J Sci Comput
DOI 10.1007/s10915-017-0403-1

Computation of Time Optimal Control Problems Governed by Linear Ordinary Differential Equations

Xiliang Lu[1,2] · **Lijuan Wang**[1,2] · **Qishu Yan**[3]

Received: 17 May 2015 / Revised: 11 February 2017 / Accepted: 25 February 2017

Abstract In this paper, a novel numerical algorithm is presented to compute the optimal time of a time optimal control problem where the governing system is a linear ordinary differential equation. By the equivalence between time optimal control problem and norm optimal control problem, computation of the optimal time can be obtained by solving a sequence of norm optimal control problems, which are transferred into their Lagrangian dual problems. The nonsmooth structure of the dual problem is approximated by the iteratively reweighted least square strategy. Several numerical tests are given to show the efficiency of the proposed algorithm.

Keywords Time optimal control · Norm optimal control · Ordinary differential equation · Bisection method · Iteratively reweighted least square

Mathematics Subject Classification 34H05 · 49K15 · 49M30

1 Introduction

Let A and B be two matrices in $\mathbb{R}^{n\times n}$ and $\mathbb{R}^{n\times m}$, respectively, where $n, m \in \mathbb{N}^+$. Consider the following controlled ordinary differential equation:

✉ Qishu Yan
yanqishu@whu.edu.cn

Xiliang Lu
xllv.math@whu.edu.cn

Lijuan Wang
ljwang.math@whu.edu.cn

1 School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China

2 Hubei Key Laboratory of Computational Science, Wuhan University, Wuhan 430072, China

3 School of Science, Hebei University of Technology, Tianjin 300401, China

Published online: 07 March 2017

$$\begin{cases} y'(t) + Ay(t) = Bu(t), \ t \in (0, +\infty), \\ y(0) = y_0. \end{cases} \tag{1}$$

Here $y_0 \in \mathbb{R}^n \setminus \{0\}$ and u is a control function in the following admissible set:

$$\mathcal{U}_M \triangleq \{v : [0, +\infty) \to \mathbb{R}^m \text{ measurable and } \|v(t)\| \leq M \text{ for almost all } t \in (0, +\infty)\},$$

where M is a positive constant. When y_0 is given, for each $u \in \mathcal{U}_M$, (1) admits a unique solution which will be denoted by $y(\cdot; u)$.

Time optimal control problem reads as follows:

$$(TP)^M \qquad \inf \ \{T > 0 : y(T; u) = 0, u \in \mathcal{U}_M\},$$

where the minimal time denoted by $t^*(M)$ is called the optimal time of $(TP)^M$; a control $u^* \in \mathcal{U}_M$ is called an optimal control of $(TP)^M$ when $y(t^*(M); u^*) = 0$ and $u^*(\cdot) = 0$ over $(t^*(M), +\infty)$. We use $\|\cdot\|$ and $\langle \cdot, \cdot \rangle$ for the usual Euclidean norm and the inner product of $\mathbb{R}^n$ or $\mathbb{R}^m$, respectively. We denote by D^{T} the transpose of a matrix D.

Time optimal control problems of ordinary differential equations have been of great interest for a long time due to their practical applications. For example, the time optimal control problems on springs, pendulums or elevators can be described as this model [8,14,20,24], see also the Examples 1, 2 and 4 in Sect. 4. This issue has been extended to partial differential equations, see, e.g., [1,7,15,21]. Among these problems, time optimal control problems of parabolic equations and hyperbolic equations are studied most extensively, which depict the diffusion problems like heat transfer problems or chemical reaction processes and the vibration problems, respectively (see, [15]).

The computation of the optimal time for a time optimal control problem is an important and interesting subject. However, the time optimal control problem is a nonconvex problem, which is difficult to be solved numerically. By changing variable $\tau = \dfrac{t}{T}$, the first order optimality system for the problem $(TP)^M$ reads as: find $(T, \hat{y}, \hat{u}, \hat{\psi})$ such that

$$\begin{cases} \hat{y}'(\tau) + T \cdot A\hat{y}(\tau) = T \cdot B\hat{u}(\tau), & \tau \in (0, 1), \\ \hat{\psi}'(\tau) = T \cdot A^{\mathrm{T}}\hat{\psi}(\tau), & \tau \in (0, 1), \\ \hat{u}(\tau) = M\dfrac{B^{\mathrm{T}}\hat{\psi}(\tau)}{\|B^{\mathrm{T}}\hat{\psi}(\tau)\|}, & \tau \in (0, 1), \\ M = T \cdot \displaystyle\int_0^1 \|B^{\mathrm{T}}\hat{\psi}(\tau)\| d\tau, \\ \hat{y}(0) = y_0, \\ \hat{y}(1) = 0, \end{cases} \tag{2}$$

where T is the optimal time to $(TP)^M$ and the function defined by

$$v(t) \triangleq \begin{cases} \hat{u}\left(\dfrac{t}{T}\right), & t \in (0, T), \\ 0, & t \in [T, +\infty) \end{cases}$$

is the optimal control to $(TP)^M$, see Proposition 1.

The bilinear structure of system (2) gives a lot of trouble for numerical computation. To avoid this difficulty, we consider norm optimal control problems, and establish the equivalence between time and norm optimal control problems. Then the optimal time for the time optimal control problem can be obtained by solving a sequence of norm optimal control problems. This treatment is inspired by the work [22]. The equivalence between these two kinds of optimal control problems has also been studied in other literatures, see for instances [7,13,25]. The norm optimal control problem $(NP)^T$ reads:

$$(NP)^T \qquad \inf \left\{ \sup_t \|u(t)\| : y(T; u) = 0, u \in L^\infty\left(0, +\infty; \mathbb{R}^m\right)\right\},$$

where the minimum norm is denoted by $M^*(T)$. A control $u^*(\cdot)$ is called an optimal control for $(NP)^T$ if

$$y(T; u^*) = 0, \ \|u^*\|_{L^\infty(0,+\infty;\mathbb{R}^m)} = M^*(T) \text{ and } u^*(t) = 0, \ \text{for } t > T.$$

The norm optimal control problem is a convex optimization problem, and its Lagrangian dual problem (after changing cost functional to $\sup_t \frac{1}{2}\|u(t)\|^2$) can be formulated as

$$\min_{\mu\in\mathbb{R}^n} J^T(\mu) = \frac{1}{2}\left(\int_0^T \|B^{\mathrm{T}}\varphi(s; T, \mu)\|\, ds\right)^2 + \langle y_0, \varphi(0; T, \mu)\rangle, \tag{3}$$

where φ satisfies the adjoint equation

$$\begin{cases} \varphi'(t) = A^{\mathrm{T}}\varphi(t), & t \in (0, T), \\ \varphi(T) = \mu. \end{cases}$$

Throughout this paper, we use $\varphi(\cdot; T, \mu)$ to denote the solution of the adjoint equation with the initial time T and the initial state μ. One can show that the norm optimal control problem is equivalent to its Lagrangian dual problem (3), see Lemma 2. In some applications the "regularity" of the adjoint variable will cause trouble in the computation of $(NP)^T$. For example in the norm optimal control problem to the heat equation, the adjoint variable is the solution of a backward heat equation, and generally blows up (See, e.g., [22]). For the norm optimal control problems of some ODEs (e.g., discrete heat equation), adjoint variable may have a very large norm, which makes the computation unstable. To avoid "regularity" problem for the adjoint variable, we can consider a class of "regularized" norm optimal control problems $(NP)^T_\varepsilon$ for $\varepsilon \geq 0$ as follows:

$$(NP)^T_\varepsilon \qquad \inf \left\{ \sup_t \|u(t)\| : \|y(T; u)\| \leq \varepsilon, u \in L^\infty(0, +\infty; \mathbb{R}^m)\right\},$$

where the minimum denoted by $M^*_\varepsilon(T)$ is called the optimal norm of $(NP)^T_\varepsilon$. A control $u^*_\varepsilon(\cdot)$ is called an optimal control for $(NP)^T_\varepsilon$ if

$$\|y(T; u^*_\varepsilon)\| \leq \varepsilon, \ \|u^*_\varepsilon\|_{L^\infty(0,T;\mathbb{R}^m)} = M^*_\varepsilon(T) \text{ and } u^*_\varepsilon(t) = 0, \ \text{for } t > T.$$

Its Lagrangian dual problem can be formulated as

$$\min_{\mu\in\mathbb{R}^n} J^T_\varepsilon(\mu) = \frac{1}{2}\left(\int_0^T \|B^{\mathrm{T}}\varphi(s; T, \mu)\|\, ds\right)^2 + \langle y_0, \varphi(0; T, \mu)\rangle + \varepsilon\|\mu\|. \tag{4}$$

When $\varepsilon = 0$, the problem $(NP)^T_\varepsilon$ is the same as the norm optimal control problem $(NP)^T$. We will not distinguish $(NP)^T_0$ and $(NP)^T$ in the later discussion. When $\varepsilon = 0$, $J^T_0(\cdot)$ is the function $J^T(\cdot)$. One can prove the convergence for the optimal norm $M^*_\varepsilon(T)$ when ε approaches zero (see Theorem 1):

$$\lim_{\varepsilon\to 0^+} M^*_\varepsilon(T) = M^*(T).$$

The problem (4) is a nonsmooth convex optimization problem. A lot of efficient algorithms have been proposed in recent years to handle the nonsmoothness, e.g., iterative thresholding method [3], alternative direction method of multiplier [5], smoothing and continuation strategy [2], iteratively reweighted least square method [4], semismooth Newton method [9], just to name a few. Due to the complicated structure of $J_\varepsilon^T(\mu)$, we will apply the idea of iteratively reweighted least square method (IRLS). More precisely, given a small positive parameter $\delta > 0$, we solve a quadratic optimization problem with weight ω_k:

$$\min_{\mu\in\mathbb{R}^n} f_\delta(\mu,\omega_k) = \frac{1}{2}\int_0^T \sqrt{\|B^{\mathrm{T}}\varphi(t;T,\omega_k)\|^2+\delta^2}\,dt \cdot \int_0^T \frac{\|B^{\mathrm{T}}\varphi(t;T,\mu)\|^2+\delta^2}{\sqrt{\|B^{\mathrm{T}}\varphi(t;T,\omega_k)\|^2+\delta^2}}\,dt$$
$$+\langle y_0,\varphi(0;T,\mu)\rangle + \frac{\varepsilon}{2}\left(\frac{\|\mu\|^2+\delta^2}{\sqrt{\|\omega_k\|^2+\delta^2}} + \sqrt{\|\omega_k\|^2+\delta^2}\right),$$

then update the weight ω_{k+1}. With the carefully chosen weight ω_k, the convergence of this approach (IRLS) can be proved, see Theorem 3.

Now we mention some existing numerical algorithms for the time optimal control problem. When the control constraint is in the form of L^∞-norm, i.e., $|u_i(t)| \le M$ for almost every $t \in (0,+\infty)$ and each $1 \le i \le m$, where $u(\cdot) = (u_1(\cdot), u_2(\cdot), \ldots, u_m(\cdot))^{\mathrm{T}}$, the optimal control has a switching structure. There are a class of numerical methods by applying this switching structure. The most popular one is the switching time optimization method (see for instance [11,12] and [16]), where switching times are taken as extra unknowns and the optimization problem is solved by nonlinear programming technique. Recently, a semismooth Newton method to a regularized problem [10] has also been proposed. For the time optimal control problem considered in this paper, it is the same for L^∞-norm and Euclidean norm when $m = 1$. However, this is not the case when $m \ge 2$. To the best of our knowledge, there is no efficient algorithm for time optimal control problem when the control constraint is Euclidean norm in the literatures due to the lack of switching structure.

Throughout this paper, we assume

(H_1) Re$(\lambda)\ge 0$, for each eigenvalue λ of A;
(H_2) rank$[B, AB, \ldots, A^{n-1}B] = n$.

Under the above conditions, for any $M > 0$, the problem $(TP)^M$ has optimal controls (see Theorem 2.6 of [6]).

The rest of the paper is organized as follows. In Sect. 2, we shall give some properties of problems $(TP)^M$ and $(NP)_\varepsilon^T$ for $0 \le \varepsilon < \|y_0\|$, including the bang-bang principle, equivalence property, the first order necessary and sufficient condition and the convergence of optimal norm $M_\varepsilon^*(T)$ for $\varepsilon \to 0^+$. A bisection based root finding algorithm, which takes the IRLS method as inner loop, is derived in Sect. 3 to compute the optimal time $t^*(M)$. In Sect. 4, several numerical examples are given to show the efficiency of the algorithm.

2 Necessary and Sufficient Condition for $(TP)^M$ and $(NP)_\varepsilon^T$

Our algorithm is based on the equivalence between the time optimal control problem and the norm optimal control problem. We will give the main results in this section and leave the proof in the "Appendix" for the completeness. Let us consider problems $(TP)^M$ and $(NP)_\varepsilon^T$ for $0 \le \varepsilon < \|y_0\|$. Define

$$T_\varepsilon^* \triangleq \inf\{T > 0 : \|y(T;0)\| \le \varepsilon\},\ \text{for } 0 \le \varepsilon < \|y_0\|,$$

and if $\{T > 0 : \|y(T;0)\| \le \varepsilon\} = \emptyset$, we denote by $T_\varepsilon^* = +\infty$.

First, one can observe that

$$T_0^* = +\infty \quad \text{and} \quad \lim_{\varepsilon\to 0^+} T_\varepsilon^* = +\infty.$$

This can be proved by contradiction. We notice that T_ε^* is monotonically decreasing with respect to ε. If $\lim_{\varepsilon\to 0^+} T_\varepsilon^* = \widetilde{T} < \infty$, then $\|y(\widetilde{T};0)\| = \lim_{\varepsilon\to 0^+} \|y(T_\varepsilon^*;0)\| = 0$, which leads to a contradiction.

We recall some properties of $(TP)^M$ and $(NP)^T$ in Lemma 1. Since the proof is rather standard (see e.g. [7,17,19,24]), we put it in the "Appendix".

Lemma 1 *Given $M > 0$, $T > 0$.*

(i) There exists a unique optimal control for $(TP)^M$ and $(NP)^T$. Moreover, the bang-bang property holds for $(TP)^M$ and $(NP)^T$, i.e., the optimal control $u_M^*(\cdot)$ to $(TP)^M$ satisfies that

$$\|u_M^*(t)\| = M \text{ for a.e. } t \in (0, t^*(M)),$$

and the optimal control $u_T^*(\cdot)$ to $(NP)^T$ satisfies

$$\|u_T^*(t)\| = M^*(T) \text{ for a.e. } t \in (0, T).$$

(ii) The function $t^*(\cdot)$ is strictly monotonically decreasing and continuous over $(0, +\infty)$,

$$\lim_{M\to+\infty} t^*(M) = 0 \text{ and } \lim_{M\to 0^+} t^*(M) = +\infty.$$

(iii) The function $M^*(\cdot)$ is strictly monotonically decreasing and continuous over $(0, +\infty)$,

$$\lim_{T\to 0^+} M^*(T) = +\infty \text{ and } \lim_{T\to+\infty} M^*(T) = 0.$$

(iv) The optimal control to $(TP)^M$ is the optimal control to $(NP)^{t^*(M)}$. Conversely, the optimal control to $(NP)^T$ is the optimal control to $(TP)^{M^*(T)}$. Moreover,

$$M^*(t^*(M)) = M \text{ for each } M > 0,$$

and

$$t^*(M^*(T)) = T \text{ for each } T > 0.$$

The next proposition gives a necessary condition to $(TP)^M$, i.e., Eq. (2), and the proof can be found in the "Appendix".

Proposition 1 *Given $M > 0$, the system (2) has a solution $\left(T, \hat{y}(\cdot), \hat{u}(\cdot), \hat{\psi}(\cdot)\right)$. Moreover, $T = t^*(M)$ and the control $v(\cdot)$, defined by*

$$v(t) \triangleq \begin{cases} \hat{u}\left(\dfrac{t}{T}\right), & t \in (0, T), \\ 0, & t \in [T, +\infty), \end{cases}$$

is the optimal control to $(TP)^M$.

The next Lemma is concerned with the norm optimal control problem $(NP)_\varepsilon^T$ and its Lagrangian dual problem (4). The proof is in the "Appendix".

Lemma 2 *Let $0 \le \varepsilon < \|y_0\|$ and $0 < T < T_\varepsilon^*$.*

(i) The function $J_\varepsilon^T(\cdot)$ *is continuous, coercive and convex, hence it admits a minimizer* μ_ε^* *in* $\mathbb{R}^n$. *Moreover,* $\mu_\varepsilon^* \neq 0$.

(ii) μ_ε^* *is a minimizer of* $J_\varepsilon^T(\cdot)$ *if and only if it satisfies*

$$\begin{aligned} &\int_0^T \|B^{\mathrm{T}}\varphi(t;T,\mu_\varepsilon^*)\|\,dt \int_0^T \frac{\langle B^{\mathrm{T}}\varphi(t;T,\mu_\varepsilon^*), B^{\mathrm{T}}\varphi(t;T,\nu)\rangle}{\|B^{\mathrm{T}}\varphi(t;T,\mu_\varepsilon^*)\|}\,dt \\ &\quad + \langle y_0, \varphi(0;T,\nu)\rangle + \varepsilon\frac{\langle \mu_\varepsilon^*, \nu\rangle}{\|\mu_\varepsilon^*\|} = 0, \ \textit{for any } \nu \in \mathbb{R}^n. \end{aligned} \tag{5}$$

(iii) Let μ_ε^* *be a minimizer of* $J_\varepsilon^T(\cdot)$. *Then the control defined by*

$$u_\varepsilon^*(t) \triangleq \begin{cases} \int_0^T \|B^{\mathrm{T}}\varphi(t;T,\mu_\varepsilon^*)\|\,dt \dfrac{B^{\mathrm{T}}\varphi(t;T,\mu_\varepsilon^*)}{\|B^{\mathrm{T}}\varphi(t;T,\mu_\varepsilon^*)\|}, & t \in (0,T), \\ 0, & t \in [T,+\infty) \end{cases} \tag{6}$$

is an optimal control to $(NP)_\varepsilon^T$, *and* $M_\varepsilon^*(T) = \int_0^T \|B^{\mathrm{T}}\varphi(t;T,\mu_\varepsilon^*)\|\,dt$.

In general, $J^T(\cdot)$ is not strictly convex (see Remark 2.2 of [25]). Consequently, the minimizer of this function may not be unique. However, for $\varepsilon \in (0, \|y_0\|)$ we have the following strictly convex property, see "Appendix" for the proof.

Proposition 2 *When* $0 < \varepsilon < \|y_0\|$ *and* $0 < T < T_\varepsilon^*$, *the function* $J_\varepsilon^T(\cdot)$ *is strictly convex, thus it admits a unique minimizer.*

Next we establish the convergence of the optimal norm and optimal control of $(NP)_\varepsilon^T$ when ε goes to zero.

Theorem 1 *For any* $T > 0$, *it holds that*

$$\lim_{\varepsilon\to 0^+} M_\varepsilon^*(T) = M^*(T).$$

Let u_ε^* *be an optimal control to* $(NP)_\varepsilon^T$ *defined by* (6) *for* $0 \le \varepsilon < \|y_0\|$. *Then*

$$u_\varepsilon^*(\cdot) \to u_0^*(\cdot) \textit{ in } L^p(0,+\infty) \textit{ as } \varepsilon \to 0^+, \quad \forall\, p \in [1,\infty).$$

Proof For fixed $T > 0$, let $\varepsilon_0 \in (0, \|y_0\|)$ be a constant such that $T < T_{\varepsilon_0}^*$. We can easily check that

$$M_{\varepsilon_0}^*(T) \le M_{\delta_1}^*(T) \le M_{\delta_2}^*(T) \le M^*(T), \ \text{for any } 0 \le \delta_2 \le \delta_1 \le \varepsilon_0. \tag{7}$$

Let $\{\varepsilon_k\}_{k\ge 1}$ be a strictly decreasing sequence with $\varepsilon_1 \le \varepsilon_0$ and $\lim_{k\to+\infty} \varepsilon_k = 0$. Let μ_k^* be the minimizer of $J_{\varepsilon_k}^T(\cdot)$ for each $k \ge 1$. Then by (ii) and (iii) in Lemma 2, we have that

$$\begin{aligned} &\int_0^T \|B^{\mathrm{T}}\varphi(t;T,\mu_k^*)\|\,dt \int_0^T \frac{\langle B^{\mathrm{T}}\varphi(t;T,\mu_k^*), B^{\mathrm{T}}\varphi(t;T,\nu)\rangle}{\|B^{\mathrm{T}}\varphi(t;T,\mu_k^*)\|}\,dt \\ &\quad + \langle y_0, \varphi(0;T,\nu)\rangle + \varepsilon\frac{\langle \mu_k^*, \nu\rangle}{\|\mu_k^*\|} = 0 \text{ for any } \nu \in \mathbb{R}^n, \end{aligned} \tag{8}$$

$$M_{\varepsilon_k}^*(T) = \int_0^T \|B^{\mathrm{T}}\varphi(t;T,\mu_k^*)\|\,dt, \tag{9}$$

and the function $u^*_{\varepsilon_k}(\cdot)$ defined by

$$u^*_{\varepsilon_k}(t) \triangleq \begin{cases} \int_0^T \|B^{\mathrm{T}}\varphi(t;T,\mu_k^*)\|\,dt \dfrac{B^{\mathrm{T}}\varphi(t;T,\mu_k^*)}{\|B^{\mathrm{T}}\varphi(t;T,\mu_k^*)\|}, & t\in(0,T),\\ 0, & t\in[T,+\infty) \end{cases} \tag{10}$$

is an optimal control to $(NP)^T_{\varepsilon_k}$.

Since $J^T(\mu_k^*) < J^T_{\varepsilon_k}(\mu_k^*) \le J^T_{\varepsilon_k}(\mu_1^*) < J^T_{\varepsilon_1}(\mu_1^*)$ for each $k \ge 1$ and $J^T(\cdot)$ is coercive (See (i) in Lemma 2), there exists a positive constant C independent of k, such that

$$\|\mu_k^*\| \le C, \quad \forall k \ge 1. \tag{11}$$

Hence there exists $\mu_0^* \in \mathbb{R}^n$ and a subsequence of $\{\mu_k^*\}_{k\ge1}$, still denoted by $\{\mu_k^*\}$, such that

$$\lim_{k\to+\infty} \mu_k^* = \mu_0^*. \tag{12}$$

By (7), passing to the limit for $k \to +\infty$ in (9), we get that

$$\int_0^T \|B^{\mathrm{T}}\varphi(t;T,\mu_0^*)\|\,dt \ge M^*_{\varepsilon_0}(T) > 0,$$

which indicates that $\mu_0^* \ne 0$. Now passing to the limit for $k \to +\infty$ in (8), we get that

$$\int_0^T \|B^{\mathrm{T}}\varphi(t;T,\mu_0^*)\|\,dt \int_0^T \frac{\langle B^{\mathrm{T}}\varphi(t;T,\mu_0^*), B^{\mathrm{T}}\varphi(t;T,\nu)\rangle}{\|B^{\mathrm{T}}\varphi(t;T,\mu_0^*)\|}\,dt$$
$$+ \langle y_0, \varphi(0;T,\nu)\rangle = 0, \text{ for any } \nu \in \mathbb{R}^n.$$

This, together with (ii) and (iii) in Lemma 2, implies that μ_0^* is a minimizer of $J^T(\cdot)$ and the function $u_0^*(\cdot)$ defined by

$$u^*_0(t) \triangleq \begin{cases} \int_0^T \|B^{\mathrm{T}}\varphi(t;T,\mu_0^*)\|\,dt \dfrac{B^{\mathrm{T}}\varphi(t;T,\mu_0^*)}{\|B^{\mathrm{T}}\varphi(t;T,\mu_0^*)\|}, & t\in(0,T),\\ 0, & t\in[T,+\infty) \end{cases} \tag{13}$$

is the optimal control to $(NP)^T$. Moreover,

$$M^*(T) = \int_0^T \|B^{\mathrm{T}}\varphi(t;T,\mu_0^*)\|\,dt. \tag{14}$$

By (9), (12) and (14), we have that $\lim_{k\to+\infty} M^*_{\varepsilon_k}(T) = M^*(T)$. Therefore,

$$\lim_{\varepsilon\to0^+} M^*_\varepsilon(T) = M^*(T).$$

Besides, from (27), (7), (10), (12) and (13), we can get that

$$u^*_{\varepsilon_k}(t) \to u_0^*(t) \text{ for a.e. } t \in (0,+\infty)$$

and for each $k \ge 1$,

$$\|u^*_{\varepsilon_k}(t) - u_0^*(t)\|^p \le 2^p(M^*(T))^p \text{ for a.e. } t \in (0,T).$$

Then by Lebesgue dominated convergence theorem, we obtain

$$\lim_{k\to+\infty} \|u^*_{\varepsilon_k} - u_0^*\|_{L^p(0,T)} = 0, \quad \forall p \in [1,\infty).$$

The convergence for the whole sequence is from the uniqueness of the optimal control u_0^*. This completes the proof. □

3 The Numerical Algorithm

In this section, we will propose an efficient method to compute the optimal time and the optimal control. The whole method can be divided into three parts. First by Lemma 1, to find the optimal time to $(TP)^M$ is equivalent to solve a nonlinear equation, hence we only need to solve a sequence of norm optimal control problems $(NP)^T$. Then the norm optimal control problem $(NP)^T$ is approximated by $(NP)^T_\varepsilon$ and its Lagrangian dual problem min J^T_ε. Last an iteratively reweighted least square method is proposed to solve the nonsmooth optimization problem min J^T_ε. The details can be found in the later discussion.

3.1 Bisection Based Algorithm

By (iii) and (iv) in Lemma 1, we have

$$M^*(t^*(M)) = M, \quad M^*(t) < M \Leftrightarrow t > t^*(M) \quad \text{and} \quad M^*(t) > M \Leftrightarrow t < t^*(M).$$

Therefore the equation $M^*(t) - M = 0$ has a unique solution $t^*(M)$. Then one can compute the optimal time $t^*(M)$ by the bisection based root finding method which is given in Algorithm 1.

Algorithm 1 Bisection method

```
Given an initial guess T_0, tolerance τ_1, τ_2;
find the initial interval [a, b] as follows:
if M*(T_0) > M, then b = T_0; do a = b, b = 2a until M*(b) < M;
if M*(T_0) < M, then a = T_0; do b = a, a = b/2 until M*(a) > M;
set initial interval [a_0, b_0] = [a, b];
for k = 1, 2 ······ do
  t_k = (a_{k-1} + b_{k-1})/2 and compute M_k = M*(t_k);
  if |M_k − M|/M < τ_1 or b_{k-1} − a_{k-1} < τ_2 then
    stop;
  else if M_k > M then
    a_k = t_k, b_k = b_{k-1};
  else
    a_k = a_{k-1}, b_k = t_k;
  end if
end for
output t_k.
```

From Algorithm 1, and the standard argument of bisection method, it follows that

$$\lim_{k \to +\infty} t_k = t^*(M).$$

Remark 1 The bisection method can be replaced by other root finding techniques, for example, the secant method and the regular falsi method(see section 6.2.2 in [18]).

3.2 Computation for Approximate Norm Optimal Control Problem $(NP)_\varepsilon^T$

During each iteration of Algorithm 1, one has to solve the norm optimal control problem to obtain $M^*(T)$. From Theorem 1, the approximate norm optimal control problem $(NP)_\varepsilon^T$ converges to $(NP)^T$. For some problems, we will compute $M_\varepsilon^*(T)$ to approximate $M^*(T)$. By Lemma 2, the problem $(NP)_\varepsilon^T$ can be reformulated as the minimization problem: $\min J_\varepsilon^T(\cdot)$, and the optimal norm is given by $M_\varepsilon^*(T) = \int_0^t \|B^{\mathrm{T}}\varphi(t;T,\mu_\varepsilon^*)\|dt$, where $\mu_\varepsilon^* = \operatorname{argmin}_x J_\varepsilon^T(x)$.

Now we fix T and ε, to find the minimizer of $J_\varepsilon^T(\cdot)$. To avoid the nonsmoothness of the cost function $J_\varepsilon^T(\cdot)$, we first approximate $J_\varepsilon^T(\cdot)$ by a smooth function, and then use the IRLS method to find the minimizer of the smooth one.

For a small positive parameter $\delta > 0$, we consider a function $g_\delta(\cdot) : \mathbb{R}^n \to \mathbb{R}$:

$$g_\delta(x) \triangleq \frac{1}{2}\left(\int_0^T \sqrt{\|B^{\mathrm{T}}\varphi(t;T,x)\|^2+\delta^2}\,dt\right)^2 + \langle y_0, \varphi(0;T,x)\rangle + \varepsilon\sqrt{\|x\|^2+\delta^2}. \tag{15}$$

This function is a smooth version of $J_\varepsilon^T(\cdot)$. Similar as in Lemma 2, we have the following Lemma for $g_\delta(\cdot)$.

Lemma 3 *The function $g_\delta(\cdot)$ is continuous, coercive and strictly convex. Thus it admits a unique minimizer. A vector x_δ^* is the minimizer of $g_\delta(\cdot)$ if and only if it satisfies*

$$\begin{aligned}&\int_0^T \sqrt{\|B^{\mathrm{T}}\varphi(t;T,x_\delta^*)\|^2+\delta^2}\,dt \int_0^T \frac{\langle B^{\mathrm{T}}\varphi(t;T,x_\delta^*), B^{\mathrm{T}}\varphi(t;T,\nu)\rangle}{\sqrt{\|B^{\mathrm{T}}\varphi(t;T,x_\delta^*)\|^2+\delta^2}}\,dt \\ &\quad + \langle y_0, \varphi(0;T,\nu)\rangle + \varepsilon\frac{\langle x_\delta^*, \nu\rangle}{\sqrt{\|x_\delta^*\|^2+\delta^2}} = 0, \text{ for any } \nu \in \mathbb{R}^n.\end{aligned} \tag{16}$$

Proof It is obvious that $g_\delta(\cdot)$ is continuous. By (4), we have that

$$g_\delta(x) \geq J_\varepsilon^T(x) \text{ for each } x \in \mathbb{R}^n,$$

which, combined with (i) in Lemma 2, indicates that $g_\delta(\cdot)$ is coercive.

For fixed $d \in \mathbb{R}^n$ with $d \neq 0$, by regular calculations, we have that

$$\begin{aligned}&d^{\mathrm{T}}\nabla^2 g_\delta(x) d \\ &= \left(\int_0^T \frac{\varphi(t;T,d)^{\mathrm{T}}BB^{\mathrm{T}}\varphi(t;T,x)}{\sqrt{\|B^{\mathrm{T}}\varphi(t;T,x)\|^2+\delta^2}}\,dt\right)^2 \\ &\quad + \int_0^T \sqrt{\|B^{\mathrm{T}}\varphi(t;T,x)\|^2+\delta^2}\,dt \int_0^T \frac{\|B^{\mathrm{T}}\varphi(t;T,d)\|^2}{\sqrt{\|B^{\mathrm{T}}\varphi(t;T,x)\|^2+\delta^2}}\,dt \\ &\quad - \int_0^T \sqrt{\|B^{\mathrm{T}}\varphi(t;T,x)\|^2+\delta^2}\,dt \int_0^T \frac{(\varphi(t;T,d)^{\mathrm{T}}BB^{\mathrm{T}}\varphi(t;T,x))^2}{(\sqrt{\|B^{\mathrm{T}}\varphi(t;T,x)\|^2+\delta^2})^3}\,dt \\ &\quad + \frac{\|d\|^2}{\sqrt{\|x\|^2+\delta^2}} - \frac{(x^{\mathrm{T}}d)^2}{(\sqrt{\|x\|^2+\delta^2})^3} \\ &\geq \left(\int_0^T \frac{\varphi(t;T,d)^{\mathrm{T}}BB^{\mathrm{T}}\varphi(t;T,x)}{\sqrt{\|B^{\mathrm{T}}\varphi(t;T,x)\|^2+\delta^2}}\,dt\right)^2\end{aligned}$$

$$+\int_0^T \sqrt{\|B^{\mathrm{T}}\varphi(t;T,x)\|^2+\delta^2}\,dt \int_0^T \frac{\delta^2\|B^{\mathrm{T}}\varphi(t;T,d)\|^2}{(\sqrt{\|B^{\mathrm{T}}\varphi(t;T,x)\|^2+\delta^2})^3}\,dt$$
$$+\frac{\delta^2\|d\|^2}{(\sqrt{\|x\|^2+\delta^2})^3} > 0.$$

Thus $\nabla^2 g_\delta(x)$ is positive definite for each $x \in \mathbb{R}^n$, which shows that $g_\delta(\cdot)$ is strictly convex. Then $g_\delta(\cdot)$ has a unique minimizer. By the same arguments as those lead to (ii) in Lemma 2, we obtain that x_δ^* is a minimizer of $g_\delta(\cdot)$ if and only (16) holds.

This completes the proof. □

In the rest of this section, we denote the minimizer of $g_\delta(\cdot)$ and $J_\varepsilon^T(\cdot)$ by x_δ^* and μ_ε^*, respectively. Then we have the following convergence property.

Theorem 2

$$\lim_{\delta\to 0^+} x_\delta^* = \mu_\varepsilon^*.$$

Proof Let J_0 be the minimum value of $J_\varepsilon^T(\cdot)$. It is obvious that

$$J_0 \le J_\varepsilon^T(x_\delta^*) \le g_\delta(x_\delta^*) \le g_\delta(\mu_\varepsilon^*) < g_{\delta_0}(\mu_\varepsilon^*), \quad \forall\, 0<\delta<\delta_0, \tag{17}$$

for some fixed $\delta_0 > 0$. By the coercivity of $J_\varepsilon^T(\cdot)$, we know that $\{x_\delta^*\}_{0<\delta<\delta_0}$ is bounded. Then there exists $\mu^* \in \mathbb{R}^n$ and a subsequence, denoted by $\{\delta_k\}_{k\ge 1}$, with $\lim\limits_{k\to\infty}\delta_k = 0$, such that

$$\lim_{k\to+\infty} x_{\delta_k}^* = \mu^*.$$

Thus

$$\lim_{k\to\infty} g_{\delta_k}(\mu_\varepsilon^*) = J_\varepsilon^T(\mu_\varepsilon^*) = J_0 \text{ and } \lim_{k\to+\infty} J_\varepsilon^T(x_{\delta_k}^*) = J_\varepsilon^T(\mu^*).$$

Passing to the limit for $k \to \infty$ in (17), we have that

$$J_0 \le J_\varepsilon^T(\mu^*) \le J_0.$$

Hence μ^* is the minimizer of $J_\varepsilon^T(\cdot)$. Due to the the uniqueness of the minimizer of $J_\varepsilon^T(\cdot)$, we have $\mu^* = \mu_\varepsilon^*$ and the convergence for $\delta \to 0^+$. □

In view of Theorem 2, in each step of Algorithm 1, we can use x_δ^* to approximate μ_ε^*. Therefore we only need to solve the minimizer x_δ^* of $g_\delta(\cdot)$. In the following we will apply the iteratively reweighed least square(IRLS) method to solve it. Figure 1 sums up the steps for the numerical computation of $t^*(M)$.

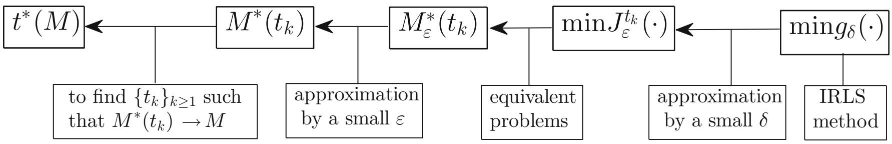

Fig. 1 Numerical computation of $t^*(M)$

To find the minimizer x_δ^* of $g_\delta(\cdot)$ defined by (15), we introduce an auxiliary functional $f_\delta(\cdot,\cdot)\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, for any $x \in \mathbb{R}^n$ and $\omega \in \mathbb{R}^n$,

$$f_\delta(x,\omega) \triangleq \frac{1}{2}\int_0^T \sqrt{\|B^{\mathrm{T}}\varphi(s;T,\omega)\|^2+\delta^2}\,ds \int_0^T \frac{\|B^{\mathrm{T}}\varphi(s;T,x)\|^2+\delta^2}{\sqrt{\|B^{\mathrm{T}}\varphi(s;T,\omega)\|^2+\delta^2}}\,ds + \langle y_0, \varphi(0;T,x)\rangle + \frac{\varepsilon}{2}\left(\frac{\|x\|^2+\delta^2}{\sqrt{\|\omega\|^2+\delta^2}} + \sqrt{\|\omega\|^2+\delta^2}\right).$$

It is clear that $f_\delta(x,x) = g_\delta(x)$. For any $\omega \in \mathbb{R}^n$, $f_\delta(\cdot,\omega)$ is a quadratical function and has a unique nonzero minimizer. By the similar arguments as those in Proposition 1, we have the following result.

Proposition 3 *Given $\omega \in \mathbb{R}^n$. The system*

$$\begin{cases} y'(t) + Ay(t) = \displaystyle\int_0^T \sqrt{\|B^{\mathrm{T}}\varphi(t;T,\omega)\|^2+\delta^2}\,dt \cdot \frac{BB^{\mathrm{T}}\psi(t)}{\sqrt{\|B^{\mathrm{T}}\varphi(t;T,\omega)\|^2+\delta^2}}, & t \in (0,T),\\ \psi'(t) = A^{\mathrm{T}}\psi(t), & t\in(0,T),\\ y(0) = y_0, \\ y(T) = -\varepsilon \dfrac{\psi(T)}{\sqrt{\|\omega\|^2+\delta^2}} \end{cases} \tag{18}$$

has a unique solution $(y(\cdot),\psi(\cdot))$. Moreover,

$$\psi(T) \neq 0 \text{ and } \psi(T) = \operatorname{argmin}_x f_\delta(x,\omega).$$

Proof Let x_0 be the minimizer of $f_\delta(\cdot,\omega)$. Then by the same arguments as those lead to (5), we have that x_0 is the minimizer of $f_\delta(\cdot,\omega)$ if and only if it satisfies that

$$\int_0^T \sqrt{\|B^{\mathrm{T}}\varphi(t;T,\omega)\|^2+\delta^2}\,dt \int_0^T \frac{\langle B^{\mathrm{T}}\varphi(t;T,x_0), B^{\mathrm{T}}\varphi(t;T,\nu)\rangle}{\sqrt{\|B^{\mathrm{T}}\varphi(t;T,\omega)\|^2+\delta^2}}\,dt + \langle y_0, \varphi(0;T,\nu)\rangle + \varepsilon\frac{\langle x_0,\nu\rangle}{\sqrt{\|\omega\|^2+\delta^2}} = 0, \text{ for any } \nu \in \mathbb{R}^n. \tag{19}$$

Consider the system

$$\begin{cases} y'(t) + Ay(t) = \displaystyle\int_0^T \sqrt{\|B^{\mathrm{T}}\varphi(t;T,\omega)\|^2+\delta^2}\,dt \cdot \frac{BB^{\mathrm{T}}\psi(t)}{\sqrt{\|B^{\mathrm{T}}\varphi(t;T,\omega)\|^2+\delta^2}}, & t \in [0,T],\\ \psi'(t) = A^{\mathrm{T}}\psi(t), & t\in[0,T],\\ y(0) = y_0. \end{cases}$$

For any $\nu \in \mathbb{R}^n$, multiplying the first equation of the above system by $\varphi(\cdot;T,\nu)$ and integrating it over $(0,T)$, we obtain that

$$\langle y(T),\nu\rangle = \langle y_0, \varphi(0;T,\nu)\rangle + \int_0^T \sqrt{\|B^{\mathrm{T}}\varphi(t;T,\omega)\|^2+\delta^2}\,dt \int_0^T \frac{\langle B^{\mathrm{T}}\psi(t), B^{\mathrm{T}}\varphi(t;T,\nu)\rangle}{\sqrt{\|B^{\mathrm{T}}\varphi(t;T,\omega)\|^2+\delta^2}}\,dt,$$

which, combined with (19), indicates that

$$\langle y(T), \nu\rangle = -\varepsilon \frac{\langle \psi(T), \nu\rangle}{\sqrt{\|\omega\|^2 + \delta^2}}, \text{ for any } \nu \in \mathbb{R}^n.$$

This show that $y(T) = -\varepsilon \frac{\psi(T)}{\sqrt{\|\omega\|^2 + \delta^2}}$. From the above arguments it follows that $(y(\cdot), \varphi(\cdot; T, x_0))$ is a solution to (18).

Next, for any solution $(y(\cdot), \psi(\cdot))$ to system (18), multiplying the first equation of (18) by $\varphi(t; T, \nu)$ and integrating it over $(0, T)$, we obtain that $\psi(T)$ satisfies (19). This shows that $\psi(T)$ is a minimizer of $f_\delta(\cdot, \omega)$. Since $f_\delta(\cdot, \omega)$ has a unique minimizer and the minimizer is not 0, we complete the proof. □

The IRLS method is an iterative process which reads

$$\begin{cases} \omega_k = \text{argmin}_\omega \, f_\delta(x_k, \omega); \\ x_{k+1} = \text{argmin}_x \, f_\delta(x, \omega_k). \end{cases} \tag{20}$$

One can verify that ω_k could be chosen to be x_k, and x_{k+1} satisfies (18) with $\omega = \omega_k$. The IRLS algorithm can be found in Algorithm 2.

Algorithm 2 IRLS for computing x_δ^*

```
1: Given x_0 ≠ 0, error tolerance τ;
2: for k = 0, 1 ······ do
3:    solve x_{k+1} from system (18) with ω = x_k;
4:    if ‖x_{k+1} − x_k‖/‖x_k‖ < τ then
5:       stop;
6:    end if
7: end for
```

The global convergence of the IRLS algorithm is obtained as following.

Theorem 3 *Given $x_0 \in \mathbb{R}^n$, the sequence $\{x_k\}_{k\geq 1}$ generated by Algorithm 2 converges to x_δ^*.*

Proof Firstly, by (20) we have that

$$L \triangleq f_\delta(x_0, \omega_0) \geq f_\delta(x_1, \omega_0) \geq f_\delta(x_1, \omega_1) \cdots \geq g_\delta(x_\delta^*), \tag{21}$$

hence the sequence $\{f_\delta(x_k, \omega_k)\}_{k\geq 1}$ converges. Denote the limit by L_0. By the coercivity of $g_\delta(\cdot)$ and the inequality

$$g_\delta(x_k) = f_\delta(x_k, x_k) \leq L,$$

we obtain that

$$\|x_k\| \leq C. \tag{22}$$

Here and throughout this proof, C denotes a generic positive constant independent of k. Then from the definition of $f_\delta(x_k, \omega_k)$ and (21), we have that

$$\begin{aligned} & f_\delta(x_k, \omega_k) - f_\delta(x_{k+1}, \omega_{k+1}) \\ & \quad \geq f_\delta(x_k, \omega_k) - f_\delta(x_{k+1}, \omega_k) \\ & \quad = \frac{1}{2} \int_0^T \sqrt{\|B^{\mathrm{T}}\varphi(t; T, \omega_k)\|^2 + \delta^2}\, dt \\ & \qquad \cdot \int_0^T \frac{\|B^{\mathrm{T}}\varphi(t; T, x_k)\|^2 - \|B^{\mathrm{T}}\varphi(t; T, x_{k+1})\|^2}{\sqrt{\|B^{\mathrm{T}}\varphi(t; T, \omega_k)\|^2 + \delta^2}}\, dt \\ & \qquad + \langle y_0, \varphi(0; T, x_k - x_{k+1})\rangle + \frac{\varepsilon}{2} \frac{\|x_k\|^2 - \|x_{k+1}\|^2}{\sqrt{\|\omega_k\|^2 + \delta^2}}. \end{aligned} \tag{23}$$

Since $x_{k+1} = \mathrm{argmin}_x\, f_\delta(x, \omega_k)$, it stands that

$$\begin{aligned} & \int_0^T \sqrt{\|B^{\mathrm{T}}\varphi(t; T, \omega_k)\|^2 + \delta^2}\, dt \int_0^T \frac{\langle B^{\mathrm{T}}\varphi(t; T, x_{k+1}), B^{\mathrm{T}}\varphi(t; T, \nu)\rangle}{\sqrt{\|B^{\mathrm{T}}\varphi(t; T, \omega_k)\|^2 + \delta^2}}\, dt \\ & \quad + \langle y_0, \varphi(0; T, \nu)\rangle + \varepsilon \frac{\langle x_{k+1}, \nu\rangle}{\sqrt{\|\omega_k\|^2 + \delta^2}} = 0, \text{ for any } \nu \in \mathbb{R}^n. \end{aligned} \tag{24}$$

Taking $\nu = x_k - x_{k+1}$ in the above equation, by (23), we get that

$$\begin{aligned} & f_\delta(x_k, \omega_k) - f_\delta(x_{k+1}, \omega_{k+1}) \\ & \quad \geq \frac{1}{2} \int_0^T \sqrt{\|B^{\mathrm{T}}\varphi(t; T, \omega_k)\|^2 + \delta^2}\, dt \cdot \int_0^T \frac{\|B^{\mathrm{T}}\varphi(t; T, x_{k+1} - x_k)\|^2}{\sqrt{\|B^{\mathrm{T}}\varphi(t; T, \omega_k)\|^2 + \delta^2}}\, dt \\ & \qquad + \frac{\varepsilon}{2} \frac{\|x_k - x_{k+1}\|^2}{\sqrt{\|\omega_k\|^2 + \delta^2}}, \end{aligned}$$

which, combined with (22) and $\omega_k = x_k$, indicates that

$$f_\delta(x_k, \omega_k) - f_\delta(x_{k+1}, \omega_{k+1}) \geq C\|x_k - x_{k+1}\|^2, \quad \forall k \geq 1.$$

Summing the above inequality from $k = 1$ to $+\infty$, we get that

$$f_\delta(x_1, \omega_1) - L_0 = \sum_{k=1}^{+\infty} (f_\delta(x_k, \omega_k) - f_\delta(x_{k+1}, \omega_{k+1})) \geq C \sum_{k=1}^{+\infty} \|x_k - x_{k+1}\|^2,$$

which indicates that

$$\lim_{k\to+\infty} \|x_k - x_{k+1}\| = 0. \tag{25}$$

By (22), there exists a subsequence of $\{x_k\}_{k\geq1}$, still denoted by itself, such that

$$\lim_{k\to+\infty} x_k = x^*.$$

Passing to the limit for $k \to \infty$ in (24), by (25), $\omega_k = x_k$ and the above equation, we get that

$$\begin{aligned} & \int_0^T \sqrt{\|B^{\mathrm{T}}\varphi(t; T, x^*)\|^2 + \delta^2}\, dt \int_0^T \frac{\langle B^{\mathrm{T}}\varphi(t; T, x^*), B^{\mathrm{T}}\varphi(t; T, \nu)\rangle}{\sqrt{\|B^{\mathrm{T}}\varphi(t; T, x^*)\|^2 + \delta^2}}\, dt \\ & \quad + \langle y_0, \varphi(0; T, \nu)\rangle + \varepsilon \frac{\langle x^*, \nu\rangle}{\sqrt{\|x^*\|^2 + \delta^2}} = 0, \text{ for any } \nu \in \mathbb{R}^n. \end{aligned}$$

From Lemma 3, we know that $x^* = x^*_\delta$. Then the uniqueness of minimizer implies the convergence for the whole sequence. □

Remark 2 In the numerical test, the choice of δ is not sensitive, see Example 1 in the next section for details. To ensure the efficiency, one may apply warm start technique to choose x_0 in Algorithm 2, i.e., x_0 is chosen as the solution of the previous subproblem.

4 Numerical Examples

In this section, we give several examples to show the efficiency of the proposed numerical algorithm. From Algorithms 1 and 2, one needs to solve the ODE system (18) extensively. The Eq. (18) is a two-point boundary value problem which we will discrete by Crank-Nicolson scheme with step length Δt. Denote $N = T/\Delta t + 1$. During this section, the optimal time and its numerical approximation are denoted by $t^*(M)$ and $\widetilde{t}^*(M)$, respectively.

Example 1 Let

$$A = \begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix}, B = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, y_0 = \begin{pmatrix} -2 \\ 0 \end{pmatrix} \text{ and } M = 1.$$

This is Example 3.1 in [24], which models the problem that an elevator reaches the given height in the shortest time with the power of its engine constrained. The optimal time $t^*(M)$ for this example is 2.8284 and the optimal control reads

$$u^*(t) = \begin{cases} 1, & 0 < t < t_s, \\ -1, & t_s < t < t^*(M), \end{cases}$$

where the transition time $t_s = 1.4142$. In Table 1, 2 and 3, we use $\widetilde{t_s}$ to represent the numerical transition time for the optimal control. Table 1 shows the convergence when ε decreases, where we choose $N = 151$ and $\delta = 10^{-10}$. In Table 2 the convergence with respect to mesh size is presented, where $\varepsilon = 10^{-4}$ and $\delta = 10^{-10}$. Table 3 presents the convergence when δ goes to zero, where $\varepsilon = 10^{-4}$ and $N = 151$. In the later computation we will fix δ to be 10^{-10}.

Example 2 Let

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, B = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, y_0 = \begin{pmatrix} -5 \\ 5 \end{pmatrix} \text{ and } M = 1.$$

Table 1 Example 1, convergence with respect to ε

ε	1	10^{-1}	10^{-2}	10^{-3}	10^{-4}	0
$\widetilde{t}^*(M)$	1.7438	2.7081	2.8163	2.8273	2.8284	2.8285
$\widetilde{t_s}$	1.1334	1.4059	1.4269	1.4324	1.4332	1.4143

Table 2 Example 1, convergence with respect to Δt

N	16	31	61	91	121	151	180
$\widetilde{t}^*(M)$	2.8345	2.8315	2.829	2.8286	2.8286	2.8284	2.8284
$\widetilde{t_s}$	1.5117	1.5101	1.4617	1.4457	1.4379	1.4332	1.4299

Table 3 Example 1, convergence with respect to δ

δ	1	10^{-1}	10^{-2}	10^{-4}	10^{-6}	10^{-8}	10^{-10}	10^{-12}
$\widetilde{t}^*(M)$	2.5267	2.8048	2.8281	2.8284	2.8284	2.8284	2.8284	2.8284
$\widetilde{t}_s$	1.2802	1.4211	1.4329	1.4332	1.4332	1.4332	1.4332	1.4332

Table 4 Example 2

ε	1	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}
$\widetilde{t}^*(M)$	9.5229	10.4854	10.5791	10.5894	10.5894	10.5894

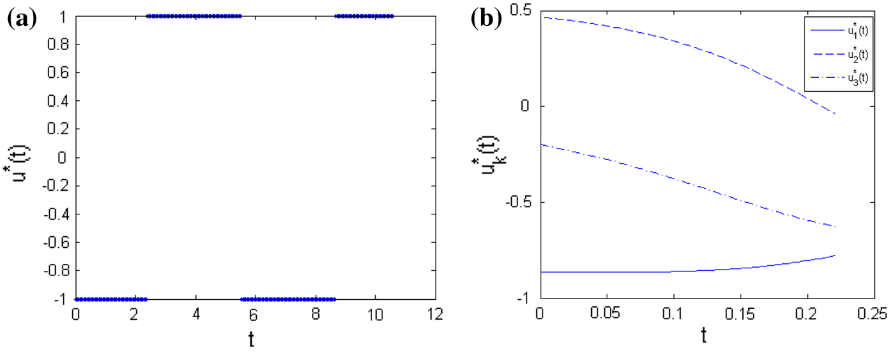

Fig. 2 Optimal controls of Examples 2 and 3. **a** Example 2, **b** Example 3

This is a classical test problem in control theory, where the exact result $t^*(M)$ is known to be 10.5831 (see section 4 of [10]). In this problem, the control could be interpreted as an exterior force acting on an oscillating weight hanging from a spring(see Page 37 of [6]). In Table 4 the approximated optimal times $\widetilde{t}^*(M)$ for different ε are given with $N = 301$. The optimal control for this example is illustrated in Fig. 2a, where $\varepsilon = 10^{-3}$. We can then find its switching structure between 1 and -1.

Example 3 Let

$$A = \begin{pmatrix} 5 & 5 & 0 \\ 5 & 10 & 0 \\ 0 & 0 & 9 \end{pmatrix}, B = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix}, y_0 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \text{ and } M = 1.$$

We choose ε to be 0 and $N = 81$ in this example. Table 5 shows the detailed results by the bisection method. After the 4th iteration in Algorithm 1, the optimal time is located in the interval [0.125, 0.25]. After 6 more iterations we get the approximated optimal time which is 0.2207. Table 6 shows the detailed results by the secant method. In Tables 5 and 6, the inner iterations is the iteration numbers of IRLS method, which is applied in each iteration of Algorithm 1.

We also give the plot of the optimal control for Examples 3 in Fig. 2b. Since the optimal control is of dimension 3 and thus it does not have the switching structure, which differs from Examples 1 and 2.

Table 5 Example 3

k	T_k	Inner iterations	$\widetilde{M}_{\varepsilon}^{*}(T_k)$
1	1	3	0.078058
2	0.5	2	0.28217
3	0.25	3	0.81934
4	0.125	3	2.479
k	t_k	Inner iterations	$\widetilde{M}_{\varepsilon}^{*}(t_k)$
1	0.1875	3	1.302
2	0.21875	2	1.0144
3	0.23438	2	0.90802
4	0.22656	2	0.95873
5	0.22266	2	0.98592
6	**0.2207**	2	1

Bold value indicates the optimal times

Table 6 Example 3

k	t_k	inner iterations	$\widetilde{M}_{\varepsilon}^{*}(t_k)$
1	0.1	5	3.4599
2	0.2	6	1.1726
3	0.20755	4	1.1043
4	0.21908	4	1.102
5	**0.22057**	3	1.0009

Bold value indicates the optimal times

Table 7 Example 4

ε	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
$\widetilde{t}^{*}(M)$	2.2629	2.3171	2.3225	2.323	2.323	2.323

Example 4 Let

$$A = -\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -0.8 & 0.3 & 0 & 0 \\ 0.3 & -1.1 & 0 & 0 \end{pmatrix}, \; B = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, \; y_0 = \begin{pmatrix} 1 \\ 0 \\ 2 \\ 0 \end{pmatrix} \text{ and } M = 1.$$

The system describes the displacements and velocities of two particles connected by a spring. Moreover, each particle is connected with a fixed end by another spring. In the system, the control denotes the external force imposed on the particles. See, Example 2.1.2 in [23]. Table 7 shows the convergence when ε decreases, where we choose $N = 151$. The optimal control of this example is presented in Fig. 3.

Example 5 Consider the discretization of one dimensional wave equation

$$\begin{cases} y_{tt} - y_{xx} = u & \text{in } (0,1) \times (0,+\infty), \\ y(0,t) = y(1,t) = 0 & \text{in } (0,+\infty), \\ y(x,0) = \sin(2\pi x) & \text{in } (0,1), \\ y_t(x,0) = 1 - \cos(2\pi x) & \text{in } (0,1). \end{cases} \tag{26}$$

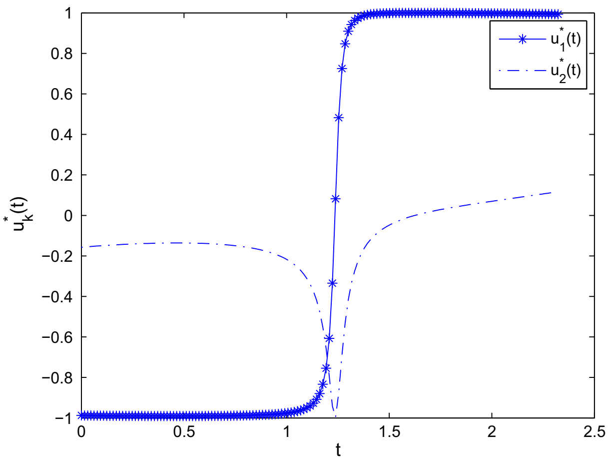

Fig. 3 Optimal control of Example 4

Table 8 Example 5

ε	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
$\widetilde{t}^*(M)$	1.7531	1.7832	1.7836	1.7836	1.7836	1.7836

Consider the finite difference discretization with the uniform mesh $\Delta x = \dfrac{1}{n+1}$ for the spatial variable, and introduce a new variable to y_t, we then obtain the ordinary differential equations (1) with

$$A = \begin{pmatrix} A_1 & A_2 \\ A_3 & A_4 \end{pmatrix}_{2n\times 2n}, \quad B = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}_{2n\times n}, \quad y_0 = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}_{2n\times 1},$$

where

$$A_1 = A_4 = \boldsymbol{O}_{n\times n}, \ A_2 = \boldsymbol{I}_{n\times n}, \ B_1 = \boldsymbol{O}_{n\times n}, \ B_2 = \boldsymbol{I}_{n\times n},$$

and

$$A_3 = (n+1)^2 \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ 0 & -1 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 2 \end{pmatrix}_{n\times n}, \ y_1 = \begin{pmatrix} \sin\frac{2\pi}{n+1} \\ \sin\frac{4\pi}{n+1} \\ \vdots \\ \sin\frac{n\pi}{n+1} \end{pmatrix}_{n\times 1}, \ y_2 = \begin{pmatrix} 1-\cos\frac{2\pi}{n+1} \\ 1-\cos\frac{4\pi}{n+1} \\ \vdots \\ 1-\cos\frac{n\pi}{n+1} \end{pmatrix}_{n\times 1}.$$

In order to observe whether the result of the ODE could approximate that of system (26), for given M and ε, we make a rascaling on them in the computation, i.e., we take $M\sqrt{n+1}$ and $\varepsilon\sqrt{n+1}$ instead of M and ε, respectively.

Let $M = 4$, $n = 29$ and $N = 101$, Table 8 shows the numerical optimal times $\widetilde{t}^*(M)$ as ε decreases. Then we fix $M = 4$, $N = 101$ and $\varepsilon = 10^{-4}$, Table 9 gives the different optimal times with respect to different spatial meshes.

Table 9 Example 5

$n+1$	10	20	30	40	50
$\widetilde{t}^*(M)$	1.7766	1.7827	1.7836	1.7836	1.7836

Table 10 Example 6, $a = 0.4$

ε	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
$\widetilde{t}^*(M)$	0.10334	0.11328	0.1187	0.12208	0.12436	0.12603	0.12733

Table 11 Example 6, $a = 0.4$

$n+1$	25	50	100
$\widetilde{t}^*(M)$	0.12603	0.12612	0.12666

Table 12 Example 6, $a = 0.8$

ε	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
$\widetilde{t}^*(M)$	0.071631	0.07373	0.074756	0.075342	0.075781	0.07583	0.07583

Example 6 The last example is the discretization to the one dimensional heat equation

$$\begin{cases} y_t - \Delta y = \chi_\omega u & \text{in } (0,1)\times(0,+\infty), \\ y(0,t) = y(1,t) = 0 & \text{in } (0,+\infty), \\ y(x,0) = \sin(2\pi x) & \text{in } (0,1), \end{cases}$$

where the control u acts into the system through the subinterval $\omega = (1-a, 1)$ with $0 < a \leq 1$. Using finite difference discretization with uniform mesh size $\Delta x = \dfrac{1}{n+1}$, and assuming $m = a(n+1)$ be an integer, we can obtain the discrete system to be the ordinary differential system (1), with

$$A = (n+1)^2 \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ 0 & -1 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 2 \end{pmatrix}_{n\times n}, B = \begin{pmatrix} \boldsymbol{O}_{(n-m)\times m} \\ \boldsymbol{I}_{m\times m} \end{pmatrix}, y_0 = \begin{pmatrix} \sin\frac{2\pi}{n+1} \\ \sin\frac{4\pi}{n+1} \\ \vdots \\ \sin\frac{n\pi}{n+1} \end{pmatrix}_{n\times 1}.$$

Similar as in Example 4, we re-scale M and ε to $M\sqrt{n+1}$ and $\varepsilon\sqrt{n+1}$, respectively.

Let $M = 2$, we consider two different control domains: $a = 0.4$ and $a = 0.8$ (the smaller a implies a stronger instability in the computation). For both cases, we first let $N = 101$, $n = 24$, and then obtain $m = 10$ and $m = 20$, respectively. Tables 10 and 12 presents the approximated optimal times for different ε, from which one may find the latter one has a better convergence. Then we consider the convergence with respect to the mesh size. For $a = 0.4$ and $a = 0.8$, we fix $\varepsilon = 10^{-8}$. In Tables 11 and 13 one can find the approximated numerical optimal times $\widetilde{t}^*(M)$ for different mesh sizes, where the convergence is observed numerically.

Table 13 Example 6, $a = 0.8$

$n+1$	25	50	100
$\widetilde{t}^*(M)$	0.07583	0.077783	0.076855

5 Conclusion

We have proposed a new algorithm to compute the time optimal control problems for linear ODEs with Euclidean constraints. This algorithm is based on the equivalence between the time optimal control problem $(TP)^M$ and the norm optimal control problem $(NP)^T$. To avoid the possible numerical blow up of $(NP)^T$ we introduced perturbation problems $(NP)^T_\varepsilon$ with $\varepsilon \geq 0$. By utilizing the IRLS method, we could compute $(NP)^T_\varepsilon$ approximately. The main advantage of this algorithm is simple and efficient, which was confirmed by several numerical examples.

There are several avenues deserving further investigations. First, the proposed algorithm can be applied to the time optimal control problem governed by the heat or the wave equations with L^2-norm constraints, but the convergence of the algorithm needs a careful study. Second, how to extend the proposed method to a general constraint for the control variables is still missing. Finally, the governing equations for some real applications of the time optimal control problems are nonlinear ODE systems. One natural research problem is to extend the analysis and numerics of the proposed approach to these more complicated problems.

Acknowledgements The authors would like to thank the referees for their valuable comments, and to thank Professor Kunisch Karl and Professor Gengsheng Wang for their interesting inputs. The work of X. Lu was partially supported by the National Natural Science Foundation of China under Grants Nos. 11471253 and 91630313, and the work of L. Wang was partially supported by the National Natural Science Foundation of China under Grant No. 11371285.

Appendix

Proof of Lemma 1 The proof is divided into four steps as follows.

Step 1.

$$\text{If } \mu \neq 0, \text{ then } B^{\mathrm{T}}\varphi(t; T, \mu) = 0 \text{ at most finite } t \in [0, T]. \tag{27}$$

It can be proved by contradiction. Assume that $B^{\mathrm{T}}\varphi(t; T, \mu) = 0$ at infinite $t \in [0, T]$. Since $B^{\mathrm{T}}\varphi(t; T, \mu)$ is analytic, we then obtain $B^{\mathrm{T}}\varphi(t; T, \mu) \equiv 0$ in $[0, T]$, which implies that

$$B^{\mathrm{T}}\mu = 0,\ B^{\mathrm{T}}A^{\mathrm{T}}\mu = 0, \ldots, B^{\mathrm{T}}(A^{\mathrm{T}})^{n-1}\mu = 0. \tag{28}$$

From (H_2) and (28), it follows that

$$\mu = 0,$$

which leads to a contradiction.

Step 2. The bang-bang property of $(TP)^M$.

By Pontryagin's maximum principle (see Theorem 3.5 in Chapter 3 of [24]), there exists $\xi \in \mathbb{R}^n$ with $\|\xi\| = 1$ such that

$$\max_{\|v\| \leq M} \langle \varphi(t; t^*(M), \xi), Bv \rangle = \langle \varphi(t; t^*(M), \xi), Bu^*_M(t) \rangle, \text{ for a.e. } t \in (0, t^*(M)).$$

This, together with (27), indicates the bang-bang property of $(TP)^M$. Then the uniqueness of the optimal control follows.

Step 3. For any $\tau \in (0, T)$, there exists a generic positive constant $C(\tau)$ dependent on τ, such that

$$\|\varphi(0; \tau, \mu)\| \leq C(\tau) \int_0^\tau \|B^{\mathrm{T}}\varphi(t; \tau, \mu)\| \, dt, \quad \forall \mu \in \mathbb{R}^n. \tag{29}$$

It can be proved by contradiction. Assume that there exists $\mu_\ell \in \mathbb{R}^n$ with $\|\mu_\ell\| = 1$ and

$$\frac{\int_0^\tau \|B^{\mathrm{T}}\varphi(t; \tau, \mu_\ell)\| \, dt}{\|\varphi(0; \tau, \mu_\ell)\|} < \frac{1}{\ell}, \quad \forall \ell \geq 1. \tag{30}$$

Since $\|\mu_\ell\| = 1$, there exists a subsequence of $\{\mu_\ell\}_{\ell \geq 1}$, still denoted by $\{\mu_\ell\}$, and $\mu_0 \in \mathbb{R}^n$, such that

$$\lim_{\ell \to \infty} \mu_\ell = \mu_0 \text{ and } \|\mu_0\| = 1.$$

Passing to the limit for $\ell \to +\infty$ in (30), we have that

$$\frac{\int_0^\tau \|B^{\mathrm{T}}\varphi(t; \tau, \mu_0)\| \, dt}{\|\varphi(0; \tau, \mu_0)\|} = 0.$$

This implies that $B^{\mathrm{T}}\varphi(t; \tau, \mu_0) = 0$ for a.e. $t \in [0, T]$ and contradicts (27).

Step 4. By (29) and the equivalence between the observability and controllability, see e.g. Theorem 2.1 in [19], we get that for any $z_0 \in \mathbb{R}^n$, there exists a control $u \in L^\infty(0, \tau; \mathbb{R}^m)$ with

$$\|u\|_{L^\infty(0,\tau;\mathbb{R}^m)} \leq C(\tau)\|z_0\|, \tag{31}$$

such that the solution $z(\cdot)$ to the equation

$$\begin{cases} z'(t) + Az(t) = Bu(t), & t \in (0, \tau), \\ z(0) = z_0 \end{cases}$$

satisfies that $z(\tau) = 0$.

By (31) and using the same argument as in [22], we have that (ii), (iii) and (iv) hold. Finally, the bang-bang property and the uniqueness of the optimal control for $(NP)^T$ follow from (iv) and Step 2. □

Proof of Proposition 1 By changing variables, it suffices to show that there exists a solution $(T, y(\cdot), u(\cdot), \psi(\cdot))$ to the the following system

$$\begin{cases} y'(t) + Ay(t) = Bu(t), & t \in (0, T), \\ \psi'(t) = A^{\mathrm{T}}\psi(t), & t \in (0, T), \\ u(t) = M\dfrac{B^{\mathrm{T}}\psi(t)}{\|B^{\mathrm{T}}\psi(t)\|}, & t \in (0, T), \\ M = \displaystyle\int_0^T \|B^{\mathrm{T}}\psi(t)\| \, dt, \\ y(0) = y_0, \\ y(T) = 0, \end{cases} \tag{32}$$

where T is the optimal time to $(TP)^M$ and $u(\cdot)$, when extended by zero to $(T, +\infty)$, is the optimal control to $(TP)^M$.

On one hand, let $y^*(\cdot)$ and $u^*(\cdot)$ be the optimal state and optimal control to $(TP)^M$, respectively and let μ^* be a minimizer of $J^{t^*(M)}(\cdot)$. By (iv) in Lemma 1 and (iii) in Lemma 2, we have that

$$\big(t^*(M), y^*(\cdot), u^*(\cdot), \varphi(\cdot; t^*(M), \mu^*)\big)$$

is a solution to (32). On the other hand, for any solution $(T, y(\cdot), u(\cdot), \psi(\cdot))$ to (32) and any $\nu \in \mathbb{R}^n$, multiplying the first equation of (32) by $\varphi(\cdot; T, \nu)$ and integrating it over $(0, T)$, we get that

$$\int_0^T \|B^{\mathrm{T}}\psi(t)\|\, dt \int_0^T \frac{\langle B^{\mathrm{T}}\psi(t), B^{\mathrm{T}}\varphi(t; T, \nu)\rangle}{\|B^{\mathrm{T}}\psi(t)\|}\, dt + \langle \varphi(0; T, \nu), y_0\rangle = 0.$$

This, together with (ii) in Lemma 2, indicates that

$$\psi(T) \text{ is a minimizer of } J^T(\cdot). \tag{33}$$

It follows from the third equality, the fourth equality in (32), (33) and (iii) in Lemma 2 that

$$u(\cdot),\ \text{when extended by zero to } (T, +\infty),\ \text{is the optimal control to } (NP)^T. \tag{34}$$

By the third equality in (32) and (34), we obtain that $M = M^*(T)$, which, combined with (iii) and (iv) in Lemma 1, implies

$$T = t^*(M). \tag{35}$$

From this equation, (34) and (iv) in Lemma 1, we obtain the conclusion. □

Proof of Lemma 2 (i) It is obvious that $J_\varepsilon^T(\cdot)$ is continuous and convex. By (29), we have that

$$\begin{aligned} J_\varepsilon^T(\mu) &\geq \frac{1}{2}\left(\int_0^T \|B^{\mathrm{T}}\varphi(s; T, \mu)\|\, ds\right)^2 - C(T)\|y_0\| \int_0^T \|B^{\mathrm{T}}\varphi(s; T, \mu)\|\, ds + \varepsilon\|\mu\| \\ &= \frac{1}{2}\left(\int_0^T \|B^{\mathrm{T}}\varphi(s; T, \mu)\|\, ds - C(T)\|y_0\|\right)^2 - \frac{1}{2}C(T)^2\|y_0\|^2 + \varepsilon\|\mu\|. \end{aligned}$$

We will prove

$$\lim_{\|\mu\|\to+\infty} \int_0^T \|B^{\mathrm{T}}\varphi(s; T, \mu)\|\, ds = +\infty \tag{36}$$

by contradiction, and the coercivity follows immediately. Assume that there exist $\{\mu_k\} \subset \mathbb{R}^n$ and a positive constant C which is independent of k, such that $\|\mu_k\| \to +\infty$ and

$$\int_0^T \|B^{\mathrm{T}}\varphi(s; T, \mu_k)\|\, ds \leq C \text{ for each } k \geq 1. \tag{37}$$

Let $v_k \triangleq \dfrac{\mu_k}{\|\mu_k\|}$, then there exists a subsequence of $\{v_k\}_{k\geq 1}$, still denoted by the same way, and $v_0 \in \mathbb{R}^n$ with $\|v_0\| = 1$ and $v_k \to v_0$. From (37), it follows that

$$\int_0^T \|B^{\mathrm{T}}\varphi(s; T, v_k)\|\, ds \leq \frac{C}{\|\mu_k\|} \text{ for each } k \geq 1.$$

Passing to the limit for $k \to \infty$ in the above inequality, we get that

$$\int_0^T \|B^{\mathrm{T}}\varphi(s; T, v_0)\|\, ds = 0.$$

This implies that $v_0 = 0$, which leads to a contradiction. By the continuity and the coercivity of $J_\varepsilon^T(\cdot)$, there exists a minimizer μ_ε^*.

Next we prove that $\mu_\varepsilon^* \neq 0$. If not, we have that for any $\mu \in \mathbb{R}^n$,

$$J_\varepsilon^T(\lambda\mu) = \frac{\lambda^2}{2}\left(\int_0^T \|B^{\mathrm{T}}\varphi(s; T, \mu)\|\, ds\right)^2 + \lambda\langle y_0, \varphi(0; T, \mu)\rangle + \lambda\varepsilon\|\mu\| \geq 0, \quad \forall\, \lambda > 0.$$

and

$$J_\varepsilon^T(-\lambda\mu) = \frac{\lambda^2}{2}\left(\int_0^T \|B^{\mathrm{T}}\varphi(s; T, \mu)\|\, ds\right)^2 - \lambda\langle y_0, \varphi(0; T, \mu)\rangle + \lambda\varepsilon\|\mu\| \geq 0, \quad \forall\, \lambda > 0.$$

From the above two inequalities, we obtain that

$$|\langle y_0, \varphi(0; T, \mu)\rangle| \leq \frac{\lambda}{2}\left(\int_0^T \|B^{\mathrm{T}}\varphi(s; T, \mu)\|\, ds\right)^2 + \varepsilon\|\mu\|, \quad \forall\, \lambda > 0.$$

Letting $\lambda \to 0^+$, we obtain that

$$|\langle y_0, \varphi(0; T, \mu)\rangle| \leq \varepsilon\|\mu\|, \quad \forall\, \mu \in \mathbb{R}^n,$$

i.e.,

$$|\langle e^{-AT} y_0, \mu\rangle| \leq \varepsilon\|\mu\|, \quad \forall\, \mu \in \mathbb{R}^n.$$

This implies $\|e^{-AT} y_0\| \leq \varepsilon$, which contradicts $T < T_\varepsilon^*$.

(ii) On one hand, if μ_ε^* is a minimizer of $J_\varepsilon^T(\cdot)$, we have that

$$\frac{J_\varepsilon^T(\mu_\varepsilon^* + \lambda\nu) - J_\varepsilon^T(\mu_\varepsilon^*)}{\lambda} \geq 0, \quad \forall\, \lambda > 0, \quad \forall\, \nu \in \mathbb{R}^n.$$

Letting $\lambda \to 0^+$ in the above inequality, one can check that (5) holds. On the other hand, since $J_\varepsilon^T(\cdot)$ is convex and the minimizer is nonzero, then $\nabla J_\varepsilon^T(\mu_\varepsilon^*) = 0$ (which is equation (5)) implies $\mu_\varepsilon^* \neq 0$ is a minimizer of $J_\varepsilon^T(\cdot)$.

(iii) Let μ_ε^* be a minimizer of $J_\varepsilon^T(\cdot)$ and u_ε^* be defined by (6). First, we can prove that u_ε^* is an admissible control for $(NP)_\varepsilon^T$. Let $y(\cdot)$ be the solution to

$$\begin{cases} y'(t) + Ay(t) = Bu_\varepsilon^*(t),\ t \in (0, T), \\ y(0) = y_0. \end{cases}$$

Multiplying the first equation of the above system by $\varphi(\cdot; T, \nu)$ and integrating it over $(0, T)$, we get that

$$\langle \nu, y(T)\rangle - \langle \varphi(0; T, \nu), y_0\rangle = \int_0^T \langle u_\varepsilon^*(t), B^{\mathrm{T}}\varphi(t; T, \nu)\rangle\, dt.$$

From the latter and (6) it follows that

$$\begin{aligned} &\langle \nu, y(T)\rangle \\ &\quad = \langle \varphi(0; T, \nu), y_0\rangle + \int_0^T \langle u_\varepsilon^*(t), B^{\mathrm{T}}\varphi(t; T, \nu)\rangle\, dt \\ &\quad = \langle \varphi(0; T, \nu), y_0\rangle + \int_0^T \|B^{\mathrm{T}}\varphi(t; T, \mu_\varepsilon^*)\|\, dt \int_0^T \frac{\langle B^{\mathrm{T}}\varphi(t; T, \mu_\varepsilon^*), B^{\mathrm{T}}\varphi(t; T, \nu)\rangle}{\|B^{\mathrm{T}}\varphi(t; T, \mu_\varepsilon^*)\|}\, dt, \end{aligned}$$

which, combined with (5), indicates that

$$|\langle \nu, y(T)\rangle| = \left| \varepsilon \frac{\langle \mu_\varepsilon^*, \nu\rangle}{\|\mu_\varepsilon^*\|} \right| \leq \varepsilon \|\nu\|, \quad \forall\, \nu \in \mathbb{R}^n.$$

This implies $\|y(T)\| \leq \varepsilon$ and u_ε^* is an admissible control for $(NP)_\varepsilon^T$.

Then, we prove that u_ε^* is an optimal control for $(NP)_\varepsilon^T$. For any $u \in L^\infty(0, +\infty; \mathbb{R}^m)$ satisfying that

$$\begin{cases} h'(t) + Ah(t) = Bu(t),\ t \in (0, T), \\ h(0) = y_0, \\ \|h(T)\| \leq \varepsilon, \end{cases}$$

multiplying the first equation of the above system by $\varphi(\cdot; T, \mu_\varepsilon^*)$ and integrating it over $(0, T)$, we get that

$$\langle \mu_\varepsilon^*, h(T)\rangle - \langle \varphi(0; T, \mu_\varepsilon^*), y_0\rangle = \int_0^T \langle u(t), B^{\mathrm{T}}\varphi(t; T, \mu_\varepsilon^*)\rangle\, dt.$$

This, together with (5), implies that

$$\langle \mu_\varepsilon^*, h(T)\rangle - \int_0^T \langle u(t), B^{\mathrm{T}}\varphi(t; T, \mu_\varepsilon^*)\rangle\, dt = -\varepsilon\|\mu_\varepsilon^*\| - \left(\int_0^T \|B^{\mathrm{T}}\varphi(t; T, \mu_\varepsilon^*)\|\, dt \right)^2.$$

Since $\|h(T)\| \leq \varepsilon$, it follows from the latter equality that

$$\int_0^T \langle u(t), B^{\mathrm{T}}\varphi(t; T, \mu_\varepsilon^*)\rangle\, dt - \left(\int_0^T \|B^{\mathrm{T}}\varphi(t; T, \mu_\varepsilon^*)\|\, dt \right)^2 = \langle \mu_\varepsilon^*, h(T)\rangle + \varepsilon\|\mu_\varepsilon^*\| \geq 0.$$

Thus

$$\|u\|_{L^\infty(0,T;\mathbb{R}^m)} \geq \int_0^T \|B^{\mathrm{T}}\varphi(t; T, \mu_\varepsilon^*)\|\, dt = \|u_\varepsilon^*\|_{L^\infty(0,T;\mathbb{R}^m)}.$$

This shows that u_ε^* is an optimal control to $(NP)_\varepsilon^T$ and

$$M_\varepsilon^*(T) = \int_0^T \|B^{\mathrm{T}}\varphi(t; T, \mu_\varepsilon^*)\|\, dt.$$

□

Proof of Proposition 2 Let $0 < \varepsilon < \|y_0\|$ and $0 < T < T_\varepsilon^*$. For any $\mu_1, \mu_2 \in \mathbb{R}^n$, $0 < \alpha < 1$, we have that

$$\begin{aligned}
&J_\varepsilon^T(\alpha\mu_1 + (1-\alpha)\mu_2) \\
&= \frac{1}{2}\left(\int_0^T \|B^{\mathrm{T}}\varphi(s; T, \alpha\mu_1 + (1-\alpha)\mu_2)\|\, ds \right)^2 \\
&\quad + \langle y_0, \varphi(0; T, \alpha\mu_1 + (1-\alpha)\mu_2)\rangle + \varepsilon\|\alpha\mu_1 + (1-\alpha)\mu_2\| \\
&\leq \frac{1}{2}\left(\int_0^T (\alpha\|B^{\mathrm{T}}\varphi(s; T, \mu_1)\| + (1-\alpha)\|B^{\mathrm{T}}\varphi(s; T, \mu_2)\|)\, ds \right)^2 \\
&\quad + \alpha\langle y_0, \varphi(0; T, \mu_1)\rangle + (1-\alpha)\langle y_0, \varphi(0; T, \mu_2)\rangle + \alpha\varepsilon\|\mu_1\| + (1-\alpha)\varepsilon\|\mu_2\| \\
&\leq \frac{\alpha}{2}\left(\int_0^T \|B^{\mathrm{T}}\varphi(s; T, \mu_1)\|\, ds \right)^2 + \frac{1-\alpha}{2}\left(\int_0^T \|B^{\mathrm{T}}\varphi(s; T, \mu_2)\|\, ds \right)^2 \\
&\quad + \alpha\langle y_0, \varphi(0; T, \mu_1)\rangle + (1-\alpha)\langle y_0, \varphi(0; T, \mu_2)\rangle + \alpha\varepsilon\|\mu_1\| + (1-\alpha)\varepsilon\|\mu_2\| \\
&= \alpha J_\varepsilon^T(\mu_1) + (1-\alpha)J_\varepsilon^T(\mu_2).
\end{aligned} \tag{38}$$

First, we notice that the first inequality in (38) becomes equality if and only if there exists a constant $k \geq 0$ and a function $\beta(\cdot) : [0, T] \to [0, +\infty)$, such that

$$\begin{cases} \mu_1 = k\mu_2, \\ B^{\mathrm{T}}\varphi(t; T, \mu_1) = \beta(t)B^{\mathrm{T}}\varphi(t; T, \mu_2), \text{ for a.e. } t \in (0, T). \end{cases}$$

These imply that the first inequality in (38) becomes equality if and only if there exists a constant $k \geq 0$, such that

$$\mu_1 = k\mu_2.$$

Then, the second inequality in (38) becomes equality if and only if

$$\int_0^T \|B^{\mathrm{T}}\varphi(t; T, \mu_1)\| \, dt = \int_0^T \|B^{\mathrm{T}}\varphi(t; T, \mu_2)\| \, dt.$$

Hence $J_\varepsilon^T(\alpha\mu_1 + (1-\alpha)\mu_2) = \alpha J_\varepsilon^T(\mu_1) + (1-\alpha)J_\varepsilon^T(\mu_2)$ if and only if $\mu_1 = \mu_2$. This shows that $J_\varepsilon^T(\cdot)$ is strictly convex and it has a unique minimizer. □

References

1. Barbu, V.: Analysis and Control of Nonlinear Infinite Dimensional Systems. Academic Press, Boston (1993)
2. Chen, X.: Smoothing methods for nonsmooth, nonconvex minimization. Math. Program. **134**, 71–99 (2012)
3. Daubechies, I., Defrise, M., De Mol, C.: An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. Commun. Pure Appl. Math. **57**, 1413–1457 (2004)
4. Daubechies, I., DeVore, R., Fornasier, M., Güntürk, C.S.: Iteratively reweighted least squares minimization for sparse recovery. Commun. Pure Appl. Math. **63**, 1–38 (2010)
5. Eckstein, J., Bertsekas, D.P.: On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math. Program. **55**, 293–318 (1992)
6. Evans, L.C.: An Introduction to Mathematical Optimal Control Theory. Lecture Notes, Department of Mathematics, University of California, Berkeley (2008)
7. Fattorini, H.O.: Infinite Dimensional Linear Control Systems, the Time Optimal and Norm Optimal Control Problems, vol. 201. Elsevier, Amsterdam (2005)
8. Hermes, H., LaSalle, J.: Functional Analysis and Time Optimal Control. Academic Press, New York (1969)
9. Hintermüller, M., Ito, K., Kunisch, K.: The primal-dual active set strategy as a semismooth Newton method. SIAM J. Optim. **13**, 865–888 (2002)
10. Ito, K., Kunisch, K.: Semismooth newton methods for time-optimal control for a class of ODES. SIAM J. Control Optim. **48**, 3997–4013 (2010)
11. Kaya, C.Y., Noakes, J.L.: Computations and time-optimal controls. Optim. Control Appl. Methods **17**, 171–185 (1996)
12. Kaya, C.Y., Noakes, J.L.: Computational methods for time-optimal switching controls. J. Optim.Theory Appl **117**, 69–92 (2003)
13. Kunisch, K., Wang, L.: Time optimal control of the heat equation with pointwise control constraints. ESAIM Control Optim. Calc. Var. **19**, 460–485 (2013)
14. Lasalle, J.P.: The Time Optimal Control Problem, Contributions to the Theory of Nonlinear Oscillations, vol. 5, pp. 1–24. Princeton University Press, Princeton (1960)
15. Li, X., Yong, J.: Optimal Control Theory for Infinite Dimensional Systems. Birkhäuser, Boston (1995)
16. Meier, E., Bryson, A.E.: Efficient algorithms for time-optimal rigid spacecraft reorientation problem. J. Guid. Control Dyn. **13**, 859–866 (1990)
17. Pontryagin, L.S., Boltyanskii, V.G., Gamkrelidze, R.V., Mishchenko, E.F.: The Mathematical Theory of Optimal Processes. Interscience Publishers, New York (1962)
18. Quarteroni, A., Sacco, R., Saleri, F.: Numerical Mathematics. Springer, Berlin (2000)
19. Sontag, E.D.: Mathematical Control Theory: Deterministic Finite Dimensional Systems. Springer, New York (1998)

20. Wang, G., Zheng, G.: An approach to the optimal time for a time optimal control problem of an internally controlled heat equation. SIAM J. Control Optim. **50**, 601–628 (2012)
21. Wang, G., Zuazua, E.: On the equivalence of minimal time and minimal norm controls for internally controlled heat equations. SIAM J. Control Optim. **50**, 2938–2958 (2012)
22. Xu, S.: The Compution of Matrices in Control Theorey (in Chinese). China Higher Education Press, Beijing (2011)
23. Yong, J., Lou, H.: A Concise Course Of Optimal Control Theory. China Higher Education Press, Beijing (2006). (in Chinese)
24. Zhang, C.: The time optimal control with constraints of the rectangular type for linear time-varying ODEs. SIAM J. Control Optim. **51**, 1528–1542 (2013)
25. Zuazua, E.: Controllability and observability of partial differential equations: some results and open problems. In: Dafermos, C.M., Feireisl, E. (eds.) Handbook of Differential Equations: Evolutionary Equations, vol. 3, pp. 527–621. Elsevier, Amsterdam (2006)