Copyright © 2014 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Guaranteed-Cost Controls of Minimal Variation: A Numerical Algorithm Based on Control Parameterization

Ryan LOXTON[1,2], LIN Qun[1], Kok Lay TEO[1]

1. Department of Mathematics and Statistics, Curtin University, Perth, Australia
E-mail: r.loxton@curtin.edu.au; q.lin@curtin.edu.au; k.l.teo@curtin.edu.au

2. Institute of Cyber-Systems and Control, Zhejiang University, Hangzhou, China
E-mail: rcloxton@gmail.com

Abstract: The optimal control literature is dominated by standard problems in which the system cost functional is expressed in the well-known Bolza form. Such Bolza cost functionals consist of two terms: a Mayer term (which depends solely on the final state reached by the system) and a Lagrange integral term (which depends on the state and control values over the entire time horizon). One limitation with the standard Bolza cost functional is that it does not consider the cost of control changes. Such costs should certainly be considered when designing practical control strategies, as changing the control signal will invariably cause wear and tear on the system's acutators. Accordingly, in this paper, we propose a new optimal control formulation that balances system performance with control variation. The problem is to minimize the total variation of the control signal subject to a guaranteed-cost constraint that ensures an acceptable level of system performance (as measured by a standard Bolza cost functional). We first apply the control parameterization method to approximate this problem by a non-smooth dynamic optimization problem involving a finite number of decision variables. We then devise a novel transformation procedure for converting this non-smooth dynamic optimization problem into a smooth problem that can be solved using gradient-based optimization techniques. The paper concludes with numerical examples in fisheries and container crane control.

Key Words: Optimal control, Total variation, Control parameterization, Nonlinear optimization

1 Introduction

The standard optimal control problem is well-known to researchers in control theory. This problem involves choosing a control input signal for a given dynamic system so that the system evolves in the best possible manner. One limitation with the standard optimal control framework is that the cost of changing the control signal is usually ignored. For example, in the classical Mayer optimal control problem, where the system cost is expressed solely as a function of the final state reached by the system, two control laws that drive the system to the same final state will have the same cost—even if one of them is constant and the other fluctuates wildly. Thus, the standard optimal control framework does not distinguish between these two controls, despite the constant control law being preferred in applications.

Whether it is wear and tear on mechanical components, losses in workforce productivity due to company policy changes, or transaction costs in investment portfolios, there is always a cost associated with changing the control action. Thus, it is important to consider such costs when designing an optimal control strategy. Indeed, an "optimal" strategy that is highly volatile will be of little use in practice.

We are only aware of several references in the optimal control literature that consider the cost of control changes. References [1, 2] discuss theoretical conditions for solving optimal control problems in which the cost functional includes a total variation term to penalize control changes. Reference [3] presents an algorithm for solving a simple class of optimal control problems in which the control signal can assume only two possible values, and there is a cost associated with changing from one value to another. A more general algorithm is developed in [4] for solving constrained optimal control problems in which the cost functional explicitly penalizes control changes. An alternative algorithm for solving the same class of problems is given in [5]. The algorithms in [4, 5] are based on the concept of control parameterization, whereby the control signal is approximated by a linear combination of basis functions [6, 7].

In references [4, 5, 7], the cost functional consists of two parts: the traditional Mayer/Bolza cost and a total variation term that measures the cost of control changes. The relative importance of each term is adjusted via a weighting factor; however, precise rules for choosing this weighting factor have yet to be developed. In this paper, we aim to circumvent this difficulty by exploring an alternative formulation in which the control variation is minimized subject to an upper bound on the traditional Bolza cost. Thus, we seek a control of minimal variation that satisfies a given performance requirement on the system (in the traditional Bolza sense).

The paper is organized as follows. In Section 2, we define the problem under consideration. In Section 3, we apply the control parameterization method to obtain a class of non-smooth approximate problems. In Section 4, we develop a transformation technique for converting the approximate problems in Section 3 into smooth problems that can be solved using conventional dynamic optimization methods. Finally, in Section 5, we apply the proposed approach to example problems in fisheries and container crane control.

2 Problem Statement

Consider the following nonlinear control system:

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}(t), \boldsymbol{u}(t)), \quad t \in [0, T], \tag{1}$$

$$\boldsymbol{x}(0) = \boldsymbol{x}^0, \tag{2}$$

where $\boldsymbol{x}(t) \in \mathbb{R}^n$ is the state at time t, $\boldsymbol{u}(t) \in \mathbb{R}^r$ is the control at time t, $\boldsymbol{x}^0 \in \mathbb{R}^n$ is a given initial state, $T > 0$ is

This work was supported by the National Natural Science Foundation of China (International Young Scientists Research Fund 11350110208) and the State Key Laboratory for Industrial Control Technology at Zhejiang University, China (Open Research Project ICT1301).

a given terminal time, and $\boldsymbol{f} : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}^n$ is a given continuously differentiable function.

Let $u_i : [0,T] \to \mathbb{R}$ denote the ith component of the vector-valued control signal $\boldsymbol{u} : [0,T] \to \mathbb{R}^r$. Then the *total variation* of u_i is defined by

$$\bigvee_0^T u_i = \sup \sum_{j=1}^m \big|u_i(t_j) - u_i(t_{j-1})\big|, \tag{3}$$

where the supremum is taken over all finite partitions $\{t_j\}_{j=0}^m \subset [0,T]$ satisfying

$$0 = t_0 < t_1 < \cdots < t_{m-1} < t_m = T.$$

Clearly, the total variation defined in (3) is always non-negative. Moreover, the total variation is zero if and only if u_i is constant.

The total variation of the vector-valued control signal $\boldsymbol{u} : [0,T] \to \mathbb{R}^r$ is defined as the sum of the total variations of its components:

$$\bigvee_0^T \boldsymbol{u} = \sum_{i=1}^r \bigvee_0^T u_i.$$

Note that the total variation measures the extent to which the control signal changes during the time horizon—the more change, the higher the total variation. If the total variation of $\boldsymbol{u}$ is finite, then we say that $\boldsymbol{u}$ is of *bounded variation*.

We impose the following *bound constraints* on $\boldsymbol{u}$:

$$a_i \le u_i(t) \le b_i, \quad t \in [0,T], \quad i = 1, \ldots, r, \tag{4}$$

where a_i and b_i, $i = 1, \ldots, r$, are given constants. Any function $\boldsymbol{u} : [0,T] \to \mathbb{R}^r$ of bounded variation that satisfies the bound constraints (4) is called an *admissible control*. Let $\mathcal{U}$ denote the class of all such admissible controls.

We assume throughout this paper that for each admissible control $\boldsymbol{u} \in \mathcal{U}$, system (1)-(2) admits a unique Carathéodory solution. Let $\boldsymbol{x}(\cdot|\boldsymbol{u})$ denote this solution.

Consider the following well-known *Bolza cost functional*:

$$g(\boldsymbol{u}) = \underbrace{\Phi(\boldsymbol{x}(T|\boldsymbol{u}))}_{\text{Final cost}} + \underbrace{\int_0^T \mathcal{L}(\boldsymbol{x}(t|\boldsymbol{u}), \boldsymbol{u}(t))dt}_{\text{Running cost}}, \tag{5}$$

where $\Phi : \mathbb{R}^n \to \mathbb{R}$ and $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}$ are given continuously differentiable functions. Such cost functionals are commonly used in the optimal control literature to evaluate system performance. The standard optimal control problem involves choosing an admissible control $\boldsymbol{u} \in \mathcal{U}$ to minimize (5). However, this standard problem does not consider the cost of control changes; thus, the "optimal" control could be a volatile control strategy that is difficult—and potentially dangerous—to implement in practice.

We propose a new approach in this paper. Instead of minimizing (5), we minimize the total variation of the control signal subject to an upper bound on (5). In other words, we seek a control of *guaranteed-cost* that has the smallest possible total variation.

Let Δ^* denote the maximum Bolza cost. Then our *guaranteed-cost constraint* is expressed as follows:

$$\Phi(\boldsymbol{x}(T|\boldsymbol{u})) + \int_0^T \mathcal{L}(\boldsymbol{x}(t|\boldsymbol{u}), \boldsymbol{u}(t))dt \le \Delta^*. \tag{6}$$

Let $\mathcal{F}$ denote the set of all $\boldsymbol{u} \in \mathcal{U}$ satisfying (6). We now state our optimal control problem as follows.

Problem P. *Choose $\boldsymbol{u} \in \mathcal{F}$ to minimize the total variation*

$$J(\boldsymbol{u}) = \bigvee_0^T \boldsymbol{u}.$$

Note that it is possible to include additional constraints within the framework of Problem P—for example, path constraints on the state variables. However, we ignore such constraints for simplicity.

3 Control Parameterization

The major difficulty with Problem P is that there is no closed-form analytical formula for computing the total variation of $\boldsymbol{u}$. In this section, we will apply the control parameterization method [6] to approximate $\boldsymbol{u}$ by a piecewise-constant function in which the heights and discontinuity points are decision variables to be selected optimally. As we will see, under this approximation scheme, the total variation reduces to a simple formula that can be computed easily.

Let $p \ge 2$ be a given integer. We approximate the control signal as follows:

$$\boldsymbol{u}(t) \approx \boldsymbol{u}^p(t) = \boldsymbol{\sigma}^k, \ \ t \in [\tau_{k-1}, \tau_k), \ \ k = 1, \ldots, p, \tag{7}$$

where $\boldsymbol{\sigma}^k$ is the approximate value of the control signal on the subinterval $[\tau_{k-1}, \tau_k)$. Both the control values $\boldsymbol{\sigma}^k$ and the subinterval end-points τ_k are decision variables. The subinterval end-points satisfy the following constraints:

$$\tau_k - \tau_{k-1} \ge \epsilon, \quad k = 1, \ldots, p, \tag{8}$$

where $\tau_0 = 0$, $\tau_p = T$, and $\epsilon > 0$ is a constant. Let $\mathcal{T}$ denote the set of all $\boldsymbol{\tau} = [\tau_1, \ldots, \tau_{p-1}]^\top \in \mathbb{R}^{p-1}$ satisfying (8).

Define

$$\boldsymbol{\sigma} = (\boldsymbol{\sigma}^1, \ldots, \boldsymbol{\sigma}^p) = \big[(\boldsymbol{\sigma}^1)^\top, \ldots, (\boldsymbol{\sigma}^p)^\top\big]^\top \in \mathbb{R}^{pr}. \tag{9}$$

In view of (4), the control values in (9) must satisfy

$$a_i \le \sigma_i^k \le b_i, \quad i = 1, \ldots, r, \quad k = 1, \ldots, p. \tag{10}$$

Let $\mathcal{S}$ denote the set of all $\boldsymbol{\sigma} = (\boldsymbol{\sigma}^1, \ldots, \boldsymbol{\sigma}^p) \in \mathbb{R}^{pr}$ satisfying the bound constraints (10).

For each $\boldsymbol{\tau} \in \mathcal{T}$, define

$$\mathcal{I}_k(\boldsymbol{\tau}) = \begin{cases} [\tau_{k-1}, \tau_k), & \text{if } k = 1, \ldots, p-1, \\ [\tau_{k-1}, \tau_k], & \text{if } k = p. \end{cases}$$

Then the approximate control in (7) can be written as

$$\boldsymbol{u}^p(t) = \boldsymbol{u}^p(t|\boldsymbol{\tau}, \boldsymbol{\sigma}) = \sum_{k=1}^p \boldsymbol{\sigma}^k \chi_{\mathcal{I}_k(\boldsymbol{\tau})}(t), \tag{11}$$

where $\boldsymbol{\tau} \in \mathcal{T}$, $\boldsymbol{\sigma} \in \mathcal{S}$, and $\chi_{\mathcal{I}_k(\boldsymbol{\tau})} : \mathbb{R} \to \mathbb{R}$ is the characteristic function defined by

$$\chi_{\mathcal{I}_k(\boldsymbol{\tau})}(t) = \begin{cases} 1, & \text{if } t \in \mathcal{I}_k(\boldsymbol{\tau}), \\ 0, & \text{otherwise.} \end{cases}$$

It follows from Theorem 3.1 in [7] that

$$\bigvee_0^T \boldsymbol{u}^p(\cdot|\boldsymbol{\tau},\boldsymbol{\sigma}) \leq \sum_{i=1}^{r}\sum_{k=1}^{p-1} \left|\sigma_i^{k+1}-\sigma_i^k\right| < \infty, \tag{12}$$

where σ_i^k denotes the ith component of $\boldsymbol{\sigma}^k$. This implies that $\boldsymbol{u}^p(\cdot|\boldsymbol{\tau},\boldsymbol{\sigma})$ is an admissible control for Problem P.

Since $\epsilon > 0$, it follows from (8) that $\{\tau_k\}_{k=0}^p$ is a valid partition of $[0,T]$ satisfying

$$0=\tau_0<\tau_1<\cdots<\tau_{p-1}<\tau_p=T.$$

Thus, in view of equation (3),

$$\begin{aligned}\bigvee_0^T u_i^p(\cdot|\boldsymbol{\tau},\boldsymbol{\sigma}) &\geq \sum_{k=1}^{p}\left|u_i^p(\tau_k)-u_i^p(\tau_{k-1})\right| \\ &= \sum_{k=1}^{p-1}\left|\sigma_i^{k+1}-\sigma_i^k\right|,\end{aligned}$$

where $u_i^p(\cdot|\boldsymbol{\tau},\boldsymbol{\sigma})$ is the ith component of $\boldsymbol{u}^p(\cdot|\boldsymbol{\tau},\boldsymbol{\sigma})$. Hence,

$$\bigvee_0^T \boldsymbol{u}^p(\cdot|\boldsymbol{\tau},\boldsymbol{\sigma}) \geq \sum_{i=1}^{r}\sum_{k=1}^{p-1}\left|\sigma_i^{k+1}-\sigma_i^k\right|. \tag{13}$$

Combining (12) and (13) yields

$$\bigvee_0^T \boldsymbol{u}^p(\cdot|\boldsymbol{\tau},\boldsymbol{\sigma}) = \sum_{i=1}^{r}\sum_{k=1}^{p-1}\left|\sigma_i^{k+1}-\sigma_i^k\right|. \tag{14}$$

Substituting the piecewise-constant control (11) into the dynamic system (1)-(2) yields

$$\dot{\boldsymbol{x}}(t) = \sum_{k=1}^{p}\boldsymbol{f}(\boldsymbol{x}(t),\boldsymbol{\sigma}^k)\chi_{\mathcal{I}_k(\boldsymbol{\tau})}(t), \quad t\in[0,T], \tag{15}$$

$$\boldsymbol{x}(0)=\boldsymbol{x}^0. \tag{16}$$

Let $\boldsymbol{x}^p(\cdot|\boldsymbol{\tau},\boldsymbol{\sigma})$ denote the solution of (15)-(16) corresponding to $(\boldsymbol{\tau},\boldsymbol{\sigma})\in\mathcal{T}\times\mathcal{S}$.

The guaranteed-cost constraint (6) becomes

$$\begin{aligned}&\Phi(\boldsymbol{x}^p(T|\boldsymbol{\tau},\boldsymbol{\sigma})) \\ &\quad+\sum_{k=1}^{p}\int_{\tau_{k-1}}^{\tau_k}\mathcal{L}(\boldsymbol{x}^p(t|\boldsymbol{\tau},\boldsymbol{\sigma}),\boldsymbol{\sigma}^k)dt \leq \Delta^*.\end{aligned} \tag{17}$$

Let Γ denote the set of all pairs $(\boldsymbol{\tau},\boldsymbol{\sigma})\in\mathcal{T}\times\mathcal{S}$ such that (17) is satisfied. Then based on equation (14), Problem P can be approximated by the following finite-dimensional optimization problem.

Problem Q. *Choose a pair* $(\boldsymbol{\tau},\boldsymbol{\sigma})\in\Gamma$ *to minimize*

$$J^p(\boldsymbol{\tau},\boldsymbol{\sigma}) = \bigvee_0^T \boldsymbol{u}^p(\cdot|\boldsymbol{\tau},\boldsymbol{\sigma}) = \sum_{i=1}^{r}\sum_{k=1}^{p-1}\left|\sigma_i^{k+1}-\sigma_i^k\right|.$$

Let $(\boldsymbol{\tau}^*,\boldsymbol{\sigma}^*)\in\Gamma$ be an optimal solution of Problem Q. Then $\boldsymbol{u}^p(\cdot|\boldsymbol{\tau}^*,\boldsymbol{\sigma}^*)$ is a suboptimal control for Problem P. Various convergence results are available to show that the cost of the suboptimal control generated by control parameterization converges to the true optimal cost as the number of subintervals approaches infinity. See [7] for the latest work in this area.

4 Problem Transformation

There are two main challenges with solving Problem Q: (i) the cost function J^p is non-smooth; and (ii) the dynamic equations in (15) have discontinuities at the variable time points τ_k, $k=1,\ldots,p-1$. It is well-known that numerical optimization algorithms struggle to optimize variable time points [6]. Thus, in this section, we will develop a transformation procedure for converting Problem Q into a new problem that is easier to solve.

Let $\mathcal{O}$ denote the set of all $\boldsymbol{\theta}=[\theta_1,\ldots,\theta_p]^\top\in\mathbb{R}^p$ such that

$$\begin{aligned}&\theta_k\geq\epsilon, \quad k=1,\ldots,p,\\ &\theta_1+\cdots+\theta_p=T,\end{aligned}$$

where $\epsilon > 0$ is the minimum duration between control switches and $T>0$ is the terminal time. Define

$$\boldsymbol{\zeta}=(\boldsymbol{\gamma},\boldsymbol{v}^1,\ldots,\boldsymbol{v}^{p-1},\boldsymbol{w}^1,\ldots,\boldsymbol{w}^{p-1})\in\mathbb{R}^{(2p-1)r}, \tag{18}$$

where $\boldsymbol{\gamma}\in\mathbb{R}^r$, $\boldsymbol{v}^k\in\mathbb{R}^r$, $\boldsymbol{w}^k\in\mathbb{R}^r$, and the round bracket notation has the same meaning as in (9). Furthermore, define functions $\boldsymbol{\psi}^k:\mathbb{R}^{(2p-1)r}\to\mathbb{R}^r$, $k=1,\ldots,p$, as follows:

$$\boldsymbol{\psi}^k(\boldsymbol{\zeta}) = \boldsymbol{\gamma}+\sum_{l=k}^{p-1}(\boldsymbol{v}^l-\boldsymbol{w}^l), \quad k=1,\ldots,p,$$

where $\boldsymbol{\zeta}$ is as defined in (18). Let $\mathcal{Z}$ denote the set of all $\boldsymbol{\zeta}$ defined by (18) with

$$v_i^k\geq 0, \quad w_i^k\geq 0, \quad i=1,\ldots,r, \quad k=1,\ldots,p-1,$$

and

$$a_i\leq\psi_i^k(\boldsymbol{\zeta})\leq b_i, \quad i=1,\ldots,r, \quad k=1,\ldots,p,$$

where $\psi_i^k(\boldsymbol{\zeta})$ denotes the ith component of $\boldsymbol{\psi}^k(\boldsymbol{\zeta})$.

Consider the following dynamic system on the new time horizon $[0,p]$:

$$\dot{\boldsymbol{y}}(s) = \sum_{k=1}^{p}\theta_k\boldsymbol{f}(\boldsymbol{y}(s),\boldsymbol{\psi}^k(\boldsymbol{\zeta}))\chi_{[k-1,k)}(s), \tag{19}$$

$$\boldsymbol{y}(0)=\boldsymbol{x}^0, \tag{20}$$

where $(\boldsymbol{\theta},\boldsymbol{\zeta})\in\mathcal{O}\times\mathcal{Z}$ is a given pair.

Let $\boldsymbol{y}(\cdot|\boldsymbol{\theta},\boldsymbol{\zeta})$ denote the solution of (19)-(20) corresponding to $(\boldsymbol{\theta},\boldsymbol{\zeta})\in\mathcal{O}\times\mathcal{Z}$. Furthermore, let Ξ denote the set of all $(\boldsymbol{\theta},\boldsymbol{\zeta})\in\mathcal{O}\times\mathcal{Z}$ satisfying the following constraint:

$$\Phi(\boldsymbol{y}(p|\boldsymbol{\theta},\boldsymbol{\zeta}))+\sum_{k=1}^{p}\int_{k-1}^{k}\theta_k\mathcal{L}(\boldsymbol{y}(s|\boldsymbol{\theta},\boldsymbol{\zeta}),\boldsymbol{\psi}^k(\boldsymbol{\zeta}))ds\leq\Delta^*.$$

We now define a new optimization problem as follows.

Problem R. *Choose a pair* $(\boldsymbol{\theta},\boldsymbol{\zeta})\in\Xi$ *to minimize*

$$G(\boldsymbol{\zeta}) = \sum_{i=1}^{r}\sum_{k=1}^{p-1}(v_i^k+w_i^k).$$

Unlike in Problem Q, the cost function in Problem R is differentiable. Moreover, the discontinuities in the dynamics

(19) occur at the fixed integers $s = 1, \ldots, p-1$, not at variable time points. Thus, Problem R is much easier to solve than Problem Q. In fact, Problem R can be solved readily using the dynamic optimization techniques described in [6]. We now prove that any solution of Problem R can be used to construct a corresponding solution of Problem Q.

First, for each $\boldsymbol{\theta} \in \mathcal{O}$, define

$$\mu(s|\boldsymbol{\theta}) = \sum_{l=1}^{\lfloor s \rfloor} \theta_l + \theta_{\lfloor s \rfloor + 1}(s - \lfloor s \rfloor), \quad s \in [0, p],$$

where θ_{p+1} is arbitrary. Clearly,

$$\mu(k|\boldsymbol{\theta}) = \sum_{l=1}^{k} \theta_l, \quad k = 0, \ldots, p. \tag{21}$$

For each $\boldsymbol{\theta} \in \mathcal{O}$, define

$$\tilde{\boldsymbol{\tau}}(\boldsymbol{\theta}) = \big[\mu(1|\boldsymbol{\theta}), \ldots, \mu(p-1|\boldsymbol{\theta})\big]^\top \in \mathbb{R}^{p-1}.$$

From (21), we obtain

$$\mu(k|\boldsymbol{\theta}) - \mu(k-1|\boldsymbol{\theta}) = \theta_k \geq \epsilon, \quad k = 1, \ldots, p.$$

This shows that the components of $\tilde{\boldsymbol{\tau}}(\boldsymbol{\theta})$ satisfy (8). Hence, $\tilde{\boldsymbol{\tau}}(\boldsymbol{\theta}) \in \mathcal{T}$ for each $\boldsymbol{\theta} \in \mathcal{O}$.

Now, for each $\boldsymbol{\zeta} \in \mathcal{Z}$, let

$$\tilde{\boldsymbol{\sigma}}(\boldsymbol{\zeta}) = (\boldsymbol{\psi}^1(\boldsymbol{\zeta}), \ldots, \boldsymbol{\psi}^p(\boldsymbol{\zeta})) \in \mathbb{R}^{pr},$$

where the round bracket notation has the same meaning as in (9). We immediately see that $\tilde{\boldsymbol{\sigma}}(\boldsymbol{\zeta}) \in \mathcal{S}$. Thus, each pair in $\mathcal{O} \times \mathcal{Z}$ generates a corresponding pair in $\mathcal{T} \times \mathcal{S}$ through the relation $(\boldsymbol{\theta}, \boldsymbol{\zeta}) \mapsto (\tilde{\boldsymbol{\tau}}(\boldsymbol{\theta}), \tilde{\boldsymbol{\sigma}}(\boldsymbol{\zeta}))$. Solving the dynamic system (15)-(16) with $\boldsymbol{\tau} = \tilde{\boldsymbol{\tau}}(\boldsymbol{\theta})$ and $\boldsymbol{\sigma} = \tilde{\boldsymbol{\sigma}}(\boldsymbol{\zeta})$ yields the state trajectory $\boldsymbol{x}^p(\cdot|\tilde{\boldsymbol{\tau}}(\boldsymbol{\theta}), \tilde{\boldsymbol{\sigma}}(\boldsymbol{\zeta}))$. Our next result reveals the relationship between $\boldsymbol{x}^p(\cdot|\tilde{\boldsymbol{\tau}}(\boldsymbol{\theta}), \tilde{\boldsymbol{\sigma}}(\boldsymbol{\zeta}))$ and the solution of the new system (19)-(20).

Theorem 4.1. *For each $(\boldsymbol{\theta}, \boldsymbol{\zeta}) \in \mathcal{O} \times \mathcal{Z}$,*

$$\boldsymbol{y}(s|\boldsymbol{\theta}, \boldsymbol{\zeta}) = \boldsymbol{x}^p(t|\tilde{\boldsymbol{\tau}}(\boldsymbol{\theta}), \tilde{\boldsymbol{\sigma}}(\boldsymbol{\zeta}))\big|_{t=\mu(s|\boldsymbol{\theta})}, \quad s \in [0, p]. \tag{22}$$

Proof. Similar to the proof of Theorem 4.1 in [5]. □

Theorem 4.1 shows how $\mu(\cdot|\boldsymbol{\theta})$ links the dynamic system in Problem Q with the dynamic system in Problem R. The next result links the feasible regions of these two problems.

Theorem 4.2. *Let $(\boldsymbol{\theta}, \boldsymbol{\zeta}) \in \mathcal{O} \times \mathcal{Z}$ be a given pair. Then $(\boldsymbol{\theta}, \boldsymbol{\zeta}) \in \Xi$ if and only if $(\tilde{\boldsymbol{\tau}}(\boldsymbol{\theta}), \tilde{\boldsymbol{\sigma}}(\boldsymbol{\zeta})) \in \Gamma$.*

Proof. We write $\boldsymbol{x}^p(\cdot)$ instead of $\boldsymbol{x}^p(\cdot|\tilde{\boldsymbol{\tau}}(\boldsymbol{\theta}), \tilde{\boldsymbol{\sigma}}(\boldsymbol{\zeta}))$, and $\boldsymbol{y}(\cdot)$ instead of $\boldsymbol{y}(\cdot|\boldsymbol{\theta}, \boldsymbol{\zeta})$. Note that $\mu(p|\boldsymbol{\theta}) = \theta_1 + \cdots + \theta_p = T$. Thus, it follows from Theorem 4.1 that

$$\Phi(\boldsymbol{x}^p(T)) = \Phi(\boldsymbol{x}^p(t))\big|_{t=\mu(p|\boldsymbol{\theta})} = \Phi(\boldsymbol{y}(p)). \tag{23}$$

Moreover, using the substitution $t = \mu(s|\boldsymbol{\theta})$,

$$\begin{aligned}
&\int_{\mu(k-1|\boldsymbol{\theta})}^{\mu(k|\boldsymbol{\theta})} \mathcal{L}(\boldsymbol{x}^p(t), \boldsymbol{\psi}^k(\boldsymbol{\zeta}))dt \\
&\quad = \int_{k-1}^{k} \theta_k \mathcal{L}(\boldsymbol{x}^p(\mu(s|\boldsymbol{\theta})), \boldsymbol{\psi}^k(\boldsymbol{\zeta}))ds \\
&\quad = \int_{k-1}^{k} \theta_k \mathcal{L}(\boldsymbol{y}(s), \boldsymbol{\psi}^k(\boldsymbol{\zeta}))ds.
\end{aligned} \tag{24}$$

Combining (23) and (24) gives

$$\begin{aligned}
&\Phi(\boldsymbol{x}^p(T)) + \sum_{k=1}^{p} \int_{\mu(k-1|\boldsymbol{\theta})}^{\mu(k|\boldsymbol{\theta})} \mathcal{L}(\boldsymbol{x}^p(t), \boldsymbol{\psi}^k(\boldsymbol{\zeta}))dt \\
&\quad = \Phi(\boldsymbol{y}(p)) + \sum_{k=1}^{p} \int_{k-1}^{k} \theta_k \mathcal{L}(\boldsymbol{y}(s), \boldsymbol{\psi}^k(\boldsymbol{\zeta}))ds.
\end{aligned}$$

The result follows immediately from this equation. □

Our next result characterizes the solution of Problem R.

Theorem 4.3. *Let $(\boldsymbol{\theta}^*, \boldsymbol{\zeta}^*) \in \Xi$ be an optimal solution of Problem R, where*

$$\boldsymbol{\zeta}^* = (\boldsymbol{\gamma}^*, \boldsymbol{v}^{1,*}, \ldots, \boldsymbol{v}^{p-1,*}, \boldsymbol{w}^{1,*}, \ldots, \boldsymbol{w}^{p-1,*}). \tag{25}$$

Then

$$v_i^{k,*} w_i^{k,*} = 0, \quad i = 1, \ldots, r, \quad k = 1, \ldots, p-1. \tag{26}$$

Proof. Suppose that (26) is violated for some i and k. Let $\mathcal{J}_1$ denote the set of pairs (i, k) such that $v_i^{k,*} w_i^{k,*} > 0$, and let $\mathcal{J}_2$ denote the set of pairs (i, k) such that $v_i^{k,*} w_i^{k,*} = 0$. Since $v_i^{k,*}$ and $w_i^{k,*}$ are both non-negative,

$$\mathcal{J}_1 \cup \mathcal{J}_2 = \{1, \ldots, r\} \times \{1, \ldots, p-1\}.$$

Define

$$\hat{v}_i^k = \begin{cases} \max\{v_i^{k,*} - w_i^{k,*}, 0\}, & \text{if } (i,k) \in \mathcal{J}_1, \\ v_i^{k,*}, & \text{if } (i,k) \in \mathcal{J}_2, \end{cases}$$

$$\hat{w}_i^k = \begin{cases} \max\{w_i^{k,*} - v_i^{k,*}, 0\}, & \text{if } (i,k) \in \mathcal{J}_1, \\ w_i^{k,*}, & \text{if } (i,k) \in \mathcal{J}_2. \end{cases}$$

Furthermore, define

$$\hat{\boldsymbol{\zeta}} = (\boldsymbol{\gamma}^*, \hat{\boldsymbol{v}}^1, \ldots, \hat{\boldsymbol{v}}^{p-1}, \hat{\boldsymbol{w}}^1, \ldots, \hat{\boldsymbol{w}}^{p-1}),$$

where $\hat{\boldsymbol{v}}^k = [\hat{v}_1^k, \ldots, \hat{v}_r^k]^\top$ and $\hat{\boldsymbol{w}}^k = [\hat{w}_1^k, \ldots, \hat{w}_r^k]^\top$. We have

$$\hat{v}_i^k - \hat{w}_i^k = v_i^{k,*} - w_i^{k,*}, \ i = 1, \ldots, r, \ k = 1, \ldots, p-1.$$

Hence, for each $k = 1, \ldots, p$,

$$\begin{aligned}
\boldsymbol{\psi}^k(\hat{\boldsymbol{\zeta}}) &= \boldsymbol{\gamma}^* + \sum_{l=k}^{p-1} (\hat{\boldsymbol{v}}^l - \hat{\boldsymbol{w}}^l) \\
&= \boldsymbol{\gamma}^* + \sum_{l=k}^{p-1} (\boldsymbol{v}^{l,*} - \boldsymbol{w}^{l,*}) = \boldsymbol{\psi}^k(\boldsymbol{\zeta}^*).
\end{aligned}$$

It follows immediately that $\hat{\boldsymbol{\zeta}} \in \mathcal{Z}$. Furthermore,

$$\boldsymbol{y}(s|\boldsymbol{\theta}^*, \hat{\boldsymbol{\zeta}}) = \boldsymbol{y}(s|\boldsymbol{\theta}^*, \boldsymbol{\zeta}^*), \quad s \in [0, p].$$

Thus, since $(\boldsymbol{\theta}^*, \boldsymbol{\zeta}^*)$ is feasible for Problem R, $(\boldsymbol{\theta}^*, \hat{\boldsymbol{\zeta}})$ is also feasible for Problem R.

Now, recall that $v_i^{k,*}$ and $w_i^{k,*}$ are both non-negative. Hence, if $(i, k) \in \mathcal{J}_1$, then $v_i^{k,*} > 0$ and $w_i^{k,*} > 0$. This implies that

$$\begin{aligned}
\hat{v}_i^k + \hat{w}_i^k &= \max\{v_i^{k,*} - w_i^{k,*}, 0\} + \max\{w_i^{k,*} - v_i^{k,*}, 0\} \\
&= \big|v_i^{k,*} - w_i^{k,*}\big| < v_i^{k,*} + w_i^{k,*}.
\end{aligned}$$

Consequently, we have the following implication:

$$(i,k) \in \mathcal{J}_1 \quad \Longrightarrow \quad \hat{v}_i^k + \hat{w}_i^k < v_i^{k,*} + w_i^{k,*}.$$

Thus, by our assumption that $\mathcal{J}_1 \neq \emptyset$,

$$\begin{aligned} G(\hat{\boldsymbol{\zeta}}) &= \sum_{i=1}^{r}\sum_{k=1}^{p-1}(\hat{v}_i^k + \hat{w}_i^k) \\ &= \sum_{(i,k)\in\mathcal{J}_1}(\hat{v}_i^k + \hat{w}_i^k) + \sum_{(i,k)\in\mathcal{J}_2}(v_i^{k,*} + w_i^{k,*}) \\ &< \sum_{i=1}^{r}\sum_{k=1}^{p-1}(v_i^{k,*} + w_i^{k,*}) = G(\boldsymbol{\zeta}^*). \end{aligned}$$

But since $(\boldsymbol{\theta}^*, \hat{\boldsymbol{\zeta}}) \in \Xi$, this contradicts the optimality of $(\boldsymbol{\theta}^*, \boldsymbol{\zeta}^*)$. Thus, our assumption that $\mathcal{J}_1 \neq \emptyset$ is false. It follows that equation (26) must hold for all i and k. □

We now prove our main result: that a solution of Problem R can be used to generate a solution of Problem Q.

Theorem 4.4. *Let $(\boldsymbol{\theta}^*, \boldsymbol{\zeta}^*) \in \Xi$ be an optimal solution of Problem R, where $\boldsymbol{\zeta}^*$ is as defined in equation* (25)*. Then $(\tilde{\boldsymbol{\tau}}(\boldsymbol{\theta}^*), \tilde{\boldsymbol{\sigma}}(\boldsymbol{\zeta}^*))$ is an optimal solution of Problem Q.*

Proof. Theorem 4.3 implies that for each index pair (i,k), either $v_i^{k,*} = 0$ or $w_i^{k,*} = 0$. If $v_i^{k,*} = 0$, then since $w_i^{k,*}$ is non-negative,

$$\left|w_i^{k,*} - v_i^{k,*}\right| = \left|w_i^{k,*}\right| = w_i^{k,*} = v_i^{k,*} + w_i^{k,*}.$$

Similarly, if $w_i^{k,*} = 0$, then

$$\left|w_i^{k,*} - v_i^{k,*}\right| = \left|-v_i^{k,*}\right| = v_i^{k,*} = v_i^{k,*} + w_i^{k,*}.$$

Thus, for each $i = 1, \ldots, r$ and $k = 1, \ldots, p-1$,

$$\left|\psi_i^{k+1}(\boldsymbol{\zeta}^*) - \psi_i^k(\boldsymbol{\zeta}^*)\right| = \left|w_i^{k,*} - v_i^{k,*}\right| = v_i^{k,*} + w_i^{k,*}.$$

It follows that

$$\begin{aligned} J^p(\tilde{\boldsymbol{\tau}}(\boldsymbol{\theta}^*), \tilde{\boldsymbol{\sigma}}(\boldsymbol{\zeta}^*)) &= \sum_{i=1}^{r}\sum_{k=1}^{p-1}\left|\psi_i^{k+1}(\boldsymbol{\zeta}^*) - \psi_i^k(\boldsymbol{\zeta}^*)\right| \\ &= \sum_{i=1}^{r}\sum_{k=1}^{p-1}(v_i^{k,*} + w_i^{k,*}) = G(\boldsymbol{\zeta}^*). \end{aligned} \tag{27}$$

Now, let $(\bar{\boldsymbol{\tau}}, \bar{\boldsymbol{\sigma}}) \in \Gamma$ be an arbitrary feasible pair for Problem Q, where $\bar{\boldsymbol{\tau}} = [\bar{\tau}_1, \ldots, \bar{\tau}_{p-1}]^\top$ and

$$\bar{\boldsymbol{\sigma}} = (\bar{\boldsymbol{\sigma}}^1, \ldots, \bar{\boldsymbol{\sigma}}^p).$$

Define $\bar{\boldsymbol{\theta}} = [\bar{\theta}_1, \ldots, \bar{\theta}_p]^\top \in \mathbb{R}^p$ as follows:

$$\bar{\theta}_k = \bar{\tau}_k - \bar{\tau}_{k-1}, \quad k = 1, \ldots, p,$$

where $\bar{\tau}_0 = 0$ and $\bar{\tau}_p = T$. Then clearly, $\bar{\boldsymbol{\theta}} \in \mathcal{O}$ and $\tilde{\boldsymbol{\tau}}(\bar{\boldsymbol{\theta}}) = \bar{\boldsymbol{\tau}}$. For each k, define $\bar{\boldsymbol{v}}^k = [\bar{v}_1^k, \ldots, \bar{v}_r^k]^\top$ and $\bar{\boldsymbol{w}}^k = [\bar{w}_1^k, \ldots, \bar{w}_r^k]^\top$ by

$$\begin{aligned} \bar{v}_i^k &= \max\{\bar{\sigma}_i^k - \bar{\sigma}_i^{k+1}, 0\}, \quad i = 1, \ldots, r, \\ \bar{w}_i^k &= \max\{\bar{\sigma}_i^{k+1} - \bar{\sigma}_i^k, 0\}, \quad i = 1, \ldots, r. \end{aligned}$$

Furthermore, define

$$\bar{\boldsymbol{\zeta}} = (\bar{\boldsymbol{\sigma}}^p, \bar{\boldsymbol{v}}^1, \ldots, \bar{\boldsymbol{v}}^{p-1}, \bar{\boldsymbol{w}}^1, \ldots, \bar{\boldsymbol{w}}^{p-1}) \in \mathbb{R}^{(2p-1)r}.$$

For each $i = 1, \ldots, r$ and $k = 1, \ldots, p-1$,

$$\begin{aligned} \bar{v}_i^k - \bar{w}_i^k &= \max\{\bar{\sigma}_i^k - \bar{\sigma}_i^{k+1}, 0\} - \max\{\bar{\sigma}_i^{k+1} - \bar{\sigma}_i^k, 0\} \\ &= \bar{\sigma}_i^k - \bar{\sigma}_i^{k+1} \end{aligned} \tag{28}$$

and

$$\begin{aligned} \bar{v}_i^k + \bar{w}_i^k &= \max\{\bar{\sigma}_i^k - \bar{\sigma}_i^{k+1}, 0\} + \max\{\bar{\sigma}_i^{k+1} - \bar{\sigma}_i^k, 0\} \\ &= \left|\bar{\sigma}_i^k - \bar{\sigma}_i^{k+1}\right| = \left|\bar{\sigma}_i^{k+1} - \bar{\sigma}_i^k\right|. \end{aligned} \tag{29}$$

Using (28), we obtain, for each $k = 1, \ldots, p$,

$$\begin{aligned} \boldsymbol{\psi}^k(\bar{\boldsymbol{\zeta}}) &= \bar{\boldsymbol{\sigma}}^p + \sum_{l=k}^{p-1}(\bar{\boldsymbol{v}}^l - \bar{\boldsymbol{w}}^l) \\ &= \bar{\boldsymbol{\sigma}}^p + \sum_{l=k}^{p-1}(\bar{\boldsymbol{\sigma}}^l - \bar{\boldsymbol{\sigma}}^{l+1}) = \bar{\boldsymbol{\sigma}}^k. \end{aligned}$$

This shows that $\tilde{\boldsymbol{\sigma}}(\bar{\boldsymbol{\zeta}}) = \bar{\boldsymbol{\sigma}}$. Hence, $\bar{\boldsymbol{\zeta}} \in \mathcal{Z}$.

Since $(\bar{\boldsymbol{\tau}}, \bar{\boldsymbol{\sigma}}) = (\tilde{\boldsymbol{\tau}}(\bar{\boldsymbol{\theta}}), \tilde{\boldsymbol{\sigma}}(\bar{\boldsymbol{\zeta}}))$ is feasible for Problem Q, it follows from Theorem 4.2 that $(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\zeta}})$ is feasible for Problem R. Thus, (29) implies

$$\begin{aligned} G(\bar{\boldsymbol{\zeta}}) &= \sum_{i=1}^{r}\sum_{k=1}^{p-1}(\bar{v}_i^k + \bar{w}_i^k) \\ &= \sum_{i=1}^{r}\sum_{k=1}^{p-1}\left|\bar{\sigma}_i^{k+1} - \bar{\sigma}_i^k\right| = J^p(\bar{\boldsymbol{\tau}}, \bar{\boldsymbol{\sigma}}). \end{aligned} \tag{30}$$

By combining (27) and (30), and recalling that $(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\zeta}})$ is feasible for Problem R, we obtain

$$J^p(\tilde{\boldsymbol{\tau}}(\boldsymbol{\theta}^*), \tilde{\boldsymbol{\sigma}}(\boldsymbol{\zeta}^*)) = G(\boldsymbol{\zeta}^*) \leq G(\bar{\boldsymbol{\zeta}}) = J^p(\bar{\boldsymbol{\tau}}, \bar{\boldsymbol{\sigma}}).$$

Since $(\bar{\boldsymbol{\tau}}, \bar{\boldsymbol{\sigma}}) \in \Gamma$ was chosen arbitrarily, this shows that $(\tilde{\boldsymbol{\tau}}(\boldsymbol{\theta}^*), \tilde{\boldsymbol{\sigma}}(\boldsymbol{\zeta}^*))$ is optimal for Problem Q. □

Theorem 4.4 indicates that a solution of Problem Q can be obtained by solving Problem R, a smooth dynamic optimization problem. The resulting solution can then be used to generate a suboptimal control for Problem P, our original optimal control problem. Note that Problem R is a standard problem that can be solved using existing techniques [6].

5 Numerical Examples

For numerical testing, we wrote a Fortran program that solves Problem R by combining the optimization software FFSQP [8] with the dynamic optimization techniques discussed in reference [6]. This program was used to solve two example problems: one in fisheries and the other in container crane control. The results are reported below.

5.1 Optimal Fishery Harvesting

Consider the fishery harvesting problem in reference [4]. The state equations for this problem are given below:

$$\dot{x}(t) = a_0\{(1 - u(t))x(t) - x(t)^2\}, \quad t \in [0,1], \tag{31}$$

$$x(0) = x_0, \tag{32}$$

where $x(t)$ denotes the fish population at time t (as a fraction of the carrying capacity of the environment), $u(t)$ denotes the harvesting effort at time t, $x_0 > 0$ denotes the initial population level, and a_0 is a given constant.

The harvesting effort (the control function for this problem) is subject to the following bound constraint:

$$0 \le u(t) \le 1, \quad t \in [0,1]. \tag{33}$$

In addition, the following state constraint is imposed to prevent overfishing:

$$x(t) \ge x_{\min}, \quad t \in [0,1], \tag{34}$$

where $x_{\min} > 0$ is a given constant.

The total revenue obtained from harvesting is given by

$$R = \int_0^1 e^{-\varsigma_1 t}\big\{b_1(1+b_2(1-e^{-\varsigma_2 t}))u(t)x(t) - c_1 u(t) - c_2 u(t)^2\big\}dt,$$

where ς_1, ς_2, b_1, b_2, c_1, and c_2 are constants. For the fishing operation to be viable, a minimum amount of revenue must be obtained. Thus, we impose the following *guaranteed-revenue constraint*:

$$\int_0^1 e^{-\varsigma_1 t}\big\{b_1(1+b_2(1-e^{-\varsigma_2 t}))u(t)x(t) - c_1 u(t) - c_2 u(t)^2\big\}dt \ge R_{\min}, \tag{35}$$

where $R_{\min}$ is the minimum revenue threshold.

Our optimal control problem is defined as follows: *Choose the harvesting function* $u : [0,1] \to \mathbb{R}$ *to minimize*

$$J(u) = \bigvee_0^1 u$$

subject to the dynamic system (31)-(32) *and the constraints* (33)-(35).

Note that (34) is a continuous inequality constraint imposed at every point in the time horizon. Such constraints were not included in the original problem formulation in Section 2. Nevertheless, by using the *constraint transcription method* [6], the techniques described in Sections 3 and 4 can be readily extended to handle problems with continuous inequality constraints. Our Fortran program is based on this approach.

We choose the following values for the model constants:

$$a_0 = 0.5,\ x_0 = 0.45,\ x_{\min} = 0.4,\ \varsigma_1 = 1,\ \varsigma_2 = 5,$$
$$b_1 = 1.4,\ b_2 = 0.25,\ c_1 = 0.2,\ c_2 = 0.1.$$

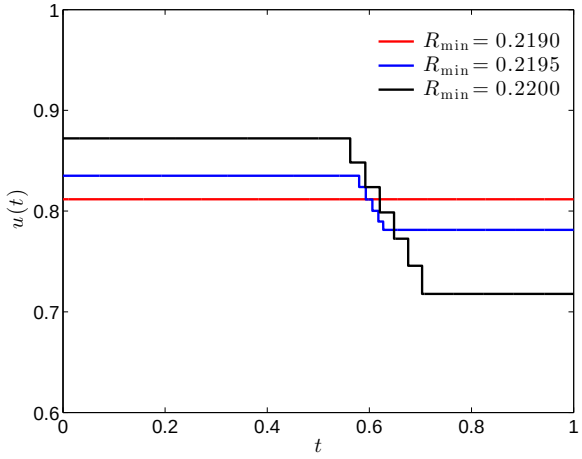
Using $p = 10$ for the number of subintervals and $\epsilon = 10^{-5}$ for the minimum subinterval duration, we solved the above optimal control problem for the following values of $R_{\min}$:

$$R_{\min} = 0.2190, \quad R_{\min} = 0.2195, \quad R_{\min} = 0.2200.$$

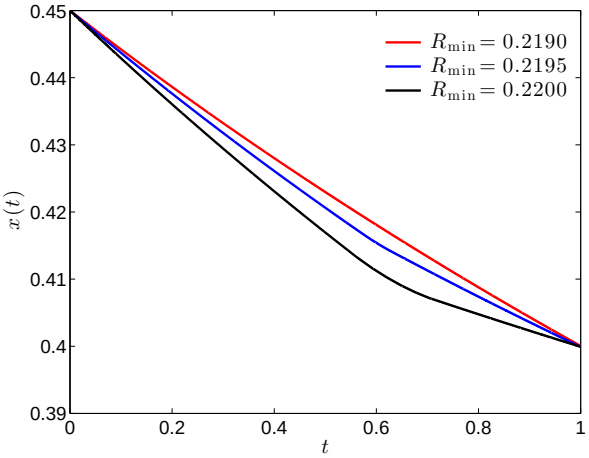
The optimal values for the total variation of u, in order of increasing $R_{\min}$, are

$$J^* = 0, \quad J^* = 0.05367, \quad J^* = 0.15442.$$

As expected, the guaranteed-revenue constraint (35) is active at each optimal solution. The optimal fishing policies and corresponding state trajectories are shown in Figure 1.



(a) Optimal fishing policy



(b) Fish population level

Fig. 1: Numerical results for Example 5.1.

5.2 Optimal Control of a Container Crane

The following dynamic equations describe the motion of a sea container being transported via crane in the Japanese port of Kobe [9, 10]:

$$\dot{x}_1(t) = x_4(t), \tag{36}$$
$$\dot{x}_2(t) = x_5(t), \tag{37}$$
$$\dot{x}_3(t) = x_6(t), \tag{38}$$
$$\dot{x}_4(t) = u_1(t) + \alpha_1 x_3(t), \tag{39}$$
$$\dot{x}_5(t) = u_2(t), \tag{40}$$
$$\dot{x}_6(t) = -\frac{u_1(t) + \alpha_2 x_3(t) + 2x_2(t)x_6(t)}{x_2(t)}, \tag{41}$$

where x_1 is the container's horizontal position, x_2 is the container's vertical position, x_3 is the container's swing angle, x_4 is the container's horizontal speed, x_5 is the container's vertical speed, and x_6 is the container's swing velocity. Furthermore, u_1 and u_2 are control functions for the crane, and $\alpha_1 = 17.2656$ and $\alpha_2 = 27.0756$ are model constants.

The initial conditions for (36)-(41) are

$$x_1(0) = 0, \quad x_2(0) = 22, \quad x_3(0) = 0, \tag{42}$$
$$x_4(0) = 0, \quad x_5(0) = -1, \quad x_6(0) = 0. \tag{43}$$

Moreover, the terminal conditions are

$$x_1(9) = 10, \quad x_2(9) = 14, \quad x_3(9) = 0, \tag{44}$$

where the terminal time here is $T = 9$. The control functions u_1 and u_2 are subject to the following bound constraints:

$$-2.83374 \leq u_1(t) \leq 2.83374, \quad t \in [0, 9], \tag{45}$$
$$-0.80865 \leq u_2(t) \leq 0.71265, \quad t \in [0, 9]. \tag{46}$$

There are also bound constraints on the container's horizontal and vertical speeds:

$$|x_4(t)| \leq 2.5, \quad |x_5(t)| \leq 1, \quad t \in [0, 9]. \tag{47}$$

In references [9, 10], the following Bolza cost functional is used to measure system performance:

$$g(u_1, u_2) = \tfrac{1}{2} \int_0^9 \{x_3(t)^2 + x_6(t)^2\} dt. \tag{48}$$

This cost functional, which penalizes large container swings, is motivated by safety considerations. Based on the Bolza cost functional (48), we impose the following guaranteed-cost constraint:

$$\tfrac{1}{2} \int_0^9 \{x_3(t)^2 + x_6(t)^2\} dt \leq \Delta^*, \tag{49}$$

where Δ^* is the maximum allowable system cost. The optimal control problem is defined as follows: *Choose control functions* $u_1 : [0, 9] \to \mathbb{R}$ *and* $u_2 : [0, 9] \to \mathbb{R}$ *to minimize*

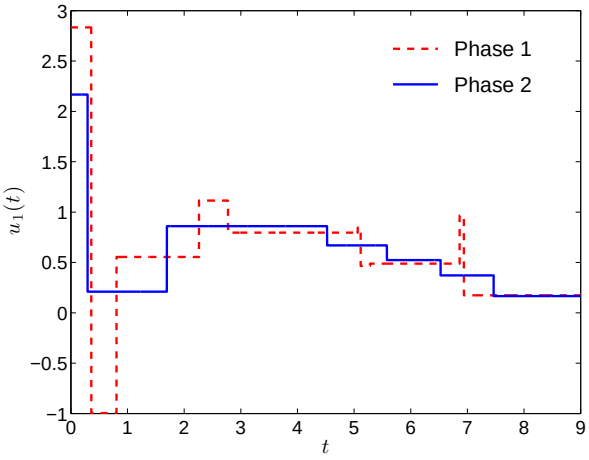
$$J(u_1, u_2) = \bigvee_0^9 u_1 + \bigvee_0^9 u_2.$$

subject to the dynamic system (36)-(43) *and the constraints* (44)-(47) *and* (49).

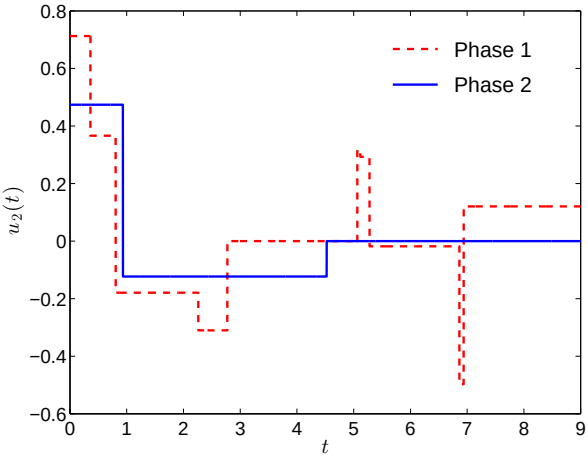
This is a challenging problem with nonlinear dynamics and multiple state constraints. Determining a solution of the corresponding Problem R proved difficult without a good initial guess for FFSQP. Thus, we applied a two-phase approach. In Phase 1, we solved the original version of the problem in which the terminal constraints (44) are appended to the Bolza cost (48) as a penalty. In Phase 2, starting with the optimal solution from Phase 1 as the initial guess, we solved the new version of the problem in which the total variation of the control is minimized.

We used $p = 10$ for the number of control subintervals and $\epsilon = 10^{-5}$ for the minimum subinterval duration. Furthermore, in Phase 2, we used $\Delta^* = 0.0032$ for the upper bound on the Bolza cost, a slightly higher value than the optimal cost in Phase 1. The idea is to sacrifice a small amount of cost in Phase 2 in exchange for a smoother control.

The optimal control functions are shown in Figure 2, and the corresponding optimal state trajectories are shown in Figure 3. Note that the bounded variation of the optimal control from Phase 1 is 11.06422. The bounded variation of the optimal control from Phase 2 is $J^* = 4.02439$.



(a) Control 1



(b) Control 2

Fig. 2: Optimal control functions for Example 5.2.

References

[1] J. M. Blatt, Optimal control with a cost of switching control, *Journal of the Australian Mathematical Society – Series B: Applied Mathematics*, 19(3): 316–332, 1976.

[2] J. Matula, On an extremum problem, *Journal of the Australian Mathematical Society – Series B: Applied Mathematics*, 28(3): 376–392, 1987.

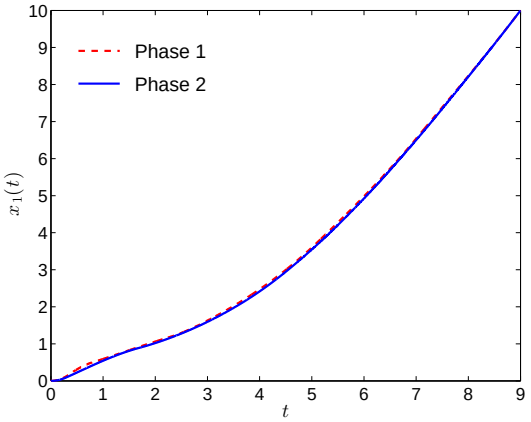
[3] D. E. Stewart, A numerical algorithm for optimal control problems with switching costs, *Journal of the Australian Mathematical Society – Series B: Applied Mathematics*, 34(2): 212–228, 1992.

[4] K. L. Teo and L. S. Jennings, Optimal control with a cost on changing control, *Journal of Optimization Theory and Applications*, 68(2): 335–357, 1991.

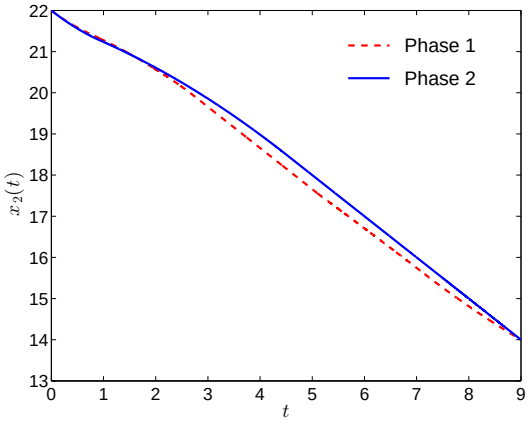
[5] R. Loxton, Q. Lin, and K. L. Teo, Minimizing control variation in nonlinear optimal control, *Automatica*, 49(9): 2652–2664, 2013.

[6] Q. Lin, R. Loxton, and K. L. Teo, The control parameterization method for nonlinear optimal control: A survey, *Journal of Industrial and Management Optimization*, 10(1): 275–309, 2014.

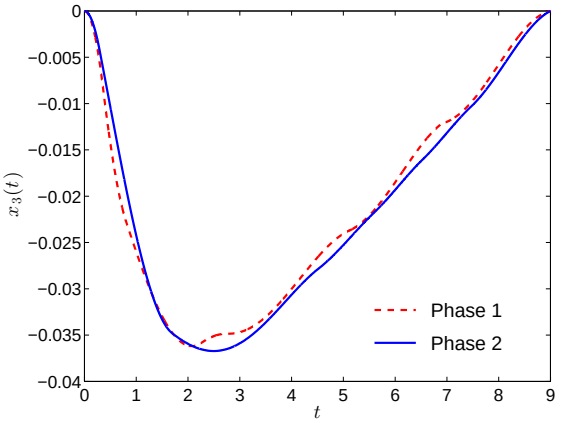
[7] R. Loxton, Q. Lin, V. Rehbock, and K. L. Teo, Control parameterization for optimal control problems with continuous inequality constraints: New convergence results, *Numerical*



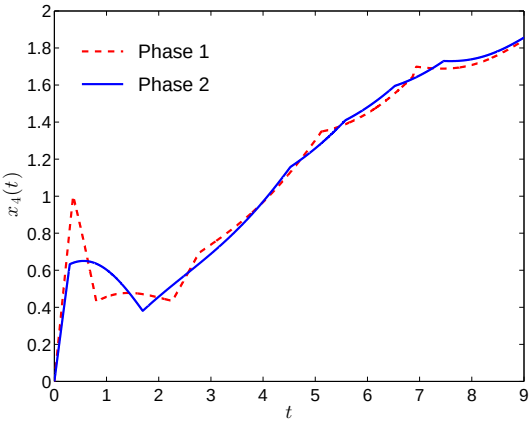
(a) State 1



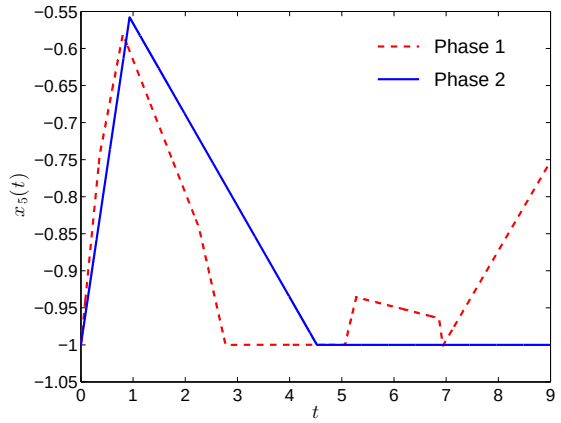
(b) State 2



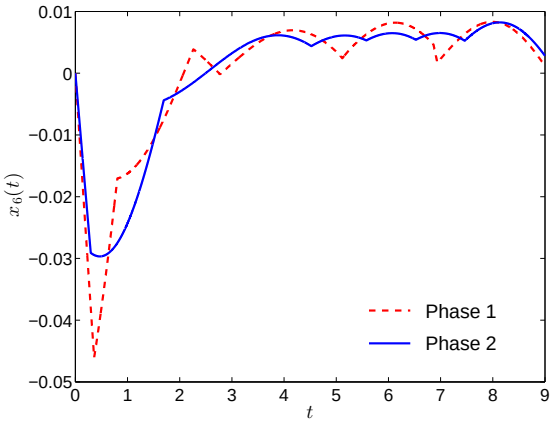
(c) State 3



(d) State 4



(e) State 5



(f) State 6

Fig. 3: Optimal state trajectories for Example 5.2.

Algebra, Control and Optimization, 2(3): 571–599, 2012.

[8] J. L. Zhou, A. L. Tits, and C. T. Lawrence, User's Guide for FFSQP Version 3.7: A FORTRAN Code for Solving Constrained Nonlinear (Minimax) Optimization Problems, Generating Iterates Satisfying All Inequality and Linear Constraints. College Park: University of Maryland, 1997.

[9] Y. Sakawa and Y. Shindo, Optimal control of container cranes, *Automatica*, 18(3): 257–266, 1982.

[10] K. L. Teo, C. J. Goh, and K. H. Wong, *A Unified Computational Approach to Optimal Control Problems*. Essex: Longman Scientific and Technical, 1991.