat{U}(\eta) + \hat{U}(2-\xi), & \eta \geq 0, \ \xi \geq 1, \\ U(\eta)\xi - U(\eta)(\eta+2) - \hat{U}(-\eta) + \hat{U}(2-\xi), & \eta < 0, \ \xi \geq 1, \\ U(\eta)\xi - U(\eta)(\eta+2) - \hat{U}(-\eta) + (2-2\xi)U(\xi) + \hat{U}(\xi), & \eta < 0, \ \xi < 1, \end{cases}
$$

where

$$U'(t) = u(t) \quad \text{and} \quad \hat{U}'(t) = U(t).$$

It follows that

(4.8) $$p_x(t,0) = 16u(t)(1-t),$$

and thus for a general control $u \in L^2(0,1)$ the image

$$q(u)(t) = -\partial_n p(t) = -p_x(t,0) = -16u(t)(1-t)$$

does not have an improved regularity $q(u) \in L^r(0,1)$ with some $r > 2$.

Remark 4.7. This lack of additional regularity is due to the nature of the wave equation. In the elliptic case as well as the parabolic case, the corresponding operator q possesses the required regularity for Dirichlet boundary control; see [22].

4.4. Dirichlet control for the strongly damped wave equation. In this section we consider Dirichlet boundary control for the strongly damped wave equation, and we will show that in this case the assumptions from section 2 for superlinear convergence of the semismooth Newton method are satisfied.

The problem under consideration is given as follows:

$$
(4.9) \quad
\begin{cases}
\min & J(y,u) = \mathcal{G}(y) + \frac{\alpha}{2}\|u\|_{L^2(\Sigma)}^2, \quad y \in L^2(Q), \quad u \in L^2(\Sigma), \\
\text{subject to} & y_{tt} - \Delta y - \rho\Delta y_t = f \quad \text{in } Q, \\
& y(0) = y_0,\ y_t(0) = y_1 \quad \text{in } \Omega, \\
& y = u \quad \text{on } \Sigma, \\
& u_a \le u \le u_b \quad \text{a.e. on } \Sigma,
\end{cases}
$$

where $\rho > 0$, $f \in L^2(L^2(\Omega))$, $y_0 \in H^1(\Omega)$, $y_1 \in L^2(\Omega)$, $u_a, u_b \in L^r(\Sigma)$ with some $r > 2$, and the state equation is understood in the sense of Theorem 3.9.

The optimality system derived with standard arguments is given by

$$
(4.10) \quad
\begin{cases}
y_{tt} - \Delta y - \rho\Delta y_t = f, \\
y(0) = y_0, \quad y_t(0) = y_1, \quad y|_\Sigma = u, \\[4pt]
p_{tt} - \Delta p + \rho\Delta p_t = -\mathcal{G}'(y), \\
p(T) = 0, \quad p_t(T) = 0, \quad p|_\Sigma = 0, \\[4pt]
\alpha u + \lambda = -\partial_n p|_\Sigma, \\
\lambda = \max(0, \lambda + c(u - u_b)) + \min(0, \lambda + c(u - u_a))
\end{cases}
$$

for $c > 0$, $\lambda \in L^2(\Sigma)$, and $p \in H^2(L^2(\Omega)) \cap C^1(H_0^1(\Omega)) \cap H^1(H^2(\Omega))$.

In the next theorem we verify Assumption 2.7 in this case.

THEOREM 4.8. *In the case of the Dirichlet boundary control problem* (4.9) *with* $\rho > 0$, *the operator q defined in* (2.2) *satisfies*

$$
q\colon L^2(\Sigma) \to L^r(\Sigma)
$$

with some $r > 2$.

Proof. By a direct comparison of the optimality systems (2.4) and (4.10) we obtain $q(u) = -\partial_n p$. From Theorem 3.4 we obtain that p in particular fulfills

$$
p \in L^2(H^2(\Omega)) \cap H^1(L^2(\Omega)).
$$

By a trace theorem (see, e.g., [15, 30]), we get

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

By interpolation estimates we obtain as in [22, Thm. 3.2]

$$
\partial_n p \in L^{\frac{2(n+1)}{n}}(\Sigma) \quad \text{for } n \ge 3
$$

and

$$
\partial_n p \in L^{3-\varepsilon}(\Sigma) \quad \text{with } \varepsilon > 0 \quad \text{for } n = 2.
$$

This completes the proof. $\square$

5. Discretization. In this section we discuss the discretization of the optimal control problems under consideration. To this end we employ appropriate finite element schemes for both the temporal and the spatial discretizations. Applying this concept, the approaches of optimize-then-discretize and discretize-then-optimize, which are different in general, coincide; see, e.g., [32, 5]. Finite element discretizations of the wave equations are analyzed, e.g., in [2, 3, 4, 12, 18, 20, 21, 25].

For the temporal discretization of the state equation we use a Petrov–Galerkin scheme with continuous piecewise linear ansatz functions and discontinuous (in time) piecewise constant test functions. For the spatial discretization we use the usual conforming (bi)linear finite elements. This type of discretization is often referred as a cG(1)cG(1) discretization. For a precise definition of our discretization we consider a partition of the time interval $\bar{I} = [0, T]$ as

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

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

For spatial discretization we will consider two- or three-dimensional shape regular meshes; see, e.g., [6]. 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 function by setting $h|_K = h_K$ with the diameter h_K of the cell K.

Let $V = H^1(\Omega)$ and $V^0 = H_0^1(\Omega)$. On the mesh $\mathcal{T}_h$ we construct conforming finite element spaces $V_h \subset V$ and $V_h^0 \subset V^0$ in the following standard way:

$$V_h = \{v \in V | v|_K \in \mathcal{Q}^1(K) \text{ for } K \in \mathcal{T}_h\},$$
$$V_h^0 = \{v \in V^0 | v|_K \in \mathcal{Q}^1(K) \text{ for } K \in \mathcal{T}_h\}.$$

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

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

Remark 5.1. The definitions of V_h and V_h^0 can be extended to the case of triangular meshes and/or to spaces of higher order in the obvious way.

We define the following space-time finite element ansatz and test spaces:

$$X_{kh} = \{v_{kh} \in C(\bar{I}, V_h) | v_{kh}|_{I_m} \in \mathcal{P}^1(I_m, V_h)\},$$
$$X_{kh}^0 = \{v_{kh} \in C(\bar{I}, V_h^0) | v_{kh}|_{I_m} \in \mathcal{P}^1(I_m, V_h^0)\},$$
$$\widetilde{X}_{kh} = \{v_{kh} \in L^2(I, V_h) | v_{kh}|_{I_m} \in \mathcal{P}^0(I_m, V_h) \text{ and } v_{kh}(0) \in V_h\},$$
$$\widetilde{X}_{kh}^0 = \{v_{kh} \in L^2(I, V_h^0) | v_{kh}|_{I_m} \in \mathcal{P}^0(I_m, V_h^0) \text{ and } v_{kh}(0) \in V_h\},$$

where $\mathcal{P}^r(I_m, V_h)$ denotes the space of polynomials up to degree r on I_m with values in V_h. Thus, the spaces X_{kh} and X_{kh}^0 consist of piecewise linear and continuous functions in time with values in the usual spatial finite element space, whereas the functions in $\widetilde{X}_{kh}$ and $\widetilde{X}_{kh}^0$ are piecewise constant in time and therefore discontinuous.

Remark 5.2. In the above definitions we used the same spatial mesh and the same finite element space for all time intervals. However, in many situations the use of different meshes $\mathcal{T}_h^m$, $m = 1, \ldots, M$, is desirable. The consideration of such dynamically changing meshes can be included in the definition of the discontinuous

spaces $\widetilde{X}_{kh}$ and $\widetilde{X}_{kh}^0$ in a natural way. The corresponding definitions of the spaces X_{kh} and X_{kh}^0 are more involved due to the continuity requirement. For details on such dynamic meshes we refer the reader to [38].

For the definition of the discrete control space in the case of boundary control, we introduce the space of traces of functions in V_h,

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

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

Based on the equivalent formulation of the state equations as first-order systems (cf. (3.2)) we introduce the Galerkin finite element formulation of the state equations. We introduce a bilinear form $a_\rho\colon X_{kh} \times X_{kh} \times \widetilde{X}_{kh} \times \widetilde{X}_{kh} \to \mathbb{R}$ by

$$a_\rho(y,\xi) = a_\rho(y^1, y^2, \xi^1, \xi^2) = (\partial_t y^2, \xi^1)_I + (\nabla y^1, \nabla \xi^1)_I + \rho(\nabla y^2, \nabla \xi^1)_I$$
$$+ (\partial_t y^1, \xi^2)_I - (y^2, \xi^2)_I + (y^2(0), \xi^1(0)) - (y^1(0), \xi^2(0))$$

with $y = (y^1, y^2)$ and $\xi = (\xi^1, \xi^2)$ and with a real parameter $\rho \geq 0$.

5.1. Distributed control. For the distributed control problem we choose the discrete control space $U_{kh}^D = X_{kh}$. The discretized optimization problem is then formulated as follows:

$$\text{Minimize} \quad J(y_{kh}^1, u_{kh})$$

for $u_{kh} \in U_{kh}^D \cap U_{\text{ad}}$ and $y_{kh} \in X_{kh}^0 \times X_{kh}$ subject to

(5.1)
$$a_0(y_{kh}, \xi_{kh}) = (u_{kh}, \xi_{kh}^1)_I + (y_1, \xi_{kh}^1(0)) - (y_0, \xi_{kh}^2(0)) \quad \text{for all} \ \xi_{kh} \in \widetilde{X}_{kh}^0 \times \widetilde{X}_{kh}.$$

5.2. Neumann control. For the Neumann boundary control problem we choose the discrete control space as

$$U_{kh}^B = \{v_{kh} \in C(\bar{I}, W_h)|v_{kh}|_{I_m} \in \mathcal{P}^1(I_m, W_h)\}.$$

The corresponding discrete optimization problem is formulated as follows:

$$\text{Minimize} \quad J(y_{kh}^1, u_{kh})$$

for $u_{kh} \in U_{kh}^B \cap U_{\text{ad}}$ and $y_{kh} \in X_{kh} \times X_{kh}$ subject to

(5.2) $a_0(y_{kh}, \xi_{kh}) = \langle u_{kh}, \xi_{kh}^1 \rangle_I + (f, \xi_{kh}^1)_I + (y_1, \xi_{kh}^1(0)) - (y_0, \xi_{kh}^2(0))$
$$\text{for all } \xi_{kh} \in \widetilde{X}_{kh} \times \widetilde{X}_{kh}.$$

5.3. Dirichlet control. For the Dirichlet boundary control problem we choose the discrete control space as in the Neumann case. For a function $u_{kh} \in U_{kh}^B$ we define an extension $\widehat{u}_{kh} \in X_{kh}$ such that

(5.3) $\gamma(\widehat{u}_{kh}(t, \cdot)) = u_{kh}(t, \cdot) \quad \text{and} \quad \widehat{u}_{kh}(t, x_i) = 0$

on all interior nodes x_i of $\mathcal{T}_h$ and for all $t \in \bar{I}$.

The discrete optimization problem is formulated as follows:

$$\text{Minimize} \quad J(y_{kh}^1, u_{kh})$$

for $u_{kh} \in U_{kh}^{\mathrm{B}} \cap U_{\mathrm{ad}}$ and $y_{kh} \in (\widehat{u}_{kh} + X_{kh}^0) \times X_{kh}$ subject to

$$(5.4) \quad a_\rho(y_{kh}, \xi_{kh}) = (f, \xi_{kh}^1)_I + (y_1, \xi_{kh}^1(0)) - (y_0, \xi_{kh}^2(0)) \text{ for all } \xi_{kh} \in \widetilde{X}_{kh}^0 \times \widetilde{X}_{kh}.$$

Remark 5.3. The employed cG(1)cG(1) discretization scheme is known to be energy conserving. For $\rho = 0$, $f = 0$, and $u = 0$ one can directly show that

$$\frac{1}{2}\|y_{kh}^2(t_m)\|_{L^2(\Omega)}^2 + \frac{1}{2}\|\nabla y_{kh}^1(t_m)\|_{L^2(\Omega)}^2 = \frac{1}{2}\|y_{kh}^2(t_{m-1})\|_{L^2(\Omega)}^2 + \frac{1}{2}\|\nabla y_{kh}^1(t_{m-1})\|_{L^2(\Omega)}^2$$

holds for all $m = 1, 2, \ldots, M$. This reflects the corresponding property of the wave equation on the continuous level.

5.4. Optimization algorithm on the discrete level. As on the continuous level, each of the discrete state equations (5.1), (5.2), and (5.4) defines the corresponding discrete solution operator S_{kh} mapping a given control u_{kh} to the first component of the state y_{kh}^1. We introduce the discrete reduced cost functional

$$(5.5) \qquad j_{kh}(u_{kh}) = J(S_{kh}(u_{kh}), u_{kh})$$

and reformulate the discrete optimization problem as follows:

$$(5.6) \qquad \text{Minimize} \quad j_{kh}(u_{kh}) \quad \text{for} \quad u_{kh} \in U_{kh} \cap U_{\mathrm{ad}},$$

where the discrete control space is $U_{kh} = U_{kh}^D$ for distributed control and $U_{kh} = U_{kh}^B$ for boundary control. This optimization problem is solved using the PDAS algorithm (semismooth Newton method) as described in section 2 for the continuous problem. For the realization of this method on the discrete level, we should discuss the structure of the operator q_{kh}, corresponding to the operator q in (2.2) on the continuous level, and the solution of the equality constrained optimization problem in step (iii) of the PDAS algorithm on the discrete level; cf. section 2. The latter problem is solved using the Newton method utilizing the derivatives $j_{kh}'(u_{kh})(\delta u_{kh})$ and $j_{kh}''(u_{kh})(\delta u_{kh}, \tau u_{kh})$ in directions $\delta u_{kh}, \tau u_{kh} \in U_{kh}$.

Remark 5.4. For quadratic functionals $\mathcal{G}(\cdot)$ the Newton method for the equality constrained optimization problem in step (iii) of the PDAS algorithm converges in one iteration.

In the case of distributed and Neumann control the required derivatives of j_{kh} can be represented as on the continuous level using adjoint and linearized (tangent) discrete equations; see [5, 32] for details. Since the case of Dirichlet boundary conditions is more involved, we discuss it in what follows. In all three cases the operator q_{kh} is defined in such a way that the derivative of the discrete reduced cost functional can be expressed by

$$j_{kh}'(u_{kh})(\delta u_{kh}) = (\alpha u_{kh} - q_{kh}(u_{kh}), \delta u_{kh})_\omega.$$

In the case of Dirichlet control the derivative $j'(u)(\delta u)$ on the continuous level is given as

$$j'(u)(\delta u) = (\alpha u + \partial_n p, \delta u)_\Sigma,$$

where p is the solution of the adjoint equation; cf. the optimality system (4.7). A direct discretization of the term $\partial_n p$ does not lead in general to the derivative of the discrete cost functional j_{kh}. Therefore, we employ another representation using a residual of the adjoint equation; cf. the discussions in [43, 22].

PROPOSITION 5.5. *Let the discrete reduced cost functional j_{kh} be defined as in (5.5), with the solution operator $S_{kh} \colon U_{kh}^B \to X_{kh}$ for the discrete state equation (5.4) in the Dirichlet case. Then the following representations hold:*

1. *The first directional derivative in direction $\delta u_{kh} \in U_{kh}^B$ can be expressed as*

$$(5.7) \qquad \begin{aligned} j_{kh}'(u_{kh})(\delta u_{kh}) &= (\mathcal{G}'(y_{kh}^1), \widehat{\delta u_{kh}})_I + (\partial_t \widehat{\delta u_{kh}}, p_{kh}^1)_I + (\nabla \widehat{\delta u_{kh}}, \nabla p_{kh}^1)_I \\ &\quad + \alpha \langle u_{kh}, \delta u_{kh} \rangle_I, \end{aligned}$$

   where $y_{kh}^1 = S_{kh}(u_{kh})$, $\widehat{\delta u_{kh}}$ is the extension of δu_{kh} defined as in (5.3), and $p_{kh} = (p_{kh}^1, p_{kh}^2) \in \widetilde{X}_{kh}^0 \times \widetilde{X}_{kh}$ is the solution to the discrete adjoint equation

$$(5.8) \qquad a(\eta, p_{kh}) = -J_y'(y_{kh}^1, u_{kh})(\eta^1) \quad \text{for all } \eta \in X_{kh}^0 \times X_{kh}.$$

2. *The second derivative of j_{kh} in directions $\delta u_{kh}, \tau u_{kh} \in U_{kh}^B$ can be expressed as*

$$\begin{aligned} j_{kh}''(u_{kh})(\delta u_{kh}, \tau u_{kh}) &= \mathcal{G}''(y_{kh}^1)(\delta y_{kh}^1, \widehat{\tau u_{kh}}) + (\partial_t \widehat{\tau u_{kh}}, \delta p_{kh}^1)_I \\ &\quad + (\nabla \widehat{\tau u_{kh}}, \nabla \delta p_{kh}^1)_I + \alpha \langle \delta u_{kh}, \tau u_{kh} \rangle_I, \end{aligned}$$

   where $\delta y_{kh} = (\delta y_{kh}^1, \delta y_{kh}^2) \in (\widehat{\delta u_{kh}} + X_{kh}^0) \times X_{kh}$ is the solution of the discrete tangent equation

$$(5.9) \qquad a(\delta y_{kh}, \xi) = 0 \quad \text{for all } \xi \in \widetilde{X}_{kh}^0 \times \widetilde{X}_{kh}$$

   and $\delta p_{kh} \in \widetilde{X}_{kh}^0 \times \widetilde{X}_{kh}$ is given by

$$(5.10) \qquad a(\eta, \delta p_{kh}) = -J_{yy}''(y_{kh}^1, u_{kh})(\delta y_{kh}^1, \eta^1) \quad \text{for all } \eta \in X_{kh}^0 \times X_{kh}.$$

Proof. Using the solution δy_{kh} of the discretized tangent equation (5.9), we obtain

$$j_{kh}'(u_{kh})(\delta u_{kh}) = J_y'(y_{kh}^1, u_{kh})(\delta y_{kh}^1) + J_u'(y_{kh}^1, u_{kh})(\delta u_{kh});$$

rewriting the first term using (5.8) and (5.9), we get

$$\begin{aligned} J_y(y_{kh}^1, u_{kh})(\delta y_{kh}) &= J_y'(y_{kh}^1, u_{kh})(\delta y_{kh}^1 - \widehat{\delta u_{kh}}) + J_y'(y_{kh}^1, u_{kh})(\widehat{\delta u_{kh}}) \\ &= -(\partial_t(\delta y_{kh}^1 - \widehat{\delta u_{kh}}), p_{kh}^1)_I - (\nabla(\delta y_{kh}^1 - \widehat{\delta u_{kh}}), \nabla p_{kh}^1)_I \\ &\quad + (\mathcal{G}'(y_{kh}^1), \widehat{\delta u_{kh}})_I \\ &= (\partial_t \widehat{\delta u_{kh}}, p_{kh}^1)_I + (\nabla \widehat{\delta u_{kh}}, \nabla p_{kh}^1)_I + (\mathcal{G}'(y_{kh}^1), \widehat{\delta u_{kh}})_I. \end{aligned}$$

This gives the desired representation (5.7). The representation of the second derivative is obtained in a similar way. $\square$

Remark 5.6.   We note that for both the state equation (5.4) and the tangent equation (5.9) the discrete solutions are continuous piecewise linear in time functions; i.e., the ansatz space is $X_{kh}^0 \times X_{kh}$ and the test space consists of discontinuous piecewise constant (in time) functions (i.e., the test space is $\widetilde{X}_{kh}^0 \times \widetilde{X}_{kh}$). For both adjoint equations (5.8) and (5.10) the ansatz and the test spaces are exchanged. The ansatz functions are discontinuous and piecewise constant (in time), and test functions are continuous piecewise linear in time. This allows for a consistent formulation; cf. the discussions in [5, 32].

5.5. Time stepping formulations. Although the discrete state equation (5.4) as well as the discrete tangent (5.9) and adjoint (5.8), (5.10) equations are formulated globally in time, they result in time stepping schemes. This is due to the fact that for all these equations either the ansatz or the test functions are discontinuous in time. Applying the trapezoidal rule piecewise for approximation of time integrals, the considered time discretization results in a Crank–Nicolson scheme; see, e.g., [32, 5]. In what follows we describe the corresponding time stepping schemes for (5.4), (5.9), (5.8), and (5.10) explicitly; cf. [22] for Dirichlet boundary control of the heat equation discretized by a discontinuous Galerkin variant of the implicit Euler scheme. Thereby, we assume that the functional $\mathcal{G}$ can be represented as

$$\mathcal{G}(y) = \int_0^T g(y(t))\, dt$$

with a functional $g \in C^2(L^2(\Omega), \mathbb{R})$.

We define for $m = 0, \ldots, M$

$$U_m = u_{kh}(t_m), \quad Y_m^1 = y_{kh}^1(t_m), \quad Y_m^2 = y_{kh}^2(t_m),$$

and for $m = 1, \ldots, M$ we define

$$P_m^1 = p_{kh}^1|_{I_m}, \quad P_m^2 = p_{kh}^2|_{I_m}$$

and

$$P_0^1 = p_{kh}^1(0), \quad P_0^2 = p_{kh}^2(0).$$

The discrete state equation for Y_0^1, $Y_0^2 \in V_h$ and $Y_m^1 \in \widehat{U}_m + V_h^0$, $Y_m^2 \in V_h$ for $m = 1, \ldots, M$ is given as follows:

$\boldsymbol{m = 0}$:

$$(Y_0^1, \varphi^1) + (Y_0^2, \varphi^2) = (y_0, \varphi^1) + (y_1, \varphi^2) \quad \text{for all } \varphi^1, \ \varphi^2 \in V_h.$$

$\boldsymbol{m = 1, \ldots, M}$:

$$(Y_m^2, \varphi^1) + (Y_m^1, \varphi^2) + \frac{k_m}{2}(\nabla Y_m^1, \nabla \varphi^1) + \rho \frac{k_m}{2}(\nabla Y_m^2, \nabla \varphi^1) - \frac{k_m}{2}(Y_m^2, \varphi^2)$$

$$= (Y_{m-1}^2, \varphi^1) + (Y_{m-1}^1, \varphi^2) - \frac{k_m}{2}(\nabla Y_{m-1}^1, \nabla \varphi^1) - \rho \frac{k_m}{2}(\nabla Y_{m-1}^2, \nabla \varphi^1)$$

$$+ \frac{k_m}{2}(Y_{m-1}^2, \varphi^2) + \frac{k_m}{2}(f(t_{m-1}), \varphi^1) + \frac{k_m}{2}(f(t_m), \varphi^1)$$

$$\text{for all } \varphi_1 \in V_h^0, \ \varphi_2 \in V_h.$$

The discrete adjoint equation for $P_0^1, P_0^2 \in V_h$ and $P_m^1 \in V_h^0$, $P_m^2 \in V_h$ for $m = 1, \ldots, M$ is given as follows:

$\boldsymbol{m = M}$:

$$(\eta^2, P_M^1) + (\eta^1, P_M^2) + \frac{k_M}{2}(\nabla \eta^1, \nabla P_M^1) - \rho \frac{k_M}{2}(\nabla \eta^2, \nabla P_M^1) - \frac{k_M}{2}(\eta^2, P_M^2)$$

$$= -\frac{k_M}{2} g'(Y_M^1)(\eta^1) \quad \text{for all } \eta^1 \in V_h^0, \ \eta^2 \in V_h.$$

$m = M - 1, \ldots, 1$:

$$(\eta^2, P_m^1) + (\eta^1, P_m^2) + \frac{k_m}{2}(\nabla\eta^1, \nabla P_m^1) - \rho\frac{k_m}{2}(\nabla\eta^2, \nabla P_m^1) - \frac{k_m}{2}(\eta^2, P_m^2)$$

$$= (\eta^2, P_{m+1}^1) + (\eta^1, P_{m+1}^2) - \frac{k_{m+1}}{2}(\nabla\eta^1, \nabla P_{m+1}^1) + \rho\frac{k_{m+1}}{2}(\nabla\eta^2, \nabla P_{m+1}^1)$$

$$+ \frac{k_{m+1}}{2}(\eta^2, P_{m+1}^2) - \frac{k_m + k_{m+1}}{2}g'(Y_m^1)(\eta^1) \quad \text{for all } \eta^1 \in V_h^0,\ \eta^2 \in V_h.$$

$m = 0$:

$$(\eta^1, P_0^1) + (\eta^2, P_0^2) = (\eta^1, P_1^1) + (\eta^2, P_1^2) - \frac{k_1}{2}(\nabla\eta^1, \nabla P_1^1) + \rho\frac{k_1}{2}(\nabla\eta^2, \nabla P_1^1)$$

$$+ \frac{k_1}{2}(\eta^2, P_1^2) - \frac{k_1}{2}g'(Y_0^1)(\eta^1) \quad \text{for all } \eta^1, \eta^2 \in V_h.$$

Next we describe (5.9) and (5.10). Therefore, we define for $i = 0, \ldots, M$,

$$\delta U_m = \delta u_{kh}(t_m), \quad \delta Y_m^1 = \delta y_{kh}^1(t_m), \quad \delta Y_m^2 = \delta y_{kh}^2(t_m),$$

and for $i = 1, \ldots, m$ we define

$$\delta P_m^1 = \delta p_{kh}^1|_{I_m}, \quad \delta P_m^2 = \delta p_{kh}^2|_{I_m}$$

and

$$\delta P_0^1 = \delta p_{kh}^1(0), \quad \delta P_0^2 = \delta p_{kh}^2(0).$$

The discrete tangent equation for $\delta Y_0^1, \delta Y_0^2 \in V_h$ and $\delta Y_m^1 \in \widehat{\delta U}_m + V_h^0,\ \delta Y_m^2 \in V_h$ for $m = 1, \ldots, M$ is given as follows:

$m = 0$:

$$\delta Y_0^1 = \delta Y_0^2 = 0.$$

$m = 1, \ldots, M$:

$$(\delta Y_m^2, \varphi^1) + (\delta Y_m^1, \varphi^2) + \frac{k_m}{2}(\nabla\delta Y_m^1, \nabla\varphi^1) + \rho\frac{k_m}{2}(\nabla\delta Y_m^2, \nabla\varphi^1) - \frac{k_m}{2}(\delta Y_m^2, \varphi^2)$$

$$= (\delta Y_{m-1}^2, \varphi^1) + (\delta Y_{m-1}^1, \varphi^2) - \frac{k_m}{2}(\nabla\delta Y_{m-1}^1, \nabla\varphi^1) - \rho\frac{k_m}{2}(\nabla\delta Y_m^2, \nabla\varphi^1)$$

$$+ \frac{k_m}{2}(\delta Y_{m-1}^2, \varphi^2) \text{ for all } \varphi^1 \in V_h^0,\ \varphi^2 \in V_h.$$

The additional adjoint equation for $\delta P_0^1, \delta P_0^2 \in V_h$ and $\delta P_m^1 \in V_h^0, \delta P_m^2 \in V_h$ for $m = 1, \ldots, M$ is given as follows:

$m = M$:

$$(\eta^2, \delta P_M^1) + (\eta^1, \delta P_M^2) + \frac{k_M}{2}(\nabla\eta^1, \nabla\delta P_M^1) - \rho\frac{k_M}{2}(\nabla\eta^2, \nabla\delta P_M^1)$$

$$- \frac{k_M}{2}(\eta^2, \delta P_M^2) = -\frac{k_M}{2}g''(Y_M)(\delta Y_M, \eta^1)$$

$$\text{for all } \eta^1 \in V_h^0, \eta^2 \in V_h.$$

$m = M - 1, \ldots, 1$:

$$(\eta^2, \delta P_m^1) + (\eta^1, \delta P_m^2) + \frac{k_m}{2}(\nabla\eta^1, \nabla\delta P_m^1) - \rho\frac{k_m}{2}(\nabla\eta^2, \nabla\delta P_M^1) - \frac{k_m}{2}(\eta^2, \delta P_m^2)$$

$$= (\eta^2, \delta P_{m+1}^1) + (\eta^1, \delta P_{m+1}^2) - \frac{k_{m+1}}{2}(\nabla\eta^1, \nabla\delta P_{m+1}^1) + \rho\frac{k_{m+1}}{2}(\nabla\eta^2, \nabla\delta P_M^1)$$

$$+ \frac{k_{m+1}}{2}(\eta^2, \delta P_{m+1}^2) - \frac{k_m + k_{m+1}}{2}g''(Y_m)(\delta Y_m, \eta^1)$$

$$\text{for all } \eta^1 \in V_h^0, \ \eta^2 \in V_h.$$

$m = 0$:

$$(\eta^2, \delta P_0^1) + (\eta^1, \delta P_0^2) = (\eta^2, \delta P_1^1) + (\eta^1, \delta P_1^2) - \frac{k_1}{2}(\nabla\eta^1, \nabla\delta P_1^1)$$

$$+ \frac{k_1}{2}(\nabla\eta^2, \nabla\delta P_1^1) + \frac{k_1}{2}(\eta^2, \delta P_1^2) - \frac{k_1}{2}g''(Y_0)(\delta Y_0, \eta^1)$$

$$\text{for all } \eta^1 \in V_h^0, \ \eta^2 \in V_h.$$

6. Numerical examples. In this section we discuss numerical examples illustrating our theoretical results for the optimal control problems under consideration. We present a comparison of the numbers of PDAS iterations for different discretization levels as well as some results illustrating the error behavior on a fixed mesh. On the discrete level (for fixed temporal and spatial meshes) the PDAS method typically converges in a finite number of steps (cf. the stopping criterion in Remark 2.12), which is better than superlinear convergence. The examples indicate superlinear convergence also before the PDAS method stops finding the optimal discrete solution.

All computations are done using the optimization library RoDoBo [37] and the finite element toolkit Gascoigne [13].

In the following we consider distributed, Neumann boundary, and Dirichlet boundary control with and without damping on the unit square $\Omega = (0,1)^2 \subset \mathbb{R}^2$. Here, we specify the functional $\mathcal{G}$ in the following way: For a given function $y_d \in L^2(Q)$ we define

$$\mathcal{G}(y) = \frac{1}{2}\|y - y_d\|_{L^2(Q)}^2.$$

6.1. Example 1: Distributed control. We compute the distributed optimal control problem (4.1) with the following data:

$$\alpha = 0.01, \quad u_a = -0.6, \quad u_b = 2, \quad T = 1,$$

$$y_d(t, x) = \begin{cases} 10x_2 & \text{if } x_1 < 0.5, \\ 1 & \text{else,} \end{cases} \quad y_0(x) = \sin(\pi x_1)\sin(\pi x_2), \quad y_1(x) = 0$$

for $t \in [0, T]$ and $x = (x_1, x_2) \in \Omega$.

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 is shown for a sequence of uniformly refined discretizations. Here, N denotes the number of cells in the spatial mesh $\mathcal{T}_h$ and M denotes the number of time intervals. The results indicate a mesh-independent behavior of the PDAS algorithm.

TABLE 6.1
PDAS method on the sequence of uniformly refined meshes for distributed control problem.

Level	N	M	PDAS steps
1	16	2	5
2	64	4	4
3	256	8	5
4	1024	16	4
5	4096	32	4
6	16384	64	5

To analyze the convergence behavior of the PDAS method we define the PDAS iteration error

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

where $u_{kh}^{(i)}$ denotes the ith iterate and u_{kh} denotes the optimal discrete solution. For a fixed discretization with $N = 16384$ cells and $M = 64$ time steps, Table 6.2 depicts the rate of convergence of the PDAS iteration. The results presented demonstrate superlinear convergence.

TABLE 6.2
Superlinear convergence of the PDAS method for distributed control.

i	1	2	3	4
e_i	$3.6 \cdot 10^{-2}$	$9.7 \cdot 10^{-4}$	$2.1 \cdot 10^{-5}$	0
e_{i+1}/e_i	$2.7 \cdot 10^{-2}$	$2.2 \cdot 10^{-2}$	0	-

6.2. Example 2: Neumann control. We consider the Neumann boundary control problem (4.4) with the following data:

$$f(t,x) = \begin{cases} 1 & \text{if } x_1 > 0.25, \\ -1 & \text{else,} \end{cases} \quad \alpha = 0.01, \quad u_a = -0.8, \quad u_b = 1, \quad T = 1,$$

$$y_d(t,x) = \begin{cases} -x_1 & \text{if } x_1 > 0.05, \\ 2 & \text{else,} \end{cases} \quad y_0(x) = \sin(\pi x_1)\sin(\pi x_2), \quad y_1(x) = 0$$

for $t \in [0,T]$ and $x = (x_1, x_2) \in \Omega$.

As in the previous example, we see in Table 6.3 that the number of PDAS iterations is mesh-independent under uniform refinement of the discretizations.

For a fixed discretization with $N = 16384$ cells and $M = 64$ time steps, Table 6.4 shows the rate of convergence of the PDAS iteration illustrating superlinear convergence.

6.3. Example 3: Dirichlet control. We consider the Dirichlet optimal control problems (4.6) and (4.9) with the following data:

$$f(t,x) = \begin{cases} 1 & x_1 > 0.5, \\ x_1 & \text{else,} \end{cases} \quad u_a = -0.18, \quad u_b = 0.2, \quad T = 1,$$

$$y_d(t,x) = \begin{cases} x_1 & x_1 > 0.5, \\ -x_1 & \text{else,} \end{cases} \quad y_0(x) = \sin(\pi x_1)\sin(\pi x_2), \quad y_1(x) = 0$$

for $t \in [0,T]$ and $x = (x_1, x_2) \in \Omega$.

TABLE 6.3

PDAS method on the sequence of uniformly refined meshes for Neumann boundary control.

Level	N	M	PDAS steps
1	16	2	5
2	64	4	5
3	256	8	3
4	1024	16	4
5	4096	32	4
6	16384	64	5

TABLE 6.4

Superlinear convergence of the PDAS method for Neumann boundary control.

i	1	2	3	4
e_i	$3.0 \cdot 10^{-2}$	$9.7 \cdot 10^{-4}$	$2.8 \cdot 10^{-5}$	0
e_{i+1}/e_i	$3.2 \cdot 10^{-2}$	$2.9 \cdot 10^{-2}$	0	-

TABLE 6.5

Numbers of PDAS iterations on the sequence of uniformly refined meshes for different parameters α and ρ.

Level	N	M	$\alpha = 10^{-4}$			$\alpha = 10^{-2}$		
			$\rho = 0$	$\rho = 0.1$	$\rho = 0.7$	$\rho = 0$	$\rho = 0.1$	$\rho = 0.7$
1	16	2	4	3	5	4	4	5
2	64	4	5	4	3	4	4	3
3	256	8	5	5	4	5	4	4
4	1024	16	6	6	6	5	7	5
5	4096	32	11	7	7	9	6	5
6	16384	64	13	9	7	10	8	5

Level	N	M	$\alpha = 1$		
			$\rho = 0$	$\rho = 0.1$	$\rho = 0.7$
1	16	2	3	3	2
2	64	4	3	3	1
3	256	8	4	3	1
4	1024	16	4	2	1
5	4096	32	3	3	1
6	16384	64	3	4	1

Table 6.5 illustrates the effect of damping introduced by the term $-\rho\Delta y_t$ on the number of PDAS steps. For $\alpha = 0.01$ and $\rho = 0$ we observe a mesh-dependence of the algorithm. Moreover, the number of PDAS steps declines for an increasing value of ρ and stays mesh-independent for $\rho > 0$. Furthermore, we consider the effect of α on the number of PDAS steps. As expected, the number of iterations declines also for increasing α.

In Tables 6.6 and 6.7 we consider the PDAS iteration error for the discretization with $N = 16384$ cells and $M = 64$ time steps, where we choose $\rho = 0$ and $\rho = 0.1$, respectively, and $\alpha = 0.01$. These tables indicate that we have only superlinear convergence for $\rho > 0$.

TABLE 6.6

Equation without damping, $\rho = 0$; PDAS iteration error.

i	1	2	3	4	5	6	7
e_i	$2.3 \cdot 10^{-2}$	$2.2 \cdot 10^{-2}$	$4.5 \cdot 10^{-3}$	$1.9 \cdot 10^{-3}$	$7.2 \cdot 10^{-4}$	$3.8 \cdot 10^{-4}$	$1.2 \cdot 10^{-4}$
e_{i+1}/e_i	$9.5 \cdot 10^{-1}$	$2.0 \cdot 10^{-1}$	$4.2 \cdot 10^{-1}$	$3.8 \cdot 10^{-1}$	$5.2 \cdot 10^{-1}$	$3.1 \cdot 10^{-1}$	$4.1 \cdot 10^{-1}$

i	7	8	9
e_i	$4.8 \cdot 10^{-5}$	$1.4 \cdot 10^{-5}$	0
e_{i+1}/e_i	$3.0 \cdot 10^{-1}$	0	-

TABLE 6.7

Equation with damping $\rho = 0.1$; PDAS iteration error.

i	1	2	3	4	5	6	7
e_i	$3.8 \cdot 10^{-1}$	$5.2 \cdot 10^{-2}$	$1.0 \cdot 10^{-2}$	$1.5 \cdot 10^{-3}$	$1.8 \cdot 10^{-4}$	$1.7 \cdot 10^{-5}$	0
e_{i+1}/e_i	$1.3 \cdot 10^{-1}$	$1.9 \cdot 10^{-1}$	$1.6 \cdot 10^{-1}$	$1.2 \cdot 10^{-1}$	$9.3 \cdot 10^{-2}$	0	-

REFERENCES

[1] R. A. Adams, *Sobolev Spaces*, Academic Press, New York, 1975.
[2] L. Bales and I. Lasiecka, *Continuous finite elements in space and time for the nonhomogeneous wave equation*, Comput. Math. Appl., 27 (1994), pp. 91–102.
[3] L. Bales and I. Lasiecka, *Negative norm estimates for fully discrete finite element approximations to the wave equation with nonhomogeneous L_2 Dirichlet boundary data*, Math. Comp., 64 (1995), pp. 89–115.
[4] W. Bangerth and R. Rannacher, *Adaptive finite element techniques for the acoustic wave equation*, J. Comput. Acoust., 9 (2001), pp. 575–591.
[5] R. Becker, D. Meidner, and B. Vexler, *Efficient numerical solution of parabolic optimization problems by finite element methods*, Optim. Methods Softw., 22 (2007), pp. 813–833.
[6] D. Braess, *Finite Elements: Theory, Fast Solvers and Applications in Solid Mechanics*, Cambridge University Press, Cambridge, UK, 2007.
[7] F. Bucci, *A Dirichlet boundary control problem for the strongly damped wave equation*, SIAM J. Control Optim., 30 (1992), pp. 1092–1100.
[8] R. Chill and S. Srivastava, *L^p-maximal regularity for second order Cauchy problems*, Math. Z., 251 (2005), pp. 751–781.
[9] C. Clason, B. Kaltenbacher, and S. Veljović, *Boundary optimal control of the Westervelt and the Kuznetsov equation*, J. Math. Anal. Appl., 356 (2009), pp. 738–751.
[10] L. C. Evans, *Partial Differential Equations*, Grad. Stud. Math. 19, AMS, Providence, RI, 1998.
[11] M. H. Farahi, J. E. Rubio, and D. A. Wilson, *The global control of a nonlinear wave equation*, Int. J. Control, 65 (1996), pp. 1–15.
[12] D. A. French and T. E. Peterson, *Continuous space-time finite elements method for the wave equation*, Math. Comp., 65 (1996), pp. 491–506.
[13] Gascoigne, *Gascoigne 3D: High Performance Adaptive Finite Element Toolkit*, http://www.gascoigne.uni-hd.de.
[14] M. Gerdts, G. Greif, and H. J. Pesch, *Numerical optimal control of the wave equation: Optimal boundary control of a string to rest in finite time*, Math. Comput. Simul., 79 (2008), pp. 1020–1032.
[15] P. Grisvard, *Commutativitè de deux foncteurs d'interpolation et applications*, J. Math. Pures Appl., 45 (1966), pp. 143–206.
[16] M. Gugat, A. Keimer, and G. Leugering, *Optimal distributed control of the wave equation subject to state constraints*, ZAMM Z. Angew. Math. Mech., 89 (2009), pp. 420–444.
[17] M. Hintermüller, K. Ito, and K. Kunisch, *The primal-dual active set strategy as a semismooth Newton method*, SIAM J. Optim., 13 (2003), pp. 865–888.
[18] G. M. Hulbert and T. J. R. Hughes, *Space-time finite element methods for second-order hyperbolic equations*, Comput. Methods Appl. Mech. Engrg., 84 (1990), pp. 327–348.
[19] K. Ito and K. Kunisch, *Semi-smooth Newton methods for state-constrained optimal control problems*, Systems Control Lett., 50 (2003), pp. 221–228.

[20] C. Johnson, *Discontinuous Galerkin finite element methods for second order hyperbolic problems*, Comput. Methods Appl. Mech. Engrg., 107 (1993), pp. 117–129.

[21] O. Karakashian and C. Makridakis, *Convergence of a continuous Galerkin method with mesh modification for nonlinear wave equation*, Math. Comp., 47 (2004), pp. 85–102.

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

[23] J. E. Lagnese and G. Leugering, *Dynamic domain decomposition in approximate and exact boundary control in problems of transmission for wave equations*, SIAM J. Control Optim., 38 (2000), pp. 503–537.

[24] J. E. Lagnese and G. Leugering, *Time-domain decomposition in optimal control of the wave equation*, Systems Control Lett., 48 (2003), pp. 229–242.

[25] S. Larsson, V. Thomee, and L. Wahlbin, *Finite element methods for a strongly damped wave equation*, IMA J. Numer. Anal., 11 (1991), pp. 115–142.

[26] I. Lasiecka and R. Triggiani, *Control Theory for Partial Differential Equations: Continuous and Approximation Theories*, Vols. 1 and 2, Encyclopedia Math. Appl. 74, 75, Cambridge University Press, Cambridge, UK, 2000.

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

[28] J.-L. Lions, *Control of Distributed Singular Systems*, Gauthier-Villars, Kent, 1985.

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

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

[31] P. Massatt, *Limiting behavior for strongly damped nonlinear wave equations*, J. Differential Equations, 48 (1983), pp. 334–349.

[32] D. Meidner, *Adaptive Space-Time Finite Element Methods for Optimization Problems Governed by Nonlinear Parabolic Systems*, Ph.D. thesis, Institut für Angewandte Mathematik, Universität Heidelberg, Heidelberg, Germany, 2008.

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

[34] S. Miyatake, *Mixed problem for hyperbolic equation of second order*, J. Math. Kyoto Univ., 13 (1973), pp. 435–487.

[35] B. S. Mordukhovich and J.-P. Raymond, *Dirichlet boundary control of hyperbolic equations in the presence of state constraints*, Appl. Math. Optim., 49 (2004), pp. 145–157.

[36] B. S. Mordukhovich and J.-P. Raymond, *Neumann boundary control of hyperbolic equations with pointwise state constraints*, SIAM J. Control Optim., 43 (2005), pp. 1354–1372.

[37] RoDoBo, *A Software Package for Solving Optimization Problems Governed by Stationary and Nonstationary PDEs with Interface to the Finite Element Toolkit Gascoigne*, http://www.rodobo.uni-hd.de.

[38] M. Schmich and B. Vexler, *Adaptivity with dynamic meshes for space-time finite element discretizations of parabolic equations*, SIAM J. Sci. Comput., 30 (2008), pp. 369–393.

[39] H. Triebel, *Interpolation Theory, Function Spaces, Differential Operators*, J. A. Barth Verlag, Heidelberg, 1975.

[40] R. Triggiani, *Regularity of some structurally damped problems with point control and boundary control*, J. Math. Anal. Appl., 161 (1991), pp. 299–331.

[41] M. Ulbrich, *Semismooth Newton methods for operator equations in function spaces*, SIAM J. Optim., 13 (2003), pp. 805–841.

[42] M. Ulbrich, *Constrained optimal control of Navier-Stokes flow by semismooth Newton methods*, Systems Control Lett., 48 (2003), pp. 297–311.

[43] B. Vexler, *Finite element approximation of elliptic Dirichlet optimal control problems*, Numer. Funct. Anal. Optim., 28 (2007), pp. 957–973.

[44] E. Zuazua, *Propagation, observation, and control of waves approximated by finite difference methods*, SIAM Rev., 47 (2005), pp. 197–243.