SpezialForschungsBereich F 32

Karl–Franzens Universität Graz
Technische Universität Graz
Medizinische Universität Graz

On Time Optimal Control of the Wave Equation and its Numerical Realization as Parametric Optimization Problem

K. Kunisch D. Wachsmuth

SFB-Report No. 2012-013 May 2012

A–8010 GRAZ, HEINRICHSTRASSE 36, AUSTRIA

Supported by the
Austrian Science Fund (FWF)

SFB sponsors:

- **Austrian Science Fund** (FWF)
- **University of Graz**
- **Graz University of Technology**
- **Medical University of Graz**
- **Government of Styria**
- **City of Graz**

On Time Optimal Control of the Wave Equation and its Numerical Realization as Parametric Optimization Problem

Karl Kunisch[1], Daniel Wachsmuth[2]

May 16, 2012

[1]Institute for Mathematics and Scientific Computing, Heinrichstraße 36, A–8010 Graz, Austria, karl.kunisch@uni-graz.at, http://www.kfunigraz.ac.at/imawww/kunisch, partially supported by 'Fonds zur Förderung der Wissenschaftlichen Forschung' under SFB 32, Mathematical Optimization and Applications in the Biomedical Sciences.

[2]University of Würzburg, Institute of Mathematics, D–97074 Würzburg, Germany, daniel.wachsmuth@mathematik.uni-wuerzburg.de, and Johann Radon Institute for Computational and Applied Mathematics (RICAM), Austrian Academy of Sciences, Altenbergerstraße 69, A–4040 Linz, Austria.q

Abstract. Time optimal control of the wave equation is analyzed on the basis of a regularized formulation which is considered as a bilevel optimization problem. For the lower level problems, which are constrained optimal control problems for the wave equation, a detailed sensitivity analysis is carried out. Further a semi-smooth Newton method is analyzed and proved to converge locally superlinearly. Numerical examples are provided.

Keywords. Time optimal control, wave equation, optimality condition, sensitivity analysis, semi-smooth Newton method.

1 Introduction

This paper is devoted to the time optimal control problem

$$(\tilde{P}) \qquad \begin{cases} \min \displaystyle\int_0^\tau dt \\ \text{subject to } \tau \geq 0, \\ y_{tt} - \Delta y = \chi_\omega u \text{ in } (0,\tau) \times \Omega, \\ y(0) = y_1,\, y_t(0) = y_2,\, y(\tau) = z_1,\, y_t(\tau) = z_2 \text{ in } \Omega, \\ y = 0 \text{ on } \Gamma, \\ \|u(t)\|_{L^2(\omega)} \leq \gamma, \text{ for a.e. } t \in (0,\tau). \end{cases}$$

Here, $\gamma > 0$ is a fixed positive constant and $\Omega \subset \mathbb{R}^n$, $n \geq 1$ is a fixed bounded domain with a C^2 boundary Γ. Further $\omega \subset \Omega$ is a measurable subset and $\chi_\omega u$ denotes the extension-by-zero operator from ω to Ω. The initial and terminal states are fixed and - unless specified otherwise - are assumed to satisfy

$$y_1 \in H_0^1(\Omega),\, z_1 \in H_0^1(\Omega),\, y_2 \in L^2(\Omega),\, z_2 \in L^2(\Omega).$$

In previous work [17] we used a regularization scheme to derive an optimality system for $(\tilde{P})$, and in particular to investigate the maximum principle and the transversality condition for $(\tilde{P})$. The present work aims at solving $(\tilde{P})$ on the basis of this regularization scheme.

Let us briefly explain the route that we will follow. The regularized time optimal problem is considered as a bilevel optimization problem. The lower level consists of constrained optimal control problems for the wave equation and the upper level problems are minimization problems with respect to τ. Alternatively we can consider our approach as a reduced approach, with the optimal time the only independent optimization variable, and the control a dependent variable. The lower level problems are parameterized optimization problems in the parameter τ. We establish Bouligand differentiability of the solutions to these lower level problems. This allows us to obtain second order directional differentiability of the cost of the parameterized optimization problems. Moreover we show that setting the first derivative of the cost equal to zero corresponds to the transversality condition in the optimality system. This structural result is not restricted to the case where the constrained equation is the wave equation. The solution of the lower level problem requires efficient solution techniques. For this purpose we analyze a semi-smooth Newton method. This necessitates to establish Newton differentiability of the projection onto the unit ball in L^2, a result which is of interest in its own right. Finally there is an outermost iteration within which the regularization parameter is driven to zero.

The development of numerical techniques for time optimal control problems has received much attention in the context of ordinary differential equations. They are frequently categorized into direct and indirect methods. Indirect methods based on multiple shooting techniques [1, 12] solve the two point boundary value problem describing first order necessary conditions. In [10] a semi-smooth Newton method was recently proposed for solving the non-smooth optimality systems. Direct methods on the other hand, consider time optimal problems as genuine nonlinear programming problems. They are used in several variants, which frequently involve re-parametrization of the controls as the unknowns. The new unknowns can be the switching times as in [19] or the arc durations as in [11].

Time optimal control for the wave equation was frequently investigated in previous work, mostly with the aim of deriving different forms of the maximum principle. We refer e.g. to Fattorini [3, 4], Gugat and Leugering [8], and Krabs [14, 15], and the references given there. In this context it was also discussed that time-optimal control problems can be addressed alternatively by solving appropriately defined dual norm-optimal control problems, which are parameterized by the time τ, see e.g. [5, 14]. If for some parameter value $\hat{\tau}$ the norm-optimal control satisfies $\|\hat{u}\|_{L^\infty(I;L^2(\omega))} = \gamma$, then $(\hat{\tau}, \hat{u})$ is a solution of $(\tilde{P})$. However, this equivalence is typically established only for the special case $\omega = \Omega$. An example in [8] shows that the equivalence of time-optimal and norm-optimal control problems cannot be expected in the general case. In [7] the relationship between norm- and time optimal problems was utilized to develop a method-of-moments-based algorithm to solve time-optimal control problems for one dimensional vibrating systems. In a recent paper Wang and Zuazua [21] analyzed the equivalence of time- and norm-optimal control problems for the heat equation for the case $\omega \neq \Omega$. The proof of equivalence uses two properties that are not available for the wave equation: unconstrained null-controllability for arbitrarily small times and the bang-bang property of time-optimal controls. Hence addressing the time-optimal control problem for the wave equation in the general situation $\omega \neq \Omega$ by means of the dual norm-optimal control problem remains to be an independent challenge.

The paper is organized as follows. In Section 2 we introduce the abstract form of the wave equation and of problem $(\tilde{P})$ and recall selected results from [17]. Section 3 is devoted to the sensitivity analysis of a parametric family of optimization problems parameterized in the variable τ. In Section 4 we give the analysis of the semi-smooth Newton method for the lower level problems. Section 5 contains a description of the numerical algorithm and selected numerical experiments.

2 Preliminaries

2.1 Abstract formulation

Let us recall some concepts for the wave equation

$$\begin{cases} y_{tt} - \Delta y = \chi_\omega u & \text{in } (0,\tau)\times\Omega, \\ y(0) = y_1,\, y_t(0) = y_2, & \text{in } \Omega \\ y = 0 & \text{on } \Gamma, \end{cases} \tag{2.1}$$

with $\tau > 0$ fixed, that will be relevant to our work. For the purpose of numerical realization it is convenient to perform a transformation of equation (2.1) to the fixed time interval

$$I := (0,1).$$

Moreover, to express (2.1) in abstract form we introduce the operators

$$\mathbf{A} := \begin{pmatrix} 0 & I \\ \Delta & 0 \end{pmatrix}, \quad \mathbf{B} := \begin{pmatrix} 0 \\ \chi_\omega \end{pmatrix},$$

and vectors

$$\mathbf{y}_0 := \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \ \mathbf{z} := \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}, \ \mathbf{y}(t) := \begin{pmatrix} \mathbf{y}_1(t) \\ \mathbf{y}_2(t) \end{pmatrix}.$$

Then the wave equation (2.1) can be expressed as the first-order evolution equation

$$\begin{aligned} \mathbf{y}_t &= \tau(\mathbf{A}\mathbf{y} + \mathbf{B}u) \quad \text{on } (0,1], \\ \mathbf{y}(0) &= \mathbf{y}_0. \end{aligned} \tag{2.2}$$

The components of the solution $\mathbf{y}$ of this equation fulfill $(\mathbf{y}_1)_t = \tau\mathbf{y}_2$ and $(\mathbf{y}_2)_t = \tau(\Delta\mathbf{y}_1 + \chi_\omega u)$. For convenience of notation we introduce the function spaces

$$Y^s = \begin{cases} H^s(\Omega) & 0 \le s < 1/2, \\ H^s(\Omega) \cap \{y : \ y|_\Gamma = 0\} & s \ge 1/2, \\ (Y^{-s})^* & s < 0, \end{cases}$$

and the associated vector-valued spaces, which take account of the regularity of the components of solutions $\mathbf{y}$ of (2.2):

$$\mathbf{Y}^s := Y^s \times Y^{s-1}.$$

By assumption we have $\mathbf{y}_0 \in \mathbf{Y}^1$. Utilizing this notation the operator $\mathbf{A}$ is a continuous linear operator in the following sense

$$\mathbf{A} \in \mathcal{L}(\mathbf{Y}^s, \ \mathbf{Y}^{s-1}).$$

Moreover, the operator $\mathbf{B}$ has the property

$$\mathbf{B} \in \mathcal{L}(L^2(\omega), \mathbf{Y}^1).$$

Next we transform $(\tilde{P})$ to the interval I and express it in abstract form. For this purpose we introduce the set of admissible controls

$$U_{ad} := \{u \in L^\infty(I; L^2(\omega)) : \ u(t) \in U \text{ a.e. on } I\},$$

with U given by

$$U = \{u \in L^2(\omega) : \ \|u\|_{L^2(\omega)} \le \gamma\}.$$

Now $(\tilde{P})$ can be expressed as

$$(\mathrm{P}) \qquad \begin{cases} \min \tau \\ \text{subject to } \tau \geq 0 \text{ and} \\ \mathbf{y}_t = \tau(\mathbf{A}\mathbf{y} + \mathbf{B}u) \text{ on } (0,1], \\ \mathbf{y}(0) = \mathbf{y}_0,\ \mathbf{y}(1) = \mathbf{z}, \\ u \in U_{ad}, \end{cases}$$

For existence and uniqueness of weak solutions of system (2.2), we have the following well-known result, see e.g. [2, 18].

Theorem 2.1. *Let $\mathbf{y}_0 \in \mathbf{Y}^0$, $u \in L^2(I; L^2(\omega))$ be given. Then the first-order equation* (2.2) *admits a unique very weak solution $\mathbf{y}$ satisfying*

$$\mathbf{y} \in C(\bar{I}; \mathbf{Y}^0).$$

If in addition $\mathbf{y}_0 \in \mathbf{Y}^1$ holds, then the first-order equation (2.2) *admits a unique weak solution $\mathbf{y}$ that satisfies*

$$\mathbf{y} \in C(\bar{I}; \mathbf{Y}^1),\ \ \mathbf{y}_t \in L^2(I; \mathbf{Y}^0)$$

If moreover $\mathbf{y}_0 \in \mathbf{Y}^2$, $u_t \in L^2(I; L^2(\omega))$, then

$$\mathbf{y} \in C(\bar{I}; \mathbf{Y}^2),\ \ \mathbf{y}_t \in C(\bar{I}; \mathbf{Y}^1).$$

In either of these cases the solution $\mathbf{y}$ depends continuously on $(u, \mathbf{y}_0)$ in the indicated norms.

Together with the primal equation (2.2), the adjoint equation will play a central role. It is defined as the evolution equation

$$-\mathbf{p}_t = \tau \mathbf{A}^* \mathbf{p} \quad \text{ on } I. \tag{2.3}$$

Here $\mathbf{A}^*$ is given as adjoint of $\mathbf{A}$:

$$\mathbf{A}^* := \begin{pmatrix} 0 & \Delta \\ I & 0 \end{pmatrix}.$$

Hence, equation (2.3) is a wave equation in the second coordinate $\mathbf{p}_2$ with $(\mathbf{p}_2)_t = -\tau\, \mathbf{p}_1$ and

$$(\mathbf{p}_2)_{tt} = \tau^2\, \Delta \mathbf{p}_2.$$

It will be convenient to introduce the notation

$$\mathbf{P}^s := Y^{s-1} \times Y^s,$$

which will be used for $s = 0, 1, 2$. The index s with $\mathbf{Y}^s$ and $\mathbf{P}^s$ denotes the regularity of the wave function for the primal state $\mathbf{y}$ and the adjoint state $\mathbf{p}$, respectively. We may note that $(\mathbf{Y}^s)^* = \mathbf{P}^{(1-s)}$.

If the adjoint equation is completed with a terminal condition $\mathbf{p}(1) = \hat{\mathbf{p}}$ with $\hat{\mathbf{p}} \in \mathbf{P}^0$, then it is uniquely solvable with solution $\mathbf{p} \in C(\bar{I}; \mathbf{P}^0)$. Moreover, one has regularity results analogous to those for the primal wave equation expressed in Theorem 2.1.

To guarantee well-posedness of (P) we need to assume that the set of feasible controls is nonempty. For this purpose we assume that

(H1) the wave equation (2.2) is null controllable,

i.e. there exists some $\tau_0 > 0$ and $c > 0$ such that

$$\|\mathbf{p}(1)\|^2_{\mathbf{P}^0} \leq c\|\mathbf{B}^*\mathbf{p}\|^2_{L^2(I;L^2(\omega))}, \tag{2.4}$$

for every solution to the adjoint equation (2.3) with $\tau = \tau_0$. With (H1) holding it can be shown that there exists a feasible solution to (P) for τ large enough [17, 20], and then existence to (P) readily follows.

2.2 A family of regularized problems

To derive an optimality system for (P), the following family of problems was considered in [17], which involves a regularization of the control and a penalization of the terminal constraint: For $\varepsilon > 0$ we consider

$$(P_\varepsilon) \qquad \begin{cases} \min J_\varepsilon(\tau, u) = \tau\left(1 + \frac{\varepsilon}{2}\|u\|^2_{L^2(I;L^2(\omega))}\right) + \frac{1}{2\varepsilon}\|\mathbf{y}(1) - \mathbf{z}\|^2_{\mathbf{Y}^0}, \\ \text{subject to } \tau \geq 0 \text{ and} \\ \mathbf{y}_t = \tau\mathbf{A}\mathbf{y} + \tau\mathbf{B}u, \text{ on } (0, 1], \\ \mathbf{y}(0) = \mathbf{y}_0, \\ u \in U_{ad}. \end{cases}$$

Here and below the norm on $\mathbf{Y}^0 = L^2(\Omega) \times H^{-1}(\Omega)$ is chosen to be

$$\|\mathbf{v}\|^2_{\mathbf{Y}^0} = \|v_1\|^2_{L^2(\Omega)} + ((-\Delta)^{-1}v_2, v_2)_{L^2(\Omega)},$$

where $w = (-\Delta)^{-1} v_2$ is the solution of $-\Delta w = v_2$ in Ω, $w = 0$ on Γ.

For any $\varepsilon > 0$ problem (P_ε) admits a solution $\{(\tau_\varepsilon, \mathbf{y}_\varepsilon, u_\varepsilon)\}$ independently of (H1). In order to guarantee convergence of these solutions to a solution of (P), it is, of course, required to assume (H1).

Theorem 2.2. *Assume that* (H1) *holds and let* $\{(\tau_\varepsilon, \mathbf{y}_\varepsilon, u_\varepsilon)\}_{\varepsilon>0}$ *denote a family of solutions of* (P_ε)*. Then we have that*

$$\tau_\varepsilon \to \tau^*, \ \text{for } \varepsilon \to 0^+,$$

and $(\mathbf{y}_\varepsilon, u_\varepsilon)$ *is uniformly bounded in* $(C(\bar{I}; \mathbf{Y}^1) \cap H^1(I; \mathbf{Y}^0)) \times L^\infty(T; L^2(\omega))$*. Moreover, for each weakly-star converging subsequence* $\{(\mathbf{y}_{\varepsilon_n}, u_{\varepsilon_n})\}$ *with*

$$\mathbf{y}_{\varepsilon_n} \rightharpoonup^* \tilde{\mathbf{y}} \text{ in } L^\infty(I; \mathbf{Y}^1) \cap H^1(I; \mathbf{Y}^0), \quad u_{\varepsilon_n} \rightharpoonup^* \tilde{u} \text{ in } L^\infty(I; L^2(\omega)),$$

the limit $(\tilde{\mathbf{y}}, \tilde{u})$ *is a solution of the original time-optimal control problem* (P)*.*

If $\tilde{u}$ *is bang-bang, then the convergence* $(\mathbf{y}_{\varepsilon_n}, u_{\varepsilon_n}) \to (\tilde{\mathbf{y}}, \tilde{u})$ *is strong in* $(C(\bar{I}; \mathbf{Y}^1) \cap H^1(I; \mathbf{Y}^0)) \times L^2(I; L^2(\omega))$*.*

This result was proved in [17]. There (P_ε) was also used to investigate the first order optimality condition for (P), involving the maximum principle and the transversality condition. In the present work we focus on the numerical realization of (P_ε).

It is based on the necessary optimality conditions for (P_ε) from [17] which we recall next.

Theorem 2.3. *Let* $(\tau_\varepsilon, \mathbf{y}_\varepsilon, u_\varepsilon)$ *be a local solution of* (P_ε)*. Then there exists* $\mathbf{p}_\varepsilon \in C(\bar{I}; \mathbf{P}^2) \cap C^1(\bar{I}; \mathbf{P}^1)$ *such that the following optimality system holds:*

$$\tag{2.5} \begin{cases} \partial_t \mathbf{y}_\varepsilon = \tau_\varepsilon \mathbf{A} \mathbf{y}_\varepsilon + \tau_\varepsilon \mathbf{B} u_\varepsilon, \quad \mathbf{y}_\varepsilon(0) = \mathbf{y}_0, \\ -\partial_t \mathbf{p}_\varepsilon = \tau_\varepsilon \mathbf{A}^* \mathbf{p}_\varepsilon, \quad \mathbf{p}_\varepsilon(1) = \dfrac{1}{\varepsilon} \begin{pmatrix} \mathbf{y}_{\varepsilon,1}(1) - z_1 \\ (-\Delta)^{-1}(\mathbf{y}_{\varepsilon,2}(1) - z_2) \end{pmatrix} \in \mathbf{P}^2, \\ (\varepsilon u_\varepsilon + \mathbf{B}^* \mathbf{p}_\varepsilon, u - u_\varepsilon)_{L^2(I;L^2(\omega))} \geq 0, \ \text{for all } u \in U_{ad}, \\ 1 + \dfrac{\varepsilon}{2} \|u_\varepsilon\|^2_{L^2(I;L^2(\omega))} + \langle \mathbf{A}\mathbf{y}_\varepsilon + \mathbf{B} u_\varepsilon, \ \mathbf{p}_\varepsilon \rangle_{L^2(I;\mathbf{Y}^0), L^2(I;\mathbf{P}^1)} = 0. \end{cases}$$

The optimal control u_ε *has the additional regularity*

$$u_\varepsilon \in C(\bar{I}; L^2(\omega)) \text{ and } \partial_t u_\varepsilon \in L^\infty(I; L^2(\omega)).$$

If moreover $\mathbf{y}_0 \in \mathbf{Y}^2$*, then*

$$\mathbf{y}_\varepsilon \in C(\bar{I}; \mathbf{Y}^2) \cap C^1(\bar{I}; \mathbf{Y}^1).$$

The last equality in (2.5) is referred to as transversality condition. Convergence of $(\tau_\varepsilon, y_\varepsilon, u_\varepsilon, p_\varepsilon)$ as $\varepsilon \to 0^+$ was already analyzed in [17].

3 Sensitivity analysis for a parametric optimization problem

Throughout this section we fix $\varepsilon > 0$, and consider, for any $\tau > 0$ the minimization problem with respect to the variable u:

$$(P_\varepsilon^\tau) \qquad \begin{cases} \min\limits_{u\in U_{ad}} \tau\left(1+\dfrac{\varepsilon}{2}\|u\|^2_{L^2(I;L^2(\omega))}\right)+\dfrac{1}{2\varepsilon}\|\mathbf{y}(1)-\mathbf{z}\|^2_{\mathbf{Y}^0} \\ \text{subject to} \\ \mathbf{y}_t=\tau\mathbf{A}\mathbf{y}+\tau\mathbf{B}u \text{ on } (0,1], \\ \mathbf{y}(0)=\mathbf{y}_0. \end{cases}$$

Clearly this problem admits a unique solution $(\mathbf{y}_\tau, u_\tau) \in (C(\bar I;\mathbf{Y}^1)\cap C^1(\bar I;\mathbf{Y}^0))\times L^2(I;L^2(\omega))$ and the necessary and sufficient optimality condition is given by

$$(3.1) \qquad \begin{cases} \partial_t\mathbf{y}_\tau=\tau\mathbf{A}\mathbf{y}_\tau+\tau\mathbf{B}u_\tau, \quad \mathbf{y}_\tau(0)=\mathbf{y}_0 \\ -\partial_t\mathbf{p}_\tau=\tau\mathbf{A}^*\mathbf{p}_\tau,\ \mathbf{p}_\tau(1)=\dfrac{1}{\varepsilon}\begin{pmatrix}\mathbf{y}_{\tau,1}(1)-z_1\\(-\Delta)^{-1}(\mathbf{y}_{\tau,2}(1)-z_2)\end{pmatrix} \\ (\varepsilon u_\tau+\mathbf{B}^*\mathbf{p}_\tau, u-u_\tau)_{L^2(I;L^2(w))}\geq 0, \text{ for all } u\in U_{ad}. \end{cases}$$

Here we use the index τ for the solution to (P_ε^τ), since our attention focuses on the dependence of the solution on this parameter. We will investigate differentiability of $\tau \to (\mathbf{y}_\tau, u_\tau, \mathbf{p}_\tau)$ as well as first and second order differentiability of the value functional in (P_ε^τ) with respect to τ. Further we shall show that the derivative of this value function is given by the left hand side of the transversality condition, which is the last equation in (2.5). Some preliminaries will be required. Let us note that these results apparently cannot be directly derived from well-known abstract sensitivity results. This is due to the fact that these results require us to express the constraint in an abstract setting, in which the partial differential equation in (P_ε^τ) is transformed to an equality constraint of the form $e(y,u)=0$. The natural domain for e is $(L^2(I;\mathbf{Y}^1)\cap H^1(I;\mathbf{Y}^0)\cap\{\mathbf{y}:(y_2)_t-\tau\Delta y_1\in L^2(\Omega)\})\times L^2(I;L^2(\omega))$, which is a Banach space when endowed with the natural norm. However, this space depends on the the parameter τ. For this reason and due to the fact that the image space of the inequality constraints is infinite dimensional, general results appear not to be readily applicable. In this section we do not need to require (H1) to hold, but throughout we assume

(H2) $$\mathbf{y}_0 \in \mathbf{Y}^2.$$

With (H2) holding the solution to (P_ε^τ) satisfies the regularity

$$\mathbf{y}_\tau \in C(\bar I; \mathbf{Y}^2) \cap C^1(\bar I; \mathbf{Y}^1). \tag{3.2}$$

This can be verified with the same techniques that provide the regularity of $\mathbf{y}_\varepsilon$ in Theorem 2.3.

Proposition 3.1. *Let* (H2) *hold. Then for every compact subset* $J \subset (0,\infty)$ *the set* $\{(\mathbf{y}_\tau, u_\tau, \mathbf{p}_\tau) : \tau \in J\}$ *is bounded in* $\big(C(\bar I; \mathbf{Y}^2) \cap C^1(\bar I; \mathbf{Y}^1)\big) \times W^{1,\infty}(I; L^2(\omega)) \times \big(C(\bar I; \mathbf{P}^2) \cap C^1(\bar I; \mathbf{P}^1)\big)$ *and the mapping* $\tau \to (\mathbf{y}_\tau, u_\tau, \mathbf{p}_\tau)$ *is globally Lipschitz continuous from* J *to* $\big(C(\bar I; \mathbf{Y}^1) \cap W^{1,2}(I; \mathbf{Y}^0)\big) \times L^2(I; L^2(\omega)) \times \big(C(\bar I; \mathbf{P}^1) \cap W^{1,2}(I; \mathbf{P}^0)\big)$.

Proof. We first show the asserted uniform bound. Since $\{u_\tau\}_{\tau>0}$ is bounded in $L^2(I; L^2(\omega))$ it follows that $\{\mathbf{y}_\tau\}_{\tau>0}$ is bounded in $C(\bar I; \mathbf{Y}^1) \cap W^{1,2}(I; \mathbf{Y}^0)$ and $\{\mathbf{p}_\tau\}$ is bounded in $C(\bar I; \mathbf{P}^2) \cap C^1(\bar I; \mathbf{P}^1)$. Using that $u_\tau = -P_{U_{ad}}(\frac{1}{\varepsilon}\mathbf{p}_{\tau,2})$ it follows from Lemma 3.2 below, that $\{u_\tau\}_{\tau>0}$ is bounded in $W^{1,\infty}(I; L^2(\omega))$. This, together with $\mathbf{y}_0 \in \mathbf{Y}^2$ implies that $\{\mathbf{y}_\tau\}$ is bounded in $C(\bar I; \mathbf{Y}^2) \cap C^1(\bar I; \mathbf{Y}^1)$.

Let $\overline{\tau} \in J$ and $\tau \in J$ and set

$$(\delta\tau, \boldsymbol{\delta}\mathbf{y}, \delta u, \boldsymbol{\delta}\mathbf{p}) = (\overline{\tau} - \tau, \mathbf{y}_{\overline{\tau}} - \mathbf{y}_\tau, u_{\overline{\tau}} - u_\tau, \mathbf{p}_{\overline{\tau}} - \mathbf{p}_\tau),$$

where $(\mathbf{y}_\tau, u_\tau, \mathbf{p}_\tau)$ is the solution to (P_ε^τ) and analogously for $(\mathbf{y}_{\overline{\tau}}, u_{\overline{\tau}}, \mathbf{p}_{\overline{\tau}})$. Then we have

$$\begin{cases} \partial_t \boldsymbol{\delta}\mathbf{y} = \tau(\mathbf{A}\boldsymbol{\delta}\mathbf{y} + \mathbf{B}\delta u) + \delta\tau(\mathbf{A}\mathbf{y}_{\overline{\tau}} + \mathbf{B}u_{\overline{\tau}}),\ \boldsymbol{\delta}\mathbf{y}(0) = 0, \\[2ex] -\partial_t \boldsymbol{\delta}\mathbf{p} = \tau\mathbf{A}^*\delta\mathbf{p} + \delta\tau\mathbf{A}^*\mathbf{p}_{\overline{\tau}},\ \boldsymbol{\delta}\mathbf{p}(1) = \dfrac{1}{\varepsilon}\begin{pmatrix} \boldsymbol{\delta}\mathbf{y}_1(1) \\ (-\Delta)^{-1}\boldsymbol{\delta}\mathbf{y}_2(1) \end{pmatrix}, \\[2ex] (\varepsilon u_\tau + \mathbf{B}^*\mathbf{p}_\tau, u - u_\tau)_{L^2(I;L^2(w))} \ge 0, \\ (\varepsilon u_{\overline{\tau}} + \mathbf{B}^*\mathbf{p}_{\overline{\tau}}, u - u_{\overline{\tau}})_{L^2(I;L^2(w))} \ge 0, \quad \text{for all } \ u \in U_{ad}. \end{cases} \tag{3.3}$$

We note that the scalar form of the first equation in (3.3) is given by

$$\begin{cases} \partial_{tt}\boldsymbol{\delta}\mathbf{y}_1 = \tau^2(\Delta\boldsymbol{\delta}\mathbf{y}_1 + \chi_\omega\delta u) + \tau\,\delta\tau(\Delta\mathbf{y}_{\overline{\tau},1} + \chi_\omega u_{\overline{\tau}}) + \delta\tau\,\partial_t\mathbf{y}_{\overline{\tau},2} \\[2ex] \boldsymbol{\delta}\mathbf{y}_1(0) = 0,\ \partial_t\boldsymbol{\delta}\mathbf{y}_1(0) = \delta\tau y_2, \end{cases} \tag{3.4}$$

and for the second component we have $\tau\boldsymbol{\delta}\mathbf{y}_2 = \partial_t\boldsymbol{\delta}\mathbf{y}_1 - \delta\tau\,\mathbf{y}_{\bar\tau,2}$. Similarly for the second equation in (3.3) we find

(3.5)
$$\begin{cases} \partial_{tt}\boldsymbol{\delta}\mathbf{p}_2 = \tau^2\Delta\boldsymbol{\delta}\mathbf{p}_2 + \tau\delta\tau\Delta\mathbf{p}_{\overline{\tau},2} - \delta\tau\,\delta_t\mathbf{p}_{\overline{\tau},1} \\ \boldsymbol{\delta}\mathbf{p}_2(1) = \dfrac{1}{\varepsilon}(-\Delta)^{-1}\boldsymbol{\delta}\mathbf{y}_2(1),\ \partial_t\boldsymbol{\delta}\mathbf{p}_2(1) = -\dfrac{\tau}{\varepsilon}\boldsymbol{\delta}\mathbf{y}_1(1) - \dfrac{\delta\tau}{\varepsilon}(\mathbf{y}_{\varepsilon,1}(1) - z_1), \end{cases}$$

and for the first component we have $\tau\boldsymbol{\delta}\mathbf{p}_1 = -\partial_t\boldsymbol{\delta}\mathbf{p}_2 - \delta\tau\,\mathbf{p}_{\bar\tau,1}$.

From Theorem 2.1, the first two equations in (3.3) and (3.4), (3.5) there exists a constant K_1 independent of $\tau \in J$ such that

$$\begin{cases} \|\boldsymbol{\delta}\mathbf{y}\|_{C(\bar I;\mathbf{Y}^1)\cap W^{1,2}(I;\mathbf{Y}^0)} \le K_1(|\delta\tau| + \|\delta u\|_{L^2(I;L^2(\omega))}) \\ \|\boldsymbol{\delta}\mathbf{p}\|_{C(\bar I;\mathbf{P}^1)\cap W^{1,2}(I;\mathbf{P}^0)} \le K_1(|\delta\tau| + \|\boldsymbol{\delta}\mathbf{y}(1)\|_{\mathbf{Y}^0}) \end{cases} \tag{3.6}$$

Setting $u = u_{\overline{\tau}}$ and $u = u_\tau$ in the two inequalities of (3.3) we obtain, after rearranging terms,

$$\varepsilon\|\delta u\|^2_{L^2(I;L^2(\omega)} \le -(\mathbf{B}^*\boldsymbol{\delta}\mathbf{p}, \delta u)_{L^2(I;L^2(\omega))}. \tag{3.7}$$

Next we take the inner product in $L^2(I;L^2(\Omega))$ of the first equation in (3.1) with $\boldsymbol{\delta}\mathbf{p}$ and of the second equation with $\boldsymbol{\delta}\mathbf{y}$. After integration by parts and subtraction of the two resulting equations we obtain, using that $\boldsymbol{\delta}\mathbf{y}(0) = 0$

$$(\boldsymbol{\delta}\mathbf{y}(1), \boldsymbol{\delta}\mathbf{p}(1))_{L^2(\Omega)} = \tau(\mathbf{B}\,\delta u, \boldsymbol{\delta}\mathbf{p}) + \delta\tau\big((\mathbf{A}\mathbf{y}_{\overline{\tau}} + \mathbf{B}u_{\overline{\tau}}, \boldsymbol{\delta}\mathbf{p}) - (\mathbf{A}^*\mathbf{p}_{\overline{\tau}}, \boldsymbol{\delta}\mathbf{y})\big),$$

where the inner products on the right hand side are taken in $L^2(I;L^2(\Omega))$. The initial condition for $\boldsymbol{\delta}\mathbf{p}(1)$ and (3.7) imply that

$$\frac{1}{\varepsilon}\|\boldsymbol{\delta}\mathbf{y}(1)\|^2_{\mathbf{Y}^0} + \tau\varepsilon\|\delta u\|^2_{L^2(I;L^2(\omega))} \le \delta\tau\big(|(\mathbf{A}\mathbf{y}_{\overline{\tau}} + \mathbf{B}u_{\overline{\tau}}, \boldsymbol{\delta}\mathbf{p})| + |(\mathbf{A}^*\mathbf{p}_{\overline{\tau}}, \boldsymbol{\delta}\mathbf{y})|\big).$$

Using (3.6) and the estimates established at the beginning of the proof, we obtain the existence of a constant K_2 independent of $\tau \in J, \overline{\tau} \in J$ such that

$$\frac{1}{\varepsilon}\|\boldsymbol{\delta}\mathbf{y}(1)\|^2_{\mathbf{Y}^0} + \tau\varepsilon\|\delta u\|^2_{L^2(I;L^2(\omega))} \le K_2|\delta\tau|(\,\|\delta u\|_{L^2(I;L^2(\omega))} + |\delta\tau|),$$

which, in turn, implies the existence of a constant K_3 independent of $\tau \in J, \overline{\tau} \in J$, such that

$$\frac{1}{\varepsilon}\|\boldsymbol{\delta}\mathbf{y}(1)\|^2_{\mathbf{Y}^0} + \varepsilon\|\delta u\|^2_{L^2(I;L^2(\omega))} \le K_3|\delta\tau|^2.$$

The conclusion now follows from (3.6). □

Lemma 3.2. *The projection $P_{U_{ad}}$ is continuous from $C^1(\bar{I}; L^2(\omega))$ to $H^1(I; L^2(\omega))$.*

Proof. The proof follows the ideas of a related result in [13] on the continuity of the operator $y \mapsto \max(y, 0)$ in $H^1(\Omega)$. Let a sequence q_n be given, with $q_n \to q$ in $C^1(\bar{I}; L^2(\omega))$. We have for the time derivative

$$\partial_t P_U(q(t)) = \begin{cases} \partial_t q(t) & \text{if } \|q(t)\|_{L^2(\omega)} \le \gamma, \\ \gamma \frac{\|q(t)\|^2_{L^2(\omega)} \partial_t q(t) - (q(t), \partial_t q(t))_{L^2(\omega)} q(t)}{\|q(t)\|^3_{L^2(\omega)}} & \text{if } \|q(t)\|_{L^2(\omega)} > \gamma, \end{cases}$$

for almost every $t \in I$. Here we use that $(q(t), \partial_t q(t)) = 0$ for a.e. $t \in I$ such that $\|q(t)\|_{L^2(\omega)} = \gamma$, cf. e.g. [13, Lemma A.4].

It follows that $\partial_t P_U(q_n(t)) \to \partial_t P_U(q(t))$ converges pointwise for all $t \in I$ such that $\|q(t)\|_{L^2(\omega)} \neq \gamma$. The pointwise a.e. convergence of $\partial_t P_U(q_n(t))$ to $\partial_t P_U(q(t))$ for $t \in I$ with $\|q(t)\|_{L^2(\omega)} = \gamma$ follows from the fact that $(q(t), \partial_t q(t)) = 0$ holds for almost all such $t \in I$.

In addition we have the pointwise bound $\|\partial_t P_U(q_n(t))\|_{L^2(\omega)} \le \|\partial_t q\|_{C(\bar{I}; L^2(\omega))}$. Hence by Lebesgue dominated convergence theorem, $\partial_t P_U(q) \to \partial_t P_U(q)$ strongly in $L^2(I; L^2(\omega))$. □

This continuity result for the projection onto the set of admissible controls allows to conclude continuity of the mapping $\tau \mapsto (\mathbf{y}_\tau, u_\tau, \mathbf{p}_\tau)$ with respect to stronger norms.

Theorem 3.3. *Let* (H2) *hold. Then the mapping $\tau \to (\mathbf{y}_\tau, u_\tau, \mathbf{p}_\tau)$ is continuous at $\tau > 0$ with values in $\big(C(\bar{I}; \mathbf{Y}^2) \cap H^1(I; \mathbf{Y}^1)\big) \times H^1(I; L^2(\omega)) \times \big(C(\bar{I}; \mathbf{P}^2) \cap H^1(I; \mathbf{P}^2)\big)$.*

Proof. Due to the result of Proposition 3.1, we have the Lipschitz continuity of $\tau \mapsto \mathbf{p}_\tau$ with values in $C(\bar{I}; \mathbf{P}^1)$. This implies the Lipschitz continuity of $\tau \mapsto \mathbf{B}^*\mathbf{p}_\tau$ with values in $C^1(\bar{I}; L^2(\omega))$. Hence, by Lemma 3.2 the mapping $\tau \mapsto u_\tau = P_{U_{ad}}(-\frac{1}{\varepsilon}\mathbf{B}^*\mathbf{p}_\tau)$ is continuous with values in $H^1(\bar{I}; L^2(\omega))$. The result now follows with the regularity result of Theorem 2.1. □

We will now prove directional differentiability of the mapping $\tau \mapsto (\mathbf{y}_\tau,\ u_\tau,\ \mathbf{p}_\tau)$ directly, using the projection representation $u_\tau = P_{U_{ad}}\big(-\frac{1}{\varepsilon}\mathbf{B}^*\mathbf{p}_\tau\big)$, following the lines of a similar result in [6]. First, we will show Bouligand differentiability of the projection onto U_{ad}.

Recall that a function $f : X \to Y$ between normed linear spaces X and Y is called Bouligand differentiable if for every $x_0 \in X$ there exists $\varepsilon > 0$ and a positively homogeneous function $f'(x_0; \cdot) : X \to Y$ such that

$$f(x) = f(x_0) + f'(x_0; x - x_0) + r(x_0; x - x_0),$$

for all $x \in X$, where $\|r(x_0; x - x_0)\|_Y / \|x - x_0\|_X \to 0$ for $\|x - x_0\|_X \to 0$. In particular this implies that f is directionally differentiable.

Proposition 3.4. *The projection P_U onto U is Bouligand differentiable from $L^2(\omega)$ to $L^2(\omega)$. That is, for all $q, h \in L^2(\omega)$ we have*

$$\|P_U(q + h) - P_U(q) - P'_U(q; h)\|_{L^2(\omega)} = o(\|h\|_{L^2(\omega)})$$

uniformly on bounded subsets of $L^2(\omega)$, where the Bouligand derivative is given by

$$P'_U(q; h) = \begin{cases} h & \text{if } \|q\|_{L^2(\omega)} < \gamma, \\ & \text{or } \|q\|_{L^2(\omega)} = \gamma,\ (q, h) < 0, \\ \gamma \left(\frac{h}{\|q\|_{L^2(\omega)}} - \frac{(q,h)_{L^2(\omega)}}{\|q\|^3_{L^2(\omega)}} q \right) & \text{if } \|q\|_{L^2(\omega)} > \gamma, \\ & \text{or } \|q\|_{L^2(\omega)} = \gamma,\ (q, h) \geq 0. \end{cases}$$

Proof. Let $q, h \in L^2(\omega)$ be given. If $\|q\|_{L^2(\omega)} \neq \gamma$ then the claim follows immediately. In the case $\|q\|_{L^2(\omega)} = \gamma$ and $(q, h) \geq 0$ it holds $\|q + h\| \geq \gamma$, and the claim is a consequence of the Fréchet differentiability of $q \mapsto \gamma \frac{q}{\|q\|_{L^2(\omega)}}$, for $q \neq 0$. Moreover, the convergence order $o(\|h\|_{L^2(\omega)})$ is uniform with respect to q in bounded subsets of $L^2(\omega)$.

It remains to consider the case $\|q\|_{L^2(\omega)} = \gamma$ and $(q, h) < 0$. If in addition $\|q + h\| \leq \gamma$ holds then the first-order remainder term is zero. This leaves to discuss the case $\|q + h\|_{L^2(\omega)} > \gamma$. Here we find

$$\gamma^2 < \|q + h\|^2_{L^2(\omega)} = \gamma^2 + 2(q, h) + \|h\|^2_{L^2(\omega)},$$

which implies the inequality $0 < -2(q, h) < \|h\|^2_{L^2(\omega)}$. The first-order remainder term is now given as

$$\begin{aligned}
P_U(q+h) - P_U(q) - P_U'(q;h) &= \gamma \frac{q+h}{\|q+h\|_{L^2(\omega)}} - q - h \\
&= (q+h)\left(\frac{\gamma}{\|q+h\|_{L^2(\omega)}} - 1\right) \\
&= \frac{\gamma(q+h)}{\|q+h\|_{L^2(\omega)}(\gamma + \|q+h\|_{L^2(\omega)})}\left(\gamma^2 - (\gamma^2 + 2(q,h) + \|h\|^2_{L^2(\omega)})\right) \\
&= \frac{\gamma(q+h)}{\|q+h\|_{L^2(\omega)}(\gamma + \|q+h\|_{L^2(\omega)})}\left(-2(q,h) - \|h\|^2_{L^2(\omega)}\right),
\end{aligned}$$

where we use $a-b = \frac{a^2-b^2}{a+b}$. Due to the inequality $0 < -2(q,h) < \|h\|^2_{L^2(\omega)}$, the L^2-norm of the right hand side of this expression is of the order $o(\|h\|_{L^2(\omega)})$ for $h \to 0$ in $L^2(\omega)$, uniformly with respect to q in bounded subsets of $L^2(\omega)$. $\square$

It shall be noted that the directional derivative P_U' is itself a projection onto a convex set:

$$P_U'(q;h) = P_{T_U(q)}(h),$$

with

$$T_U(q) := \begin{cases} L^2(\omega) & \text{if } \|q\|_{L^2(\omega)} < \gamma, \\ \{h \in L^2(\Omega) :\ (h,q) \le 0\} & \text{if } \|q\|_{L^2(\omega)} = \gamma, \\ \{h \in L^2(\Omega) :\ (h,q) = 0\} & \text{if } \|q\|_{L^2(\omega)} > \gamma. \end{cases} \tag{3.8}$$

Proposition 3.5. *The projection $P_{U_{ad}}$ is Bouligand differentiable from $C(\bar{I}; L^2(\omega))$ to $L^2(I; L^2(\omega))$ with directional derivative*

$$(P_{U_{ad}}'(q;h))(t) = P_U'(q(t); h(t)) \quad \forall t \in \bar{I}.$$

Proof. Let $q \in C(\bar{I}; L^2(\omega))$ and a sequence $h_k \in C(\bar{I}; L^2(\omega))$ be given with $h_k \to 0$ in $C(\bar{I}; L^2(\omega))$. Let us define the sequence of first-order remainders

$$r_k := P_{U_{ad}}(q + h_k) - P_{U_{ad}}(q) - P_{U_{ad}}'(q; h_k).$$

We have to show $\|r_k\|_{L^2(I;L^2(\omega))} = o(\|h_k\|_{C(\bar{I};L^2(\omega))})$ as $\|h_k\|_{C(\bar{I};L^2(\omega))} \to 0$. Due to the result of Proposition 3.4, we have $\|r_k(t)\|_{L^2(\omega)} = o(\|h_k(t)\|_{L^2(\omega)})$ as $k \to \infty$. Moreover, due to the Lipschitz continuity of projections it holds that $\|r_k(t)\|_{L^2(\omega)} \le 2\|h_k(t)\|_{L^2(\omega)}$.

By the Lebesgue dominated convergence theorem, we have that the sequence q_k defined by $q_k(t) := \|h_k(t)\|^{-1}_{L^2(\omega)} \|r_k(t)\|_{L^2(\omega)}$ converges to zero in $L^2(I)$ for all $p \in [1, \infty)$. The claim follows with the inequality

$$\|r_k\|_{L^2(I;L^2(\omega))} \le \|h_k\|_{C(\bar{I};L^2(\omega))} \left(\int_I \|h_k(t)\|^{-2}_{L^2(\omega)} \|r_k(t)\|^2_{L^2(\omega)} \, dt \right)^{1/2}.$$

□

Theorem 3.6. *Let* (H2) *hold. Then the mapping* $\tau \mapsto (\mathbf{y}_\tau, u_\tau, \mathbf{p}_\tau)$ *is Bouligand-differentiable from* $\mathbb{R}^+$ *to* $C(\bar{I}; \mathbf{Y}^1) \times L^2(I; L^2(\omega)) \times C(\bar{I}; \mathbf{P}^1)$. *The directional derivative in direction* $\delta\tau$ *is given as the unique solution* $(\dot{\mathbf{y}}, \dot{u}, \dot{\mathbf{p}}) \in C(\bar{I}; \mathbf{Y}^1) \times L^2(I; L^2(\omega)) \times C(\bar{I}; \mathbf{P}^1)$ *of the system*

$$(3.9) \qquad \begin{cases} \partial_t \dot{\mathbf{y}} = \tau(\mathbf{A}\dot{\mathbf{y}} + \mathbf{B}\dot{u}) + \delta\tau(\mathbf{A}\mathbf{y}_\tau + \mathbf{B}u_\tau), \ \dot{\mathbf{y}}(0) = 0 \\ -\partial_t \dot{\mathbf{p}} = \tau \mathbf{A}^* \dot{\mathbf{p}} + \delta\tau \mathbf{A}\mathbf{p}_\tau, \ \dot{\mathbf{p}}(1) = \dfrac{1}{\varepsilon} \begin{pmatrix} \dot{\mathbf{y}}_1(1) \\ (-\Delta)^{-1}\dot{\mathbf{y}}_2(1) \end{pmatrix} \\ \dot{u} \in T_{U_{ad}(-\frac{1}{\epsilon}\mathbf{B}^*\mathbf{p}_\tau)} \\ (\varepsilon \dot{u} + \mathbf{B}^*\dot{\mathbf{p}}, v - \dot{u})_{L^2(I;L^2(\omega))} \ge 0 \quad \forall v \in T_{U_{ad}(-\frac{1}{\epsilon}\mathbf{B}^*\mathbf{p}_\tau)} \end{cases}$$

where $T_{U_{ad}(-\frac{1}{\epsilon}\mathbf{B}^*\mathbf{p}_\tau)}$ *is given by*

$$T_{U_{ad}(-\frac{1}{\epsilon}\mathbf{B}^*\mathbf{p}_\tau)} = \left\{ u \in L^2(I; L^2(\omega)) : u(t) \in T_U \left(-\frac{1}{\epsilon}\mathbf{B}^*\mathbf{p}_\tau(t) \right) \ \text{a.e. on } I \right\}.$$

with T_U *defined in* (3.8).

Proof. Let us define for abbreviation $T := T_{U_{ad}(-\frac{1}{\epsilon}\mathbf{B}^*\mathbf{p}_\tau)}$.

Let $\tau > 0$, $\delta\tau \in (-1, 1)$ with $\tau + \delta\tau > 0$ be given. At first, let us prove existence and uniqueness of solutions of the sensitivity system (3.9). This system is the first-order necessary optimality condition of the optimization problem

$$\min \tau \frac{\varepsilon}{2} \|u\|^2_{L^2(I;L^2(\omega))} + \frac{1}{2\varepsilon} \|\mathbf{y}(1)\|^2_{\mathbf{Y}^1} + \delta\tau \langle \mathbf{A}^*\mathbf{p}_\tau, \mathbf{y} \rangle_{L^2(I;\mathbf{P}^1), L^2(I;\mathbf{Y}^0)}$$

subject to $u \in T$ and

$$\partial_t \mathbf{y} = \tau(\mathbf{A}\mathbf{y} + \mathbf{B}u) + \delta\tau(\mathbf{A}\mathbf{y}_\tau + \mathbf{B}u_\tau), \ \mathbf{y}(0) = 0.$$

Since this problem is strictly convex with respect to u, it admits a unique solution. This proves that there exists a unique solution $(\dot{\mathbf{y}}, \dot{u}, \dot{\mathbf{p}})$ of the system (3.9).

Let us define

$$(\boldsymbol{\delta}\mathbf{y}, \delta u, \boldsymbol{\delta}\mathbf{p}) = (\mathbf{y}_{\tau+\delta\tau} - \mathbf{y}_\tau, u_{\tau+\delta\tau} - u_\tau, \mathbf{p}_{\tau+\delta\tau} - \mathbf{p}_\tau).$$

Then $(\boldsymbol{\delta}\mathbf{y}, \delta u, \boldsymbol{\delta}\mathbf{p})$ solves the system

$$\begin{cases} \partial_t \boldsymbol{\delta}\mathbf{y} = \tau(\mathbf{A}\boldsymbol{\delta}\mathbf{y} + \mathbf{B}\delta u) + \delta\tau(\mathbf{A}\mathbf{y}_{\tau+\delta\tau} + \mathbf{B}u_{\tau+\delta\tau}),\ \boldsymbol{\delta}\mathbf{y}(0) = 0, \\ -\partial_t \boldsymbol{\delta}\mathbf{p} = \tau\mathbf{A}^*\delta\mathbf{p} + \delta\tau\mathbf{A}^*\mathbf{p}_{\tau+\delta\tau},\ \boldsymbol{\delta}\mathbf{p}(1) = \dfrac{1}{\varepsilon}\begin{pmatrix} \boldsymbol{\delta}\mathbf{y}_1(1) \\ (-\Delta)^{-1}\boldsymbol{\delta}\mathbf{y}_2(1) \end{pmatrix}, \\ \delta u = P_{U_{ad}}\left(-\dfrac{1}{\varepsilon}\mathbf{B}^*\mathbf{p}_{\tau+\delta\tau}\right) - P_{U_{ad}}\left(-\dfrac{1}{\varepsilon}\mathbf{B}^*\mathbf{p}_\tau\right). \end{cases} \tag{3.10}$$

By the result of Proposition 3.5, we have

$$\delta u = P_{U_{ad}}\left(-\frac{1}{\varepsilon}\mathbf{B}^*\mathbf{p}_{\tau+\delta\tau}\right) - P_{U_{ad}}\left(-\frac{1}{\varepsilon}\mathbf{B}^*\mathbf{p}_\tau\right) = P_T\left(-\frac{1}{\varepsilon}\mathbf{B}^*\boldsymbol{\delta}\mathbf{p}\right) + r$$

with $\|r\|_{L^2(I;L^2(\omega))} = o(\|\boldsymbol{\delta}\mathbf{p}\|_{C(\bar{I};\mathbf{P}^0)})$ as $\|\boldsymbol{\delta}\mathbf{p}\|_{C(\bar{I};\mathbf{P}^0)} \to 0$. With Proposition 3.1 we have $\|r\|_{L^2(I;L^2(\omega))} = o(|\delta\tau|)$. Let us write this representation of δu as the variational inequality

$$(\varepsilon(\delta u - r) + \mathbf{B}^*\boldsymbol{\delta}\mathbf{p}, v - (\delta u - r))_{L^2(I;L^2(\omega))} \geq 0 \quad \forall v \in T.$$

Setting $v := \dot{u}$ in this inequality and $v := \delta u - r$ in the fourth relation of (3.9) and adding the resulting inequalities gives

$$\varepsilon\|\dot{u} - \delta u + r\|^2_{L^2(I;L^2(\omega))} \leq (\mathbf{B}^*(\dot{\mathbf{p}} - \boldsymbol{\delta}\mathbf{p}),\, \delta u - \dot{u})_{L^2(I;L^2(\omega))} - (\mathbf{B}^*(\dot{\mathbf{p}} - \boldsymbol{\delta}\mathbf{p}),\, r)_{L^2(I;L^2(\omega))},$$

which implies

$$\begin{aligned} \frac{\varepsilon}{2}\|\dot{u} - \delta u\|^2_{L^2(I;L^2(\omega))} &\leq \varepsilon\|r\|^2_{L^2(I;L^2(\omega))} \\ &\quad - (\mathbf{B}^*(\dot{\mathbf{p}} - \boldsymbol{\delta}\mathbf{p}),\, r)_{L^2(I;L^2(\omega))} + (\mathbf{B}^*(\dot{\mathbf{p}} - \boldsymbol{\delta}\mathbf{p}),\, \delta u - \dot{u})_{L^2(I;L^2(\omega))}. \end{aligned}$$

Proceeding as in the proof of Proposition 3.1 to estimate the last term we find

$$(3.11)\quad \tau\frac{\varepsilon}{2}\|\dot{u}-\delta u\|^2_{L^2(I;L^2(\omega))}+\frac{1}{\varepsilon}\|\dot{\mathbf{y}}(1)-\boldsymbol{\delta}\mathbf{y}(1)\|^2_{\mathbf{Y}^0}$$
$$\leq \tau\|r\|^2_{L^2(I;L^2(\omega))}-\tau(\mathbf{B}^*(\dot{\mathbf{p}}-\boldsymbol{\delta}\mathbf{p}),\, r)_{L^2(I;L^2(\omega))}$$
$$+\,\delta\tau\langle\mathbf{A}\boldsymbol{\delta}\mathbf{y}+\mathbf{B}\delta u,\,\dot{\mathbf{p}}-\boldsymbol{\delta}\mathbf{p}\rangle_{L^2(I;\mathbf{Y}^0),L^2(I;\mathbf{P}^1)}$$
$$+\,\delta\tau\langle\mathbf{A}^*\boldsymbol{\delta}\mathbf{p},\,\dot{\mathbf{y}}-\boldsymbol{\delta}\mathbf{y}\rangle_{L^2(I;\mathbf{P}^0),L^2(I;\mathbf{Y}^1)}.$$

We shall return to this estimate below. Now we introduce $\mathbf{w}=\boldsymbol{\delta}\mathbf{y}-\dot{\mathbf{y}}$ and note that

$$\partial_t\mathbf{w}=\tau(\mathbf{A}\mathbf{w}+\mathbf{B}(\delta u-\dot{u}))+\delta\tau(\mathbf{A}\boldsymbol{\delta}\mathbf{y}+\mathbf{B}\,\delta u),\ \boldsymbol{\delta}\mathbf{w}(0)=0.$$

As in the proof of Proposition 3.1, see (3.4), we next write this equation in its coordinates:

$$\begin{cases}\partial_{tt}\mathbf{w}_1=\tau^2(\Delta\mathbf{w}_1+\chi_\omega(\delta u-\dot{u}))+\tau\,\delta\tau(\Delta(\boldsymbol{\delta}\mathbf{y}_1)+\chi_\omega\delta u)+\delta\tau\,\partial_t(\boldsymbol{\delta}\mathbf{y})_2\\ \mathbf{w}_1(0)=0,\ \partial_t\mathbf{w}_1(0)=0,\end{cases}$$

and

$$\tau\mathbf{w}_2=\partial_t\mathbf{w}_1-\delta\tau(\boldsymbol{\delta}\mathbf{y})_2.$$

By Theorem 2.1, there is a constant $K>0$ independent of $\delta\tau$ such that

$$\|\mathbf{w}_1\|_{C(\bar{I};H^1_0(\Omega))}+\|\partial_t\mathbf{w}_1\|_{C(\bar{I};L^2(\Omega))}\leq K\big(\|\delta u-\dot{u}\|_{L^2(I;L^2(\omega))}$$
$$+\,\delta\tau(\|(\boldsymbol{\delta}\mathbf{y})_1\|_{L^2(I;H^2(\Omega))}+\|(\boldsymbol{\delta}\mathbf{y})_2\|_{H^1(I;L^2(\Omega))}+\|\delta u\|_{L^2(I;L^2(\omega))})\big).$$

This further implies that

$$\|\mathbf{w}_2\|_{C(\bar{I};H^1_0(\Omega))}\leq K\big(\|\delta u-\dot{u}\|_{L^2(I;L^2(\omega))}$$
$$+\ \delta\tau(\|\boldsymbol{\delta}\mathbf{y}\|_{L^2(I;\mathbf{Y}^2)}+\|\boldsymbol{\delta}\mathbf{y}\|_{H^1(I;\mathbf{Y}^1)}+\|\delta u\|_{L^2(I;L^2(\omega))})\big),$$

and hence

$$\|\boldsymbol{\delta}\mathbf{y}-\dot{\mathbf{y}}\|_{C(\bar{I};\mathbf{Y}^1)}=\|\mathbf{w}\|_{C(\bar{I};\mathbf{Y}^1)}\leq K\big(\|\delta u-\dot{u}\|_{L^2(I;L^2(\omega))}$$
$$+\ \delta\tau(\|\boldsymbol{\delta}\mathbf{y}\|_{L^2(I;\mathbf{Y}^2)}+\|\boldsymbol{\delta}\mathbf{y}\|_{H^1(I;\mathbf{Y}^1)}+\|\delta u\|_{L^2(I;L^2(\omega))})\big).$$

Applying the continuity result of Theorem 3.3 to the right hand side results in

$$\|\boldsymbol{\delta}\mathbf{y} - \dot{\mathbf{y}}\|_{C(\bar{I};\mathbf{Y}^1)} \le K\|\delta u - \dot{u}\|_{L^2(I;L^2(\omega))} + o(|\delta\tau|), \tag{3.12}$$

for $\delta\tau \to 0$. In an analogous manner we estimate

$$\begin{aligned} \|\boldsymbol{\delta}\mathbf{p} - \dot{\mathbf{p}}\|_{C(\bar{I};\mathbf{P}^1)} &\le K(\varepsilon^{-1}\|\boldsymbol{\delta}\mathbf{y} - \dot{\mathbf{y}}\|_{C(\bar{I};\mathbf{Y}^0)} + \delta\tau(\|\boldsymbol{\delta}\mathbf{p}\|_{L^2(I;\mathbf{P}^2)} + \|\boldsymbol{\delta}\mathbf{p}\|_{H^1(I;\mathbf{P}^1)})) \\ &\le K\|\delta u - \dot{u}\|_{L^2(I;L^2(\omega))} + o(|\delta\tau|) \end{aligned} \tag{3.13}$$

with $K > 0$ independent of $\delta\tau$. We continue to estimate the critical expression on the right-hand side of (3.11) by using (3.12), (3.13), and the Lipschitz continuity result of Proposition 3.1:

$$\begin{aligned} &\big|\langle \mathbf{A}\boldsymbol{\delta}\mathbf{y} + \mathbf{B}\delta u,\, \dot{\mathbf{p}} - \boldsymbol{\delta}\mathbf{p}\rangle_{L^2(I;\mathbf{Y}^0),L^2(I;\mathbf{P}^1)} + \langle \mathbf{A}^*\boldsymbol{\delta}\mathbf{p},\, \dot{\mathbf{y}} - \boldsymbol{\delta}\mathbf{y}\rangle_{L^2(I;\mathbf{P}^0),L^2(I;\mathbf{Y}^1)}\big| \\ &\qquad \le (\|\boldsymbol{\delta}\mathbf{y}\|_{L^2(I;\mathbf{Y}^1)} + \|\delta u\|_{L^2(I;L^2(\omega))})\|\boldsymbol{\delta}\mathbf{p} - \dot{\mathbf{p}}\|_{L^2(\bar{I};\mathbf{P}^1)} \\ &\qquad\quad + \|\boldsymbol{\delta}\mathbf{p}\|_{L^2(\bar{I};\mathbf{P}^1)}\|\boldsymbol{\delta}\mathbf{y} - \dot{\mathbf{y}}\|_{L^2(\bar{I};\mathbf{Y}^1)} \\ &\qquad \le K|\delta\tau|\,\|\delta u - \dot{u}\|_{L^2(I;L^2(\omega))} + o(|\delta\tau|^2). \end{aligned} \tag{3.14}$$

By (3.11), (3.13), (3.14), we get the estimate

$$\|\dot{u} - \delta u\|^2_{L^2(I;L^2(\omega))} \le K|\delta\tau|^2\,\|\delta u - \dot{u}\|_{L^2(I;L^2(\omega))} + o(|\delta\tau|^2),$$

with K independent of $\delta\tau$ but dependent on ε. This proves $\|\dot{u}-\delta u\|_{L^2(I;L^2(\omega))} = o(|\delta\tau|)$ as $\delta\tau \to 0$. Hence $\|u_{\tau+\delta\tau} - u_\tau - \dot{u}\|_{L^2(I;L^2(\omega))} = o(|\delta\tau|)$, holds, and by (3.12) and (3.13) we obtain analogous estimates for the state and adjoint variables. This ends the proof. ◻

Corollary 3.7. *For any $\tau \in (0, \infty)$ and $\delta\tau$ the Bouligand derivative $\dot{u}$ satisfies*

$$(\varepsilon u_\tau + \mathbf{B}^*\mathbf{p}_\tau, \dot{u})_{L^2(I;L^2(\omega))} = 0. \tag{3.15}$$

Proof. Setting $u = u_{\tau+\sigma\delta\tau}$ in the first inequality of (3.3), dividing by $\sigma > 0$ and taking $\sigma \to 0^+$, we find $(\varepsilon u_\tau + \mathbf{B}^*\mathbf{p}_\tau, \dot{u})_{L^2(I;L^2(\omega))} \ge 0$. From the second inequality with u_τ replaced by $u_{\bar{\tau}}$ and $u = u_{\tau+\sigma\delta\tau}$ it follows that $(\varepsilon u_\tau + \mathbf{B}^*\mathbf{p}_\tau, \dot{u})_{L^2(I;L^2(\omega))} \le 0$. Combined this implies (3.15). ◻

Let $\mathcal{V}$ denote the value functional associated to (P_ε^τ), i.e.

$$\mathcal{V}(\tau) = \tau\left(1 + \frac{\varepsilon}{2}\|u_\tau\|^2_{L^2(I;L^2(\omega))}\right) + \frac{1}{2\varepsilon}\|\mathbf{y}_\tau(1) - \mathbf{z}\|_{\mathbf{Y}^0}, \tag{3.16}$$

and set $\mathbf{L}^2(\Omega) = L^2(\Omega) \times L^2(\Omega)$. The following result establishes smoothness of $\mathcal{V}$ and the relationship of $\frac{d}{d\tau}\mathcal{V}(\tau)$ to the transversality condition.

Theorem 3.8. *Let* (H1) *hold. Then the mapping $\tau \to \mathcal{V}(\tau)$ is continuously differentiable on* $(0, \infty)$ *and*

$$\frac{d}{d\tau}\mathcal{V}(\tau) = 1 + \frac{\varepsilon}{2}\|u_\tau\|^2_{L^2(I;L^2(\omega))} + (\mathbf{A}\mathbf{y}_\tau + \mathbf{B}u_\tau, \mathbf{p}_\tau)_{L^2(I;\mathbf{L}^2(\Omega))}.$$

Moreover, it admits a second order directional derivative given by

$$(\frac{d}{d\tau})^2\mathcal{V}(\tau)\delta\tau = \langle \mathbf{A}\mathbf{y}_\tau + \mathbf{B}u_\tau,\, \dot{\mathbf{p}}\rangle_{L^2(I;\mathbf{Y}^0),L^2(I;\mathbf{P}^1)} + \langle \mathbf{A}\dot{\mathbf{y}},\, \mathbf{p}_\tau\rangle_{L^2(I;\mathbf{Y}^0),L^2(I;\mathbf{P}^1)}.$$

Proof. Let $\tau \in (0, \infty)$ and let $(\dot{\mathbf{y}}, \dot{u}, \dot{\mathbf{p}})$ Bouligand derivative at τ in direction $\delta\tau$. To save notation it will be convenient to set

$$b = \delta\tau(1 + \frac{\varepsilon}{2}\|u_\tau\|^2_{L^2(I;L^2(\omega))}) + \tau\varepsilon(u_\tau, \dot{u})_{L^2(I;L^2(\omega))}.$$

We compute

$$\begin{aligned}
\frac{d}{d\tau}\mathcal{V}(\tau)\delta\tau &= \delta\tau b + \frac{1}{\varepsilon}(\mathbf{y}_\tau(1) - \mathbf{z}, \dot{\mathbf{y}}(1))_{\mathbf{Y}^0} \\
&= b + (\mathbf{p}_\tau(1), \dot{\mathbf{y}}(1))_{\mathbf{L}^2(\Omega)} \\
&= b + \int_0^1 \frac{d}{dt}(\mathbf{p}_\tau(t), \dot{\mathbf{y}}(t))_{L^2(\Omega)} + (\mathbf{p}_\tau(0), \dot{\mathbf{y}}(0))_{L^2(\Omega)} \\
&= b - \tau\int_0^1 (\mathbf{A}^*\mathbf{p}_\tau(t), \dot{\mathbf{y}}(t))_{\mathbf{L}^2(\Omega)} + \int_0^1 (\mathbf{p}_\tau(t), \tau\mathbf{A}\dot{\mathbf{y}}(t) + \tau\mathbf{B}\dot{u}(t) + \delta\tau\mathbf{A}\mathbf{y}_\tau(t) \\
&\qquad + \delta\tau\mathbf{B}u_\tau(t))_{L^2(\Omega)} \\
&= \delta\tau\big[(1 + \frac{\varepsilon}{2}\|u_\tau\|^2_{(I;L^2(\Omega))}) + \tau(\varepsilon u_\tau + \mathbf{B}^*\mathbf{p}_\tau, \dot{u})_{(I;L^2(\Omega))} + \int_0^1 (\mathbf{p}_\tau(t), \mathbf{A}\mathbf{y}_\tau(t) \\
&\qquad + \mathbf{B}u_\tau(t))_{\mathbf{L}^2(\Omega)}\big].
\end{aligned}$$

By (3.15) we have $(\varepsilon u_\tau + \mathbf{B}^* \mathbf{p}_\tau, \dot{u})_{(I;L^2(\Omega))} = 0$ and hence

$$\frac{d}{dt}\mathcal{V}(\tau) = 1 + \frac{\varepsilon}{2}\|u_\tau\|^2_{L^2(I;L^2(\omega))} + (\mathbf{p}_\tau, \mathbf{A}\mathbf{y}_\tau + \mathbf{B}u_\tau)_{L^2(I;\mathbf{L}^2(\omega))}.$$

This provides the expression for $\frac{d}{dt}\mathcal{V}(\tau)$. Continuity of $\tau \to \frac{d}{dt}\mathcal{V}(\tau)$ follows from Proposition 3.1 using the fact that $(\mathbf{p}_\tau, \mathbf{A}\mathbf{y}_\tau + \mathbf{B}u_\tau)_{L^2(I;\mathbf{L}^2(\omega))} = \langle \mathbf{p}_\tau, \mathbf{A}\mathbf{y}_\tau + \mathbf{B}u_\tau \rangle_{L^2(I;\mathbf{P}^1),L^2(I;\mathbf{Y}^0)}$ implies the first claim.

Using it, we obtain for the second-order directional derivative

$$\left(\frac{d}{d\tau}\right)^2 \mathcal{V}(\tau)\,\delta\tau = (\varepsilon u_\tau + \mathbf{B}^*\mathbf{p}_\tau, \dot{u}) \\ + \langle \mathbf{A}\mathbf{y} + \mathbf{B}u, \dot{\mathbf{p}} \rangle_{L^2(I;\mathbf{Y}^0),L^2(I;\mathbf{P}^1)} + \langle \mathbf{A}\dot{\mathbf{y}}, \mathbf{p} \rangle_{L^2(I;\mathbf{Y}^0),L^2(I;\mathbf{P}^1)}.$$

Using once again (3.15) implies the claim. □

Remark 3.9. The expression for $\frac{d}{dt}\mathcal{V}$ coincides with the left hand side of the transversality condition. Note that due to extra regularity requirement that $\mathbf{y}_0 \in \mathbf{Y}^2$ we have that $\mathbf{A}\mathbf{y} \in L^2(I; \mathbf{L}^2(\Omega))$, so that we can avoid the duality pairing that we used in the transversality condition in (2.5).

We conclude this section with an asymptotic estimate of $\mathcal{V}(\tau)$ as $\tau \to \infty$.

Theorem 3.10. *Let as assume that* (H1) *is satisfied, which is the null controllability of wave equation in time* $\tau_0 > 0$. *Then there exist constants* $\varepsilon_0 > 0$ *and* $c > 0$ *independent of* τ *such that for all* $\tau > \tau_0$ *and* $\varepsilon \in (0, \varepsilon_0)$ *we have*

$$|\mathcal{V}(\tau) - \tau| \le c\,\epsilon\,(\tau - \tau_0)^{-1},$$

$$\left|\frac{d}{d\tau}\mathcal{V}(\tau) - 1\right| \le c\,(\tau - \tau_0)^{-1/2}.$$

Proof. Let $\sigma > 0$ and a positive integer N be given. From [17, Proposition 2.3] it follows that if

$$N \ge 2c(\tau_0)\sigma^{-1}\max(\|\mathbf{y}_0\|_{\mathbf{Y}^1}, \|\mathbf{z}\|_{\mathbf{Y}^1}) \tag{3.17}$$

holds there exists controls $u_{N,0}$ and $u_{N,\mathbf{z}}$ with the property $\|u_{N,0}\|_{L^\infty(I;L^2(\omega))}$, $\|u_{N,\mathbf{z}}\|_{L^\infty(I;L^2(\omega))} \le \sigma/2$ that drive the wave equation from $\mathbf{y}_0$ to 0 and from 0 to $\mathbf{z}$, respectively. Let now $\tau > \tau_0$ be given, and define the positive integer N such that $\tau \in [N\tau_0, (N+1)\tau_0)$. Set $\sigma := 2c(\tau_0)\max(\|\mathbf{y}_0\|_{\mathbf{Y}^1}, \|\mathbf{z}\|_{\mathbf{Y}^1})N^{-1}$, hence (3.17) holds. Then by linearity of the wave equation the controls $u_{N,0}$ and $u_{N,\mathbf{z}}$ can be used to construct a control $u_{\tau,0,\mathbf{z}}$ that drives the system from 0 to $\mathbf{z}$ in time τ. Moreover, it holds

$$\|u_{\tau,0,\mathbf{z}}\| \leq \sigma = C\,N^{-1} \leq C\,(\tau - \tau_0)^{-1},$$

with a constant C independent of τ. Then for τ sufficiently large, $u_{\tau,0,\mathbf{z}}$ is admissible for (P_ε^τ) and hence

$$\tau\frac{\varepsilon}{2}\|u_\tau\|^2_{L^2(I;L^2(\omega))} + \frac{1}{2\varepsilon}\|\mathbf{y}_\tau(1) - \mathbf{z}\|^2_{\mathbf{Y}^0} \leq \tau\frac{\varepsilon}{2}\|u_{\tau,0,\mathbf{z}}\|^2_{L^2(I;L^2(\omega))} \leq c\,\varepsilon\,(\tau - \tau_0)^{-1}.$$

This proves the first claim.

This estimate further implies that

$$\|\mathbf{p}_\tau(1)\|_{\mathbf{P}^1} = \frac{1}{\epsilon}\|\mathbf{y}_\tau(1) - \mathbf{z}\|_{\mathbf{Y}^0} \leq c\,(\tau - \tau_0)^{-1/2}, \tag{3.18}$$

where c is used as a generic constant, independent of τ.

Testing the adjoint equations, $-\mathbf{p}'_{\tau,1} = \tau\Delta\mathbf{p}_2$ and $-\mathbf{p}_{\tau,2} = \tau\mathbf{p}_{\tau,1}$ by $\mathbf{p}'_{\tau,2}$ and $-\mathbf{p}'_{\tau,1}$, respectively, subtracting the resulting equations, and integrating on $(t,1)$, $t \in (0,1)$, gives the energy equation

$$\|\mathbf{p}_{\tau,1}(t)\|^2_{L^2(\Omega)} + \|\nabla\mathbf{p}_{\tau,2}(t)\|^2_{L^2(\Omega)} = \|\mathbf{p}_{\tau,1}(1)\|^2_{L^2(\Omega)} + \|\nabla\mathbf{p}_{\tau,2}(1)\|^2_{L^2(\Omega)}$$

for all $t \in (0,1)$ and $\tau > 0$. Together with (3.18) this proves

$$\|\mathbf{p}_\tau\|_{L^\infty(I;\mathbf{P}^1)} \leq c\,(\tau - \tau_0)^{-1/2}.$$

Analogously one obtains the estimate

$$\|\mathbf{y}_\tau\|_{L^\infty(\mathbf{Y}^1)} \leq c\left(\|\mathbf{y}_0\|_{\mathbf{Y}^1} + \|u_\tau\|_{L^2(I;L^2(\omega))}\right),$$

with $c > 0$ independent of τ, which shows that $\mathbf{A}\mathbf{y}_\tau$ is bounded in $L^\infty(I;\mathbf{Y}^0)$ uniformly with respect to τ. This implies that

$$\sup_{t\in I}\langle \mathbf{A}\mathbf{y}_\tau(t) + \mathbf{B}u_\tau(t),\, \mathbf{p}_\tau(t)\rangle_{\mathbf{Y}^0,\mathbf{P}^1} \leq c\,(\tau - \tau_0)^{-1/2},$$

which, together with Theorem 3.8 and (3.18), proves the second claim. ∎

4 Semi-smooth Newton algorithm for the regularized optimality system associated to (P_ε^τ)

The algorithm for solving (P_ε), which will be described in the following section, relies on an efficient numerical method to solve (P_ε^τ). For this purpose we use a semi-smooth Newton method. In this section we verify its local superlinear convergence.

In view of (3.1) a control u, which depends on τ and ε, is a solution to (P_ε^τ), if and only if

$$\mathcal{F}u = u + P_{U_{ad}}\left(\frac{1}{\varepsilon}\chi_\omega \mathbf{p}_{\tau,2}\right) = 0, \tag{4.1}$$

where $\mathbf{p}_\tau = \mathbf{p}_\tau(u)$ is defined through the primal and adjoint equations in (3.1). Here $\mathcal{F}$ is considered as an operator from $L^2(I; L^2(\omega))$ to itself. We shall verify below that $P_{U_{ad}} : C(\bar{I}; L^2(\omega)) \to L^2(I; L^2(\omega))$ is Newton differentiable [9] with Newton derivative denoted by $DP_{U_{ad}}$. Since $u \to \mathbf{y}_\tau \to \mathbf{p}_{\tau,2}$ is a continuous linear mapping from $L^2(I; L^2(\omega)) \to C(\bar{I}; L^2(\omega))$, it follows that $u \to \mathcal{F}u$ is Newton differentiable. To carry out the Newton iteration, given u_0, we compute $u_{k+1} = u_k + \delta u$ by solving

$$D\mathcal{F}(u_k)\,\delta u = -\mathcal{F}(u_k), \tag{4.2}$$

where $D\mathcal{F}(u) = I + DP_{U_{ad}}$.

To analyze this algorithm we first address Newton differentiability of radial projections in $L^2(\omega)$ and $L^2(I; L^2(\omega))$. The Newton derivative differently is — in contrast to the Bouligand derivative, which was considered in Proposition 3.4 — a linear operator.

Proposition 4.1. *The projection $P_U : L^2(\omega) \to L^2(\omega)$ given by $P_U(q) = q\min(1, \frac{\gamma}{\|q\|_{L^2(\omega)}})$ is Newton differentiable with Newton derivative given by*

$$DP_U(q)h = \begin{cases} h & \text{if } \|q\|_{L^2(\omega)} \le \gamma \\ \dfrac{\gamma h}{\|q\|_{L^2(\omega)}} - \dfrac{\gamma q(q,\, h)_{L^2(\omega)}}{\|q\|^3_{L^2(\omega)}} & \text{if } \|q\|_{L^2(\omega)} > \gamma. \end{cases} \tag{4.3}$$

Proof. We need to argue that

$$\lim_{\|h\|\to 0} \frac{1}{\|h\|}\|P(q+h) - P_U(q) - DP_U(q+h)h\| = 0$$

for any $q \in L^2(\omega)$, see e.g. [9, page 261]. Throughout the proof $\|\cdot\|$ stands for $\|\cdot\|_{L^2(\omega)}$.

Let us first consider the mapping $G : L^2(\omega) \to \mathbb{R}$ given by

$$G(q) = \min(1, \frac{\gamma}{\|q\|}).$$

Then

$$DG(q)(h) = \begin{cases} 0 & \text{if } \|q\| \le \gamma \\ -\frac{\gamma}{\|q\|^3}(q, h)_{L^2} & \text{if } \|q\| > \gamma \end{cases} \tag{4.4}$$

is a Newton derivative for G at any $q \in L^2(\omega)$. Since G is Fréchet differentiable for every u with $\|q\| \neq \gamma$ it suffices to consider the case $\|q\| = \gamma$. Let $\{h_n\}$ be an arbitrary sequence in $L^2(\omega)$ converging to 0. It has subsequences h_n^1 and h_n^2 with $\|q + h_n^1\| \le \gamma$ and $\|q + h_n^2\| > \gamma$. For the first we have

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{\|h_n^1\|} |G(q + h_n^1) - G(q) - DG(q + h_n^1)\, h_n^1| \\ &= \lim_{n\to\infty} \frac{1}{\|h_n^1\|} (1 - 1 - 0) = 0. \end{aligned} \tag{4.5}$$

For the second one we find

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{\|h_n^2\|} |G(q + h_n^2) - G(q) - DG(q + h_n^2) h_n^2| \\ &= \lim_{n\to\infty} \frac{\gamma}{\|h_n^2\|} \left(\frac{1}{\|q + h_n^2\|} - \frac{1}{\|q\|} + \frac{1}{\|q\|^3}(q, h_n^2) - \frac{1}{\|q\|^3}(q, h_n^2) + \frac{1}{\|q + h_n^2\|}(q + h_n^2, h_n^2) \right) \\ &= \lim_{n\to\infty} \frac{\gamma}{\|h_n^2\|} \left(o\,(\|h_n^2\|) - (\frac{1}{\|q\|^3} - \frac{1}{\|q + h_n^2\|^3})(q, h_n^2) + \frac{1}{\|q + h_n^2\|^3}(h_n^2, h_n^2) \right) = 0. \end{aligned} \tag{4.6}$$

Here we used that for $q \neq 0$

$$\frac{1}{\|q + h\|} - \frac{1}{\|q\|} + \frac{1}{\|q\|^3}(q, h) = o\,(\|h\|^2). \tag{4.7}$$

In the proof of the following proposition we shall use the fact that o in (4.7) is uniform with respect to q in sets of the form $\{q : \rho \le \|q\| \le R\}$, where $0 < \rho < R$.

With (4.6) Newton differentiability of G holds. Newton differentiability of P_U is now an easy consequence. In fact, we have

(4.8)

$$
\begin{aligned}
& P_U\,(q+h) - P_U\,(q) - DP_U\,(q+h)\,h \\
&= (q+h)\,\min\left(1,\,\frac{\gamma}{\|q+h\|}\right) - q\,\min\left(1,\,\frac{\gamma}{\|q+h|}\right) + q\,\big(\min(1,\,\frac{\gamma}{\|q+h\|}) \\
&\qquad - \min\,(1,\,\frac{\gamma}{\|u\|})\big) - h\,\min\left(1,\,\frac{\gamma}{\|q+h\|}\right) - q\,DG\,(q+h)\,h \\
&= q\,(G(q+h) - G(q) - DG\,(q+h)\,h) = o\,(\|h\|).
\end{aligned}
$$

□

Remark 4.2. Note that in (4.4) we assigned the value 0 to $DG(q)$ for q at the critical value $\|q\| = \gamma$. This corresponds to the equality sign in (4.3) which appears in the first, rather than the second line.

Similarly to considerations above, one can prove that the mapping

$$
\tilde{D}G(q)(h) = \begin{cases} 0 & \text{if } \|q\| < \gamma \\ -\frac{\gamma}{\|q\|^3}(q,\,h)_{L^2} & \text{if } \|q\| \geq \gamma \end{cases}
$$

is a Newton derivative of G too. Here, the proof has to be modified in the case $\|q\| = \gamma$ only for a subsequence h_n^3 with $\|q + h_n^3\| = \gamma$. Then one has

$$
0 = \|q + h_n^3\|^2 - \|q\|^2 = \|h_n^3\|^2 + 2(q, h_n^3),
$$

which implies $(q, h_n^3) = -\frac{1}{2}\|h_n^3\|^2$. Then (4.5) has to be replaced by

$$
\begin{aligned}
& \lim_{n\to\infty} \frac{1}{\|h_n^3\|} |G(q+h_n^3) - G(q) - DG(q+h_n^3)\,h_n^3| \\
&= \lim_{n\to\infty} \frac{1}{\|h_n^3\|}(1 - 1 + \frac{1}{\gamma^2}(q + h_n^3, h_n^3)) = \frac{1}{2\gamma^2} \lim_{n\to\infty} \frac{1}{\|h_n^3\|}(h_n^3, h_n^3) = 0.
\end{aligned}
$$

For our purpose the projection operates pointwise in time on elements $q \in C(\bar{I};\, L^2(\omega))$. We denote it by the same symbol.

Proposition 4.3. *The projection* $P_{U_{ad}} : C(\bar{I};\, L^2(\omega)) \to L^2(I;\, L^2(\omega))$ *given by* $P_{U_{ad}}(q)(t) = q(t)\min\,(1,\,\frac{\gamma}{\|q(t)\|_{L^2(\omega)}})$ *is Newton differentiable with Newton derivative given by* $(DP_U)q(t)$ *as in* (4.3).

Proof. Let $q \in C(\bar{I};\, L^2(\omega))$ Further let $\{h_n\}$ be a sequence in $C(I;\, L^2(\omega))$ with $h_n \to 0$ in $C(I;\, L^2(\omega))$. Let

$$K_1 = \{t : \|q(t)\| < \gamma\},\ K_2 = \{t : \|q(t)\| = \gamma\},\ \text{and } K_3 = \{t : \|q(t)\| \geq \gamma\}.$$

We need to estimate

$$\lim_{n\to\infty}(\int_0^1 \|P_U(q(t)+h_n(t)) - P_U(q(t)) - DP_U(q(t)+h_n(t))h_n(t)\|^2 dt)^{\frac{1}{2}}$$

$$\leq \lim_{n\to\infty}\big((\int_{K_1} q(t)dt)^{\frac{1}{2}} + (\int_{K_2} q(t)dt)^{\frac{1}{2}} + (\int_{K_3} q(t)dt)^{\frac{1}{2}}\big),$$

where $z(t)$ denotes the integrand of the integral on the left hand side. By Lebesgue's bounded convergence theorem

$$\lim_{n\to\infty}\int_{K_1} z(t)dt = \int_{K_1}\lim_{n\to\infty}\|q(t)+h_n(t)-q(t)-h_n(t)\|^2 dt = 0$$

and hence $\lim_{n\to\infty}\frac{1}{\|h_n\|_{C(I;L^2(\omega))}}\left(\int_{K_1} z(t)dt\right)^{\frac{1}{2}} = 0$. Similarly by (4.7)

$$\lim_{n\to\infty}\frac{1}{\|h_n\|_{C(I;L^2(\omega))}}\left(\int_{K_3} z(t)dt\right)^{\frac{1}{2}} = \lim_{n\to\infty}\frac{1}{\|h_n\|_{C(I;L^2(\omega))}}\left(\int_{K_3} o\left(\|h_n(t)\|^4\right)dt\right)^{\frac{1}{2}} = 0.$$

To obtain the estimate on K_2 we express h_n as $h_n(t) = h_n^1(t) + h_n^2(t)$, where

$$h_n^1(t) = \begin{cases} h_n(t) \text{ if } \|q(t)+h_n(t)\| \leq \gamma \\ 0 \text{ if } \|q(t)+h_n(t)\| > \gamma, \end{cases} \qquad h_n^2(t) = \begin{cases} 0 \text{ if } \|q(t)+h_n(t)\| \leq \gamma \\ h_n(t) \text{ if } \|q(t)+h_n(t)\| > \gamma. \end{cases}$$

By (4.8) we find using a splitting similar to (4.6)

$$\left(\int_{K_2} z(t)dt\right)^{\frac{1}{2}} = \left(\int_{K_2}\Big(G(q(t)+h_n(t)) - G(q(t)) - DG(q(t)+h_n(t))h_n(t)\Big)^2 \|q(t)\|^2 dt\right)^{\frac{1}{2}}$$

$$= \gamma\left(\int_{K_2}\Big(G(q(t)+h_n^2(t)) - G(q(t)) - DG(q_n(t)+h_n^2(t))h_n^2(t)\Big)^2 dt\right)^{\frac{1}{2}}$$

$$\leq \gamma^2\left(\int_{K_2}\Big(\frac{1}{\|q(t)+h_n^2(t)\|} - \frac{1}{\|q(t)\|} + \frac{1}{\|q(t)\|^3}(q(t),\, h_n^2(t))\Big)^2 dt\right)^{\frac{1}{2}}$$

$$+\gamma^2\left(\int_{K_2}\left(\left(\frac{1}{\|q(t)\|^3} - \frac{1}{\|q(t)+h_n^2(t)\|^3}\right)(q(t),\, h_n^2(t))\right)^2 dt\right)^{\frac{1}{2}}$$

$$+\gamma^2\left(\int_{K_2}\frac{1}{\|q(t)+h_n^2(t)\|^6}\|h_n^2(t)\|^4 dt\right)^{\frac{1}{2}}$$

$$\leq \gamma^2 \left(\int_{K_2} o\left(h_n^2(t)\right)^2 dt \right)^{\frac{1}{2}} + \frac{1}{\gamma^3} \left(\int_{K_2} \|h_n^2(t)\|^4 \left(\|q(t)\|^2 + \|q(t)\| \cdot \|h_n^2(t)\| + \|h_n^2(t)\|^2 \right)^2 dt \right)^{\frac{1}{2}}$$

$$+ \frac{1}{\gamma} \left(\int_{K_2} \|h_n^2(t)\|^4 dt \right)^{\frac{1}{2}} = o\left(\|h_n\|_{C(I;L^2(\omega))}\right),$$

and consequently $\lim_{n\to\infty} \frac{1}{\|h_n\|_{C(I;L^2(\omega))}} \left(\int_{K_2} z(t)dt \right)^{\frac{1}{2}} = 0$. This concludes the proof. □

Let us reconsider (4.2) and set $\mathbf{p}_k = \mathbf{p}\,(u_k)$, the solution to the first two equations in (3.1) with u_τ replaced by u_k. Let us further note that the Fréchet derivative of $u \to \mathbf{p}\,(u)$ at u_k in direction δ_u, denoted by $\mathbf{p}' = \mathbf{p}'(u_k)\,\delta_u$ satisfies

$$\text{(4.9)} \qquad \begin{cases} \partial_t \mathbf{y}' = \tau \mathbf{A} \mathbf{y}' + \tau \mathbf{B}\,\delta u, & \mathbf{y}'(0) = 0 \\ -\,\partial_t \mathbf{p}' = \tau \mathbf{A}^* \mathbf{p}', & \mathbf{p}'(1) = \dfrac{1}{\varepsilon} \begin{pmatrix} \mathbf{y}_1'(1) \\ (-\Delta)^{-1} \mathbf{y}_2'(1) \end{pmatrix}. \end{cases}$$

We further set

$$\mathcal{I} = \{t \in I : \|\mathbf{p}_{k,2}(t)\|_{L^2(\omega)} \leq \varepsilon\gamma\} \text{ and } \mathcal{A} = \{t \in I : \|\mathbf{p}_{k,2}(t)\|_{L^2(\omega)} > \varepsilon\gamma\}.$$

Then (4.2) can equivalently be expressed as

$$\delta u + \frac{\gamma \chi_{\mathcal{A}} \chi_\omega}{\|\mathbf{p}_{k,2}\|} \left(\mathbf{p}_2' - \frac{\mathbf{p}_{k,2}}{\|\mathbf{p}_{k,2}\|} \left(\frac{\mathbf{p}_{k,2}}{\|\mathbf{p}_{k,2}\|}, \mathbf{p}_2' \right)_{L^2(\mathcal{A};L^2(\omega))} \right) + \frac{1}{\varepsilon} \chi_{\mathcal{I}} \chi_\omega \mathbf{p}_2' = -u_k - P_{U_{ad}} \left(\frac{1}{\varepsilon} \chi_\omega \mathbf{p}_{k,2} \right),$$

where the norms are taken in $L^2(I; L^2(\omega))$. Setting $u_{k+1} = u_k + \delta u$ the Newton update can be expressed as

(4.10)

$$u_{k+1} + \frac{1}{\varepsilon} \chi_{\mathcal{I}} \chi_\omega \mathbf{p}_{k+1,2} + \frac{\gamma \chi_{\mathcal{A}} \chi_\omega}{\|\mathbf{p}_{k,2}\|} \left(\mathbf{p}_{k+1,2} - \frac{\mathbf{p}_{k,2}}{\|\mathbf{p}_{k,2}\|} \left(\frac{\mathbf{p}_{k,2}}{\|\mathbf{p}_{k,2}\|}, \mathbf{p}_{k+1,2} - \mathbf{p}_{k,2} \right)_{L^2(I;L^2(\omega))} \right) = 0.$$

Proposition 4.4. *The Newton update u_{k+1} is the unique solution to*

(4.11)

$$\begin{cases} \displaystyle\min_{u \in U_{ad}} \tilde{J}(\mathbf{y}, u) = \frac{\tau}{2} \int_0^1 \max\left(\frac{\|\mathbf{p}_{k,2}(t)\|_{L^2(\omega)}}{\gamma}, \varepsilon \right) \|u(t)\|^2 \, dt + \frac{1}{2\varepsilon} \|\mathbf{y}(1) - \mathbf{z}\|^2_{\mathbf{Y}^0} \\ \textit{subject to} \\ \mathbf{y}_t = \tau \mathbf{A} \mathbf{y} + \tau \mathbf{B} u \ \textit{on } (0,1], \quad \mathbf{y}(0) = \mathbf{y}_0, \\ (u(t), \mathbf{p}_{k,2}(t))_{L^2(\omega)} = -\gamma \|\mathbf{p}_{k,2}(t)\|_{L^2(\omega)} \ \textit{for a.e. } t \in \mathcal{A}. \end{cases}$$

Proof. Existence of an unique solution to (4.11) follows from the linear quadratic structure of the problem. To derive the necessary and sufficient optimality system for (4.11) we consider the Lagrangian associated to (4.11), which is given by

$$\mathcal{L}(\mathbf{y},u,\mathbf{p},\mu) = \tilde{J}(\mathbf{y},u)+\langle \mathbf{p}, -\mathbf{y}_t + \tau\mathbf{A}\mathbf{y} + \tau\mathbf{B}u\rangle \\ + \big(\mu(\cdot),\, (u(\cdot), \mathbf{p}_{k,2}(\cdot))_{L^2(\omega)} + \gamma\|\mathbf{p}_{k,2}(\cdot)\|_{L^2(\omega)}\big)_{L^2(\mathcal{A},\mathbb{R})}.$$

It follows that the adjoint equation for (4.11) is given by

$$-\partial_t \mathbf{p} = \tau A^* \mathbf{p}, \quad \mathbf{p}(1) = \begin{pmatrix} \mathbf{y}_1(1) - z_1 \\ (-\Delta)^{-1}(\mathbf{y}_2(1) - z_2) \end{pmatrix}$$

and that

$$\tau \max\left(\frac{\|\mathbf{p}_{k,2}(t)\|_{L^2(\omega)}}{\gamma}, \varepsilon\right) u + \tau\mathbf{p}_2 + \chi_{\mathcal{A}}\, \mu\, \mathbf{p}_{k,2} = 0 \text{ on } (0,1)\times\omega.$$

Consequently

$$u + \frac{1}{\varepsilon}\mathbf{p}_2 = 0 \text{ on } \mathcal{I}\times\omega, \tag{4.12}$$

and

$$\frac{\|\mathbf{p}_{k,2}\|}{\gamma} u + \mathbf{p}_2 + \frac{1}{\tau}\mu\mathbf{p}_{k,2} = 0 \text{ on } \mathcal{A}\times\omega. \tag{4.13}$$

The latter equation implies that

$$\frac{\|\mathbf{p}_{k,2}(t)\|_{L^2(\omega)}}{\gamma}(u(t), \mathbf{p}_{k,2}(t))_{L^2(\omega)} + (\mathbf{p}_2(t), \mathbf{p}_{k,2}(t))_{L^2(\omega)} + \frac{1}{\tau}\mu(t)\|\mathbf{p}_{k,2}(t)\|^2_{L^2(\omega)} = 0.$$

By the last equality in (4.11)

$$-\|\mathbf{p}_{k,2}\|^2_{L^2(\omega)} + (\mathbf{p}_2(t), \mathbf{p}_{k,2}(t))_{L^2(\omega)} + \frac{1}{\tau}\mu(t)\|\mathbf{p}_{k,2}(t)\|^2_{L^2(\omega)} = 0$$

and hence

$$\mu(t) = \tau - \frac{\tau}{\|\mathbf{p}_{k,2}(t)\|^2_{L^2(\omega)}}(\mathbf{p}_2(t), \mathbf{p}_{k,2}(t))_{L^2(\omega)} \text{ for } t \in \mathcal{A}.$$

By (4.13) we find for a.e. $t \in \mathcal{A}$

$$u(t) + \frac{\gamma}{\|\mathbf{p}_{k,2}(t)\|_{L^2(\omega)}} \left(\mathbf{p}_2(t) + \mathbf{p}_{k,2}(t) \left(1 - \frac{(\mathbf{p}_2(t), \mathbf{p}_{k,2}(t))_{L^2(\omega)}}{\|\mathbf{p}_{k,2}(t)\|^2_{L^2(\omega)}} \right) \right) = 0,$$

or equivalently for $t \in \mathcal{A}$

(4.14)

$$u(t)+\frac{\gamma \chi_\omega}{\|\mathbf{p}_{k,2}(t)\|_{L^2(\omega)}} \left(\mathbf{p}_2(t) - \frac{\mathbf{p}_{k,2}(t)}{\|\mathbf{p}_{k,2}(t)\|_{L^2(\omega)}} \left(\mathbf{p}_2(t) - \mathbf{p}_{k,2}(t), \frac{\mathbf{p}_{k,2}(t)}{\|\mathbf{p}_{k,2}(t)\|_{L^2(\omega)}} \right)_{L^2(\omega)} \right).$$

Combined, (4.12) and (4.14) imply that (4.10) is satisfied with $(u_{k+1}, \mathbf{p}_{k+1}) = (u, \mathbf{p})$. ☐

Uniform boundedness of the inverses of the generalized derivatives of F is addressed next.

Proposition 4.5. *For every $u \in L^2(I; L^2(\omega))$*

$$\|D\mathcal{F}(u)^{-1}\|_{\mathcal{L}(I;L^2(\omega))} \leq 1.$$

Proof. Let $\mathbf{p} = \mathbf{p}(u)$ denote the solution to the adjoint equation as in (P_ε^τ) with u_τ replaced by u and let $\mathbf{p}' = \mathbf{p}'(h)$ denote the linearization of $u \to \mathbf{p}(u)$ at u in direction $h \in L^2(I; L^2(\omega))$, i.e. $\mathbf{p}'$ satisfies (4.9) with $\delta u = h$.
We set

$$\mathcal{A} = \{t \in I : \|\mathbf{p}_2(t)\|_{L^2(\omega)} > \varepsilon\gamma\}, \quad \mathcal{I} = \{t \in I : \|\mathbf{p}_2(t)\|_{L^2(\omega)} \leq \varepsilon\gamma\},$$

and we define

$$q(t) = \frac{\mathbf{p}_2(t)}{\|\mathbf{p}_2(t)\|_{L^2(\omega)}} \chi_\omega \chi_\mathcal{A}.$$

We have the orthogonal decomposition

$$h = h_1 + h_2 = (h, q)_{L^2(I;L^2(\omega))} q + (h - (h, q)_{L^2(I;L^2(\omega))} q).$$

According to Proposition 4.3 we find

$$D\mathcal{F}(u)h = h + \frac{1}{\varepsilon} \chi_\omega \chi_\mathcal{I} \mathbf{p}_2' + \frac{\gamma \chi_\omega \chi_\mathcal{A}}{\|\mathbf{p}_2'\|_{L^2(\omega)}} (\mathbf{p}_2' - \frac{\mathbf{p}_2 (\mathbf{p}_2, \mathbf{p}_2')_{L^2(\omega)}}{\|\mathbf{p}_2\|^2_{L^2(\omega)}}),$$

where dependence on t is suppressed. We note that $(\mathbf{p}_2' - \frac{\mathbf{p}_2(\mathbf{p}_2,\mathbf{p}_2')_{L^2(\omega)}}{\|\mathbf{p}_2\|_{L^2(\omega)}}, h_1)_{L^2(\omega)} = 0$. Consequently we have

$$(DF(u)h, h)_{L^2(I,L^2(\omega))} = |h|^2_{L^2(I;L^2(\omega))} + \frac{1}{\varepsilon}(\mathbf{p}_2'(h), h)_{L^2(I;L^2(\omega))}. \tag{4.15}$$

We find

$$\begin{aligned}\tau(\mathbf{p}_2'(h), h)_{L^2(I;L^2(\omega))} &= \langle \mathbf{p}(h), \partial_t \mathbf{y}' - \tau \mathbf{A}\mathbf{y}'\rangle \\ &= (\mathbf{p}'(h)(1), \mathbf{y}'(1)) + \langle \partial_t \mathbf{p}'(h), \mathbf{y}'\rangle = 2(\mathbf{p}'(h)(1), \mathbf{y}'(1)) \\ &= \frac{2}{\varepsilon}(\|\mathbf{y}_1'(1)\|^2_{L^2(\Omega)} + \|\mathbf{y}_2'(1)\|^2_{H^{-1}(\Omega)}) \geq 0\end{aligned}$$

Together with (4.15) this implies the claim. □

We are now prepared to verify local convergence of the semi-smooth Newton iteration (4.2).

Theorem 4.6. *Assume that $\|u_0 - u_\tau\|_{L^2(I;L^2(\omega))}$ is sufficiently small. Then the semi-smooth Newton iteration* (4.2) *converges superlinearly to u_τ in $L^2(I; L^2(\omega))$.*

Proof. The mapping $\mathcal{F}$ is continuous and by Proposition 4.3 is also Newton differentiable at every $u \in L^2(I; L^2(\omega))$. By Proposition 4.5 the inverses of $D\mathcal{F}(u)$ are uniformly bounded. The conclusion therefore follows from well-known results on semi-smooth Newton methods, cf. [9, page 238]. □

5 A semi-smooth Newton algorithm

In this section a Newton-type algorithm for solving (P_ε) is proposed. Minimization with respect to $(\tau, \mathbf{y}, u, \mathbf{p})$ is carried out in two nested loops, an outer τ- and an inner $(\mathbf{y}, u, \mathbf{p})$-loop. Key ingredients of the method are provided by Theorem 3.8 stating that the derivative of the value functional $\mathcal{V}$ is given by the transversality condition, and that the second order directional derivative can be obtained from the sensitivity system. Moreover for each given τ, the optimal control problem (P_ε^τ) can be solved efficiently by a semi-smooth Newton method. Around these two loops which solve (P_ε) there is an outer loop, within which the regularization parameter ϵ is driven to 0. Thus we consider (P_ε) for a sequence $\{\varepsilon_i\}$ with $\lim_{i\to\infty} \varepsilon_i = 0$. Of course, this regularization also plays the role of a penalty parameter with respect to realizing the endpoint condition, and in practice these two parameters can be decoupled.

5.1 Description of the algorithm

The basic building block is a Newton algorithm for the minimization of the functional $\mathcal{V}$, defined in (3.16) by

$$\mathcal{V}(\tau) = \tau\left(1 + \frac{\varepsilon}{2}\|u_\tau\|^2_{L^2(I;L^2(\omega))}\right) + \frac{1}{2\varepsilon}\|\mathbf{y}_\tau(1) - \mathbf{z}\|^2_{\mathbf{Y}^0},$$

where $(\mathbf{y}_\tau, u_\tau)$ is the unique solution of the parametric optimization problem (P^τ_ε). The first and second derivatives of $\mathcal{V}$ are computed on the basis of Theorems 3.6 and 3.8. Each evaluation of $\mathcal{V}$ is realized by a semi-smooth Newton step to solve (P^τ_ε) as explained in section 4. The resulting Newton-type algorithm is embedded in an outer loop that drives the parameter ϵ to zero. The overall solution procedure is depicted in Algorithm 1.

Algorithm 1 Outline of the solution algorithm

Initial guess: ϵ_0, τ_0.
Parameters: $\theta \in (0,1)$, $r_{\text{tol}} > 0$.
Initialization: $k := 0$, $i = 0$, $r_0 := +\infty$.
repeat {Solve (P) - ε-loop}
  Solve (P^τ_ε), solution $(\mathbf{y}_{\tau_k}, u_{\tau_k}, \mathbf{p}_{\tau_k})$.
  repeat {Solve (P_ε) - τ-loop}
    Solve sensitivity system (3.9)
    Compute $s_k := \frac{d}{d\tau}\mathcal{V}(\tau_k)$ and $h_k := \frac{d^2}{d\tau^2}\mathcal{V}(\tau_k)$
    Set $\sigma := 2$
    repeat {Line-search in τ}
      Set $\sigma := \sigma/2$.
      Set $\tilde{\tau}_k := \tau_k - \sigma h_k^{-1} s_k$.
      Solve (P^τ_ε), solution $(\mathbf{y}_{\tilde{\tau}_k}, u_{\tilde{\tau}_k}, \mathbf{p}_{\tilde{\tau}_k})$.
    until $\mathcal{V}(\tau_k) - \mathcal{V}(\tilde{\tau}_k) \geq \theta\sigma|s_k|^2$
    Set $(\tau_{k+1}, \mathbf{y}_{\tau_{k+1}}, u_{\tau_{k+1}}, \mathbf{p}_{\tau_{k+1}}) := (\tilde{\tau}_k, \mathbf{y}_{\tilde{\tau}_k}, u_{\tilde{\tau}_k}, \mathbf{p}_{\tilde{\tau}_k})$.
    Set $k := k+1$.
  until $|\frac{d}{d\tau}\mathcal{V}(\tau_k)| < \varepsilon_i$.
  Compute $r_{i+1} := \|\mathbf{y}_{\tau_k}(1) - \mathbf{z}\|_{\mathbf{Y}^0}$.
  Set $\varepsilon_{i+1} := \varepsilon_i/2$.
  Set $i = i+1$.
until $r_t < r_{\text{tol}}$

Some remarks are in order. For the innermost semi-smooth Newton iteration that solves the parametric optimization problem (P^τ_ε) an Armijo-type linesearch is used. The iteration is stopped as soon as the residual at the current iterate $u_{\tau_k}^i$ satisfies

$$\|\mathcal{F}\left(u_{\tau_k}^i\right)\|_{L^2(I;L^2(\omega))} \leq \max\left(10^{-2}\left|\frac{d}{d\tau}\mathcal{V}(\tau_{k-1})\right|,\, 10^{-6}\right),$$

see (4.1) for the definition of $\mathcal{F}$.

To apply a non-smooth Newton's method for the minimization of $\mathcal{V}$, the second order directional derivative of $\mathcal{V}$ is computed by solving the sensitivity system (3.9). We observed that the computational effort to solve (3.9) is comparable to the effort of computing one Newton step for (P_ε^τ). Once the solution of (3.9) is available, the second order directional derivative of $\mathcal{V}$ is obtained from

$$(\frac{d}{d\tau})^2\mathcal{V}(\tau)\delta\tau = \langle \mathbf{A}\mathbf{y} + \mathbf{B}u, \dot{\mathbf{p}}\rangle_{L^2(I;\mathbf{Y}^0),L^2(I;\mathbf{P}^1)} + \langle \mathbf{A}\dot{\mathbf{y}}, p\rangle_{L^2(I;\mathbf{Y}^0),L^2(I;\mathbf{P}^1)}.$$

As stopping criteria for the solution of the regularized problem (P_{ε_i}) we chose $|\frac{d}{d\tau}\mathcal{V}(\tau_k)| < \varepsilon_i$. That is, the precision in which the middle loop is solved is directly coupled to the regularization parameter. This prevents the algorithm to spend too much time to solve problems with relatively large regularization parameter.

Turning to the discretization scheme, the state-, adjoint-, and control variables were discretized by finite elements. The amplitude components $\mathbf{y}_1$ and $\mathbf{p}_2$ are discretized by P1-elements, while P0- elements are used for the velocity unknowns $\mathbf{y}_2$ and $\mathbf{p}_1$, and for the controls.

For time discretization, we used a cg(1)dG(0)-scheme as described in [16], which corresponds to a Crank-Nicolson time-stepping procedure.

5.2 Numerical experiments

Let us report on the outcome of our computational experiments. We modified the cost functional of the penalized problem (P_ε) to

$$J_\varepsilon(\tau, u) = \tau\left(1 + \frac{\varepsilon}{2}\|u\|^2_{L^2(I;L^2(\omega))}\right) + \frac{1}{2\varepsilon_y}\|\mathbf{y}(1) - \mathbf{z}\|^2_{\mathbf{Y}^0}$$

with

$$\varepsilon_y = 0.1 \cdot \varepsilon.$$

That is, the violation of the terminal constraint is penalized with larger weight.

We chose $\Omega = (0,1)^2$ for our computations. The control bound was set to $\gamma = 3$. The target state was $z_1 = z_2 = 0$ and the initial state was given as $y_1(x_1, x_2) = x_1 x_2 (1 - x_1)(1 - x_2)$.

The spatial domain was discretized using a uniform triangulation, and the time interval was split into equidistant subintervals. We will report on the results for the following hierarchy of discretizations: $(N, M) = (50, 10)$, $(200, 20)$, $(800, 40)$, and $(3200, 80)$, where N is the numbers of triangles and M the numbers of time intervals. The resulting mesh size h is $h = 2/\sqrt{N}$, the resulting length of the temporal subintervals $\Delta t = 1/M$.

The parameters for the algorithm as described in the previous section were chosen as

$$\varepsilon_0 = 0.1, \quad \tau_0 = 1.9, \quad \theta = 10^{-3}.$$

The algorithm was stopped as soon as the terminal residual satisfies $\|\mathbf{y}_\varepsilon(1) - \mathbf{z}\|_{\mathbf{Y}^0} \le 10^{-3}$, which corresponds to the choice $r_{\text{tol}} = 10^{-3}$.

In the examples which follow, we also pay special attention to the behavior of the boundedness of $\{\|\mathbf{p}_{\varepsilon,h}(1)\|_{\mathbf{P}^1}\}$ and $\{\|\mathbf{p}_{\varepsilon,h}(1)\|_{\mathbf{P}^0}\}$ with respect to ε. As shown in [17], if $\{\|\mathbf{p}_\varepsilon(1)\|_{\mathbf{P}^1}\}$ is bounded, we can pass to the limit in the point-wise form of the last equation of (2.5) to obtain for cluster points $u, \mathbf{p}$ of $\{u_\varepsilon\}$, $\{\mathbf{p}_\varepsilon\}$

$$1 + \langle \mathbf{A}\mathbf{z} + \mathbf{B}u(1), \mathbf{p}(1) \rangle_{(\mathbf{P}^0)^*, \mathbf{P}^0} = 0, \tag{5.1}$$

where u is a solution of the unregularized problem (P). If $\{\|\mathbf{p}_\varepsilon(1)\|_{\mathbf{P}^0}\}$ is bounded and u is bang-bang, then again (5.1) holds. In case that $\{\|\mathbf{p}_\varepsilon(1)\|_{\mathbf{P}^0}\}$ is unbounded, then for weak limits $\tilde{\mathbf{p}}$ of the normalized adjoints $\tilde{\mathbf{p}}_\varepsilon = \frac{\mathbf{p}_\varepsilon}{\|\mathbf{p}_\varepsilon(1)\|_{\mathbf{P}^0}}$ we have $\langle \mathbf{A}\mathbf{z} + \mathbf{B}u(1),\, \tilde{\mathbf{p}}(1) \rangle_{(\mathbf{P}^0)^*, \mathbf{P}^0} = 0$. For the details we refer to [17].

Let us briefly comment on our experience with combining the minimization with respect to τ and to $(\mathbf{y}, u)$ in one semi-smooth Newton iteration. It turned out that the bilevel optimization analyzed here is clearly more robust for the problems that we tested. For this reason we gave preference to it in this work.

Example 1

Here the initial velocity was $y_2 = 0$. That is, the control objective was to steer the system from a given deflection into zero. The control domain was chosen to be $\omega := \Omega$.

The most delicate issue is the behavior of the algorithm with respect to the asymptotic behavior $\varepsilon \to 0^+$. So let us report on it first.

In Figure 1 we show the convergence of τ_ε for different discretizations. Moreover, we depict the evolution of $|\tau_\varepsilon - \tau^*|$ for the finest discretization, where we use as value for τ^* the optimal time for the smallest ε_i, i.e. $\tau^* = \tau_{\varepsilon_{i_{\min}}}$. In Table 1 we report on the convergence of τ_ε and $\|\mathbf{y}_\varepsilon(1) - \mathbf{z}\|_{\mathbf{Y}^0}$ for the finest discretization. We observe the convergence rate

$$|\tau_\varepsilon - \tau^*| = O(\varepsilon).$$

ε	τ_ε	$\|\tau_\varepsilon - \tau^*\|$	$\|\mathbf{y}_\varepsilon(1) - \mathbf{z}\|_{\mathbf{Y}^0}$
$1.0000 \cdot 10^{-1}$	$3.68 \cdot 10^{-1}$	$1.52 \cdot 10^{-2}$	$2.49 \cdot 10^{-2}$
$5.0000 \cdot 10^{-2}$	$3.77 \cdot 10^{-1}$	$7.03 \cdot 10^{-3}$	$1.03 \cdot 10^{-2}$
$2.5000 \cdot 10^{-2}$	$3.80 \cdot 10^{-1}$	$3.56 \cdot 10^{-3}$	$4.48 \cdot 10^{-3}$
$1.2500 \cdot 10^{-2}$	$3.82 \cdot 10^{-1}$	$1.82 \cdot 10^{-3}$	$2.12 \cdot 10^{-3}$
$6.2500 \cdot 10^{-3}$	$3.83 \cdot 10^{-1}$	$6.67 \cdot 10^{-4}$	$1.02 \cdot 10^{-3}$
$3.1250 \cdot 10^{-3}$	$3.84 \cdot 10^{-1}$		$5.10 \cdot 10^{-4}$

Table 1: Convergence history for Example 1

As argued in Corollary 3.3. of [17] this implies that

$$\|\mathbf{y}_\varepsilon(1) - \mathbf{z}\|_{\mathbf{Y}^0} = O(\varepsilon),$$

which can be seen in Table 1 as well. Moreover, since by (2.5) we have $\|\mathbf{p}_\varepsilon(1)\|_{\mathbf{P}^1} = \frac{1}{\varepsilon}\|\mathbf{y}_\varepsilon(1) - \mathbf{z}\|_{\mathbf{Y}^0}$ this convergence rate implies that $\{\mathbf{p}_\varepsilon(1)\}$ is bounded in $\mathbf{P}^1$, where $\mathbf{p}_\varepsilon$ are the solutions of the undiscretized problem, see also [17, Remark 4.5]. To address the effect of discretization we plotted in Figure 2 the evolution of $\|\mathbf{p}_{\varepsilon,h}(1)\|_{\mathbf{P}^1}$ for the solutions of the discrete problems for the 4 different discretizations, and we observed that the $\mathbf{P}^1$-norms of $\mathbf{p}_{\varepsilon,h}(1)$ are bounded uniformly with respect to ε and with respect to the discretization. This suggests that the continuous norms $\{\|\mathbf{p}_\varepsilon(1)\|_{\mathbf{P}^1}\}$ are bounded as well.

Table 2 depicts the behavior of the number of iterations of the middle and the innermost loops of Algorithm 1, i.e. the τ and the semi-smooth Newton loops, as a function of a decreasing sequence of ε_i values. Except for the solution for the initial ε- value, the number of τ iterations increases as ε_i decreases. The number of semi-smooth Newton iterations necessary for each fixed value of ε and τ remains almost constant for the example shown here.

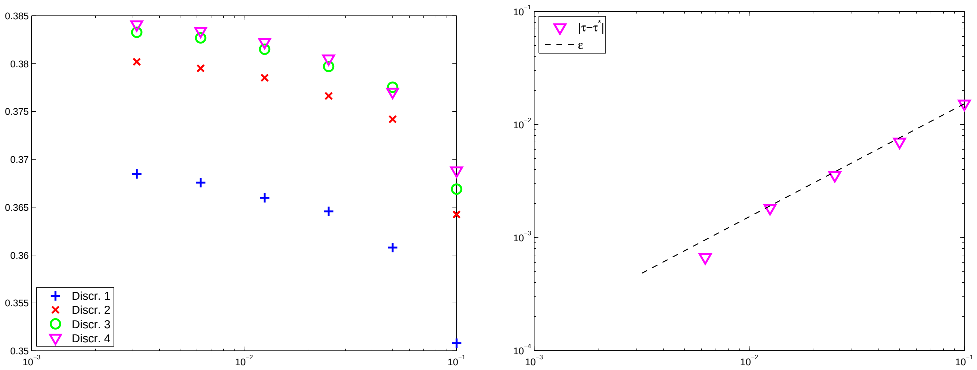

Figure 1: (Example 1) τ_ε vs. ε for different discretizations; $|\tau_\varepsilon - \tau^*|$ vs. ε for finest discretization

ε	τ-iterations	ssn-iterations
$1.0000 \cdot 10^{-1}$	9	31
$5.0000 \cdot 10^{-2}$	1	10
$2.5000 \cdot 10^{-2}$	2	17
$1.2500 \cdot 10^{-2}$	25	241
$6.2500 \cdot 10^{-3}$	20	266
$3.1250 \cdot 10^{-3}$	30	470

Table 2: (Example 1) Iteration numbers

Example 2

In our second example we chose the initial velocity to be $y_2(x_1, x_2) = x_1^2 + y_2^2$. All other data are the same as in the previous example. The convergence history can be found in Table 3. We observed the same convergence rates as in the previous example, i.e. $\|\mathbf{y}_\varepsilon(1) - \mathbf{z}\|_{\mathbf{Y}^0} = O(\varepsilon)$. Moreover, $\{\mathbf{p}_\varepsilon(1)\}$ is bounded in $\mathbf{P}^1$, which indicates that strict transversality holds for the continuous original problem.

Example 3

Here again the initial velocity was $y_2 = 0$, however, the control domain ω was chosen to be a proper subset of Ω: $\omega = \Omega \setminus [0, 0.5]^2$. For this control domain,

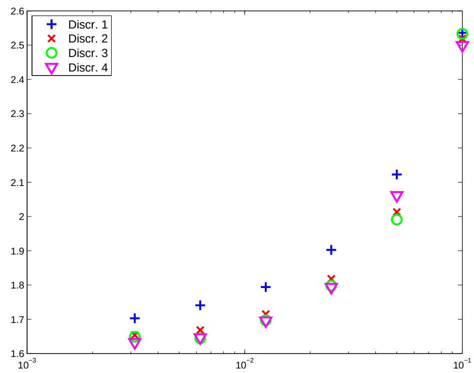

Figure 2: (Example 1) $\|\mathbf{p}_\varepsilon(1)\|_{\mathbf{P}^1}$ vs. ε for different discretizations

ε	τ_ε	$\lvert\tau_\varepsilon - \tau^*\rvert$	$\|\mathbf{y}_\varepsilon(1) - \mathbf{z}\|_{\mathbf{Y}^0}$
$1.0000 \cdot 10^{-1}$	$4.00 \cdot 10^{-1}$	$3.15 \cdot 10^{-2}$	$2.87 \cdot 10^{-2}$
$5.0000 \cdot 10^{-2}$	$4.18 \cdot 10^{-1}$	$1.45 \cdot 10^{-2}$	$1.21 \cdot 10^{-2}$
$2.5000 \cdot 10^{-2}$	$4.24 \cdot 10^{-1}$	$7.81 \cdot 10^{-3}$	$5.68 \cdot 10^{-3}$
$1.2500 \cdot 10^{-2}$	$4.28 \cdot 10^{-1}$	$3.72 \cdot 10^{-3}$	$2.71 \cdot 10^{-3}$
$6.2500 \cdot 10^{-3}$	$4.31 \cdot 10^{-1}$	$1.34 \cdot 10^{-3}$	$1.33 \cdot 10^{-3}$
$3.1250 \cdot 10^{-3}$	$4.32 \cdot 10^{-1}$		$7.00 \cdot 10^{-4}$

Table 3: Convergence history for Example 2

we have controllability for times $T \geq \sqrt{2}$. The optimal time is found to be $\tau^* \approx 0.43$, which is not within the range of controllability.

The convergence rates are worse than in Example 1. In fact, we observe $|\tau_\varepsilon - \tau^*| = O(\sqrt{\varepsilon})$.

Another interesting observation is that on the finest discretization level a smaller value of ε and thus more iterations to reach the prescribed accuracy for the terminal residual are needed, see Figure 3.

Moreover, the norms $\{\|\mathbf{p}_{\varepsilon,h}(1)\|_{\mathbf{P}^1}\}_{\varepsilon>0}$ are not bounded uniformly with respect to the discretization, see Figure 4, and $\{\|\mathbf{p}_{\varepsilon(1),h}\|_{\mathbf{P}^0}\}_{\varepsilon>0}$ is not bounded either. This suggests that for the original system the transversality condition holds in the form $\langle \mathbf{B}u(1), \tilde{\mathbf{p}}(1)\rangle_{(\mathbf{P}^0)^*,\mathbf{P}^0} = 0$, which was also observed numerically. In Remark 4.5 of [17] it was observed that if $|\tau_\varepsilon - \tau^*| = O(\varepsilon)$, then $\{\|\mathbf{p}_{\varepsilon,h}(1)\|_{\mathbf{P}^1}\}_{\varepsilon>0}$ is bounded. Since $|\tau_\varepsilon - \tau^*| = O(\sqrt{\varepsilon})$ only, unboundedness

ε	τ_ε	$\lvert\tau_\varepsilon - \tau^*\rvert$	$\lVert\mathbf{y}_\varepsilon(1) - \mathbf{z}\rVert_{\mathbf{Y}^0}$
$1.0000 \cdot 10^{-1}$	$3.98 \cdot 10^{-1}$	$2.25 \cdot 10^{-1}$	$3.97 \cdot 10^{-2}$
$5.0000 \cdot 10^{-2}$	$4.45 \cdot 10^{-1}$	$1.77 \cdot 10^{-1}$	$2.41 \cdot 10^{-2}$
$2.5000 \cdot 10^{-2}$	$4.86 \cdot 10^{-1}$	$1.37 \cdot 10^{-1}$	$1.50 \cdot 10^{-2}$
$1.2500 \cdot 10^{-2}$	$5.21 \cdot 10^{-1}$	$1.02 \cdot 10^{-1}$	$9.25 \cdot 10^{-3}$
$6.2500 \cdot 10^{-3}$	$5.47 \cdot 10^{-1}$	$7.64 \cdot 10^{-2}$	$5.81 \cdot 10^{-3}$
$3.1250 \cdot 10^{-3}$	$5.67 \cdot 10^{-1}$	$5.58 \cdot 10^{-2}$	$3.74 \cdot 10^{-3}$
$1.5625 \cdot 10^{-3}$	$5.84 \cdot 10^{-1}$	$3.95 \cdot 10^{-2}$	$2.52 \cdot 10^{-3}$
$7.8125 \cdot 10^{-4}$	$5.98 \cdot 10^{-1}$	$2.50 \cdot 10^{-2}$	$1.73 \cdot 10^{-3}$
$3.9063 \cdot 10^{-4}$	$6.12 \cdot 10^{-1}$	$1.15 \cdot 10^{-2}$	$1.21 \cdot 10^{-3}$
$1.9531 \cdot 10^{-4}$	$6.23 \cdot 10^{-1}$		$9.02 \cdot 10^{-4}$

Table 4: Convergence history for Example 3

of $\{\lVert\mathbf{p}_{\varepsilon,h}(1)\rVert_{\mathbf{P}^1}\}_{\varepsilon>0}$ should be expected.

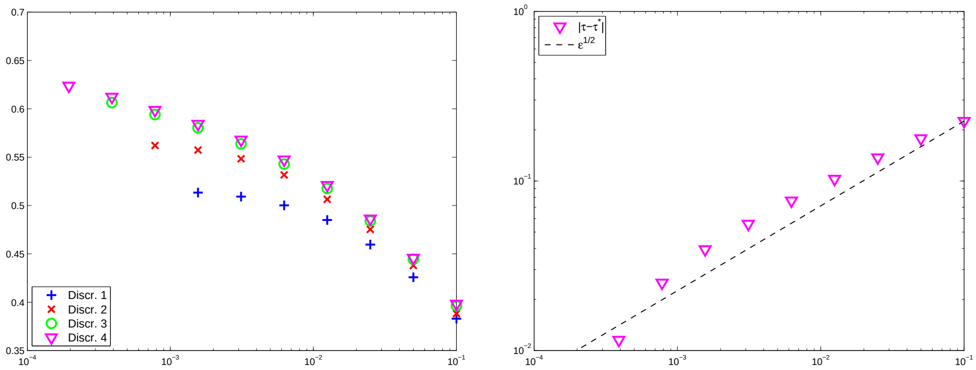

Figure 3: (Example 3) τ_ε vs. ε for different discretizations; $|\tau_\varepsilon - \tau^*|$ vs. ε for finest discretization

6 Concluding remarks

Clearly many interesting questions can be addressed as extensions of the present research. Here we used an $L^2(\omega)$-norm constraint on the controls

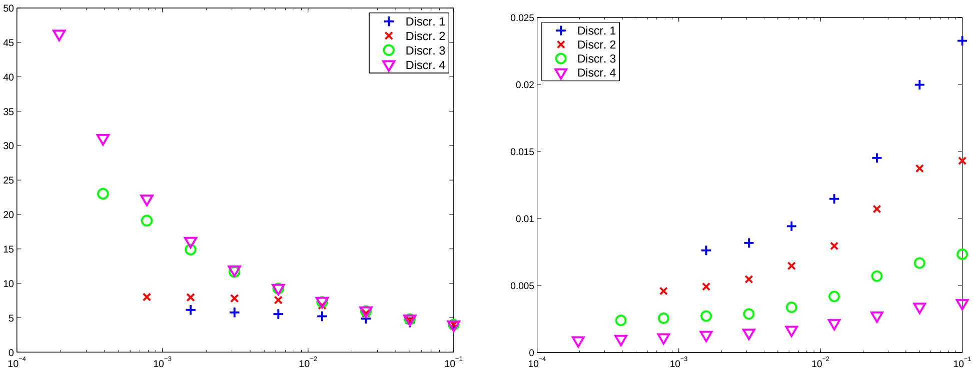

Figure 4: (Example 3) $\|\mathbf{p}_\varepsilon(1)\|_{\mathbf{P}^1}$ and $\frac{\|\mathbf{p}_\varepsilon(1)\|_{\mathbf{P}^1}}{\|\mathbf{p}_\varepsilon(1)\|_{\mathbf{P}^0}}$ vs. ε for different discretizations

spatially. Pointwise controls are equally important and require, in part, different treatment. Boundary control problems, as well, are a natural problem to be addressed. Various aspects of discretization would also be of interest.

References

[1] R. Bulirsch. Die Mehrzielmethode zur numerischen Lösung von nichtlinearen Randwertproblemen und Aufgaben der optimalen Steuerung. *DLR, Oberpfaffenhofen, Germany, Report of the Carl-Cranz-Gesellschaft (1971)*. Reprint: Department of Mathematics, Munich University of Technology (1985).

[2] L. C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1998.

[3] H. O. Fattorini. The time optimal problem for distributed control of systems described by the wave equation. In *Control theory of systems governed by partial differential equations (Conf. Naval Surface Weapons Center, Silver Spring, Md., 1976)*, pages 151–175. Academic Press, New York, 1977.

[4] H. O. Fattorini. *Infinite-dimensional optimization and control theory*, volume 62 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1999.

[5] H. O. Fattorini. *Infinite dimensional linear control systems*, volume 201 of *North-Holland Mathematics Studies*. Elsevier Science B.V., Amsterdam, 2005. The time optimal and norm optimal problems.

[6] R. Griesse, T. Grund, and D. Wachsmuth. Update strategies for perturbed nonsmooth equations. *Optim. Methods Softw.*, 23(3):321–343, 2008.

[7] M. Gugat. Penalty techniques for state constrained optimal control problems with the wave equation. *SIAM J. Control Optim.*, 48(5):3026–3051, 2009/10.

[8] M. Gugat and G. Leugering. L^∞-norm minimal control of the wave equation: on the weakness of the bang-bang principle. *ESAIM Control Optim. Calc. Var.*, 14(2):254–283, 2008.

[9] K. Ito and K. Kunisch. *Lagrange multiplier approach to variational problems and applications*, volume 15 of *Advances in Design and Control*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2008.

[10] K. Ito and K. Kunisch. Semismooth Newton methods for time-optimal control for a class of ODEs. *SIAM J. Control Optim.*, 48(6):3997–4013, 2010.

[11] C. Y. Kaya and J. L. Noakes. Computational method for time-optimal switching control. *J. Optim. Theory Appl.*, 117(1):69–92, 2003.

[12] H. B. Keller. *Numerical methods for two-point boundary-value problems*. Blaisdell Publishing Co. Ginn and Co., Waltham, Mass.-Toronto, Ont.-London, 1968.

[13] D. Kinderlehrer and G. Stampacchia. *An introduction to variational inequalities and their applications*. Academic Press, New York, 1980.

[14] W. Krabs. On time-minimal distributed control of vibrating systems governed by an abstract wave equation. *Appl. Math. Optim.*, 13(2):137–149, 1985.

[15] W. Krabs. On time-minimal distributed control of vibrations. *Appl. Math. Optim.*, 19(1):65–73, 1989.

[16] A. Kröner, K. Kunisch, and B. Vexler. Semismooth Newton methods for optimal control of the wave equation with control constraints. *SIAM Journal Contr. Opt.*, 49:830–858, 2011.

[17] K. Kunisch and D. Wachsmuth. On time optimal control of the wave equation, its regularization and optimality system. *ESAIM Control Optim. Calc. Var.*, 2011. To appear.

[18] J.-L. Lions. *Control of distributed singular systems.* Gauthier-Villars, Kent, 1985.

[19] E. Meier and A. E. Bryson, Jr. Efficient algorithm for time-optimal control of a two-link manipulator. *J. Guidance Control Dynam.*, 13(5):859–866, 1990.

[20] K. D. Phung, G. Wang, and X. Zhang. On the existence of time optimal controls for linear evolution equations. *Discrete Contin. Dyn. Syst. Ser. B*, 8(4):925–941, 2007.

[21] G. Wang and E. Zuazua. On the equivalence of minimal time and minimal norm controls for heat equations. Submitted, 2011.