ESAIM: M2AN 54 (2020) 79–103
https://doi.org/10.1051/m2an/2019046

ESAIM: Mathematical Modelling and Numerical Analysis
www.esaim-m2an.org

TIME-OPTIMALITY BY DISTANCE-OPTIMALITY FOR PARABOLIC CONTROL SYSTEMS

Lucas Bonifacius[1,*] and Karl Kunisch[2]

Abstract. The equivalence of time-optimal and distance-optimal control problems is shown for a class of parabolic control systems. Based on this equivalence, an approach for the efficient algorithmic solution of time-optimal control problems is investigated. Numerical examples are provided to illustrate that the approach works well in practice.

Mathematics Subject Classification. 49K20, 49M15.

Received September 6, 2018. Accepted June 21, 2019.

1. Introduction

This article is devoted to time optimal control problems for parabolic systems. Specifically, we propose a formulation which is equivalent to the original time optimal control formulation and amenable for numerical realization. We consider the problem

$$\text{Minimize } T \quad \text{subject to} \quad \begin{cases} T > 0, & \\ u \in U_{ad}(0,T), & \\ \partial_t y + Ay = Bu, & \text{in } (0,T), \\ y(0) = y_0, & \\ \|y(T) - y_d\|_H \le \delta_0, & \end{cases} \tag{P}$$

where y denotes the state, u the control, and T the terminal time. Here, the set of admissible controls is

$$U_{ad}(0,T) \coloneqq \left\{ u \in L^2((0,T); L^2(\omega)) \colon u_a \le u(t) \le u_b \ a.e.\ t \in (0,T) \right\}$$

for $u_a, u_b \in L^\infty(\omega)$ the control constraints, where ω is a measurable set. Moreover, A is an unbounded operator satisfying Gårding's inequality and B is the (bounded) control operator; see also Section 2 for the precise assumptions. The goal is to steer the system into a ball centered at y_d with radius δ_0 in the shortest time possible. Note that the state equation is posed on a variable time horizon which causes a nonlinear dependency of the state y with respect to the terminal time T and the control u. For this reason, (P) is a nonlinear and

Keywords and phrases. Time-optimal controls, bang-bang controls, distance optimal controls, parabolic control systems.

[1] Fakultät für Mathematik, Technische Universität München, München, Germany.

[2] Institute for Mathematics and Scientific Computing, University of Graz, Graz, Austria.

*Corresponding author: lucas.bonifacius@tum.de

nonconvex optimization problem subject to control as well as state constraints. Additionally we emphasize that the objective functional does not contain control costs which complicates the algorithmic solution of (P) compared to the situation with an L^2 term in the objective; *cf.*, *e.g.*, [15, 19].

In this article, we propose an equivalent reformulation in terms of minimal distance problems that can be used algorithmically to solve the time-optimal control problem. For $\delta > 0$ consider the perturbed time-optimal control problem defined as

$$\inf_{\substack{T>0 \\ u\in U_{ad}(0,T)}} T \quad \text{subject to} \quad \|y[u](T) - y_d\|_H \le \delta_0 + \delta, \tag{P_δ}$$

where $y[u](T)$ denotes the state associated with the control u evaluated at T. Moreover, we consider the *minimum distance control problem*

$$\min \|y[u](T) - y_d\|_H - \delta_0 \quad \text{subject to} \quad u \in U_{ad}(0,T). \tag{δ_T}$$

Under weak assumptions we show that the associated value functions defined by

$$T(\delta) = \inf\,(P_\delta) \quad \text{and} \quad \delta(T) = \inf\,(\delta_T)$$

are inverse to each other; see Proposition 3.4. Furthermore, we prove that (P_δ) and (δ_T) are equivalent. Precisely, if $\delta > 0$ is given and $(T, \bar u)$ is optimal for (P_δ), then $\bar u$ is also distance-optimal for (δ_T). Conversely, if $T > 0$ is given and $\bar u$ is distance optimal, then $(T, \bar u)$ is time-optimal with $\delta = \delta(T)$; see Theorem 3.1 for details. Hence, instead of solving the time-optimal control problem directly, we can search for a root of the $\delta(\cdot)$-value function. A similar equivalence first appeared in [35] (see also [37], Sect. 5.4) for the situation where one aims at delaying the activation of the control as long as possible. However, to the best of our knowledge it has never been considered for an algorithmic approach. In this regard, we also mention a similar approach used in [11] for time-optimal control of a one-dimensional vibrating system with controls in a subspace of L^2 determined by certain moment equations.

We show that the $\delta(\cdot)$-value function is continuously differentiable for many important control scenarios; see Section 5. If in addition qualified optimality conditions hold for the original problem, then the derivative of $\delta(\cdot)$ is nonvanishing near the optimal solution; see Proposition 6.2. This justifies to use a Newton method for the calculation of a root of $\delta(\cdot)$. Moreover, under an additional assumption we show that the derivative of the value function is Lipschitz continuous which guarantees fast local convergence of the Newton method. In fact, in all our numerical examples, we observe quadratic order of convergence, even if the additional assumption does not hold. For the solution of the resulting minimal distance problem with simple control constraints, the literature offers a wide spectrum of algorithms.

Time optimal control problems are among the most studied problems of optimal control, and thus it comes at no surprise that diverse techniques have been proposed for their solution. In the following let us briefly describe some of them. An approach, which is conceptually close, rests on an equivalent reformulation utilizing minimum norm problems. In contrast, to the perturbations in the terminal constraint as in (P_δ), perturbations in the control constraint are introduced. To explain the approach, we consider the time optimal control problem

$$\inf_{\substack{T>0 \\ u\in L^2((0,T)\times\omega)}} T \quad \text{subject to} \quad \|y[u](T) - y_d\|_H \le \delta_0, \quad \|u\|_{L^\infty((0,T);L^2(\omega))} \le \rho, \tag{P_ρ}$$

which is related to the *minimal norm problem* defined as

$$\inf_{u\in L^2((0,T)\times\omega)} \|u\|_{L^\infty((0,T);L^2(\omega))} \quad \text{subject to} \quad \|y[u](T) - y_d\|_H \le \delta_0. \tag{N_T}$$

To follow much of the literature, we adapted the control constraints to be chosen in $L^\infty((0,T);L^2(\omega))$ rather than $L^\infty((0,T)\times\omega)$. Under appropriate controllability assumptions these two problems have been shown to be equivalent; see [8,10,17,30,36] for parabolic equations, [38] for time-varying ordinary differential equations, Chapter 5 of [37] for abstract evolution equations, and [39] for the Schrödinger equation. Note that typically these publications consider the case of exact controllability, *i.e.* $\delta_0 = 0$, or exact null controllability, *i.e.* $\delta_0 = 0$ and $y_d = 0$. The solution to (N_T) can be determined by solving an unconstrained optimization problem given by

$$\inf_{\varphi_T \in H} \frac{1}{2}\left(\int_0^T \|B^*\varphi(t)\|_{L^2(\omega)}\,\mathrm{d}t\right)^2 + (\varphi(0), y_0) + \delta_0\|\varphi_T\|_H - (y_d, \varphi_T), \tag{1.1}$$

where φ is the solution to the adjoint state equation

$$-\partial_t\varphi + A^*\varphi = 0, \quad \varphi(T) = \varphi_T;$$

see Section 4 of [30], compare also Section 1.7 of [9]. If $\bar{\varphi}_T$ is the minimizer of (1.1), then the minimum norm control is given by

$$\bar{u}(t) = \left(\int_0^T \|B^*\bar{\varphi}(t)\|_{L^2(\omega)}\,\mathrm{d}t\right)\frac{(B^*\bar{\varphi})(t)}{\|(B^*\bar{\varphi})(t)\|_{L^2(\omega)}}, \quad a.e.\ t \in (0,T),$$

where $\bar{\varphi}$ is the adjoint state with terminal value $\bar{\varphi}_T$. Turning to the numerical realization, as far as we know, the only algorithmic studies based on this equivalence are [21] for time-optimal control problems subject to ordinary differential equations and [24] for problems subject to partial differential equations employing an optimal design approach. A direct numerical realization of (N_T) is impeded by the difficulties related to the appearance of the state constraint and the fact that the minimization is carried out over a non-reflexive Banach space. In contrast (1.1) does not contain state constraints. Turning to the realization of (N_T) by means of (1.1), one has to cope with the non-smoothness of the L^1-norm, whereas (δ_T) involves the minimization of a Hilbert space norm. In addition, (1.1) can be considered as an inverse source problem for the initial condition of the adjoint problem. Such problems are inherently ill-posed. For the specific context of (1.1) this was analyzed in [25].

An alternative approach to solve time-optimal control problems for finite or infinite dimensional systems is based on solving the optimality system for (P) after adding a regularization term of the form $\alpha\|u\|^2$ to the cost functional. In an additional outer loop the regularization parameter α can be driven to zero; see [15,18,19]. This is a flexible method, but one has to cope with the difficulties of the asymptotic behavior as the regularization parameters tends to zero. We compare our approach with the regularization approach in one numerical example and observe that even for a fixed regularization parameter our algorithm performs roughly five to ten times faster in terms of the required number of solves for the partial differential equation; see Table 3.

Yet another approach, which has mostly been investigated for time-optimal control problems subject to ordinary differential equations, rests on the reformulation of (P) as an optimization problem with respect to the switching points of the optimal controls; see *e.g.*, [16,23]. This approach cannot be extended to the distributed control setting in a straightforward way.

This paper is organized as follows: In Section 2 we introduce the notation and main assumptions. The equivalence of time and distance optimal controls is proved in Section 3. Section 4 is devoted to general properties of the time-optimal control problem. Differentiability of the value function associated to the minimal distance problems is proved in Section 5. The algorithm is presented in Section 6. Various numerical examples in Section 7 show that our approach is efficient in practice. Last, in Section 8 we conclude with some open problems.

2. Notation and main assumptions

Let V and H be real Hilbert spaces forming a Gelfand triple, *i.e.* $V \hookrightarrow_c H \cong H^* \hookrightarrow V^*$, where $\hookrightarrow$ denotes the continuous embedding and $\hookrightarrow_c$ the continuous and compact embedding. We abbreviate the duality pairing between V and V^* as well as the inner product and norm in H by

$$\langle\cdot,\cdot\rangle = \langle\cdot,\cdot\rangle_{V^*,V}, \quad (\cdot,\cdot) = (\cdot,\cdot)_H, \quad \|\cdot\| = \|\cdot\|_H.$$

Assumption 2.1. *Let $a\colon V\times V\to\mathbb{R}$ be a continuous bilinear form, which satisfies the Gårding inequality (also referred to as weak coercivity): There are constants $\alpha_0>0$ and $\omega_0\geq 0$ such that*

$$a(v,v)+\omega_0\|v\|^2\geq\alpha_0\|v\|_V^2\quad\text{for all } v\in V. \tag{2.1}$$

We denote by $A\colon V\subset V^*\to V^*$ the unique linear operator with

$$\langle Ay,v\rangle=a(y,v)\quad\text{for all } v\in V.$$

The Gårding inequality implies that $-A$ generates an analytic semigroup on V^* denoted $\mathrm{e}^{-\cdot A}$; see *e.g.*, Section 1.4 of [27].

Assumption 2.2. *Let (ω,ϱ) be a measure space. We assume that the control operator $B\colon L^2(\omega,\varrho)\to V^*$ is linear and continuous. Moreover, $y_d\in H$ is the desired state and $\delta_0>0$.*

The abstract measure space allows for one consistent notation for different control scenarios. For example, in case of a distributed control on a subset ω of the spatial domain $\Omega\subset\mathbb{R}^d$ we take ω equipped with the Lebesgue measure. If no ambiguity arises, we drop the measure ϱ and simply write ω in the following. The space of admissible controls is defined as

$$U_{ad}\coloneqq\left\{u\in L^2(\omega)\colon u_a\leq u\leq u_b\ \ a.e.\text{ in }\ \omega\right\}\subset L^\infty(\omega)$$

for $u_a,u_b\in L^\infty(\omega)$ with $u_a<u_b$ almost everywhere. In addition, for $T>0$ we set $U(0,T)\coloneqq L^2((0,T)\times\omega)$ and

$$U_{ad}(0,T)\coloneqq\{u\in U(0,T)\colon u(t)\in U_{ad}\ a.e.\ t\in(0,T)\}\subset L^\infty((0,T)\times\omega),$$

where $(0,T)\times\omega$ is equipped with the completion of the product measure. For $T>0$ we use $W(0,T)$ to abbreviate $H^1((0,T);V^*)\cap L^2((0,T);V)$, endowed with the canonical norm and inner product. The symbol $i_T\colon W(0,T)\to H$ denotes the trace mapping $i_Ty=y(T)$. For any two Banach spaces X and Y, let $\mathcal{L}(X,Y)$ denote the space of linear and bounded operators from X to Y. The symbol $\mathcal{B}_r(x)$ stands for the ball centered at $x\in X$ with radius $r>0$ in X. Moreover, $\mathbb{R}_+$ abbreviates the open interval $(0,+\infty)$.

Last, to ensure the existence of optimal controls we require the following.

Assumption 2.3. *There exist a finite time $T>0$ and a feasible control $u\in U_{ad}(0,T)$ such that the solution to the state equation of (P) satisfies $\|y(T)-y_d\|\leq\delta_0$. To exclude the trivial case, we in addition assume $\|y_0-y_d\|>\delta_0$.*

3. Equivalence of time and distance optimal controls

Instead of solving the time-optimal control problem directly, we propose to solve an equivalent reformulation in terms of minimal distance control problems. The reformulation leads to a bilevel optimization problem, where we search for a root of a certain value function in the outer loop and solve convex optimization problems in the inner loop. We start by proving the equivalence of minimal time and minimal distance controls.

For any $\delta\geq 0$ we consider the perturbed time-optimal control problem

$$\begin{aligned}\inf T\quad\text{subject to}\quad &T\in\mathbb{R}_+,\ u\in U_{ad}(0,T),\\ &\|y[u](T)-y_d\|\leq\delta_0+\delta.\end{aligned}\tag{P_δ}$$

Moreover, for fixed $T>0$ we consider the *minimal distance control problem*

$$\min\ \|y[u](T)-y_d\|-\delta_0\quad\text{subject to}\quad u\in U_{ad}(0,T).\tag{δ_T}$$

Note that (P_δ) is a nonlinear and nonconvex optimization problem subject to control as well as state constraints, whereas (δ_T) is a convex problem subject to control bounds only.

We define the value functions $T\colon [0,\infty) \to [0,\infty]$ and $\delta\colon [0,\infty) \to [0,\infty)$ as

$$T(\delta) = \inf\,(P_\delta) \quad \text{and} \quad \delta(T) = \inf\,(\delta_T),$$

where we set $\delta(0) = \|y_0 - y_d\| - \delta_0$. Let us formulate the main result of this section.

Theorem 3.1. *Let $T(\cdot)$ be left-continuous. If $T \in (0, T(0)]$ and $u \in U_{ad}(0,T)$ is distance-optimal for (δ_T), then (T,u) is also time-optimal for $(P_{\delta(T)})$. Conversely, if $\delta \in [0,\delta^\bullet)$ and $(T,u) \in \mathbb{R}_+ \times U_{ad}(0,T)$ is time-optimal for (P_δ), then u is also distance-optimal for (δ_T), where $\delta^\bullet = \|y_0 - y_d\| - \delta_0$.*

The proof of Theorem 3.1 will be given in the following. We first note that due to boundedness of U_{ad}, linearity of the control-to-state mapping (for fixed $T > 0$), and weak lower semicontinuity of the norm, the problem (δ_T) is well-posed, and for this reason the value function $\delta(\cdot)$ is well-defined. In contrast, to verify well-posedness of (P_δ) we require Assumption 2.3; *cf.* also Proposition 4.1.

Proposition 3.2. *The value function T is finite, i.e. $T(\cdot) < \infty$ on $[0,\infty)$.*

Proof. Let $(T,u) \in \mathbb{R}_+ \times U_{ad}(0,T)$ be the feasible point from Assumption 2.3, *i.e.* $\|y[u](T) - y_d\| \le \delta_0$. Clearly, $(T,u) \in \mathbb{R}_+ \times U_{ad}(0,T)$ is also feasible for (P_δ) for any $\delta > 0$. Thus, $T(\delta) \le T < \infty$. ☐

Proposition 3.3. *Set $\delta^\bullet = \|y_0 - y_d\| - \delta_0$. The function $T\colon [0,\delta^\bullet] \to [0,\infty)$ is strictly monotonically decreasing and right-continuous.*

Proof. Step 1: T is strictly decreasing. Clearly, T is monotonically decreasing. To show strict monotonicity, let $\delta_1 > \delta_2 \ge 0$. We have to show $T(\delta_1) < T(\delta_2)$. Suppose $T(\delta_1) = T(\delta_2)$ and let $(T(\delta_i), u_i) \in \mathbb{R}_+ \times U_{ad}(0, T(\delta_i))$ be optimal solutions to (P_{δ_i}), $i = 1, 2$. Since

$$\|y[u_2](T(\delta_2)) - y_d\| - \delta_0 = \delta_2 < \delta_1,$$

we infer that $(T(\delta_2), u_2)$ is also feasible for (P_{δ_1}). Note that in the problem formulation we can equivalently use $\|y[u](T) - y_d\| \le \delta$ and $\|y[u](T) - y_d\| = \delta$. From continuity of $y[u_2]\colon [0, T(\delta_2)] \to H$ and $T(\delta_1) = T(\delta_2)$ we deduce that $(T(\delta_1), u_1)$ cannot be optimal for the time-optimal problem (P_{δ_1}). This contradicts the assumption and we conclude $T(\delta_1) < T(\delta_2)$.

Step 2: T is right-continuous. Consider a sequence $\delta_1 \ge \delta_2 \ge \ldots \ge \delta_n \to \delta$. We have to show $\lim_{n\to\infty} T(\delta_n) = T(\delta)$. Assume that $\lim_{n\to\infty} T(\delta_n) \ne T(\delta)$. Then, due to monotonicity of T, there is $\varepsilon > 0$ such that

$$\lim_{n\to\infty} T(\delta_n) = T(\delta) - \varepsilon.$$

Let $u_n = u_n(\delta_n, T(\delta_n)) \in U_{ad}(0, T(\delta_n))$ denote an optimal control to (P_{δ_n}). We can extend each u_n to the time-interval $(0, T(\delta))$ so that $u_n \in U_{ad}(0, T(\delta))$ for all $n \in \mathbb{N}$. Due to boundedness of $U_{ad}(0, T(\delta))$, there is a subsequence denoted in the same way such that $u_n \rightharpoonup u$ in $L^s((0, T(\delta)) \times \omega)$ with $u \in U_{ad}(0, T(\delta))$ and some $s > 2$. Now, continuity of $y[u]\colon [0, T(\delta)] \to H$ and the triangle inequality imply

$$\begin{aligned}
\lim_{n\to\infty} \|y[u_n](T(\delta_n)) - y_d\| &\ge \lim_{n\to\infty} \|y[u](T(\delta_n)) - y_d\| - \lim_{n\to\infty} \|y[u](T(\delta_n)) - y[u_n](T(\delta_n))\| \\
&\ge \lim_{n\to\infty} \|y[u](T(\delta_n)) - y_d\| - \lim_{n\to\infty} \sup_{t\in[0,T(\delta)]} \|y[u](t) - y[u_n](t)\| \\
&= \lim_{n\to\infty} \|y[u](T(\delta_n)) - y_d\|,
\end{aligned}$$

where in the last step we have used compactness of the control-to-state mapping from $\mathbb{R} \times L^s((0, T(\delta)) \times \omega)$ to $C([0, T(\delta)]; H)$; see Proposition A.19 of [2]. Therefore,

$$\delta + \delta_0 = \lim_{n\to\infty} \delta_n + \delta_0 = \lim_{n\to\infty} \|y[u_n](T(\delta_n)) - y_d\| \ge \|y[u](T(\delta) - \varepsilon) - y_d\|.$$

Thus, $(T(\delta) - \varepsilon, u)$ is admissible for (P_δ), contradicting optimality of $T(\delta)$. ☐

Proposition 3.4. *Let $T(\cdot)$ be left-continuous. Then $\delta\colon [0,T(0)] \to [0,\infty)$ is continuous and strictly monotonically decreasing. Moreover,*

$$T(\delta(T')) = T' \quad \text{for all } T' \in [0,T(0)] \tag{3.1}$$

and

$$\delta(T(\delta')) = \delta' \quad \text{for all } \delta' \in [0,\delta^\bullet]. \tag{3.2}$$

Proof. First, since T is strictly decreasing, its inverse T^{-1} is continuous. Moreover, as T is right-continuous according to Proposition 3.3, the assumption implies that T is continuous. Hence, T^{-1} is defined everywhere on $[0,T(0)]$; see *e.g.*, Theorem III.5.7 of [1].

Let $T > 0$. Then there exists $u \in U_{ad}(0,T)$ such that $\|y[u](T) - y_d\| - \delta_0 = \delta(T)$. Hence, $T(\delta(T)) \le T$ holds. Suppose that $T(\delta(T)) < T$. Then by continuity of T there exists $\delta' < \delta(T)$ such that $T(\delta') = T$. Let $u' \in U_{ad}(0,T)$ be an optimal control to $(P_{\delta'})$. Then

$$\delta' < \delta(T) \le \|y[u'](T) - y_d\| - \delta_0 \le \delta',$$

a contradiction, which proves (3.1).

Moreover, (3.1) implies that $T(\delta(T(\delta'))) = T(\delta')$ for all $\delta' \in [0,\delta^\bullet]$. Strict monotonicity of T therefore yields (3.2). For these reasons, $\delta = T^{-1}$ and we conclude that δ is continuous and strictly monotonically decreasing. $\square$

After this preparation we can now prove the equivalence of time and distance optimal controls.

Proof of Theorem 3.1. Let $T > 0$ and $u \in U_{ad}(0,T)$ be distance-optimal for (δ_T), *i.e.* $\delta(T) = \|y[u](T) - y_d\| - \delta_0$. Due to (3.1) we have $T(\delta(T)) = T$. Thus, (T,u) is also time-optimal for $(P_{\delta(T)})$.

Conversely, let $\delta \ge 0$ and $(T,u) \in \mathbb{R}_+ \times U_{ad}(0,T)$ be time-optimal for (P_δ). In particular, this gives $\|y[u](T) - y_d\| - \delta_0 = \delta$. Using (3.2) we infer that

$$\delta(T(\delta)) = \delta = \|y[u](T) - y_d\| - \delta_0,$$

i.e. u is also distance-optimal for (δ_T). $\square$

Since monotone functions have at most countably many discontinuities, see *e.g.*, Proposition III.5.6 of [1], it is unlikely that we accidentally hit a point where T is not left-continuous. However, for the algorithm to be presented later we are interested in continuity of the value function in a neighborhood of the optimal value. To this end, we state two sufficient conditions. Note that the second condition even guarantees Lipschitz continuity from the left of the value function $T(\cdot)$. The setting considered in [36] for example automatically satisfies the assumptions of Proposition 3.6, since $y_d = 0$ and $0 \in U_{ad}$.

Proposition 3.5. *Let the following controllability condition hold: For all $\delta \in (0,\delta^\bullet]$ there exists $T' > 0$ and a control $u \in U_{ad}(0,T')$ such that the trajectory $y[u, y^\delta(T(\delta))]$ with initial value $y^\delta(T(\delta))$ satisfies*

$$\|y[u, y^\delta(T(\delta))](t) - y_d\| < \delta_0 + \delta \quad \textit{for all } t \in (0,T'],$$

where $y^\delta \in W(0,T(\delta))$ denotes an optimal trajectory for (P_δ). Then, T is left-continuous.

Proof. Let $\delta > 0$ and let $u \in U_{ad}(0,T(\delta))$ be an optimal control to (P_δ). According to the controllability assumption there exists $T' > 0$ and an extended control $u' \in U_{ad}(0,T(\delta)+T')$, *i.e.* $u' = u$ on $(0,T(\delta))$, such that

$$\|y[u', y_0](T(\delta)+t) - y_d\| < \delta_0 + \delta \quad \text{for all } t \in (0,T'].$$

Set $f(t) := \|y[u'](T(\delta)+t) - y_d\| - \delta_0$. Since f is continuous, $f(0) = \delta$, and $f(t) < \delta$ for $t \in (0,T']$, for all $\delta_n > 0$ sufficiently small there exists $t_n \in (0,T']$ such that $f(t_n) = \delta - \delta_n$. Hence, $(T(\delta)+t_n, u')$ is feasible for $(P_{\delta-\delta_n})$ and we have

$$T(\delta - \delta_n) \le T(\delta) + t_n.$$

Moreover, if $\delta_n \to 0$ then $t_n \to 0$ due to $f(t) < \delta$ for $t \in (0, T']$. Passing to the limit $\delta_n \to 0$ yields the result. □

Proposition 3.6. *Let $y_d \in V$ and assume that Gårding's inequality (2.1) holds with $\omega_0 = 0$. If there exists a control $\breve{u} \in U_{ad}$ such that $\|B\breve{u} - Ay_d\|_{V^*} < \alpha_0\delta_0$, then $T\colon [0, \delta^\bullet] \to \mathbb{R}$ is Lipschitz continuous from the left.*

Proof. We argue similarly as in Theorem 4.5 of [3]. First, the assumptions of Proposition 3.6 ensure that (3.3) of [3] holds with $h_0 = \alpha_0\delta_0 - \|B\breve{u} - Ay_d\|_{V^*} > 0$; see the proof of Proposition 5.3 from [3].

Let $\delta \in [0, \delta^\bullet]$, $\delta' \in [0, \delta]$, and (T, u, y) be a solution to problem (P_δ). Consider the auxiliary problem $\partial_t\breve{y} + A\breve{y} = B\breve{u}$ with initial condition $\breve{y}(0) = y(T)$ and an auxiliary control $\breve{u}\colon [0, \infty) \to U_{ad}$. Employing Lemma 3.9 of [3] we can choose $\breve{u}$ such that

$$\max\{\,0, \|\breve{y}(t) - y_d\| - \delta_0\,\} \leq \max\{\,0,\ \delta - h_0 t\,\} \quad \text{for } t \geq 0$$

holds, since $d_U(y) = \max\{\,0, \|y - y_d\| - \delta_0\,\}$, where $d_U(\cdot)$ denotes the distance function to the set $U = \mathcal{B}_{y_d}(\delta_0)$. Choose $\tau = (\delta - \delta')/h_0$. Then, $u \in U_{ad}(0, T + \tau)$ defined by

$$u(t) = \begin{cases} u(t) & \text{if } t \leq T, \\ \breve{u}(t - T) & \text{if } t > T, \end{cases}$$

is admissible for $(P_{\delta'})$ and we find

$$T(\delta') \leq T(\delta) + \tau = T(\delta) + (\delta - \delta')/h_0$$

concluding the proof. □

Instead of solving the time-optimal control problem, we can equivalently search for a root of the value function $\delta(\cdot)$ by virtue of Theorem 3.1. However, this still might be a difficult task, as $\delta(\cdot)$ can have several roots and we do not know the approximate region of the root we are looking for. If in addition the target set $\mathcal{B}_{\delta_0}(y_d)$ is weakly invariant under (A, BU_{ad}), *i.e.* for every $y_0 \in \mathcal{B}_{\delta_0}(y_d)$ there exists a control $u\colon [0, \infty) \to U_{ad}$ such that the solution y to

$$\partial_t y + Ay = Bu, \quad y(0) = y_0,$$

satisfies $y(t) \in \mathcal{B}_{\delta_0}(y_d)$ for all $t \geq 0$, then there is only one root where $\delta(\cdot)$ changes from a strictly positive value to a nonpositive value. We also refer to Theorem 3.8 from [3] for a characterization of weak invariance.

Proposition 3.7. *Let $T \in \mathbb{R}_+$ such that $\delta(T) = 0$. Suppose that the target set $\mathcal{B}_{\delta_0}(y_d)$ is weakly invariant under (A, BU_{ad}). Then, $\delta(T + t) \leq 0$ for all $t \in \mathbb{R}_+$.*

Proof. This immediately follows from the definition of weak invariance. □

Hence, if $T(\cdot)$ is continuous from the left and the target set $\mathcal{B}_{\delta_0}(y_d)$ is weakly invariant under (A, BU_{ad}), then an iterative procedure is able to find the global optimal solution to the time-optimal control problem, provided that the initial value T_0 for the minimization of $\delta(\cdot)$ satisfies $\delta(T_0) > 0$. If the latter condition is violated, then the procedure has to be restarted with a smaller initial value. Repeating the steps above will lead to an optimal solution.

4. The time-optimal control problem

We introduce a change of variables to discuss first order necessary optimality conditions for (P). In particular, we consider optimality conditions in qualified form that will be essential for the Newton method (introduced later) to be well-defined.

4.1. Change of variables

In order to deal with the variable time horizon of (P), we transform the state equation to the fixed reference time interval $(0, 1)$. For $\nu \in \mathbb{R}_+$ we set $T_\nu(t) = \nu t$ and obtain the transformed state equation

$$\partial_t y + \nu A y = \nu B u, \quad y(0) = y_0.$$

We generally abbreviate $I = (0, 1)$. Gårding's inequality guarantees that for each pair $(\nu, u) \in \mathbb{R}_+ \times U(0, 1)$ there exists a unique solution $y \in W(I)$ to the transformed state equation; see *e.g.*, Theorem 2, Chapter XVIII, Section 3 from [6]. Hence, it is justified to introduce the control-to-state mapping $S\colon \mathbb{R}_+ \times U_{ad}(0, 1) \to W(I)$ with $(\nu, u) \mapsto y = S(\nu, u)$. The transformed optimal control problem reads as

$$\min \nu \text{ subject to } \begin{cases} (\nu, u) \in \mathbb{R}_+ \times U_{ad}(0,1), \\ \|i_1 S(\nu, u) - y_d\| \leq \delta_0. \end{cases} \tag{$\hat{P}$}$$

We emphasize that the problem (P) and the transformed problem ($\hat{P}$) are equivalent; see Proposition 4.6 of [3].

Since there exists at least one feasible control due to Assumption 2.3, well-posedness of ($\hat{P}$) is obtained by the direct method; *cf.*, *e.g.*, Proposition 4.1 of [3]. We note that the optimal solution must fulfill the terminal constraint with equality (otherwise, a control with a shorter time is admissible, while having a smaller objective value).

Proposition 4.1. *Problem ($\hat{P}$) admits a solution $(\bar{\nu}, \bar{u}) \in \mathbb{R}_+ \times U_{ad}(0, 1)$ with associated state $\bar{y} = S(\bar{\nu}, \bar{u})$. Moreover, $\|\bar{y}(1) - y_d\| = \delta_0$ holds.*

4.2. First order optimality conditions

Next, we derive general necessary optimality conditions.

Lemma 4.2. *Let $(\bar{\nu}, \bar{u}) \in \mathbb{R}_+ \times U_{ad}(0, 1)$ be a solution to ($\hat{P}$). Then there exist $\bar{\mu} > 0$ and $\bar{\mu}_0 \in \{0, 1\}$ such that*

$$\int_0^1 \langle B\bar{u}(t) - A\bar{y}(t), \bar{p}(t)\rangle \,\mathrm{d}t = -\bar{\mu}_0, \tag{4.1}$$

$$\int_0^1 (B^*\bar{p}(t), u(t) - \bar{u}(t))_{L^2(\omega)} \,\mathrm{d}t \geq 0 \quad \text{for all } u \in U_{ad}(0, 1), \tag{4.2}$$

$$\|\bar{y}(1) - y_d\| = \delta_0, \tag{4.3}$$

where the adjoint state $\bar{p} \in W(0, 1)$ *is determined by*

$$-\partial_t \bar{p}(t) + \bar{\nu} A^* \bar{p}(t) = 0, \quad t \in (0, 1) \quad \bar{p}(1) = \bar{\mu}(\bar{y}(1) - y_d). \tag{4.4}$$

If $\bar{\mu}_0 = 1$, then the optimality conditions are called qualified.

Proof. Since the terminal set has finite codimension, see *e.g.*, Definition 4.1.5 of [20], we can argue as in Theorem 4.13 of [3] to obtain $\bar{\mu} > 0$ and $\bar{\mu}_0 \in \{0, 1\}$ such that

$$\min_{u \in U_{ad}(0,1)} \int_0^1 \langle Bu - A\bar{y}, \bar{p}\rangle \,\mathrm{d}t = \int_0^1 \langle B\bar{u} - A\bar{y}, \bar{p}\rangle \,\mathrm{d}t = -\bar{\mu}_0.$$

Now, (4.1) follows from the second equality and the first equality is equivalent to (4.2). □

Last, for the terminal set considered in this article, we cite the following criterion from [3] that guarantees qualified optimality conditions. It is worth mentioning that this condition can be checked *a priori* without knowing an optimal solution.

Proposition 4.3. *Adapt the assumptions of Proposition 3.6. Then qualified optimality conditions hold.*

Proof. This follows from Proposition 5.3 of [3] and Theorem 4.12 of [3]. □

5. Properties of the minimal distance value function

We discuss differentiability of the value function associated with the minimal distance control problems that will later be used for a Newton method. To this end, we first study optimality conditions and uniqueness of solutions to the minimal distance problems.

5.1. Minimal distance control problems

As in Section 4.1 the minimal distance control problem (δ_T) is transformed to the reference time interval $I = (0,1)$. Moreover, for fixed $\nu \in \mathbb{R}_+$ we define $\bar{u}(\nu)$ as

$$\bar{u}(\nu) = \underset{u \in U_{ad}(0,1)}{\operatorname{argmin}} \, \|i_1 S(\nu, u) - y_d\|. \tag{5.1}$$

Note that $\bar{u}(\nu)$ is not necessarily unique and for this reason $\nu \mapsto \bar{u}(\nu)$ is in general a set-valued mapping. However, the observation $i_1 S(\nu, u)$ is unique, because in (5.1) we can equivalently consider the squared norm that is strictly convex. For the following arguments we introduce $f\colon \mathbb{R}_+ \times U(0,1) \to \mathbb{R}$ defined by

$$f(\nu, u) = \|i_1 S(\nu, u) - y_d\| - \delta_0.$$

The minimal distance value function $\delta(\cdot)$ and the functional f are related via

$$\delta(\nu) = f(\nu, u), \quad u \in \bar{u}(\nu).$$

Differentiability of the control to state mapping, see Proposition 4.7 of [3], and the chain rule immediately imply that f is continuously differentiable for all $\nu \in \mathbb{R}_+$ such that $\delta(\nu) > -\delta_0$. Furthermore, introducing an adjoint state, we have the representation

$$\partial_\nu f(\nu, u) = \int_0^1 \langle Bu - Ay, p \rangle \,\mathrm{d}t, \quad \partial_u f(\nu, u) = \nu \int_0^1 (B^* p, \cdot)_{L^2(\omega)} \,\mathrm{d}t,$$

where $y = S(\nu, u)$ and $p \in W(0,1)$ is the associated adjoint state determined by

$$-\partial_t p + \nu A^* p = 0, \quad p(1) = (y(1) - y_d) / \|y(1) - y_d\|. \tag{5.2}$$

Note that the adjoint state p is independent of the concrete optimal control $u \in \bar{u}(\nu)$, due to uniqueness of the observation $y(1)$. Since both the objective functional in (5.1) and $U_{ad}(0,1)$ are convex, the following necessary and sufficient optimality condition holds: Given $\nu \in \mathbb{R}_+$ such that $\delta(\nu) > -\delta_0$, a control $u \in \bar{u}(\nu) \subset U_{ad}(0,1)$ is optimal for (5.1) if and only if

$$\int_0^1 (B^* p, u' - u)_{L^2(\omega)} \geq 0 \quad \text{for all } u' \in U_{ad}(0,1), \tag{5.3}$$

where $p \in W(0,1)$ solves (5.2) with $y = S(\nu, u)$; see *e.g.*, Lemma 2.21 of [32]. From the variational inequality (5.3) we deduce that an optimal control $u \in \bar{u}(\nu)$ satisfies

$$u(t,x) = \begin{cases} u_a(x) & \text{if } (B^* p)(t,x) > 0 \\ u_b(x) & \text{if } (B^* p)(t,x) < 0. \end{cases} \tag{5.4}$$

Hence, u is bang-bang, if the set where $B^* p$ vanishes has zero measure. Indeed, the latter condition ensures uniqueness of the control.

Proposition 5.1. *Let $\nu \in \mathbb{R}_+$ such that $\delta(\nu) > -\delta_0$ and $u \in \bar{u}(\nu)$. Moreover, suppose that the associated adjoint state p determined by (5.2) satisfies*

$$|\{ (t,x) \in I \times \omega \colon (B^*p)(t,x) = 0 \}| = 0, \tag{5.5}$$

where $|\cdot|$ denotes the measure associated with $I \times \omega$. Then u is bang-bang and $\bar{u}(\nu)$ is a singleton.

Proof. Due to uniqueness of the observation $y(1)$, the associated adjoint state p is unique. Hence, from (5.4) and (5.5) we conclude that u is bang-bang and unique. □

Condition (5.5) can be deduced from a unique continuation property. Let p denote the adjoint state with terminal value $p_1 \in H$. The system satisfies the *unique continuation property*

$$\text{if } B^*p = 0 \text{ on some } \Lambda \subset I \times \omega \text{ with } |\Lambda| \neq 0, \text{ then } p = 0. \tag{5.6}$$

Proposition 5.2. *If the unique continuation property (5.6) is satisfied, then (5.5) holds. In particular, $\bar{u}(\nu)$ is a singleton.*

Proof. Assume that condition (5.5) is violated. Then there exists a subset $\Lambda \subset I \times \omega$ with nontrivial measure such that $B^*p = 0$ on Λ. From the unique continuation property (5.6) we deduce $p = 0$. This contradicts $p(1) = (y(1) - y_d) / \|y(1) - y_d\| \neq 0$. □

Remark 5.3. The unique continuation property (also referred to as backward uniqueness property) is guaranteed to hold in the following situations.

(i) In the case of purely time-dependent controls, *i.e.* $Bu(t) = \sum_{i=1}^{N_c} e_i u_i(t)$, $u \in L^2(I;\mathbb{R}^{N_c})$, for $e_i \in V^*$, the unique continuation property is equivalent to normality of (A,B); see and Theorem 11.2.1, Definition 6.1.1 of [33]. A system (A,B) is called normal, if (A,B_i) is approximately controllable for all $i = 1,2,\ldots,N_c$; *cf.* also Section II.16 of [13] or Section III.3 of [22].
(ii) For the linear heat-equation on a bounded domain with a distributed control acting on an open subset of the spatial domain, the unique continuation property is known to hold; see Theorem 4.7.12 of [8] using Theorem 1.1 of [12].

5.2. Differentiability of $\delta(\cdot)$

Next, we present the central differentiability result of this section. After its proof, we discuss specific situations where the directional derivative of $\delta(\cdot)$ can be strengthened to a classical derivative.

Theorem 5.4. *Let $\nu \in \mathbb{R}_+$ such that $\delta(\nu) > -\delta_0$. Then the value function $\delta(\cdot)$ is directionally differentiable at ν and the expression*

$$\mathrm{d}^{\pm}\delta(\nu) = \min_{u \in \bar{u}(\nu)} \pm \int_0^1 \langle Bu - Ay, p\rangle \, \mathrm{d}t, \tag{5.7}$$

holds, where $p \in W(0,1)$ satisfies

$$-\partial_t p + \nu A^* p = 0, \quad p(1) = (y(1) - y_d) / \|y(1) - y_d\|,$$

and $y = S(\nu, u)$. If additionally the value of the integral in (5.7) is independent of the concrete minimizer $u \in \bar{u}(\nu)$, then $\delta(\cdot)$ is continuously differentiable at ν. Here, d^+ and d^- denote the right and left directional derivatives.

For the proof we require

Proposition 5.5. *Let $\nu \in \mathbb{R}_+$. Then*

$$\lim_{n\to\infty} \sup_{u\in U_{ad}(0,1)} \|i_1 S(\nu_n, u) - i_1 S(\nu, u)\| = 0$$

for all sequences $\nu_n \in \mathbb{R}_+$ such that $\nu_n \to \nu$.

Proof. Set $y_n = y(\nu_n)$ and $y = y(\nu)$. Then the difference $w = y - y_n$ satisfies

$$\partial_t w + \nu A w = (\nu - \nu_n)\left(-A y_n + B u\right), \quad w(0) = 0.$$

Hence, the assertion follows by standard energy estimates as well as the embedding $H^1(I; V^*) \cap L^2(I; V) \hookrightarrow C([0,1]; H)$. □

Proof of Theorem 5.4. Let $\nu \in \mathbb{R}_+$ and $\tau_n \in \mathbb{R}$ such that $\tau_n \to 0$. Set $\nu_n = \nu + \tau_n$ and $u_n \in \bar{u}(\nu_n)$. Due to boundedness of $u_n \in U_{ad}(0,1)$, there exists a subsequence denoted in the same way such that $u_n \rightharpoonup u$ in $L^s(I \times \omega)$ for some $s > 2$ as $n \to \infty$ with $u \in U_{ad}(0,1)$. Let $\tilde{u} \in U_{ad}(0,1)$ denote a minimizer of (5.1). Affine linearity of $u \mapsto S(\nu, u)$ for fixed ν, weak lower semi continuity of $\|\cdot\|$, and optimality of (ν_n, u_n) imply

$$\begin{aligned} f(\nu, u) &\le \liminf_{n\to\infty} f(\nu, u_n) \le \limsup_{n\to\infty} f(\nu_n, u_n) + \limsup_{n\to\infty} \left[f(\nu, u_n) - f(\nu_n, u_n)\right] \\ &\le \limsup_{n\to\infty} f(\nu_n, \tilde{u}) = f(\nu, \tilde{u}), \end{aligned}$$

where we have used Proposition 5.5 in the second last step. Hence, the weak limit u is also a minimizer of (5.1), *i.e.* $u \in \bar{u}(\nu)$.

Optimality of the tuples (ν, u) and (ν_n, u_n) leads to

$$\begin{aligned} f(\nu_n, u_n) - f(\nu, u_n) &\le f(\nu_n, u_n) - f(\nu, u_n) + f(\nu, u_n) - f(\nu, u) \\ &= \delta(\nu_n) - \delta(\nu) \le f(\nu_n, u) - f(\nu, u). \end{aligned}$$

Without restriction suppose that $\tau_n > 0$ for all $n \in \mathbb{N}$. Dividing the above chain of inequalities by τ_n, we infer that

$$\tau_n^{-1}\left[f(\nu_n, u_n) - f(\nu, u_n)\right] \le \tau_n^{-1}\left[\delta(\nu_n) - \delta(\nu)\right] \le \tau_n^{-1}\left[f(\nu_n, u) - f(\nu, u)\right]. \tag{5.8}$$

The right-hand side of (5.8) converges to $\partial_\nu f(\nu, u)$ due to differentiability of the control-to-state mapping. Concerning the left-hand side, we first observe that

$$\tau_n^{-1}\left[f(\nu_n, u_n) - f(\nu, u_n)\right] = \partial_\nu f(\nu + \theta_n, u_n) = \int_0^1 \langle B u_n - A y_n, p_n\rangle \,\mathrm{d}t$$

with $\theta_n \to 0$, $y_n = S(\nu + \theta_n, u_n)$, and p_n the associated adjoint state with terminal value $y_n(1) - y_d$. Convergence of $\nu_n \to \nu$, weak convergence of $u_n \rightharpoonup u$, and compactness of $(\nu, u) \mapsto S(\nu, u)$ from $\mathbb{R}_+ \times L^s(I \times \omega)$ to $C([0,1]; H)$, see Proposition A.19 of [2], yields $p_n \to p$ in $W(0,1)$. Hence

$$\lim_{n\to\infty} \int_0^1 \langle B u_n - A y_n, p_n\rangle \,\mathrm{d}t = \int_0^1 \langle B u - A y, p\rangle \,\mathrm{d}t.$$

In summary, this proves

$$\lim_{n\to\infty} \tau_n^{-1}\left[\delta(\nu_n) - \delta(\nu)\right] = \int_0^1 \langle B u - A y, p\rangle \,\mathrm{d}t. \tag{5.9}$$

We have to argue that the limit is independent of the chosen subsequence. To this end, we first observe that

$$\delta(\nu_n) - \delta(\nu) \le f(\nu_n, \tilde{u}) - f(\nu, \tilde{u})$$

for any minimizer $\tilde{u}$ of (5.1). Hence, dividing the inequality above by τ_n and passing to the limit implies the additional estimate

$$\lim_{n\to\infty} \tau_n^{-1} \left[\delta(\nu_n) - \delta(\nu)\right] \le \int_0^1 \langle B\tilde{u} - A\tilde{y}, p\rangle \,\mathrm{d}t, \tag{5.10}$$

where $\tilde{y} = S(\nu, \tilde{u})$. Recall that the adjoint state p is unique due to uniqueness of the observation. Let u'_n denote another subsequence of u_n with weak limit $u' \in U_{ad}(0,1)$ and associated times $\nu'_n = \nu + \tau'_n$. Repeating the arguments above we obtain

$$\int_0^1 \langle Bu' - Ay', p'\rangle \,\mathrm{d}t = \lim_{n'\to\infty} \tau_n'^{-1} \left[\delta(\nu'_n) - \delta(\nu)\right] \le \int_0^1 \langle B\tilde{u} - A\tilde{y}, p\rangle \,\mathrm{d}t \tag{5.11}$$

for any minimizer $\tilde{u}$ of (5.1). Now, combining (5.9), as well as (5.10) with $\tilde{u} = u'$, and (5.11) with $\tilde{u} = u$ yields

$$\begin{aligned}\int_0^1 \langle Bu - Ay, p\rangle \,\mathrm{d}t = \lim_{n\to\infty} \tau_n^{-1} \left[\delta(\nu_n) - \delta(\nu)\right] &\le \int_0^1 \langle Bu' - Ay', p\rangle \,\mathrm{d}t \\ = \lim_{n'\to\infty} \tau_n'^{-1} \left[\delta(\nu'_n) - \delta(\nu)\right] &\le \int_0^1 \langle Bu - Ay, p\rangle \,\mathrm{d}t.\end{aligned}$$

Hence, equality must hold and we conclude that the limit is independent of the chosen subsequence. Taking the infimum in the inequalities above implies

$$\mathrm{d}^{\pm}\delta(\nu) = \inf_{u\in\bar{u}(\nu)} \pm \int_0^1 \langle Bu - Ay, p\rangle \,\mathrm{d}t.$$

By standard arguments we can show that the infimum exists and we conclude (5.7).

Clearly, if the integral expression in (5.7) is independent of u, then $\delta(\cdot)$ is differentiable at ν. To show continuity of $\delta'(\cdot)$, let $\nu_n \in \mathbb{R}_+$ with $\nu_n \to \nu$. Moreover, let $u_n \in \bar{u}(\nu_n)$ such that u_n minimizes the expression (5.7) for $\nu = \nu_n$. As in the beginning of the proof, there exists a subsequence converging weakly to $u \in U_{ad}(0,1)$ that is a minimizer of (5.1). Compactness of the control-to-state mapping from $\mathbb{R}_+ \times L^s(I\times\omega)$ to $C([0,1];H)$, see Proposition A.19 of [2], as before leads to

$$\lim_{n\to\infty} \mathrm{d}^{\pm}\delta(\nu_n) = \lim_{n\to\infty} \pm \int_0^1 \langle Bu_n - Ay_n, p_n\rangle \,\mathrm{d}t = \pm \int_0^1 \langle Bu - Ay, p\rangle \,\mathrm{d}t$$

where $y_n = S(\nu_n, u_n)$ and $y = S(\nu, u)$ with p_n and p denoting the associated adjoint states. Hence, we conclude that $\delta'(\cdot)$ is continuous. □

If $\bar{u}(\nu)$ is a singleton, which can be guaranteed under the unique continuation property (see Prop. 5.2), then we immediately deduce that $\delta(\cdot)$ is continuously differentiable.

Corollary 5.6. *If the unique continuation property* (5.6) *holds, the integral expression in* (5.7) *is independent of $u \in \bar{u}(\nu)$. In particular, $\delta(\cdot)$ is continuously differentiable.*

Moreover, in the case of purely time-dependent controls, the expression for the derivative is independent of the concrete minimizer $u \in \bar{u}(\nu)$, even for multiple optimal controls.

Proposition 5.7. *In the case of purely time-dependent controls (i.e. $\omega = \{1, 2, \ldots, N_c\}$ equipped with the counting measure in Assumption 2.2), the integral expression in* (5.7) *is independent of $u \in \bar{u}(\nu)$. In particular, $\delta(\cdot)$ is continuously differentiable.*

Proof. We consider the splitting

$$\int_0^1 \langle Bu - Ay, p\rangle \,\mathrm{d}t = \int_0^1 \langle Bu, p\rangle - \langle Ay_1, p\rangle - \langle Ay_2, p\rangle \,\mathrm{d}t$$

with $y_1 = S(\nu, 0)$ and $y_2 = y - y_1$. Recall that the adjoint state p is independent of u, due to uniqueness of the observation. Hence, the optimality condition (5.4) for $u \in \bar{u}(\nu)$ implies that the first summand is independent of u. Moreover, the second summand is independent of u, because y_1 depends on the initial state y_0 and the time ν, only. For the remaining summand, the variation of constants formula yields

$$\begin{aligned}\langle Ay_2(t), p(t)\rangle &= \nu \left\langle A \int_0^t \mathrm{e}^{-\nu(t-s)A} Bu(s)\,\mathrm{d}s, \mathrm{e}^{-\nu(1-t)A^*} p(1)\right\rangle \\ &= \nu \int_0^t \langle Bu(s), A^* \mathrm{e}^{-\nu(t-s)A^*} \mathrm{e}^{-\nu(1-t)A^*} p(1)\rangle \,\mathrm{d}s = \int_0^t \langle Bu(s), \nu A^* p(s)\rangle \,\mathrm{d}s,\end{aligned}$$

where we have used the identity $(\mathrm{e}^{-\cdot A})^* = \mathrm{e}^{-\cdot A^*}$, see Corollary 1.10.6 of [28], the fact that the semigroup commutes with its generator, see Theorem 1.2.4 of [28], and the semigroup property. Hence, Fubini's theorem and the definition of p imply

$$\begin{aligned}\int_0^1 \langle Ay_2(t), p(t)\rangle \,\mathrm{d}t &= \int_0^1 \int_0^1 \mathbb{1}_{[0,t]}(s)\langle Bu(s), \nu A^* p(s)\rangle \,\mathrm{d}s\,\mathrm{d}t \\ &= \int_0^1 \int_0^1 \mathbb{1}_{[s,1]}(t)\langle Bu(s), \nu A^* p(s)\rangle \,\mathrm{d}t\,\mathrm{d}s \\ &= \int_0^1 (1-s)\langle Bu(s), \nu A^* p(s)\rangle \,\mathrm{d}s = \int_0^1 (1-s)\langle u(s), B^* \partial_t p(s)\rangle \,\mathrm{d}s.\end{aligned}$$

Since ω is discrete, we can identify $(B^*p)(i) = (B^*p)_i$ for $i \in \{1, 2, \ldots, N_c\}$. If $(B^*p)_i$ vanishes on a set with nonzero measure for some $i \in \{1, 2, \ldots, N_c\}$, then it has to vanish on $(0, 1)$ due to analyticity of the semigroup generated by $-A^*$. Thus, $(B^*\partial_t p)_i = \partial_t (B^*p)_i = 0$ on $(0, 1)$. Due to uniqueness of the adjoint state p and the fact that only those components of u are not uniquely determined where B^*p vanishes (see optimality condition (5.4)) we conclude that the above expression is independent of u. Last, the second assertion follows from the first and Theorem 5.4. ☐

5.3. Lipschitz continuity of $\delta'(\cdot)$

Last, we consider a sufficient condition for Lipschitz continuity of $\delta'(\cdot)$, which in turn guarantees fast local convergence of the Newton method. Let $\nu \in \mathbb{R}_+$, $u \in \bar{u}(\nu) \subset U_{ad}(0, 1)$, and let p denote the corresponding adjoint state. We say that the structural assumption holds at $\bar{u}(\nu)$, if there exists a $C > 0$ such that

$$|\{ (t, x) \in I \times \omega \colon -\varepsilon \le (B^*p)(t, x) \le \varepsilon \}| \le C\varepsilon \tag{5.12}$$

for all $\varepsilon > 0$. Since (5.12) implies that $\bar{u}(\nu)$ is a singleton, see Proposition 5.1, it is justified to say that (5.12) holds at $\bar{u}(\nu)$.

Proposition 5.8. *Let $\nu \in \mathbb{R}_+$ and suppose that (5.12) holds at $\{ u \} = \bar{u}(\nu)$. Then*

$$\partial_u f(\nu, u)(u' - u) \ge \nu c_0 \|u' - u\|^2_{L^1(I\times\omega)} \quad \text{for all } u' \in U_{ad}(0, 1), \tag{5.13}$$

where $c_0 = (2\|u_b - u_a\|_{L^\infty(\omega)} C)^{-1}$.

Proof. The proof can be obtained along the lines of Proposition 2.7 from [5]. ☐

For the following considerations, we assume that the adjoint states $p(\nu_1)$ and $p(\nu_2)$ associated with the time transformations ν_1 and ν_2 and the states $i_1S(\nu_1,u)$ and $i_1S(\nu_2,u)$ satisfy

$$\|B^*\left(p(\nu_1)-p(\nu_2)\right)\|_{L^\infty(I\times\omega)} \le c|\nu_1-\nu_2| \tag{5.14}$$

for all $\nu_1,\nu_2 \in [\nu_{\min},\nu_{\max}]$ and all $u\in U_{ad}(0,1)$, where $0<\nu_{\min}<\nu_{\max}$ are constants. The stability estimate (5.14) holds in case of purely-time dependent controls under the general conditions of this article. Moreover, the estimate can be shown in case of a distributed control for fairly general elliptic operators and spatial domains; see *e.g.*, Proposition A.3 of [4].

Proposition 5.9. *Suppose that* (5.14) *is valid and let* $\bar\nu\in\mathbb{R}_+$ *with* $\{\,\bar u\,\}=\bar u(\bar\nu)$*. If* (5.13) *holds at* $(\bar\nu,\bar u)$ *for some constant* $c_0>0$*, then there is* $\delta>0$ *such that*

$$\|u-\bar u\|_{L^1(I\times\omega)} \le c|\nu-\bar\nu| \quad \textit{for all } \nu\in\mathbb{R}_+, |\nu-\bar\nu|\le\delta,\ \textit{and } u\in\bar u(\nu),$$

with $c>0$ *a constant independent of* ν *and* u*.*

Proof. Let $u\in\bar u(\nu)$ and let $p(\nu,u)$ denote the associated adjoint state. Employing Proposition 5.8 with $u'=u$ and the first order necessary optimality condition (5.3) for u yield

$$\begin{aligned}\bar\nu c_0\|u-\bar u\|^2_{L^1(I\times\omega)} &\le \partial_u f(\bar\nu,\bar u)(u-\bar u) \le \partial_u f(\bar\nu,\bar u)(u-\bar u) - \bar\nu(B^*p(\nu,u),u-\bar u)_{L^2(I\times\omega)}\\ &= \bar\nu(B^*\left(\bar p - p(\bar\nu,u)\right),u-\bar u)_{L^2(I\times\omega)} + \bar\nu(B^*\left(p(\bar\nu,u)-p(\nu,u)\right),u-\bar u)_{L^2(I\times\omega)},\end{aligned}$$

where $p(\bar\nu,u)$ denotes the adjoint state associated with $\bar\nu$ and u. Concerning the first term on the right-hand side we observe

$$\bar\nu(B^*\left(\bar p-p(\bar\nu,u)\right),u-\bar u)_{L^2(I\times\omega)} = -\bar\nu\|i_1\left(\partial_t+\bar\nu A\right)^{-1}B(u-\bar u)\|^2 \le 0,$$

where $\left(\partial_t+\bar\nu A\right)^{-1}$ denotes the solution operator to the linear parabolic state equation with zero initial value. Thus, Hölder's inequality implies

$$c_0\|u-\bar u\|_{L^1(I\times\omega)} \le \|B^*\left(p(\bar\nu)-p(\nu)\right)\|_{L^\infty(I\times\omega)}.$$

Finally, we apply the stability estimate (5.14) to conclude the proof. □

Applying Proposition 5.9 twice, we immediately infer the following Lipschitz type estimate.

Corollary 5.10. *There are* $\delta>0$ *and* $c>0$ *such that*

$$\|u_1-u_2\|_{L^1(I\times\omega)} \le c|\nu_1-\nu_2| \quad \textit{for all } u_1\in\bar u(\nu_1) \textit{ and } u_2\in\bar u(\nu_2), \tag{5.15}$$

and all $\nu_1\in[\bar\nu-\delta,\bar\nu]$ *and* $\nu_2\in[\bar\nu,\bar\nu+\delta]$*.*

Moreover, if $B\in\mathcal{L}(L^1(\omega),H)$ then the control-to-state mapping is continuous from $L^1(I\times\omega)$ to $C([0,1];H)$ and we infer the following result.

Corollary 5.11. *If* $B\in\mathcal{L}(L^1(\omega),H)$*, then there are* $\delta>0$ *and* $c>0$ *such that*

$$|\delta'(\nu_1)-\delta'(\nu_2)| \le c|\nu_1-\nu_2|$$

for all $\nu_1\in[\bar\nu-\delta,\bar\nu]$ *and* $\nu_2\in[\bar\nu,\bar\nu+\delta]$*.*

6. Algorithm

We now turn to the algorithmic solution of (P). Throughout the rest of this article we assume that $T(\cdot)$ is left-continuous. In view of Theorem 3.1, we are interested in finding a root of the value function $\delta(\cdot)$ in order to solve the time-optimal control problem (P). This will generally lead to a bi-level optimization problem: The outer loop finds the optimal T and the inner loop determines for each given T a control such that the associated state has a minimal distance to the target set. It is worth mentioning that this procedure will find a global solution to (P) provided that we initiate the outer optimization with a time smaller than the optimal one.

6.1. Newton method for the outer minimization

To find a root of the value function, we apply the Newton method. As this requires $\delta(\cdot)$ to be continuously differentiable, we require the following assumption. Recall that Assumption 6.1 automatically holds in the case of purely time-dependent controls and for the linear heat-equation on a bounded domain with distributed control; see Corollary 5.6 and Proposition 5.7.

Assumption 6.1. *Suppose that the integral expression in (5.7) is independent of the concrete minimizer $u \in \bar{u}(\nu)$ for all $\nu \in \mathbb{R}_+$ with $\delta(\nu) > -\delta_0$.*

For well-posedness of the method, we have to guarantee that $\delta'(\bar{\nu}) \neq 0$. The following result underlines the practical relevance of qualified optimality conditions for ($\hat{P}$) in the context of its algorithmic solution.

Proposition 6.2. *Let $(\bar{\nu}, \bar{u}) \in \mathbb{R}_+ \times U_{ad}(0,1)$ be a solution to ($\hat{P}$). The first order optimality conditions of Lemma 4.2 hold in qualified form if and only if $\delta'(\bar{\nu}) \neq 0$.*

Proof. According to the general form of the optimality conditions of Lemma 4.2 there exist $\bar{\mu} > 0$ and $\bar{\mu}_0 \in \{0, 1\}$ such that

$$\int_0^1 \langle B\bar{u} - A\bar{y}, \bar{p}\rangle \, \mathrm{d}t = -\bar{\mu}_0,$$

where $\bar{p} \in W(0,1)$ is the adjoint state with terminal value $\bar{\mu}\,(\bar{y}(1) - y_d)$. Hence, linearity of the expression above and (5.7) imply $\delta'(\bar{\nu}) = -\bar{\mu}^{-1}\bar{\mu}_0 \|\bar{y}(1) - y_d\|^{-1}$. Thus, qualified optimality conditions (*i.e.* $\bar{\mu}_0 = 1$) hold if and only if $\delta'(\bar{\nu}) \neq 0$. □

The resulting Newton method is summarized in Algorithm 1. By means of Theorem 5.4 and well-known properties of the Newton method, see *e.g.*, Theorem 11.2 of [26], we obtain the following convergence result.

Proposition 6.3. *Let $\bar{\nu} \in \mathbb{R}_+$ and suppose that Assumption 6.1 holds. If $\delta'(\bar{\nu}) \neq 0$, then the sequence ν_n generated by Algorithm 1 converges locally q-superlinearly to $\bar{\nu}$.*

Algorithm 1: Newton method for solution of minimal distance problem (outer loop).

```
Choose ν_0 > 0;
for n = 0 to n_max do
    Calculate u ∈ ū(ν_n) using Algorithm 2 and y = S(ν_n, u);
    if δ(ν_n) < ε_tol then
        return;
    end
    Evaluate δ'(ν_n) using (5.7);
    Set ν_{n+1} = ν_n − δ(ν_n)δ'(ν_n)^{-1};
end
```

If we in addition assume that the control operator is bounded from L^1 into H, then the variation of constants formula implies that the control-to-state mapping is linear and continuous from $L^1(I \times \omega)$ to $C([0,1]; H)$ for any fixed $\nu \in \mathbb{R}_+$. Hence, if the structural assumption (5.12) on the adjoint state holds, we immediately obtain the following fast convergence result.

Proposition 6.4. *Let $\bar{\nu} \in \mathbb{R}_+$ and suppose that Assumption 6.1 holds. Moreover, assume that (5.12) holds at $\bar{u}(\bar{\nu})$ and that $B \in \mathcal{L}(L^1(\omega), H)$. If $\delta'(\bar{\nu}) \neq 0$, then the sequence ν_n generated by Algorithm 1 converges locally q-quadratically to $\bar{\nu}$.*

Proof. First, Proposition 6.3 guarantees q-linear convergence of the sequence ν_n. The improved convergence rate follows from Lipschitz continuity of $\delta'(\cdot)$, see Corollary 5.11, and well-known properties of the Newton method; see *e.g.*, Theorem 11.2 of [26]. Note that the Lipschitz type estimate of Corollary 5.11 is sufficient for the proof of Theorem 11.2 from [26]. □

Remark 6.5. For convenience we summarize that under Assumption 6.1 (which implies that $\delta(\cdot)$ is continuously differentiable) and if qualified optimality conditions hold for ($\hat{P}$) (which implies that $\delta'(\cdot)$ is nonzero near the optimal solution), the Newton method for finding a root of $\delta(\cdot)$ is well-defined. If in addition, $T(\cdot)$ is left-continuous, then the root of $\delta(\cdot)$ is the optimal time for the time-optimal control problem ($\hat{P}$).

6.2. Conditional gradient method for the inner optimization

For the algorithmic solution of the inner problem, *i.e.* the determination of $\bar{u}(\nu)$ in (5.1), we employ the conditional gradient method; see *e.g.*, [7]. We abbreviate

$$f(u) = \|i_1 S(\nu, u) - y_d\|$$

neglecting the ν dependence for a moment. Clearly, we are interested in minimizing f over $U_{ad}(0,1)$. As in Section 5.1, we have

$$f'(u)^* = \nu B^* p,$$

where $p \in W(0,1)$ solves (5.2) with $y = S(\nu, u)$. Given $u_n \in U_{ad}(0,1)$, we take

$$u_{n+1/2} = \begin{cases} u_a, & \text{if } B^* p_n > 0, \\ u_b, & \text{if } B^* p_n < 0, \\ (u_a + u_b)/2, & \text{else}, \end{cases} \tag{6.1}$$

almost everywhere. This choice guarantees that

$$f'(u_n)(u_{n+1/2} - u_n) = \min_{u \in U_{ad}(0,1)} f'(u_n)(u - u_n).$$

The next iterate u_{n+1} is defined by the optimal convex combination of u_n and $u_{n+1/2}$. Precisely, we take $u_{n+1} = (1 - \lambda^*) u_n + \lambda^* u_{n+1/2}$ with

$$\lambda^* = \operatorname*{argmin}_{0 \le \lambda \le 1} f((1 - \lambda) u_n + \lambda u_{n+1/2}). \tag{6.2}$$

This expression can be analytically determined, employing the fact that $u \mapsto S(\nu, u)$ is affine linear. Using the convexity of f and the definition of $u_{n+1/2}$, we immediately derive the following a posteriori error estimator

$$0 \le f(u_n) - f(\bar{u}) \le f'(u_n)(u_n - \bar{u}) \le \max_{u \in U_{ad}(0,1)} f'(u_n)(u_n - u) = f'(u_n)(u_n - u_{n+1/2}).$$

The expression on the right-hand side can be efficiently evaluated using the adjoint representation and serves as a termination criterion for the conditional gradient method. The algorithm for the inner optimization is summarized in Algorithm 2.

The conditional gradient method has the following convergence properties.

Algorithm 2: Conditional gradient method for solution of (5.1).

Let $\nu > 0$ be given. Choose $u_0 \in U_{ad}(0,1)$;
for $n = 0$ **to** $n_{\max}$ **do**
  Calculate $y_n = S(\nu, u_n)$ and p_n;
  Choose $u_{n+1/2}$ as in (6.1);
  if $f'(u_n)(u_n - u_{n+1/2}) < \varepsilon_{tol}$ **then**
    return;
  end
  Calculate λ^* as in (6.2);
  Set $u_{n+1} = (1-\lambda^*)u_n + \lambda^* u_{n+1/2}$;
end

Proposition 6.6. *Let $(u_n)_n$ be a sequence generated by the conditional gradient method. Then $f(u_n)$ decreases monotonically and*

$$0 \le f(u_n) - f(\bar{u}) \le \frac{f(u_0) - f(\bar{u})}{1 + cn}, \quad n \ge 0,$$

with a constant c exclusively depending on the Lipschitz constant of f' on $U_{ad}(0,1)$, the initial residuum, and U_{ad}.

Proof. This follows from Theorem 3.1 (i) of [7], since both f and $U_{ad}(0,1)$ are convex. □

If the control operator B defines a bounded operator from $L^1(\omega)$ to H, then under the structural assumption (5.12) on the adjoint state, the objective values converges q-linearly.

Proposition 6.7. *Suppose that $B \in \mathcal{L}(L^1(\omega), H)$. If (5.12) holds at $\bar{u}(\nu)$, then there is $\lambda \in [1/2, 1)$ such that*

$$0 \le f(u_n) - f(\bar{u}) \le [f(u_0) - f(\bar{u})]\,\lambda^n, \quad n \ge 0. \tag{6.3}$$

The constant λ exclusively depends on C, u_a, u_b, ω, and the Lipschitz constant of f' on $U_{ad}(0,1)$. Moreover, for a constant $c > 0$ we have

$$\|u_n - \bar{u}\|_{L^1(I\times\omega)} \le c\lambda^{n/2}, \quad n \ge 0. \tag{6.4}$$

Proof. Since $B\colon L^1(\omega) \to H$, the variation of constants formula implies that the control-to-state mapping is linear and continuous from $L^1(I \times \omega)$ to $C([0,1]; H)$. Hence, f as a mapping defined on $L^1(I \times \omega)$ is (infinitely often) continuously differentiable. Furthermore, Proposition 5.8 implies

$$f'(\bar{u})(u - \bar{u}) \ge c_0 \nu \|u - \bar{u}\|^2_{L^1(I\times\omega)} \quad \text{for all } u \in U_{ad}(0,1),$$

for some constant $c_0 > 0$. Therefore, (6.3) follows from Theorem 3.1 (iii) of [7]. Finally, convexity of f and the inequality above yield (6.4). □

6.3. Accelerated conditional gradient method for the inner optimization

Since the criterion from Proposition 6.7 guaranteeing q-linear convergence of the conditional gradient method is not satisfied for many examples and we in fact observe slow convergence in practice, we employ an acceleration strategy that is described in the following: Instead of minimizing the convex combination of the last iterate u_n and the new point $u_{n+1/2}$ in (6.2), we search for the best convex combination of all previous iterates plus the new point $u_{n+1/2}$. Concretely, instead of (6.2) we determine λ^* as

$$\lambda^* = \operatorname*{argmin}_{\lambda \in \mathbb{P}_{n+2}} f\left(\sum_{i=0}^{n} \lambda_i u_i + \lambda_{n+1} u_{n+1/2}\right), \tag{6.5}$$

where $\mathbb{P}_{n+2} = \{ \lambda \in \mathbb{R}^{n+2} \colon \lambda_i \geq 0 \text{ and } \sum_{i=0}^{n+1} \lambda_i = 1 \}$ denotes the probability simplex in $\mathbb{R}^{n+2}$. The next iterate is then defined as $u_{n+1} = \sum_{i=0}^{n} \lambda_i^* u_i + \lambda_{n+1}^* u_{n+1/2}$. In order to derive an efficient algorithm for the determination of λ^*, we first reformulate the optimality condition associated with (6.5) employing the normal map due to Robinson [31]. To this end, let us abbreviate $h(\lambda) \coloneqq f(\sum_{i=0}^{n} \lambda_i u_i + \lambda_{n+1} u_{n+1/2})$. For any $c > 0$ we define Robinson's normal map as

$$G(\eta) = c(\eta - \Pi(\eta)) + \nabla h(\Pi(\eta)),$$

where Π denotes the projection onto $\mathbb{P}_{n+2}$. Due to convexity of h, which follows immediately from the convexity of f, an optimal solution of (6.5) can be characterized by means of the normal map as follows; *cf.* Proposition 3.5 of [29].

Proposition 6.8. *$\lambda^* \in \mathbb{P}_{n+2}$ is optimal for (6.5) if and only if there exists $\eta \in \mathbb{R}^{n+2}$ such that $G(\eta) = 0$ and $\lambda^* = \Pi(\eta)$.*

Proof. First of all, λ^* is optimal for (6.5) if and only if $h'(\lambda^*)(\lambda - \lambda^*) \geq 0$ for all $\lambda \in \mathbb{P}_{n+1}$, because h is convex. Let $\lambda^* \in \mathbb{P}_{n+2}$ be optimal and set $\eta = \lambda^* - c^{-1} \nabla h(\lambda^*)$. Then we have

$$(\eta - \lambda^*, \lambda - \lambda^*)_{\mathbb{R}^{n+2}} = -c^{-1} h'(\lambda^*)(\lambda - \lambda^*) \leq 0$$

for all $\lambda \in \mathbb{P}_{n+2}$. Hence, $\lambda^* = \Pi(\eta)$. Moreover, by construction $G(\eta) = 0$. Conversely, let η be given such that $G(\eta) = 0$ and $\lambda^* = \Pi(\eta)$. Then

$$-c(\eta - \lambda^*, \lambda - \lambda^*)_{\mathbb{R}^{n+2}} = h'(\lambda^*)(\lambda - \lambda^*) \geq 0$$

for all $\lambda \in \mathbb{P}_{n+2}$. Thus, λ^* is optimal. $\square$

In view of Proposition 6.8, to determine λ^* defined in (6.5), we can equivalently solve the nonlinear equation $G(\eta) = 0$ for η and obtain the optimal solution λ^* by projecting η onto the probability simplex, *i.e.* $\lambda^* = \Pi(\eta)$. We propose to solve the equation $G(\eta) = 0$ by means of a semi-smooth Newton method. Note that the projection Π and its derivative $D\,\Pi$ can be efficiently evaluated (with cost $\mathcal{O}(n \log n)$), see Algorithm 4, where we have extended the algorithm from [34] by adding the derivative $D\,\Pi$. The resulting semi-smooth Newton method is summarized in Algorithm 3.

Algorithm 3: Semi-smooth Newton method for solution of (6.5).

Choose $c > 0$ and $\eta \in \mathbb{R}^{n+2}$;
Set $\lambda = \Pi(\eta)$;
while $\|c(\eta - \lambda) + \nabla h(\lambda)\|_{\mathbb{R}^{n+2}} > \varepsilon_{tol}$ **do**
  Calculate $\xi = \big(c(\mathrm{Id} - D\,\Pi(\eta)) + \nabla^2 h(\lambda) D\,\Pi(\eta)\big)^{-1} (c(\eta - \lambda) + \nabla h(\lambda))$;
  Set $\eta = \eta - \xi$ and $\lambda = \Pi(\eta)$;
end

The accelerated conditional gradient method is exactly Algorithm 2 except for the last two lines: The parameter λ^* is determined using Algorithm 3 (in contrast to the standard conditional gradient method where we could calculate λ^* explicitly). Moreover, in the last line we set $u_{n+1} = \sum_{i=0}^{n} \lambda_i^* u_i + \lambda_{n+1}^* u_{n+1/2}$. We note that the accelerated version is at least as fast as the standard conditional gradient method, because the feasible set from (6.2) is contained in (6.5). In practice we observe that the acceleration strategy significantly improves the performance.

Algorithm 4: Projection Π onto the probability simplex $\mathbb{P}_n$ and its derivative.

Input: $y \in \mathbb{R}^n$
Sort y such that $y_{\pi(1)} \geq y_{\pi(2)} \geq \ldots \geq y_{\pi(n)}$;
Find $\rho = \max\{\, 1 \leq j \leq n \colon y_{\pi(j)} + \frac{1}{j}\left(1 - \sum_{i=1}^{j} y_{\pi(i)}\right) > 0 \,\}$;
Define $\lambda = \frac{1}{\rho}\left(1 - \sum_{i=1}^{\rho} y_{\pi(i)}\right)$;
Set $\Gamma = (\gamma_{i,j})_{i,j} \in \mathbb{R}^{n\times n}$ with $\gamma_{i,i} = 1$ if $x_i + \lambda > 0$ and $\gamma_{i,j} = 0$ otherwise for $1 \leq i, j \leq n$;
Set $x = (x_i)_i \in \mathbb{R}^n$ with $x_i = \max\{\, y_i + \lambda, 0 \,\}$ for $1 \leq i \leq n$;
Set $\Lambda = (\lambda_{i,j})_{i,j} \in \mathbb{R}^{n\times n}$ with $\lambda_{i,j} = -1/\rho$ if $j = \pi(k)$ for some $1 \leq k \leq \rho$ and $\lambda_{i,j} = 0$ otherwise for $1 \leq i, j \leq n$;
Output: $\Pi(y) = x$ and $D\,\Pi(y) = \Gamma\,(\mathrm{Id} + \Lambda)$

7. Numerical examples

As a proof of concept, we implement numerical examples illustrating that the proposed algorithm can be realized in practice. We begin with one example governed by an ordinary differential equation, even though our main focus are systems subject to partial differential equations.

Since the value function $\delta(\cdot)$ can be non-convex, we consider the damped Newton method. If $\delta(\nu_{n+1}) < -\varepsilon_{\mathrm{tol}}$, then the Newton step is iteratively multiplied by the damping factor $\gamma = 0.9$ until $\delta(\nu_{n+1}) > -\varepsilon_{\mathrm{tol}}$. Note that this strategy does not require the inner problem to be solved with high accuracy. If a feasible control with sufficiently negative value for $\delta(\cdot)$ is known, then the conditional gradient method can be restarted with a smaller Newton step.

Moreover, we have implemented the acceleration strategy from Section 6.3 for the conditional gradient method. To keep the memory requirements moderate, points that are associated with small coefficients in the convex combination are being removed from the list of former iterates. In our examples, this strategy significantly improves the convergence.

7.1. Linearized pendulum

We first consider a time-optimal control example subject to an ordinary differential equation from Example 17.2 of [13]. The operators A and B are given by the matrices

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Hence, we set $V = H = V^* = \mathbb{R}^2$ and $Q = \mathbb{R}^1$. Moreover, the control constraints are $u_a = -1$ and $u_b = 1$, and the desired state is $y_d = 0$. The corresponding state equation describes a harmonic oscillator, precisely the linearized pendulum $\ddot{x} + x = u$ with forcing term u. Note that the system is normal, so (5.1) possesses a unique minimizer; see Proposition 5.2 and Remark 5.3. As shown in Example 17.2 from [13], the optimal trajectories for $\delta_0 = 0$ can be constructed geometrically. For example, if

$$y_0 = -r\,(\cos(\pi/3 - \theta_0), \sin(\pi/3 - \theta_0))^T + (1, -3)^T, \quad \theta_0 = \arcsin(1/r), \quad r = \sqrt{17},$$

then the optimal trajectory consists of three semi circles with $\theta = \pi/3$ and center $(1,0)^T$, $\theta = \pi$ and center $(-1,0)^T$, and $\theta = \pi/2$ and center $(1,0)^T$. In addition, the optimal time is $T = \pi(1/3 + 1 + 1/2) = 11\pi/6$, and the unique optimal control is given by

$$\bar{u}(t) = \begin{cases} 1 & \text{if } 0 \leq t \leq \pi/3, \\ -1 & \text{if } \pi/3 < t \leq 4\pi/3, \\ 1 & \text{if } 4\pi/3 < t \leq 11\pi/6. \end{cases}$$

The ordinary differential equation is discretized by means of the discontinuous Galerkin method with piecewise constant functions (corresponding to the implicit Euler method) for an equidistant time grid with M denoting

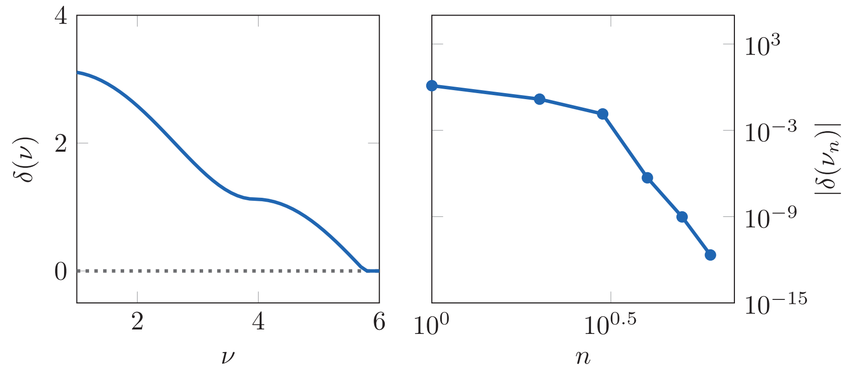

FIGURE 1. Value function δ (*left*) and absolute value of $\delta(\nu_n)$ (*right*) for ν_n the iterates generated by the Newton method (Algorithm 1) for Example 7.1 with $M = 10\,000$ time steps for the implicit Euler method.

TABLE 1. Computed optimal times, absolute errors, number of Newton steps in outer loop (number of damped steps in brackets), and number of conditional gradient steps in inner loop for Example 7.1 with M denoting the number of time steps. Moreover, $\delta_0 = 10^{-6}$ and the inital value for the Newton method is $\nu_0 = 0.6\,T$.

M	T_k	$\lvert T_k - T\rvert$	Newton steps	cG steps
100	5.501553	2.5803_{-1}	6 (1)	44
1000	5.730029	2.9557_{-2}	6 (1)	52
10 000	5.756636	2.9504_{-3}	6 (1)	58
100 000	5.759346	2.4051_{-4}	6 (1)	56
1 000 000	5.759618	3.1244_{-5}	6 (1)	53

the number of time intervals. To solve the problem with our approach, we consider a relaxation of the terminal constraint by taking $\delta_0 = 10^{-6}$. Since the solution is stable with respect to perturbations in the constraint, the relaxation has no significant influence on the optimal solution, as long as the error due to the discretization of the state equation dominates the overall error.

As depicted in Figure 1 we observe fast convergence of the Newton method. Moreover, the number of Newton steps in the outer loop and the number of iterations of the conditional gradient method in the inner loop seem to be essentially independent of the discretization of the state equation; see Table 1.

7.2. Linear heat-equation with distributed control

Next, we consider the following problem subject to the linear heat-equation. Let

$$\begin{aligned}\Omega = (0,1)^2, \quad \omega = (0.25, 0.75)^2, \quad \delta_0 = 1/10,\\ y_0(x) = 4\sin(\pi x_1^2)\sin(\pi x_2)^3, \quad y_d(x) = -2\min\{\, x_1, 1-x_1, x_2, 1-x_2 \,\},\\ U_{ad}(0,1) = \{\, u \in L^2(I\times\omega)\colon\ -5 \le u \le 0 \,\}.\end{aligned}$$

Moreover, $A = -0.03\Delta$ with $-\Delta$ the Laplace operator equipped with homogeneous Dirichlet boundary conditions. The control operator B is the extension by zero operator. Hence, we take $V = H_0^1(\Omega)$, $H = L^2(\Omega)$, $V^* = H^{-1}(\Omega)$, and $U = L^2(\omega)$. Note that the control acts on a subset $\omega \subset \Omega$, only. Concerning the practical implementation, we consider a discontinuous Galerkin method in time and a continuous Galerkin method in

TABLE 2. Computed optimal times, absolute errors, number of Newton steps in outer loop (number of damped steps in brackets), and number of conditional gradient steps in inner loop for Example 7.2 with M denoting the number of time steps and N the number of nodes for the spatial discretization. Moreover, the inital value for the Newton method is $\nu_0 = 0.8$.

M	N	T_k	$\lvert T_k - T\rvert$	Newton steps	cG steps
20	4225	1.573876	9.0814_{-2}	6 (0)	206
40	4225	1.525617	4.2554_{-2}	6 (0)	210
80	4225	1.501899	1.8836_{-2}	6 (0)	209
160	4225	1.490127	7.0645_{-3}	6 (0)	210
320	4225	1.484246	1.1830_{-3}	6 (0)	196
640	81	1.291744	1.9132_{-1}	6 (0)	133
640	289	1.423571	5.9492_{-2}	6 (0)	162
640	1089	1.468758	1.4304_{-2}	6 (0)	188
640	4225	1.481302	1.7608_{-3}	6 (0)	190

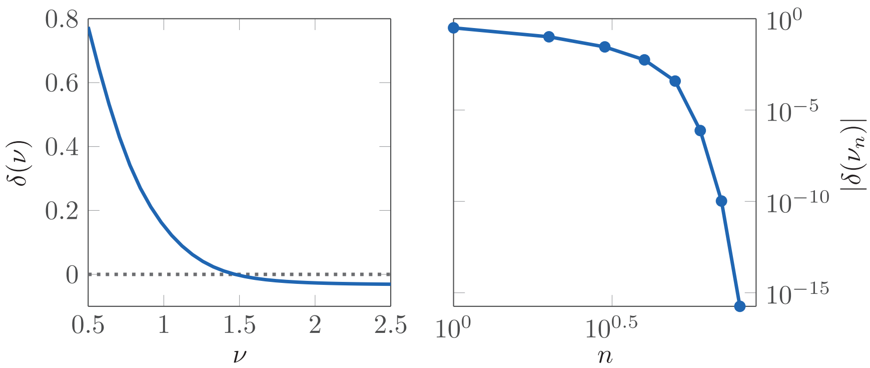

FIGURE 2. Value function δ (*left*) and absolute value of $\delta(\nu_n)$ (*right*) for ν_n the iterates generated by the Newton method (Algorithm 1) for Example 7.2.

space. The state and adjoint state equations are discretized by means of piecewise constant functions in time (corresponding to the implicit Euler method) and continuous and cellwise linear functions in space.

As in the first example, we observe fast convergence of the Newton method, see Table 2 and Figure 2. Moreover, we observe quadratic order of convergence with respect to the spatial discretization and linear order convergence with respect to the temporal discretization. For further details and *a priori* discretization error estimates we also refer to [4].

Before turning to the next example, we would like to compare the algorithm from Section 6 to an alternative approach, where the time-optimal control problem ($\hat{P}$) is solved directly after adding a regularization term to the objective functional, precisely the L^2-norm of the control variable. Clearly, we are interested in steering the regularization parameter to zero. The terminal constraint in ($\hat{P}$) is treated algorithmically by means of the augmented Lagrange method. The resulting optimization problems are solved by means of a semi-smooth Newton method in a monolithic way, *i.e.* we consider the tuple (ν, u) as a joint optimization variable; *cf.* [19] and Section 4.1 of [2]. We observe that our approach requires roughly four to ten times less solves of the PDE than the regularization approach for any fixed regularization parameter in the range from $\alpha = 0.001$ to $\alpha = 10$; see Table 3. Employing a path-following strategy, where one iteratively decreases the regularization parameter starting with a moderate value of α and uses the solution of the former iteration as the initial value for the next

Table 3. Number of PDE solves for the algorithm from Section 6 based on solving minimal distance problems and the augmented Lagrange method with L^2-regularization for the control and α the regularization parameter. M denotes the number of intervals for the temporal discretization and N the number of nodes for the spatial discretization of the PDE. The initial parameters for the augmented Lagrange method are $c_0 = 2 \times 10^4$ and $\mu_0 = 80$.

		Min. dist.	Augmented Lagrange method with regularization				
M	N		$\alpha = 0.001$	$\alpha = 0.01$	$\alpha = 0.1$	$\alpha = 1$	$\alpha = 10$
20	4225	260	2681	1598	1208	1375	1813
40	4225	250	2703	1447	1177	1440	1675
80	4225	250	2687	1424	1130	1434	1685
160	4225	252	3017	1697	1154	1388	1705
320	4225	222	3068	1587	1100	1356	1709
640	81	214	1879	1091	924	1074	1215
640	289	198	1824	1070	774	1180	1479
640	1089	212	2495	1412	1028	1372	1789
640	4225	246	3597	1647	1050	1364	1781

optimization (see *e.g.*, [14]), one could avoid the high computational costs for small α. However, this strategy requires at least one solution without warm start, so that our approach is (in this example) at least five times faster.

7.3. Linear heat-equation with Neumann boundary control

Last, we consider the following problem subject to the linear heat-equation with Neumann boundary control. Concretely, let

$$\Omega = (0,1)^2, \quad \omega = \partial\Omega, \quad \delta_0 = 1/10,$$
$$y_0(x) = 4\sin(\pi x_1^2)\sin(\pi x_2)^3, \quad y_d(x) = 0,$$
$$U_{ad}(0,1) = \{\, u \in L^2(I \times \omega)\colon\ -5 \le u \le 5 \,\}.$$

Moreover, $A = -0.03\Delta$ with $-\Delta$ the Laplace operator. The control operator B is the adjoint of the trace operator, *i.e.* $B = \mathrm{Tr}^*\colon L^2(\partial\Omega) \to (H^1(\Omega))^*$. Hence, we take $V = H^1(\Omega)$, $H = L^2(\Omega)$, $V^* = (H^1(\Omega))^*$, and $U = L^2(\omega)$. We consider the same discretization scheme for the state and adjoint state equation as before. Moreover, the control is discretized by edge-wise constant functions on the boundary.

The optimal control obtained numerically is depicted in Figure 3, where the boundary of the square domain has been unrolled. Note that switching hyperplanes of the control seem to accumulate towards the end of the time horizon. As in the preceding examples, we observe fast convergence of the Newton method for the outer loop; see Figure 4 and Table 4.

8. Open problems

We conclude with some open problems.

(i) To prove the equivalence of time-optimal and distance-optimal controls, we required that $T(\cdot)$ is left-continuous; see Theorem 3.1. We stated two sufficient conditions; see Propositions 3.5 and 3.6. The latter can be checked *a priori* without knowing an optimal solution, whereas the first depends on a certain controllability condition under pointwise control constraints that is difficult to verify. It would be desirable to know further sufficient conditions that can be easily verified for concrete problems.

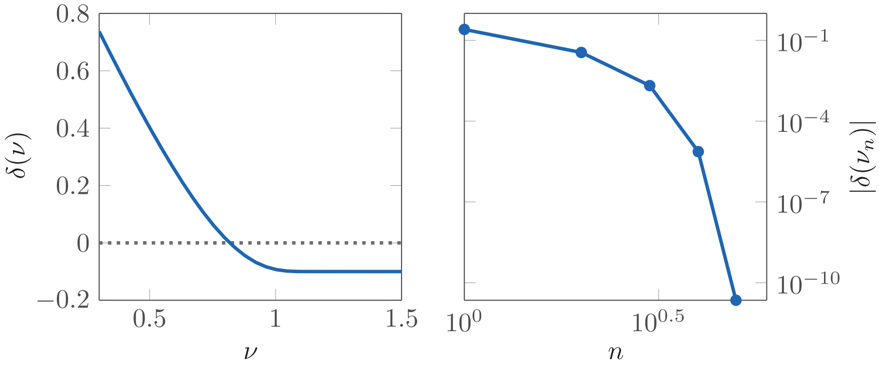

Figure 3. Optimal control for Example 7.3. Black denotes the upper bound and white the lower bound of the control constraints.

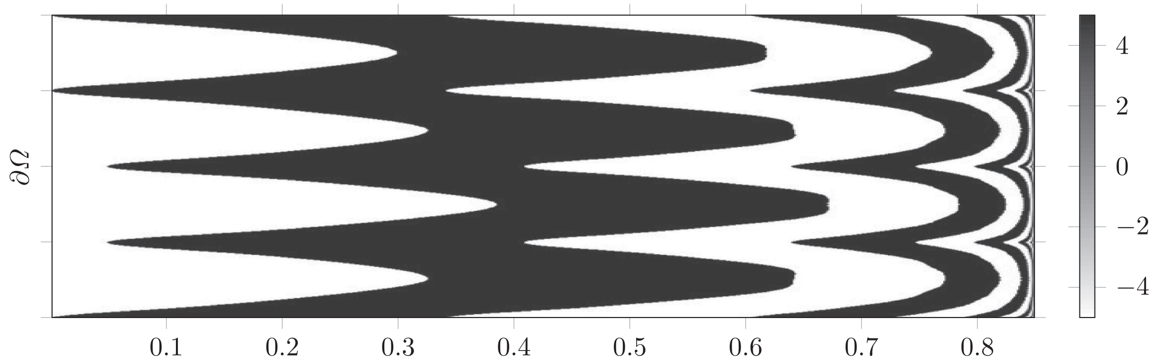

Figure 4. Value function δ (*left*) and absolute value of $\delta(\nu_n)$ (*right*) for ν_n the iterates generated by the Newton method (Algorithm 1) for Example 7.3.

Table 4. Computed optimal times, absolute errors, number of Newton steps in outer loop (number of damped steps in brackets), and number of conditional gradient steps in inner loop for Example 7.3 with M denoting the number of time steps and N the number of nodes for the spatial discretization. Moreover, the initial value for the Newton method is $\nu_0 = 0.6$.

M	N	T_k	$\lvert T_k - T\rvert$	Newton steps	cG steps
20	4225	8.8971_{-1}	4.0493_{-2}	4 (0)	3302
40	4225	8.6784_{-1}	1.8619_{-2}	4 (0)	4141
80	4225	8.5681_{-1}	7.5882_{-3}	4 (0)	4979
160	4225	8.5134_{-1}	2.1231_{-3}	4 (0)	6294
320	81	6.9746_{-1}	1.5176_{-1}	3 (0)	1938
320	289	8.1011_{-1}	3.9113_{-2}	4 (0)	5265
320	1089	8.4096_{-1}	8.2584_{-3}	4 (0)	8029
320	4225	8.4865_{-1}	5.7155_{-4}	4 (0)	8432

(ii) Moreover, to strengthen the directional derivative of $\delta(\cdot)$ to a classical derivative, one has to ensure that the integral expression in (5.7) is independent of the control variable. This is guaranteed for purely time-dependent controls (see Prop. 5.7) or if a backwards uniqueness property holds (see Cor. 5.6). Clearly, the backwards uniqueness property of other control scenarios is of independent interest and would also lead to more applications for our approach.

(iii) Last, Lipschitz continuity of $\delta'(\cdot)$ yields fast local convergence of the Newton method, which further justifies to use the equivalence of time-optimal and distance-optimal controls for numerical realization. Here, we only stated one sufficient condition that relies on the structural assumption of the adjoint state (5.12); see Proposition 6.4. This condition does not seem to be sufficient as we observe Lipschitz continuity of $\delta'(\cdot)$ in the numerical examples even if (5.12) is violated. Different techniques to show Lipschitz continuity of $\delta'(\cdot)$ would require a second order sufficient optimality condition. However, such a condition cannot be expected to hold in the case of bang-bang controls.

Acknowledgements. The author acknowledges support from the International Research Training Group IGDK1754, funded by the DFG and FWF and the ERC advanced grant 668998 (OCLOC) under the EU's H2020 research program.

References

[1] H. Amann and J. Escher, Analysis. I. Translated from the 1998 German original by Gary Brookfield. Birkhäuser Verlag, Basel (2005).

[2] L. Bonifacius, *Numerical analysis of parabolic time-optimal control problems*. Ph.D. thesis, Technische Universität München (2018).

[3] L. Bonifacius and K. Pieper, Strong stability of linear parabolic time-optimal control problems. *ESAIM: COCV* **25** (2019) 35.

[4] L. Bonifacius, K. Pieper and B. Vexler, Error estimates for space-time discretization of parabolic time-optimal control problems with bang-bang controls. *SIAM J. Control Optim.* **57** (2019) 1730–1756.

[5] E. Casas, D. Wachsmuth and G. Wachsmuth, Sufficient second-order conditions for bang-bang control problems. *SIAM J. Control Optim.* **55** (2017) 3066–3090.

[6] R. Dautray and J.-L. Lions, Mathematical Analysis and Numerical Methods for Science and Technology. Evolution problems. I, With the collaboration of Michel Artola, Michel Cessenat and Hélène Lanchon, Translated from the French by Alan Craig. Springer-Verlag, Berlin (1992).

[7] J.C. Dunn, Convergence rates for conditional gradient sequences generated by implicit step length rules. *SIAM J. Control Optim.* **18** (1980) 473–487.

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

[9] R. Glowinski, J.-L. Lions and J. He, Exact and approximate controllability for distributed parameter systems. A numerical approach. In: Vol. 117 of *Encyclopedia of Mathematics and its Applications.* Cambridge University Press, Cambridge (2008).

[10] F. Gozzi and P. Loreti, Regularity of the minimum time function and minimum energy problems: the linear case. *SIAM J. Control Optim.* **37** (1999) 1195–1221.

[11] M. Gugat, A Newton method for the computation of time-optimal boundary controls of one-dimensional vibrating systems. Control of partial differential equations (Jacksonville, FL, 1998). *J. Comput. Appl. Math.* **114** (2000) 103–119.

[12] Q. Han and F.-H. Lin, Nodal sets of solutions of parabolic equations. II. *Comm. Pure Appl. Math.* **47** (1994) 1219–1238.

[13] H. Hermes and J.P. LaSalle, Functional analysis and time optimal control. In: Vol. 56 of *Mathematics in Science and Engineering.* Academic Press, New York-London (1969).

[14] M. Hintermüller and K. Kunisch, Path-following methods for a class of constrained minimization problems in function space. *SIAM J. Optim.* **17** (2006) 159–187.

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

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

[17] W. Krabs, Optimal control of processes governed by partial differential equations. I. Heating processes. *Z. Oper. Res. Ser. A-B* **26** (1982) A21–A48.

[18] K. Kunisch and D. Wachsmuth, On time optimal control of the wave equation, its regularization and optimality system. *ESAIM: COCV* **19** (2013) 317–336.

[19] K. Kunisch, K. Pieper and A. Rund, Time optimal control for a reaction diffusion system arising in cardiac electrophysiology – a monolithic approach. *ESAIM: M2AN* **50** (2016) 381–414.

[20] X.J. Li and J.M. Yong, Optimal control theory for infinite-dimensional systems. In: Systems & Control: Foundations & Applications. Birkhäuser Boston, Inc., Boston, MA (1995).

[21] X. Lu, L. Wang and Q. Yan, Computation of time optimal control problems governed by linear ordinary differential equations. *J. Sci. Comput.* **73** (2017) 1–25.

[22] J.W. Macki and A. Strauss, Introduction to optimal control theory. In: *Undergraduate Texts in Mathematics.* Springer-Verlag, New York-Berlin (1982).

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

[24] A. Münch and F. Periago, Numerical approximation of bang-bang controls for the heat equation: an optimal design approach. *Syst. Control Lett.* **62** (2013) 643–655.

[25] A. Münch and E. Zuazua, Numerical approximation of null controls for the heat equation: ill-posedness and remedies. *Inverse Prob.* **26** (2010) 085018, 39.

[26] J. Nocedal and S.J. Wright, Numerical optimization, 2nd edition. In: *Springer Series in Operations Research and Financial Engineering.* Springer, New York, NY (2006).

[27] E.M. Ouhabaz, Analysis of heat equations on domains. In: Vol. 31 of *London Mathematical Society Monographs Series.* Princeton University Press, Princeton, NJ (2005).

[28] A. Pazy, Semigroups of linear operators and applications to partial differential equations. In: Vol. 44 of *Applied Mathematical Sciences.* Springer-Verlag, New York (1983).

[29] K. Pieper, *Finite element discretization and efficient numerical solution of elliptic and parabolic sparse control problems.* Ph.D. thesis, Technische Universität München (2015).

[30] S. Qin and G. Wang, Equivalence between minimal time and minimal norm control problems for the heat equation. *SIAM J. Control Optim.* **56** (2018) 981–1010.

[31] S.M. Robinson, Normal maps induced by linear transformations. *Math. Oper. Res.* **17** (1992) 691–714.

[32] F. Trltzsch, Optimal control of partial differential equations. Theory, methods and applications, Translated from the 2005 German original by Jürgen Sprekels. In: Vol. 112 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI (2010).

[33] M. Tucsnak and G. Weiss, Observation and Control for Operator Semigroups. Birkhäuser Advanced Texts: Basler Lehrbücher. [Birkhäuser Advanced Texts: Basel Textbooks]. Birkhäuser Verlag, Basel (2009).

[34] W. Wang and M.Á. Carreira-Perpiñán, Projection onto the probability simplex: an efficient algorithm with a simple proof, and an application. Preprint arXiv:1309.1541 (2013).

[35] G. Wang and Y. Xu, Equivalence of three different kinds of optimal control problems for heat equations and its applications. *SIAM J. Control Optim.* **51** (2013) 848–880.

[36] G. Wang and E. Zuazua, On the equivalence of minimal time and minimal norm controls for internally controlled heat equations. *SIAM J. Control Optim.* **50** (2012) 2938–2958.

[37] G. Wang, L. Wang, Y. Xu and Y. Zhang, Time optimal control of evolution equations. In: *Progress in Nonlinear Differential Equations and Their Applications.* Springer International Publishing (2018).

[38] C. Zhang, The time optimal control with constraints of the rectangular type for linear time-varying ODEs. *SIAM J. Control Optim.* **51** (2013) 1528–1542.

[39] Y. Zhang, Two equivalence theorems of different kinds of optimal control problems for Schrödinger equations. *SIAM J. Control Optim.* **53** (2015) 926–947.