Electronic Research Archive

Research article

A new computational method for sparse optimal control of cyber-physical systems with varying delay

Sida Lin[1], Dongyao Yang[1], Jinlong Yuan[1,*], Changzhi Wu[2], Tao Zhou[3], An Li[4], Chuanye Gu[5], Jun Xie[6] and Kuikui Gao[7]

[1] School of Science, Dalian Maritime University, Dalian 116026, China

[2] Chongqing National Center for Applied Mathematics, Chongqing Normal University, Chongqing 404087, China

[3] Institute for Intelligent Systems Research and Innovation (IISRI), Deakin University, Geelong 3217, Australia

[4] School of Mathematical Sciences, Xiamen University, Xiamen 361005, China

[5] School of Management, Guangzhou University, Guangzhou 510006, China

[6] Department of Basics, PLA Dalian Naval Academy, Dalian 116018, China

[7] Ayata Incorporation, 2700 Post Oak Blvd, 21st Floor, Houston, TX 77056, USA

* **Correspondence:** Email: yuanjinlong@dlmu.edu.cn.

Abstract: In practice, network operators tend to choose sparse communication topologies to cut costs, and the concurrent use of a communication network by multiple users commonly results in feedback delays. Our goal was to obtain the optimal sparse feedback control matrix K. For this, we proposed a sparse optimal control (SOC) problem governed by the cyber-physical system with varying delay, to minimize $\|K\|_0$ subject to a maximum allowable compromise in system cost. A penalty method was utilized to transform the SOC problem into a form that was constrained solely by box constraints. A smoothing technique was used to approximate the nonsmooth element in the resulting problem, and an analysis of the errors introduced by this technique was subsequently conducted. The gradients of the objective function concerning the feedback control matrix were obtained by solving the state system and a variational system simultaneously forward in time. An optimization algorithm was devised to tackle the resulting problem, building on the piecewise quadratic approximation. Finally, we have presented of simulations.

Keywords: cyber-physical system; l_0-norm of a matrix; sparse feedback matrix; varying delay; smoothing technique

1. Introduction

A cyber-physical system (CPS) is a sophisticated, multi-layered system that combines computing, networking, and the physical environment [1]. By leveraging the integrated collaboration features of computation, communication, and control (3C) technology, it is possible to achieve real-time monitoring, control, and information services for large-scale engineering systems [2]. The applications of a CPS are wide-ranging. In study [3], the existing research on insider threat detection in a CPS is thoroughly reviewed and discussed. In [4], the problem of observer-based adaptive resilient control for a class of nonlinear CPSs is studied, taking the sensors that are vulnerable to deception attacks into account. The study in [5] focuses on the challenge of security control for CPSs subjected to aperiodic denial-of-service attacks. To reduce the need for explicit communication, the semantic knowledge within a CPS is leveraged, especially the use of physical radio resources to transmit potential informative data [6]. As discussed in [7], CPSs are vulnerable to numerous attacks and their attack surface continues to expand. To summarize, there are two key features of a CPS [8]: (i) large-scale, intricate systems focused on physical, biological, and engineering domains; (ii) a network core that includes communication networks and computing resources for monitoring, coordinating, and controlling these physical systems. A CPS closely integrates these two essential components, enabling analysis and design within a unified framework.

Over the past two decades, extensive research has been conducted on control theory related to the CPS. Traditional CPS control designs, however, often produce dense feedback matrices, with the optimal controller relying on all the information within the feedback matrix [9]. In extensive networks, implementation costs can be considerable, and the computational load required for communication between the controller and the dynamical system can be heavy [10]. Two idealistic assumptions are embedded in traditional CPS control designs: communication costs are infinite, and the communication network is solely reserved for control purposes [11]. In reality, however, network operators typically prefer sparse communication topologies to reduce costs, and the shared usage of the communication network by various users commonly leads to feedback delays [12]. To address this issue, we propose a CPS system with a static state feedback controller $u(t) = -Kx(t - \tau)$, where $K \in \mathbb{R}^{m \times n}$ and τ represents the delay, which can be either constant or variable, and is introduced by the communication process between the state x and the computation of the input u [2]. The network control design presented in this paper aims to achieve a balance between two primary objectives: (i) system performance, represented by the traditional cost function $J^0(K)$, and (ii) the sparsity of the communication network. Thus, in this paper, we seek to solve the following problem: for a given CPS system, identify a feedback matrix K that balances system performance with controller sparsity.

Sparsity refers to a situation where the majority of elements in a vector or matrix are zero. The sparsity of a vector or matrix is characterized by its l_0-norm. Sparsity is crucial in large-scale optimization problems, such as sound field reconstruction [13]. Employing sparsity not only minimizes storage requirements but also cuts transmission costs through vector compression. It streamlines a complex problem by extracting and using only the essential information from large datasets. In the absence of consideration for network topology, maximizing the sparsity of the feedback matrix typically has several reasons: (i) A sparse feedback matrix means that the controller only focuses on a small subset of variables in the system. This can significantly reduce the computational load and improve the responsiveness of real-time control, especially in large-scale

systems. (ii) Sparse control strategies can decrease the number of required sensors and actuators, thus reducing the overall system cost and energy consumption. This is particularly important in resource-constrained environments. (iii) Sparse matrices often lead to simpler and more understandable control decisions. This can help designers more easily analyze and comprehend control strategies, making debugging and optimization processes more effective. (iv) Sparsity may enhance the system's tolerance to failures of certain components. If some sensors or actuators fail, the system can still maintain functionality through other effective connections. (v) In some cases, sparse control strategies can exhibit better robustness to noise and disturbances, as they rely only on key parts of the system, reducing sensitivity to the overall system state.

Currently, sparse optimization has been extensively applied in various fields such as blade tip timing [14], robotic machining systems [15], and perimeter control [16, 17]. Sparse optimization models can be generally categorized into two types [18]: (i) l_0-regularization optimization problems modifying the traditional objective function by incorporating the l_0-norm into a new objective function, and (ii) sparse constrained optimization problems including the l_0-norm within the constraints. However, both types of problems are NP-hard. In earlier studies, methods for solving the l_0-norm minimization problem are typically categorized into model transformation techniques and direct processing techniques. The common feature of the model transformation method is to approximate the l_0-norm with the l_1-norm [19]. In terms of algorithms, several methods have been studied, including the iterative hard-thresholding algorithm (IHTA) [20], fast iterative shrinkage-thresholding algorithm (FISTA) [21], augmented method (ALM) [22], and alternating direction method of multipliers (ADMM) [23].

IHTA has two advantages as follows [20]: (i) It is straightforward to implement, making it accessible for various applications; and (ii) it effectively promotes sparsity in solutions, which is beneficial in many signal processing and statistical tasks. There exists two disadvantages for IHTA as follows [20]: (i) It can converge slowly, especially for large-scale problems; and (ii) it may struggle with nonsmooth functions, limiting its applicability in some optimization scenarios. FISTA offers two key benefits [21]: (i) It significantly accelerates the convergence compared to IHT by using Nesterov's acceleration, making it suitable for large datasets; and (ii) it can handle a variety of loss functions and regularization terms, providing versatility in applications. FISTA has two drawbacks, outlined below [21]: (i) The implementation is more complex than IHT, requiring careful tuning of parameters; and (ii) it may require more memory for storing additional variables, which could be a concern for very large problems. ALM has two notable benefits, listed below [22]: (i) It is effective for problems with constraints, making it a good choice for constrained optimization; and (ii) it generally exhibits robust convergence properties, especially for nonconvex problems. ALM presents two notable drawbacks, detailed below [22]: (i) The performance heavily depends on the choice of parameters, which can be challenging to optimize; and (ii) the method can be computationally intensive, particularly for high-dimensional problems. ADMM offers two key advantages, as outlined below [23]: (i) It allows large problems to be decomposed into smaller subproblems, which can be solved more easily; and (ii) it handles a wide range of objective functions and constraints, making it versatile for various applications. ADMM comes with two significant downsides, as outlined below [23]: (i) While it has good convergence properties, it can sometimes converge slowly compared to other methods; and (ii) like ALM, the performance can be sensitive to the choice of parameters, requiring careful tuning.

Optimal sparse control theory is now well developed. In [24], it explores the optimal control problem involving sparse controls for a Timoshenko beam, including its numerical approximation using the finite element method and its numerical solution through nonsmooth techniques. In [25], the study is focused on sparse optimal control for continuous-time stochastic systems using a dynamic programming approach, analyzing the optimal control through the value function. In [26], the study aims to develop a sparse tube-based robust economic model predictive control scheme. In [27], a novel sparse control strategy for acidic wastewater sulfidation is presented to ensure the continuous and safe HSS process. In [28], the construction of an eigenfunction vector of the Koopman operator is based on the sparse control strategy. However, for the CPS system with varying delay, these methods in [24–28] are not sufficient to determine an optimal sparse control policy. In this paper, we first demonstrate the existence of the partial derivatives of the system state with respect to the elements of the feedback matrix, and then use this to show that the gradient of the cost function can be computed by solving the state system and a variational system forward in time. In this case, our optimal control policy would be more straightforward and efficient for real-world applications compared to the methods in [24, 25] depending on the numerical approximation, which could introduce a gap between the real and the approximated control policy.

Many existing optimal sparse control theories for the CPS assume constant delays, which can oversimplify the dynamics of a CPS where delays are often variable and unpredictable. This can lead to suboptimal control strategies that do not account for the true nature of system behavior. Sparse control with feedback delay plays a vital role in the CPS. (i) Effective control strategy design is essential in the CPS. Sparse control can optimize input signals, particularly when there are limited sensors and actuators or when energy efficiency is a priority. It is also important to account for feedback delays in the control algorithms to prevent potential instability. (ii) The CPS generally requires immediate responses, but feedback delays necessitate that control strategies be equipped to mitigate their effects. By simplifying control signals, sparse control can improve responsiveness while ensuring that performance remains stable despite these delays. (iii) Additionally, sparse control can help minimize the need for computational power and communication bandwidth in the CPS environments. The optimal control of communication networks described previously only considers the system cost. However, network operators frequently opt for sparse communication topologies to lower costs. With this in mind, to determine the optimal feedback control matrix K, we propose a sparse optimal control (SOC) problem based on the CPS system with varying delay, aiming to minimize the sparsity of the controller subject to a maximum allowable compromise in system cost. A penalty method is employed to convert the SOC problem into a problem that is constrained only by box constraints. A smoothing technique is applied to approximate the nonsmooth component in the resulting problem. Subsequently, an analysis of the errors introduced by the smoothing technique is conducted. The gradients of the objective function with respect to the feedback control matrix are determined by solving both the state system and a variational system forward in time. Building upon the piecewise quadratic approximation [29], an optimization algorithm is developed to address the resulting problem. Finally, the paper provides the outcomes of the simulations.

The rest of the paper is organized as follows. We first describe the SOC problem based on the CPS system in Section 2. In Section 3, we develop an optimization algorithm to solve the SOC problem. Finally, a numerical example is given in Section 4.

2. Problem formulation

2.1. CPS modeling

Let I_n be the set of $\{1, 2, ..., n\}$.

2.1.1. Linear time-invariant system

We consider the following linear time-invariant (LTI) system [2]:

$$\dot{x}(t) = Ax(t) + Bu(t),$$

where $x \in \mathbb{R}^n$ is the state and $u \in \mathbb{R}^m$ is the control input. $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{n \times m}$. Assume that (A, B) is controllable.

2.1.2. Feedback control system with varying delay $\tau(\|K\|_0)$

Two idealistic hypotheses are involved in conventional CPS control designs: that communication costs are unlimited and that the communication network is exclusively allocated for control purposes. In practice, however, network operators often favor sparse communication topologies to minimize costs, and the shared use of the same communication network by multiple users frequently results in feedback delay.

After the sparsity level of matrix $K \in \mathbb{R}^{m \times n}$ is achieved, the bandwidth c is equally redistributed among the remaining links. Assume that the communication network follows frequency division multiplexing. Then, delay τ can be defined by [2]:

$$\tau(\|K\|_0) = \tau_t + \tau_p = \mathcal{Z}(\|K\|_0, c, \tau_p) := \kappa\left(\|K\|_0/c\right) + \tau_p, \tag{2.1}$$

where $\|K\|_0$ denotes the number of nonzero elements in K, and $\kappa : \mathbb{R} \to \mathbb{R}$ is a positive function. Eq (2.1) implies that τ will change as $\|K\|_0$ changes. This change is captured by the function $\mathcal{Z}(\cdot) : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$. The transmission of state x_l for the computation of input u_i is anticipated to encounter a delay denoted as τ_{il} (expressed in seconds), $i \in I_m, l \in I_n$. This delay comprises two distinct components: $\tau_{il} = \tau_{p_{il}} + \tau_{t_{il}}$, where $\tau_{p_{il}}$ denotes the propagation delay, and $\tau_{t_{il}}$ represents the transmission delay. The parameter $\tau_{p_{il}}$ is characterized as the quotient of the link length divided by the speed of light, assumed to possess a uniform value denoted as τ_p across all pairs i, l. Our assumption posits an equal allocation of bandwidth for the communication link connecting any l_{th} sensor to any i_{th} actuator. Consequently, this implies that $\tau_{t_{il}}$ maintains a uniform value across all i, l pairs [2]. Henceforth, we denote τ_{il} uniformly as τ across all pairs i, l. In practice, potential deviations of τ_{il} from the designated τ due to variations in traffic and uncertainties within the network are acknowledged.

This controller will be deployed in a distributed manner utilizing a communication network, as depicted in Figure 1. This figure presents the CPS represented by a closed-loop system architecture. The ith control input is written as $u_i(t) = -\sum_{l=1}^{n} K_{il} x_l(t - \tau(\|K\|_0)), i \in I_m$. Then accordingly the closed-loop system is written as:

$$\begin{aligned} \dot{x}(t) &= f(x(t), \tilde{x}(t), K) = Ax(t) - BKx(t - \tau(\|K\|_0)), \\ x(t) &= v, t \leq 0, \end{aligned} \tag{2.2}$$

where $v \in \mathbb{R}^n$ is a given vector, where each of its elements is assumed, without loss of generality, to be 0.5; and $\tilde{x}(t)$ denotes $x(t - \tau(\|K\|_0))$. Let $x(\cdot|K)$ be the solution of system (2.2) corresponding to the feedback matrix $K \in \mathbb{R}^{m \times n}$.
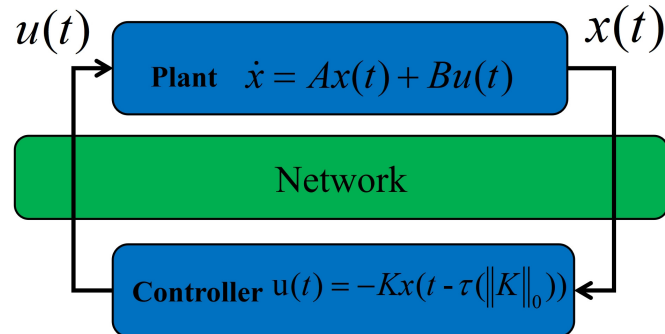

Figure 1. Closed-loop CPS representation [2]. The cyber network layer is to receive an input signal, denoted as $x(t)$, which is then transmitted through the network to generate an output signal, represented as $u(t)$. Clearly, this process will admit a varying delay, denoted as $\tau(\|K\|_0)$. After the computation of $u(t)$, the resultant signal is transmitted to the actuators for further action.

2.2. Problem statement

2.2.1. Traditional optimal control problem

With reference to the delay mentioned above, we introduce the corresponding system cost as given below [30]:

$$J^0(K) = (x(T|K))^\top S x(T|K) + \int_0^T [(x(t|K))^\top Q x(t|K) + (u(t))^\top W u(t)] \mathrm{d}t, \tag{2.3}$$

where T is the final time, the matrix $W \in \mathbb{R}^{m \times m}$ is symmetric positive definite, the feedback controller $u = -Kx(t - \tau(\|K\|_0))$, and the matrices $S \in \mathbb{R}^{n \times n}$ and $Q \in \mathbb{R}^{n \times n}$ are symmetric positive semidefinite.

We now present the feedback optimal control problem as follows.

$$\text{Problem } \mathbf{P_1}: \quad \min_{K \in \mathbb{R}^{m \times n}} \quad J^0(K)$$
$$s.t. \quad \dot{x}(t) = f(x(t), \tilde{x}(t), K),$$
$$x(t) = v, t \leq 0,$$

where $J^0(K)$ is given by (2.3).

2.2.2. Sparse optimal control problem

Gradient-based optimization methods [31] can be used to solve Problem $\mathbf{P_1}$. Let $K_1^* \in \mathbb{R}^{m \times n}$ be the optimal feedback matrices for Problem $\mathbf{P_1}$. However, these matrices tend to be rather dense, and

for large networks, the implementation cost will be expensive. Furthermore, the computation burden of the controller will be high because the state information is required to be transmitted through the communication network. Thus, we introduce the following Problem $\mathbf{P_2}$ given by

$$\text{Problem } \mathbf{P_2} : \quad \min_{K \in \mathbb{R}^{m \times n}} \quad \|K\|_0 \tag{2.4}$$
$$s.t. \quad \dot{x}(t) = f(x(t), \tilde{x}(t), K),$$
$$x(t) = v, t \leq 0,$$
$$|J^0(K) - J^0(K_1^*)| \leq \varepsilon,$$

where $\|K\|_0$ denotes the number of nonzero entries of the feedback matrix $K \in \mathbb{R}^{m \times n}$, $J^0(K)$ is given by (2.3), and $J^0(K_1^*)$ is the benchmark optimal system cost obtained through solving Problem $\mathbf{P_1}$. ε is a small number that is used to ensure that the system cost is not greatly affected during the sparsity process of the feedback matrix $K \in \mathbb{R}^{m \times n}$.

Obviously, Problem $\mathbf{P_2}$ balances system performance and the sparse level of the the feedback matrix $K \in \mathbb{R}^{m \times n}$.

3. Computational approaches

3.1. Preconditioning algorithm: the approximation of $\|K\|_0$

The feedback matrix K is decomposed as n column vectors, i.e., $K = (K^1, K^2, \ldots, K^n) \in \mathbb{R}^{m \times n}$. Note that $\|K^l\|_0, l \in I_n$, regularization is NP-hard. Thus, it is difficult to solve. In the past two decade, many approximation methods, such as $\|K^l\|_1$ and $\|K^l\|_q^q$ ($0 < q < 1$) have been proposed. In [29], the l_0-norm of the vector is approximated by a piecewise quadratic approximation (PQA) method. In this paper, we shall extend PQA to spare the feedback matrix K.

Remark 1. *We shall illustrate that $P(K^l), l \in I_n$, performs better than other common approximations of $\|K^l\|_0, l \in I_n$, on $[-e, e]$, $e = \{1, 1, \ldots, 1\} \in \mathbb{R}^m$.*

For $l \in I_n$, Figure 2 shows the approximation effects of $\|K^l\|_1$, $\|K^l\|_{1/2}^{1/2}$, $\|K^l\|_{1/3}^{1/3}$, $\|K^l\|_1 - \|K^l\|_2$, and $P(K^l)$ for the one-dimensional case in $[-1,1]$ [29]. Obviously, for $l \in I_n$, $P(K^l)$ is superior to $\|K^l\|_1$ for approximating the l_0-norm when $|K_i^l| \leq 1, i \in I_m$. For $l \in I_n$, when $0.38 \leq |K_i^l| \leq 1, i \in I_m$, $P(K^l)$ gives a better approximation for $\|K^l\|_0$, and when $0.61 \leq |K_i^l| \leq 1, i \in I_m$, $P(K^l)$ is better than $\|K^l\|_{1/3}^{1/3}$. Also, for $l \in I_n$, $P(K^l)$ is superior to $\|K^l\|_1 - \|K^l\|_2$, which is identically equal to 0 and has a large gap with $\|K^l\|_0$ in $[-1,1]$.

On this basis, we use the piecewise quadratic function [29] to approximate $\|K^l\|_0$ over $[-e, e]$.

$$P(K^l) = -(K^l)^\top K^l + 2\|K^l\|_1, K^l \in \mathbb{R}^m, l \in I_n.$$

If we choose $f(K^l) = -\|K^l\|_2^2$ and $g(K^l) = 2\|K^l\|_1$, then

$$F(K) = \sum_{l=1}^{n} P(K^l) = \sum_{l=1}^{n} \left[f(K^l) + g(K^l) \right],$$

where g is a proper closed convex and possibly nonsmooth function; f is a smooth nonconvex function of the type $C_{L_f}^{1,1}(\mathbb{R}^n)$, i.e., continuously differentiable with Lipschitz continuous gradient

$$\|\nabla f(K^l) - \nabla f(y^l)\| \leq L_f \|K^l - y^l\|, \quad K^l \in \mathbb{R}^m, y^l \in \mathbb{R}^m, l \in I_n,$$

with $L_f > 0$ denoting the Lipschitz constant of ∇f.

Based on the piecewise quadratic approximation [29], Problem $\mathbf{P_2}$ can be approximated as given below:

$$\text{Problem } \mathbf{P_3} : \quad \min_{K \in \mathbb{R}^{m \times n}} \quad F(K)$$
$$s.t. \quad \dot{x}(t) = f(x(t), \tilde{x}(t), K),$$
$$x(t) = \nu, t \leq 0,$$
$$|J^0(K) - J^0(K_1^*)| \leq \varepsilon,$$

where $J^0(K)$, $J^0(K_1^*)$, and ε are defined in Problem $\mathbf{P_2}$.
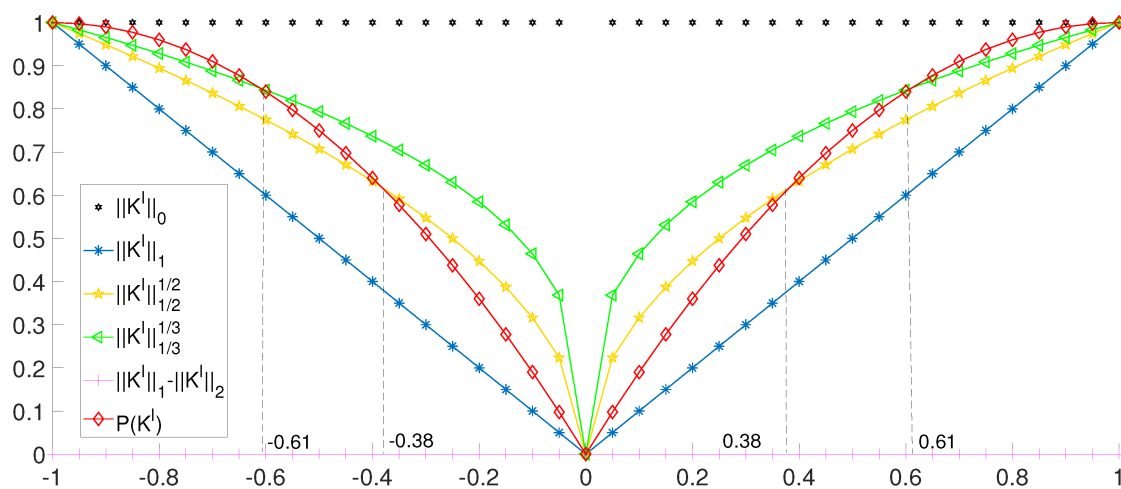

Figure 2. Various approximations for the one-dimensional case in $[-1, 1]$ [29].

3.2. Smoothing the objective function $F(K)$

Since the objective function $F(K)$ is nonsmooth, it is difficult to solve Problem $\mathbf{P_3}$ by using gradient-based algorithms. To overcome this difficulty, we aim to find a smooth function for the objective function $F(K)$ described in Problem $\mathbf{P_3}$.

First, we introduce the following notation: for $x, y, z \in \mathbb{R}^m$,

$$x = \max\{y, z\} \Leftrightarrow x_i = \max\{y_i, z_i\}, \forall i \in I_m.$$

Lemma 1. *If we define that $p(K_i^l) = \max\{K_i^l, 0\}$ and $q(K_i^l) = \max\{-K_i^l, 0\}$, then the following properties are satisfied:*

(1) $K_i^l = p(K_i^l) - q(K_i^l)$, $i \in I_m, l \in I_n$; and

(2) $\|K^l\|_1 = \sum_{i=1}^{m} [p(K_i^l) + q(K_i^l)]$, $l \in I_n$.

Proof. (1) For $l \in I_n$, and $i \in I_m$, we have

$$p(K_i^l) - q(K_i^l) = \max\{K_i^l, 0\} - \max\{-K_i^l, 0\},$$

$$K_i^l \geq 0 \Rightarrow p(K_i^l) - q(K_i^l) = K_i^l - 0 = K_i^l,$$

$$K_i^l < 0 \Rightarrow p(K_i^l) - q(K_i^l) = 0 - (-K_i^l) = K_i^l,$$

which imply that $K_i^l = p(K_i^l) - q(K_i^l), i \in I_m, l \in I_n$.

(2) For $l \in I_n$, and $i \in I_m$, we get

$$p(K_i^l) + q(K_i^l) = \max\{K_i^l, 0\} + \max\{-K_i^l, 0\},$$

$$K_i^l \geq 0 \Rightarrow p(K_i^l) + q(K_i^l) = K_i^l - 0 = |K_i^l|,$$

$$K_i^l < 0 \Rightarrow p(K_i^l) + q(K_i^l) = 0 - (-K_i^l) = |K_i^l|,$$

which prove that $\|K^l\|_1 = \sum_{i=1}^{m} |K_i^l| = \sum_{i=1}^{m} [p(K_i^l) + q(K_i^l)]$.

Based on Lemma 1, we obtain

$$F(K) = \sum_{l=1}^{n} [f(K^l) + g(K^l)] = \sum_{l=1}^{n} \left(-\|K^l\|_2^2 + 2\|K^l\|_1 \right) = \sum_{l=1}^{n} \left(-\|K^l\|_2^2 + 2\sum_{i=1}^{m} [p(K_i^l) + q(K_i^l)] \right)$$

$$= \sum_{l=1}^{n} \left(-\|K^l\|_2^2 + 2\sum_{i=1}^{m} [\max\{K_i^l, 0\} + \max\{-K_i^l, 0\}] \right).$$

Then Problem $\mathbf{P_3}$ is equivalent to the following problem:

$$\text{Problem } \mathbf{P_4}: \quad \min_{K \in \mathbb{R}^{m \times n}} \quad F_1(K)$$
$$s.t. \quad \dot{x}(t) = f(x(t), \tilde{x}(t), K),$$
$$x(t) = v, t \leq 0,$$
$$|J^0(K) - J^0(K_1^*)| \leq \varepsilon,$$

where

$$F_1(K) = \sum_{l=1}^{n} [f(K^l) + g(K^l)] = \sum_{l=1}^{n} \left(-\|K^l\|_2^2 + 2\|K^l\|_1 \right) = \sum_{l=1}^{n} \left(-\|K^l\|_2^2 + 2\sum_{i=1}^{m} [p(K_i^l) + q(K_i^l)] \right)$$

$$= \sum_{l=1}^{n} \left(-\|K^l\|_2^2 + 2\sum_{i=1}^{m} [\max\{K_i^l, 0\} + \max\{-K_i^l, 0\}] \right).$$

Clearly, the function $\max\{x, 0\}$ is nondifferentiable with respect to x at $x = 0$. Thus, in order to smooth it, a smooth function $\phi(x, \sigma)$ [32] is introduced as follows:

$$\phi(x, \sigma) = \frac{2\sigma^2}{\sqrt{x^2 + 4\sigma^2} - x},$$

where $\sigma > 0$ is an adjustable smooth parameter.

Lemma 2. *For any $x \in R$ and $\sigma > 0$, the smooth function $\phi(x, \sigma)$ has the following properties:*

(1) $\lim_{\sigma \to 0^+} \phi(x, \sigma) = \max\{x, 0\}$,

(2) $\phi(x, \sigma) > 0$,

(3) $0 < \phi'(x, \sigma) = \dfrac{1}{2}\left(\dfrac{x}{\sqrt{x^2 + 4\sigma^2}} + 1\right) < 1$,

(4) $0 < \phi(x, \sigma) - \max\{x, 0\} \leq \sigma$.

Lemma 2 shows that the function $\phi(x, \sigma)$ is an effective smooth approximation for the function $\max\{x, 0\}$, and the approximation level can be controlled artificially by adjusting the value of smooth parameter σ.

Therefore, Problem $\mathbf{P_4}$ can be approximated by the following problem:

$$\text{Problem } \mathbf{P_5}: \quad \min_{K \in \mathbb{R}^{m \times n}} \quad F_2(K)$$
$$s.t. \quad \dot{x}(t) = f(x(t), \tilde{x}(t), K),$$
$$x(t) = v, t \leq 0,$$
$$|J^0(K) - J^0(K_1^*)| \leq \varepsilon,$$

where

$$F_2(K) = \sum_{l=1}^{n}\left(-\|K^l\|_2^2 + 2\sum_{i=1}^{m}[\phi(K_i^l, \sigma) + \phi(-K_i^l, \sigma)]\right).$$

Note that the function $F_2(K)$ is continuously differentiable. Thus, Problem $\mathbf{P_5}$ is a constrained optimal parameter selection problem, which can be solved efficiently by using any gradient-based algorithm.

3.3. The relationship between Problem $\mathbf{P_4}$ and Problem $\mathbf{P_5}$

Theorem 1 presents that the optimal solution of Problem $\mathbf{P_5}$ is also the optimal solution of Problem $\mathbf{P_4}$ as long as $\sigma \to 0$ and an error estimation between the solutions of Problem $\mathbf{P_4}$ and Problem $\mathbf{P_5}$ is given.

Lemma 3. *For any $\sigma > 0$, one has*

$$0 < F_2(K) - F_1(K) \leq 4mn\sigma.$$

Proof. By using Lemma 2, one has

$$0 < \phi(x, \sigma) - \max\{x, 0\} \leq \sigma,$$

and

$$F_2(K) - F_1(K) = 2\sum_{l=1}^{n}\sum_{i=1}^{m}\left(\phi(K_i^l, \sigma) + \phi(-K_i^l, \sigma) - \max\{K_i^l, 0\} - \max\{-K_i^l, 0\}\right).$$

Thus,

$$0 < F_2(K) - F_1(K) \leq 4mn\sigma.$$

This completes the proof of Lemma 3.

Theorem 1. *Let K_4 and K_5 be the optimal solutions of Problem $\mathbf{P_4}$ and Problem $\mathbf{P_5}$, respectively. Then, we have*

$$0 < F_2(K_5) - F_1(K_4) \leq 4mn\sigma.$$

Proof. By using Lemma 3, one has

$$0 < F_2(K_4) - F_1(K_4) \leq 4mn\sigma,$$

$$0 < F_2(K_5) - F_1(K_5) \leq 4mn\sigma.$$

Note that K_4 is the optimal solution of Problem $\mathbf{P_4}$. Then, we have

$$F_1(K_5) \geq F_1(K_4),$$

which indicates that

$$F_2(K_5) - F_1(K_5) \leq F_2(K_5) - F_1(K_4).$$

Note that K_5 is the optimal solution of Problem $\mathbf{P_5}$. Then, we have

$$F_2(K_4) \geq F_2(K_5),$$

which indicates that

$$F_2(K_5) - F_1(K_4) \leq F_2(K_4) - F_1(K_4).$$

Then, we have $0 < F_2(K_5) - F_1(K_5) \leq F_2(K_5) - F_1(K_4) \leq F_2(K_4) - F_1(K_4) \leq 4mn\sigma$, which implies that

$$0 < F_2(K_5) - F_1(K_4) \leq 4mn\sigma.$$

This completes the proof of Theorem 1.

Remark 2. *Theorem 1 shows that the optimal solution of Problem $\mathbf{P_5}$ is an approximate optimal solution of Problem $\mathbf{P_4}$, as long as the adjustable parameter σ is sufficiently small.*

3.4. Penalty function method

The inequality constraint $|J^0(K) - J^0(K_1^*)| \leq \varepsilon$ is equivalent to

$$\max\{|J^0(K) - J^0(K_1^*)| - \varepsilon, 0\} = 0,$$

and equivalent to

$$\max\{J^0(K) - J^0(K_1^*) - \varepsilon, 0\} = 0, \tag{3.1}$$

and

$$\max\{-J^0(K) + J^0(K_1^*) - \varepsilon, 0\} = 0. \tag{3.2}$$

Then, by using the idea of the penalty function method described by [31], the equality constraints (3.1) and (3.2) are appended to the objective function of Problem $\mathbf{P_5}$ to form an augmented objective function. Thus, a penalty problem can be defined as follows:

$$\text{Problem } \mathbf{P_6} : \quad \min_{K \in \mathbb{R}^{m \times n}} \quad G^0(K)$$

$$s.t. \quad \dot{x}(t) = f(x(t), \tilde{x}(t), K),$$
$$x(t) = v, t \leq 0,$$

where

$$
\begin{aligned}
G^0(K) &= F_2(K) + \gamma_1 \max\{J^0(K) - J^0(K_1^*) - \varepsilon, 0\} + \gamma_2 \max\{-J^0(K) + J^0(K_1^*) - \varepsilon, 0\} \\
&= \sum_{l=1}^{n} \left(-\|K^l\|_2^2 + 2 \sum_{i=1}^{m} [\phi(K_i^l, \sigma) + \phi(-K_i^l, \sigma)] \right) \\
&\quad + \gamma_1 \max\{J^0(K) - J^0(K_1^*) - \varepsilon, 0\} + \gamma_2 \max\{-J^0(K) + J^0(K_1^*) - \varepsilon, 0\},
\end{aligned}
$$

with γ_1 and γ_2 being the penalty parameters. It is important to note that violations of the equality constraints in Eqs (3.1) and (3.2) are addressed through the integral term $G^0(K)$ in Problem $\mathbf{P_6}$. It can be demonstrated that by selecting sufficiently high values for γ_1 and γ_2, any minimizer of $G^0(K)$ in Problem $\mathbf{P_6}$ within the region defined by $\gamma_1 > \gamma^*$ and $\gamma_2 > \gamma^*$ (where γ^* denotes the threshold for the penalty parameters) will also satisfy the feasibility conditions of Problem $\mathbf{P_5}$. Therefore, a feasible solution to Problem $\mathbf{P_5}$ can be effectively found by minimizing $G^0(K)$ in Problem $\mathbf{P_6}$ with appropriately chosen penalty values for γ_1 and γ_2.

By adopting the function $\phi(x, \sigma)$ to approximate the function $\max\{x, 0\}$ again, Problem $\mathbf{P_6}$ can be written as the following problem:

$$
\begin{aligned}
\text{Problem } \mathbf{P_7}: \quad &\min_{K \in \mathbb{R}^{m \times n}} \quad G(K) \\
&s.t. \quad \dot{x}(t) = f(x(t), \tilde{x}(t), K), \\
&\qquad\quad x(t) = v, t \leq 0,
\end{aligned}
$$

where

$$
\begin{aligned}
G(K) &= \sum_{l=1}^{n} \left(-\|K^l\|_2^2 + \sum_{i=1}^{m} [\phi(K_i^l, \sigma) + \phi(-K_i^l, \sigma)] \right) \\
&\quad + \gamma_1 \phi(J^0(K) - J^0(K_1^*) - \varepsilon, \sigma) + \gamma_2 \phi(-J^0(K) + J^0(K_1^*) - \varepsilon, \sigma).
\end{aligned}
$$

Note that the function $G(K)$ is continuously differentiable. Thus, Problem $\mathbf{P_7}$ is an unconstrained optimal parameter selection problem, which can be solved efficiently by using any gradient-based algorithm.

3.5. The relationship of Problem $\mathbf{P_5}$, Problem $\mathbf{P_6}$, and Problem $\mathbf{P_7}$

Theorem 2 shows that the optimal solution of Problem $\mathbf{P_7}$ is also the optimal solution of Problem $\mathbf{P_6}$ as long as $\sigma \to 0$ and an error estimation between the solutions of Problem $\mathbf{P_5}$ and Problem $\mathbf{P_7}$ is given.

Definition 1. *A control input K_7 is σ-feasible to Problem $\mathbf{P_7}$, if the control input K_7 satisfies the following inequality constraint:*

$$|J^0(K_7) - J^0(K_1^*)| - \varepsilon \leq \sigma.$$

Lemma 4 and Theorem 2 are the variations of Lemma 3 and Theorem 1, respectively.

Lemma 4. *For any $\sigma > 0$, one has*

$$0 < G_0(K) - G(K) \leq (\gamma_1 + \gamma_2)\sigma,$$

where $0 < \gamma_1$ and $\gamma_2 < 1$ are the penalty parameters.

Proof. By using Lemma 2, we have

$$0 < \phi(x, \sigma) - \max\{x, 0\} \leq \sigma,$$

and

$$
\begin{aligned}
G_0(K) - G(K) \;=\; & \gamma_1\Big(\phi(J^0(K) - J^0(K_1^*) - \varepsilon, \sigma) - \max\{J^0(K) - J^0(K_1^*) - \varepsilon, 0\}\Big) \\
& + \gamma_2\Big(\phi(-J^0(K) + J^0(K_1^*) - \varepsilon, \sigma) - \max\{J^0(K) - J^0(K_1^*) - \varepsilon, 0\}\Big).
\end{aligned}
$$

Thus, we have

$$0 < G_0(K) - G(K) \leq (\gamma_1 + \gamma_2)\sigma.$$

This completes the proof of Lemma 4.

Theorem 2. *Let K_6 and K_7 be the optimal solutions of Problem $\mathbf{P_6}$ and Problem $\mathbf{P_7}$, respectively. Then, we have*

$$0 < G(K_7) - G_0(K_6) \leq (\gamma_1 + \gamma_2)\sigma.$$

Proof. By using Lemma 4, one has

$$0 < G_0(K_6) - G(K_6) \leq (\gamma_1 + \gamma_2)\sigma,$$

$$0 < G_0(K_7) - G(K_7) \leq (\gamma_1 + \gamma_2)\sigma.$$

Note that K_6 is the optimal solution of Problem $\mathbf{P_6}$. Then, we have

$$G_0(K_7) \geq G_0(K_6),$$

which indicates that

$$G(K_7) - G_0(K_7) \leq G(K_7) - G_0(K_6).$$

Note that K_7 is the optimal solution of Problem $\mathbf{P_7}$. Then, we have

$$G(K_6) \geq G(K_7),$$

which indicates that

$$G(K_7) - G_0(K_6) \leq G(K_6) - G_0(K_6).$$

Then, $0 < G(K_7) - G_0(K_7) \leq G(K_7) - G_0(K_6) \leq G(K_6) - G_0(K_6) \leq (\gamma_1 + \gamma_2)\sigma$, which implies that

$$0 < G(K_7) - G_0(K_6) \leq (\gamma_1 + \gamma_2)\sigma.$$

This completes the proof of Theorem 2.

Remark 3. *Theorem 2 shows that the optimal solution of Problem $\mathbf{P_7}$ is an approximate optimal solution of Problem $\mathbf{P_6}$, as long as the adjustable parameter σ is sufficiently small.*

Then, we can obtain the following theorem.

Theorem 3. *Let K_6 and K_7 be the optimal solutions of Problem $\mathbf{P_6}$ and Problem $\mathbf{P_7}$, respectively. If K_6 is feasible to Problem $\mathbf{P_5}$ and K_7 is σ−feasible to Problem $\mathbf{P_5}$, then we obtain*

$$-(\gamma_1 + \gamma_2)\sigma < F_2(K_7) - F_2(K_6) \leq (\gamma_1 + \gamma_2)\sigma.$$

Proof. Note that K_6 is feasible to Problem $\mathbf{P_5}$. Then, one has

$$\max\{J^0(K) - J^0(K_1^*) - \varepsilon, 0\} = 0,$$

and

$$\max\{-J^0(K) + J^0(K_1^*) - \varepsilon, 0\} = 0.$$

Then, we have $J^0(K) - J^0(K_1^*) - \varepsilon \leq 0$ and $-J^0(K) + J^0(K_1^*) - \varepsilon \leq 0$. Note that K_7 is σ-feasible to Problem $\mathbf{P_5}$ in Definition 1 and the smooth function $\phi(x, \sigma)$ is strictly monotone increasing (the first derivative is positive, see Lemma 2(3)). Thus, we obtain

$$\phi_{max}(J^0(K) - J^0(K_1^*) - \varepsilon, \sigma) = \phi(\sigma, \sigma) = \frac{1}{2}(\sqrt{5} + 1)\sigma,$$

and

$$\phi_{max}(-J^0(K) + J^0(K_1^*) - \varepsilon, \sigma) = \phi(\sigma, \sigma) = \frac{1}{2}(\sqrt{5} + 1)\sigma.$$

By using Lemma 2(2) that $\phi(x, \sigma) > 0$, we obtain

$$0 < \gamma_1\phi(J^0(K) - J^0(K_1^*) - \varepsilon, \sigma) + \gamma_2\phi(-J^0(K) + J^0(K_1^*) - \varepsilon, \sigma) \leq \frac{1}{2}(\sqrt{5} + 1)(\gamma_1 + \gamma_2)\sigma.$$

Based on Theorem 2, we have

$$0 < G(K_7) - G_0(K_6) = F_2(K_7) - F_2(K_6) + \gamma_1\Big(\phi(J^0(K) - J^0(K_1^*) - \varepsilon, \sigma) - \max\{J^0(K) - J^0(K_1^*) - \varepsilon, 0\}\Big)$$

$$+ \gamma_2\Big(\phi(-J^0(K) + J^0(K_1^*) - \varepsilon, \sigma) - \max\{-J^0(K) + J^0(K_1^*) - \varepsilon, 0\}\Big) \leq (\gamma_1 + \gamma_2)\sigma.$$

Thus, we get

$$-\frac{1}{2}(\sqrt{5} + 1)(\gamma_1 + \gamma_2)\sigma \leq F_2(K_7) - F_2(K_6) \leq (\gamma_1 + \gamma_2)\sigma.$$

This completes the proof of Theorem 3.

Remark 4. *If γ_1 and γ_2 are greater than the threshold value γ^*, the optimal solution of Problem $\mathbf{P_6}$ is the exact optimal solution of Problem $\mathbf{P_7}$ [31]. Furthermore, Theorem 3 provides an error estimation between the solutions of Problem $\mathbf{P_5}$ and Problem $\mathbf{P_7}$ as long as $\gamma_1 > \gamma^*$ and $\gamma_2 > \gamma^*$. Thus, the approximate optimal solution of Problem $\mathbf{P_5}$ can be achieved by solving Problem $\mathbf{P_7}$. Note that we have proved that the optimal solution of Problem $\mathbf{P_5}$ is also the optimal solution of Problem $\mathbf{P_4}$ as long as the adjustable smooth parameter σ is sufficiently small, and Problem $\mathbf{P_4}$ and Problem $\mathbf{P_3}$ are equivalent. Thus, the approximate optimal solution of Problem $\mathbf{P_3}$ can be achieved by solving Problem $\mathbf{P_7}$.*

3.6. Gradient formulae

Note that the cost function $G(K)$ in Problem $\mathbf{P_7}$ is continuously differentiable. Thus, Problem $\mathbf{P_7}$ is an unconstrained optimal parameter selection problem, which can be solved efficiently by using any gradient-based algorithm. Since the functionals depend implicitly on the feedback matrix K via system (2.2), it is important to derive an effective computational procedure for calculating the gradient of the cost function $G(K)$ in Problem $\mathbf{P_7}$.

3.6.1. The gradients of $x(\cdot|K)$ with respect to the feedback matrix

In this subsection, we investigate the gradients of $x(\cdot|K)$ with respect to the feedback matrix. Define

$$\Theta := [-1 - K_i^l, 1 - K_i^l].$$

Then, $0 \in \Theta$ and $\epsilon \in \Theta \Leftrightarrow K + \epsilon E_i^l \in \Theta$. For each $\epsilon \in \Theta$, define

$$\varphi^\epsilon(t) := x^\epsilon(t) - x(t), t \le T,$$

and

$$\theta^\epsilon := x^\epsilon(t - \tau(F_2(K))) - x(t - \tau(F_2(K))), t \le T.$$

Clearly,

$$\varphi^\epsilon(t - \tau(F_2(K))) = \theta^\epsilon(t), t \le T.$$

Then we have the following lemmas.

Lemma 5. *There exists a positive real number $L_1 > 0$ such that for all $\epsilon \in \Theta$, we have*

$$|x^\epsilon(t)| \le L_1, t \in [-\tau, T].$$

Lemma 6. *There exists a positive real number $L_2 > 0$ such that for all $\epsilon \in \Theta$, we have*

$$|\varphi^\epsilon(t)| \le L_2|\epsilon|, \left| \theta^\epsilon + \frac{\partial \tilde{x}}{\partial \tau} \frac{\partial \tau(F_2(K))}{\partial F_2} \frac{\partial F_2(K)}{\partial K_i^l} \epsilon \right| \le L_2|\epsilon|, i \in I_m, l \in I_n, t \in [0, T].$$

The partial derivatives of the system state with respect to the feedback matrix are given in Theorem 4.

Theorem 4. *Let $t \in (0, T]$ be a fixed time point. Then $x(t|\cdot)$ is differentiable with respect to K_i^l on $[-1,1]$. Moreover, for each K_i^l, we have*

$$\frac{\partial x(t|K)}{\partial K_i^l} = \Lambda_i^l(t|K), t \in [0, T], i \in I_m, l \in I_n,$$

where $\Lambda_i^l(\cdot|K), i \in I_m, l \in I_n$, is the solution of the following auxiliary system:

$$\dot{\Lambda}_i^l(t) = \frac{\partial f(x(t), \tilde{x}(t), K)}{\partial x} \Lambda_i^l(t) + \frac{\partial f(x(t), \tilde{x}(t), K)}{\partial \tilde{x}} \left[\Lambda_i^l(t - \tau(F_2(K))) + \frac{\partial \tilde{x}}{\partial \tau} \frac{\partial \tau(F_2(K))}{\partial F_2} \frac{\partial F_2(K)}{\partial K_i^l} \right] + \frac{\partial f(x(t), \tilde{x}(t), K)}{\partial K_i^l},$$

$$\Lambda_i^l(t) = 0, t \le 0.$$

Proof. Let $K_i^l \in [-1, 1], i \in I_m, l \in I_n$, be arbitrary but fixed. For each $\epsilon \in \Theta$, define the following functions:

$$
\bar{f}^\epsilon(s, \eta) = f(x(s) + \eta\varphi^\epsilon, \tilde{x}(s) + \eta\theta^\epsilon(s), K_i^l + \eta\epsilon E_i^l), \eta \in [0, 1],
$$

$$
\Delta_1^\epsilon = \int_0^1 \left\{ \frac{\partial \tilde{f}^\epsilon(s, \eta)}{\partial x} - \frac{\partial \tilde{f}^\epsilon(s, 0)}{\partial x} \right\} \varphi^\epsilon(s) \mathrm{d}\eta, s \in [0, t],
$$

$$
\Delta_2^\epsilon = \int_0^1 \left\{ \frac{\partial \tilde{f}^\epsilon(s, \eta)}{\partial \tilde{x}} - \frac{\partial \tilde{f}^\epsilon(s, 0)}{\partial \tilde{x}} \right\} \left[\theta^\epsilon(s) + \frac{\partial \tilde{x}(s)}{\partial \tau} \frac{\partial \tau(F_2(K))}{\partial F_2} \frac{\partial F_2(K)}{\partial K_i^l} \epsilon \right] \mathrm{d}\eta, s \in [0, t],
$$

$$
\Delta_3^\epsilon = \int_0^1 \epsilon \left\{ \frac{\partial \tilde{f}^\epsilon(s, \eta)}{\partial K_i^l} - \frac{\partial \tilde{f}^\epsilon(s, 0)}{\partial K_i^l} \right\} \mathrm{d}\eta, s \in [0, t],
$$

where E_i^l denotes the matrix in which the elements in row i and column l are 1 and the rest are 0.

Based on Lemmas 5 and 6 and the continuous differentiability of the functions f, we can obtain that there exist two constants $M_1 > 0$ and $M_2 > 0$ such that

$$
\|\frac{\partial \bar{f}^\epsilon(s, 0)}{\partial x}\| \le M_1, \|\frac{\partial \bar{f}^\epsilon(s, 0)}{\partial \tilde{x}}\| \le M_2,
$$

where $\| \cdot \|$ denotes the natural matrix norm on $\mathbb{R}^{n \times n}$. In addition, by Lemmas 5 and 6, the following limits exist uniformly with respect to $\eta \in [0, 1]$:

$$
\lim_{\epsilon \to 0} \{x(s) + \eta\varphi^\epsilon(s)\} = x(s),
$$

$$
\lim_{\epsilon \to 0} \{\tilde{x}(s) + \eta\theta^\epsilon(s)\} = \tilde{x}(s).
$$

Thus, for each $\delta > 0$, there exists an $\epsilon^1 > 0$ such that for all ϵ satisfying $|\epsilon| < \epsilon^1$,

$$
\left\| \frac{\partial \bar{f}^\epsilon(s, \eta)}{\partial x} - \frac{\partial \bar{f}^\epsilon(s, 0)}{\partial x} \right\| < \delta, \eta \in [0, 1],
$$

$$
\left\| \frac{\partial \bar{f}^\epsilon(s, \eta)}{\partial \tilde{x}} - \frac{\partial \bar{f}^\epsilon(s, 0)}{\partial \tilde{x}} \right\| < \delta, \eta \in [0, 1],
$$

and

$$
\left\| \frac{\partial \bar{f}^\epsilon(s, \eta)}{\partial K_i^l} - \frac{\partial \bar{f}^\epsilon(s, 0)}{\partial K_i^l} \right\| < \delta, \eta \in [0, 1].
$$

Thus, it follows from Lemma 6 that

$$
|\Delta_1^\epsilon(s)| \le L_2 \delta |\epsilon|, |\Delta_2^\epsilon(s)| \le L_2 \delta |\epsilon|, |\Delta_3^\epsilon(s)| \le \delta |\epsilon|. \tag{3.3}
$$

Now, let $\delta > 0$ be arbitrary but fixed and choose $\epsilon \in \Theta$ such that $0 < |\epsilon| < \epsilon^1$. Then, by the chain rule, we have

$$
\frac{\partial \bar{f}^\epsilon(s, \eta)}{\partial \eta} = \frac{\partial \bar{f}^\epsilon(s, \eta)}{\partial x} \cdot \frac{\partial x}{\partial \eta} + \frac{\partial \bar{f}^\epsilon(s, \eta)}{\partial \tilde{x}} \left[\frac{\partial \tilde{x}}{\partial \eta} + \frac{\partial \tilde{x}}{\partial \tau} \frac{\partial \tau(F_2(K))}{\partial F_2} \frac{\partial F_2(K)}{\partial K_i^l} \frac{\partial K_i^l}{\partial \eta} \right] + \frac{\partial \bar{f}^\epsilon(s, \eta)}{\partial K_i^l} \cdot \frac{\partial K_i^l}{\partial \eta}
$$

$$
= \frac{\partial \bar{f}^\epsilon(s, \eta)}{\partial x} \cdot \varphi^\epsilon(s) + \frac{\partial \bar{f}^\epsilon(s, \eta)}{\partial \tilde{x}} \cdot \left[\theta^\epsilon(s) + \frac{\partial \tilde{x}}{\partial \tau} \frac{\partial \tau(F_2(K))}{\partial F_2} \frac{\partial F_2(K)}{\partial K_i^l} \epsilon \right] + \frac{\partial \bar{f}^\epsilon(s, \eta)}{\partial K_i^l} \cdot \epsilon.
$$

Recall that $t \in (0, T]$ is a fixed time point. Then, by the fundamental theorem of calculus, we get

$$\varphi^\epsilon(t) = x^\epsilon(t) - x(t) = \int_0^t \bar{f}^\epsilon(s, 1) - \bar{f}^\epsilon(s, 0)\mathrm{d}s = \int_0^t \left(\int_0^1 \frac{\partial \bar{f}^\epsilon(s, \eta)}{\partial \eta} \mathrm{d}\eta \right) \mathrm{d}s.$$

Thus, the chain rule follows:

$$\varphi^\epsilon(t) = \int_0^t \left(\int_0^1 \frac{\partial \bar{f}^\epsilon(s, \eta)}{\partial x} \cdot \varphi^\epsilon(s)\mathrm{d}\eta \right) \mathrm{d}s + \int_0^t \left\{ \int_0^1 \frac{\partial \bar{f}^\epsilon(s, \eta)}{\partial \tilde{x}} \left[\theta^\epsilon(s) + \frac{\partial \tilde{x}}{\partial \tau} \frac{\partial \tau(F_2(K))}{\partial F_2} \frac{\partial F_2(K)}{\partial K_i^l} \epsilon \right] \mathrm{d}\eta \right\} \mathrm{d}s$$

$$+ \int_0^t \left(\int_0^1 \frac{\partial \bar{f}^\epsilon(s, \eta)}{\partial K_i^l} \cdot \epsilon \mathrm{d}\eta \right) \mathrm{d}s. \tag{3.4}$$

Note that we have

$$\int_0^1 \frac{\partial \bar{f}^\epsilon(s, \eta)}{\partial x} \varphi^\epsilon(s)\mathrm{d}\eta = \Delta_1^\epsilon(s) + \frac{\partial \bar{f}^\epsilon(s, 0)}{\partial x} \varphi^\epsilon(s), \tag{3.5}$$

$$\int_0^1 \frac{\partial \bar{f}^\epsilon(s, \eta)}{\partial \tilde{x}} \theta^\epsilon(s)\mathrm{d}\eta = \Delta_2^\epsilon(s) + \frac{\partial \bar{f}^\epsilon(s, 0)}{\partial \tilde{x}} \left[\theta^\epsilon(s) + \frac{\partial \tilde{x}}{\partial \tau} \frac{\partial \tau(F_2(K))}{\partial F_2} \frac{\partial F_2(K)}{\partial K_i^l} \epsilon \right], \tag{3.6}$$

$$\int_0^1 \epsilon \frac{\partial \bar{f}^\epsilon(s, \eta)}{\partial K_i^l} \mathrm{d}\eta = \Delta_3^\epsilon(s) + \epsilon \frac{\partial \bar{f}^\epsilon(s, 0)}{\partial K_i^l}. \tag{3.7}$$

Substituting (3.5)–(3.7) into (3.4) gives

$$\varphi^\epsilon(t) = \int_0^t \left(\Delta_1^\epsilon(s) + \frac{\partial \bar{f}^\epsilon(s, 0)}{\partial x} \varphi^\epsilon(s) \right) \mathrm{d}s + \int_0^t \left\{ \Delta_2^\epsilon(s) + \frac{\partial \bar{f}^\epsilon(s, 0)}{\partial \tilde{x}} \left[\theta^\epsilon(s) + \frac{\partial \tilde{x}}{\partial \tau} \frac{\partial \tau(F_2(K))}{\partial F_2} \frac{\partial F_2(K)}{\partial K_i^l} \epsilon \right] \right\} \mathrm{d}s$$

$$+ \int_0^t \left(\Delta_3^\epsilon(s) + \epsilon \frac{\partial \bar{f}^\epsilon(s, 0)}{\partial K_i^l} \right) \mathrm{d}s. \tag{3.8}$$

Note that

$$\Lambda_i^l(t|K) = \int_0^t \left\{ \frac{\partial f(x(s), \tilde{x}(s), K)}{\partial x} \Lambda_i^l(s) + \frac{\partial f(x(s), \tilde{x}(s), K)}{\partial \tilde{x}} \left[\Lambda_i^l(s - \tau(F_2(K))) + \frac{\partial \tilde{x}}{\partial \tau} \frac{\partial \tau(F_2(K))}{\partial F_2} \frac{\partial F_2(K)}{\partial K_i^l} \right] \right.$$

$$\left. + \frac{\partial f(x(s), \tilde{x}(s), K)}{\partial K_i^l} \right\} \mathrm{d}s. \tag{3.9}$$

Then multiplying (3.8) by ϵ^{-1}, subtracting (3.9), taking the norm of both sides, and finally applying (3.3) yields

$$|\epsilon^{-1}\varphi^\epsilon(t) - \Lambda_i^l(t|K)| \le (L_2 + L_2 + 1)\delta T + M_1 \int_0^t |\epsilon^{-1}\varphi^\epsilon(s) - \Lambda_i^l(s)|\mathrm{d}s$$

$$+ M_2 \int_0^t |\epsilon^{-1}\varphi^\epsilon(s - \tau(F_2(K))) - \Lambda_i^l(s - \tau(F_2(K)))|\mathrm{d}s. \tag{3.10}$$

Also, we know that $\varphi^\epsilon(t) = 0, t \le 0$, $\Lambda_i^l(s) = 0, s \le 0$, and then the last integral term on the right-hand side of (3.10) can be simplified as follows:

$$\int_{-\tau}^{t-\tau} M_2 |\epsilon^{-1}\varphi^\epsilon(s) - \Lambda_i^l(s|K)|\mathrm{d}s \le \int_0^t M_2 |\epsilon^{-1}\varphi^\epsilon(s) - \Lambda_i^l(s|K)|\mathrm{d}s.$$

Thus, (3.10) becomes

$$|\epsilon^{-1}\varphi^{\epsilon}(s) - \Lambda_i^l(s|K)| \le (2L_2 + 1)\delta T + (M_1 + M_2) \int_0^t |\epsilon^{-1}\varphi^{\epsilon}(s) - \Lambda_i^l(s|K)|\mathrm{d}s.$$

By the Gronwall-Bellman lemma, it follows that

$$|\epsilon^{-1}\varphi^{\epsilon}(s) - \Lambda_i^l(s|K)| \le (2L_2 + 1)\delta T \exp[(M_1 + M_2)T],$$

which holds whenever $0 < |\epsilon| < \epsilon^1$. Since δ is arbitrarily chosen, we conclude that $\epsilon^{-1}\varphi^{\epsilon}(t) \to \Lambda_i^l(t|K)$ as $\epsilon \to 0$.

Then, we get

$$\lim_{\epsilon \to 0} \frac{\varphi^{\epsilon}(t)}{\epsilon} = \Lambda_i^l(t|K).$$

Because of

$$\lim_{\epsilon \to 0} \frac{\varphi^{\epsilon}(t)}{\epsilon} = \lim_{\epsilon \to 0} \frac{x^{\epsilon}(t) - x(t)}{\epsilon} = \Lambda_i^l(t|K), i \in I_m, l \in I_n,$$

we obtain

$$\frac{\partial x(t|K)}{\partial K_i^l} = \Lambda_i^l(t|K), i \in I_m, l \in I_n,$$

thereby completing the proof. $\qquad\qquad\blacksquare$

We now present the following algorithm for computing the cost function of Problem $\mathbf{P_7}$ and its gradient at a given controller K.

3.6.2. The cost function with respect to the feedback matrix

Based on Theorem 4 and the chain rule, we have:

Theorem 5. *The gradients of the cost function $G(K)$ with respect to $K_i^l, i \in I_m, l \in I_n$, is given by*

$$\begin{aligned}
\frac{\partial G(K)}{\partial K_i^l} &= \sum_{l=1}^{n}\Bigg\{ -2K_i^l + \sum_{i=1}^{m}\Big[\frac{2\sigma^2}{(\sqrt{(K_i^l)^2 + 4\sigma^2} - K_i^l)^2}(\frac{2K_i^l}{\sqrt{(K_i^l)^2 + 4\sigma^2}} - 1) \\
&\quad + \frac{2\sigma^2}{(\sqrt{(-K_i^l)^2 + 4\sigma^2} + K_i^l)^2}(\frac{2K_i^l}{\sqrt{(K_i^l)^2 + 4\sigma^2}} + 1)\Big]\Bigg\} \\
&\quad + \gamma_1 \frac{2\sigma^2}{(\sqrt{J^0(K) - J^0(K_1^*) - \varepsilon + 4\sigma^2} - J^0(K) + J^0(K_1^*) + \varepsilon)^2}\Big[\frac{\partial J^0(K)/\partial K_i^l}{\sqrt{J^0(K) - J^0(K_1^*) - \varepsilon + 4\sigma^2}} - \frac{\partial J^0(K)}{\partial K_i^l}\Big] \\
&\quad + \gamma_2 \frac{2\sigma^2}{(\sqrt{(-J^0(K) + J^0(K_1^*) - \varepsilon) + 4\sigma^2} + J^0(K) - J^0(K_1^*) + \varepsilon)^2}\Big[\frac{-\partial J^0(K)/\partial K_i^l}{\sqrt{J^0(K) - J^0(K_1^*) - \varepsilon + 4\sigma^2}} + \frac{\partial J^0(K)}{\partial K_i^l}\Big],
\end{aligned}$$

where

$$\frac{\partial J^0(K)}{\partial K_i^l} = 2(x(T|K))^\top S \Lambda_i^l(t|K) + \int_0^T [(x(t|K))^\top Q\Lambda_i^l(t|K) + (u(t))^\top Wu(t)]dt, i \in I_m, l \in I_n.$$

Then, Problem $\mathbf{P_7}$ can be solved efficiently by using any gradient-based algorithm based on Theorem 5.

4. Numerical results

4.1. Experiment design

In this section, all computational experiments are carried out in MATLAB R2021a on a computer with a 3.70 GHz Intel Core i9-10900K CPU243 and 32.0GB RAM. The Euler method is used to solve system (2.2) with a step size of 1/10, and the initial time and terminal time are 0 and 1, respectively. We consider the 10th-order.

We use the gradient-based methods described in [31] for solving Problem $\mathbf{P_1}$ to obtain the optimal dense feedback matrix denoted by K_1^*. Then, through the application of Algorithm 1, we solve Problem $\mathbf{P_7}$ to obtain the optimal sparse feedback matrix K_2^*.

To indicate the sparse level of the feedback matrix K, we define the following indicator:

$$r = \frac{\text{Number of nonzero elements in } K}{\text{Number of elements in } K},$$

which represents the proportion of nonzero elements in the matrix. Obviously, a smaller value of r means a better sparse level of K.

4.2. Experiment 1

Based on empirical data, we consider a 10-th order LTI system with a state matrix

$$A = \begin{bmatrix} -8 & -10 & 0 & -1 & -9 & -2 & -8 & -4 & -4 & -6 \\ -10 & -4 & -9 & -7 & -5 & -6 & -4 & -3 & -1 & 0 \\ -4 & -6 & -4 & -7 & -3 & -4 & -4 & -8 & -1 & -9 \\ -6 & -9 & -7 & 0 & -4 & -6 & -5 & -7 & -5 & -1 \\ -2 & -1 & -5 & -9 & 0 & -2 & -5 & -5 & -10 & -7 \\ -5 & -5 & -10 & -2 & -6 & -8 & -5 & -3 & -3 & 0 \\ -4 & -6 & 0 & -9 & -9 & -4 & -5 & -10 & -9 & -9 \\ -3 & -3 & -4 & -2 & -3 & -6 & -6 & -7 & -10 & -5 \\ -9 & -7 & -8 & -6 & -9 & -4 & -10 & -2 & -7 & -9 \\ 0 & -5 & -9 & -3 & 0 & 0 & -1 & -4 & -8 & -2 \end{bmatrix}.$$

Assume that $B = R = Q = E_{10}$, $S = \mathbf{0} \in \mathbb{R}^{10 \times 10}$, and $\varepsilon = 0.1$. The cost function is given by [30]:

$$J^0(K) = \int_0^1 [(x(t|K))^\top x(t|K) + u(t)^\top u(t)]\mathrm{d}t,$$

where $u = -Kx(t - \tau(\|K\|_0))$.

4.3. Experiment 2

Based on empirical data, we consider another 10th-order LTI system with a state matrix

$$A = \begin{bmatrix} 3 & -10 & 0 & -1 & -9 & -2 & -8 & -4 & -4 & -6 \\ -10 & -4 & 6 & -7 & -5 & 6 & -4 & 2 & -1 & 0 \\ -4 & -6 & 7 & -7 & -3 & -4 & -4 & -8 & -1 & -9 \\ -6 & -9 & -7 & 0 & -4 & -6 & -5 & -7 & -5 & -1 \\ -2 & -1 & -5 & -9 & 0 & -2 & -1 & -5 & -10 & -7 \\ -5 & -5 & -10 & -2 & -6 & -8 & -5 & -3 & -3 & 0 \\ -4 & -6 & 0 & -9 & 4 & -4 & -5 & -10 & -9 & -9 \\ -3 & -3 & -4 & -2 & -3 & 9 & -6 & -7 & 2 & -5 \\ 3 & -7 & -8 & -6 & -9 & -4 & 1 & -2 & -7 & -9 \\ 0 & -5 & -9 & -3 & 0 & 0 & -1 & -4 & -8 & -2 \end{bmatrix}.$$

Assume that $B = R = Q = E_{10}$, $S = \mathbf{0} \in \mathbb{R}^{10 \times 10}$, $\varepsilon = 0.1$. The cost function is given by [30]:

$$J^0(K) = \int_0^1 [(x(t|K))^\top x(t|K) + u(t)^\top u(t)]\mathrm{d}t,$$

where $u = -Kx(t - \tau(\|K\|_0))$.

4.4. Experiment analysis

The distributions of nonzero components in the feedback matrices K_1^* and K_2^* and their corresponding state at each moment are displayed in Figures 3 and 4, where *nz* means the number of nonzero elements. Their corresponding optimal cost and sparsity levels are given in Tables 1 and 2. Figures 3 and 4 indicate that the feedback matrix K_1^* exhibits a high degree of density, while the feedback matrix K_2^* displays a notable level of sparsity. Furthermore, from Tables 1 and 2, it follows that the value of the cost function, specifically $J^0(K_2^*)$, slightly exceeds that of $J^0(K_1^*)$.

We can see that the number of zero components in K_2^* increases rapidly with only a small increase in cost. Therefore, we can conclude that Algorithm 1 proposed in this paper can produce a better quality solution which balances the system performance and sparsity.

Table 1. The optimal feedback matrix, the corresponding optimal cost function, and the sparsity indicator *r* for Experiment 1.

the optimal feedback matrix	the optimal cost function	*r*
K_1^*	$J^0(K_1^*) = 2.59$	0.88
K_2^*	$J^0(K_2^*) = 2.68$	0.06

(a) Distribution of nonzero components in K_1^* (b) Distribution of nonzero components in K_2^*

(c) Changes in state values x of K_1^* (d) Changes in state values x of K_2^*
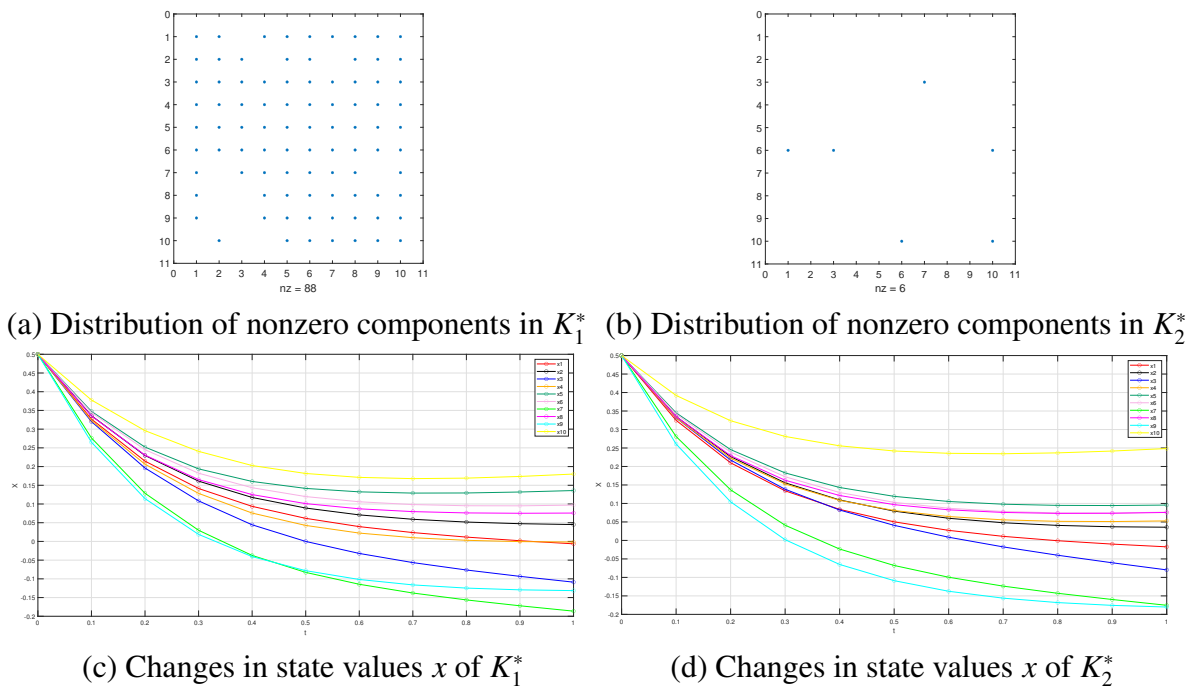
Figure 3. Distribution of the nonzero components of K_1^* without considering sparsity and K_2^* with considering sparsity, and their corresponding change in x for Experiment 1.
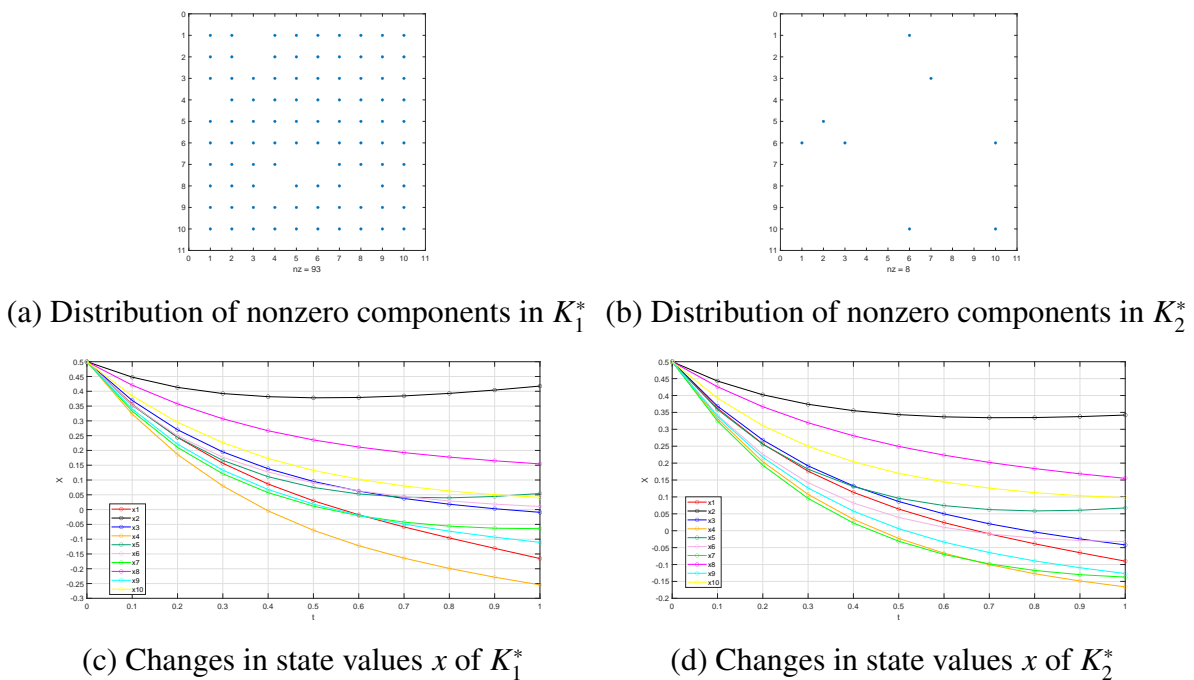

(a) Distribution of nonzero components in K_1^* (b) Distribution of nonzero components in K_2^*

(c) Changes in state values x of K_1^* (d) Changes in state values x of K_2^*

Figure 4. Distribution of the nonzero components of K_1^* without considering sparsity and K_2^* with considering sparsity, and their corresponding change in x for Experiment 2.

Table 2. The optimal feedback matrix, the corresponding optimal cost function and sparsity indicator r for Experiment 2.

the optimal feedback matrix	the optimal cost function	r
K_1^*	$J^0(K_1^*) = 2.41$	0.93
K_2^*	$J^0(K_2^*) = 2.57$	0.08

Algorithm 1 The gradients calculation of the cost function in Problem $\mathbf{P_7}$

1: **Step 1**: Obtain $x(t|K)$, $\Lambda_i^l(t)$, $i \in I_m, l \in I_n$, by solving the enlarged time-delay system consisting of the original system (2.2) and the auxiliary system in Theorem 4.

2: **Step 2**: Use the state values $x(t|K)$ to compute $G(K)$ in Problem $\mathbf{P_7}$.

3: **Step 3**: Use $x(t|K)$, $G(K)$ and $\Lambda_i^l(t)$ to compute $\frac{\partial G(K)}{\partial K_i^l}$, $i \in I_m, l \in I_n$.

---

5. Conclusions

In practical scenarios, network operators often choose sparse communication topologies to minimize costs, but the concurrent use of the network by multiple users often leads to feedback delays. Our goal is to determine the optimal sparse feedback control matrix K. To achieve this, we formulate a SOC problem tailored to the CPS with variable delays, with the aim of minimizing $\|K\|_0$ while adhering to a specified maximum compromise in system costs. We employ a penalty method to transform the SOC problem into one governed solely by box constraints. To handle the nonsmooth aspects of the resulting problem, we utilize a smoothing technique and subsequently analyze the errors it may introduce. The gradients of the objective function concerning the feedback control matrix are computed by simultaneously solving the state system and the associated variational system over time. An optimization algorithm is designed to tackle the resulting challenge, utilizing a piecewise quadratic approximation. The paper concludes with a discussion of the simulation results.

The innovation of the paper is stated as follows: (i) The innovative use of a penalty method transforms the problem into one constrained by box limits, simplifying the optimization process while still enforcing necessary control constraints; (ii) the application of smoothing techniques to approximate nonsmooth components enables the derivation of gradients efficiently, facilitating the use of gradient-based optimization algorithms in sparse settings; (iii) the method incorporates simultaneous solving of the state and variational systems, enhancing accuracy in gradient calculations and improving the overall performance of the control strategy; and (iv) the development of an optimization algorithm based on piecewise quadratic approximation offers a computationally efficient way to navigate the optimization landscape, making the approach feasible for real-time applications.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

Conflict of interest

The authors declare there is no conflicts of interest.

References

1. M. Alowaidi, S. K. Sharma, A. AlEnizi, S. Bhardwaj, Integrating artificial intelligence in cyber security for cyber-physical systems, *Electron. Res. Arch.*, **31** (2023), 1876–1896. http://doi.org/10.3934/era.2023097

2. N. Negi, A. Chakrabortty, Sparsity-promoting optimal control of cyber-physical systems over shared communication networks, *Automatica*, **122** (2020), 109217. http://doi.org/10.1016/j.automatica.2020.109217

3. M. N. Al-Mhiqani, T. Alsboui, T. Al-Shehari, K. H. Abdulkareem, R. Ahmad, M. A. Mohammed, Insider threat detection in cyber-physical systems: a systematic literature review, *Comput. Electr. Eng.*, **119** (2024), 109489. http://doi.org/10.1016/j.compeleceng.2024.109489

4. S. Liu, X. Wang, B. Niu, X. Song, H. Wang, X. Zhao, Adaptive resilient output feedback control against unknown deception attacks for nonlinear cyber-physical systems, *IEEE Trans. Circuits Syst. II: Express Briefs*, **71** (2024). 3855–3859. http://doi.org/10.1109/TCSII.2024.3372413

5. M. Zhao, W. Qin, J. Yang, G. Lu, Security control for cyber-physical systems under aperiodic denial-of-service attacks: A memory-event-triggered active approach, *Neurocomputing*, **600** (2024), 128159. http://doi.org/10.1016/j.neucom.2024.128159

6. P. E. G. Silva, P. H. J. Nardelli, A. S. de Sena, H. Siljak, N. Nevaranta, N. Marchetti, et al., Semantic-functional communications in cyber-physical systems, *IEEE Network*, **38** (2024), 241–249. http://doi.org/10.1109/MNET.2023.3329192

7. L. M. Castiglione, E. C. Lupu, Which attacks lead to hazards combining safety and security analysis for cyber-physical systems, *IEEE Trans. Dependable Secure Comput.*, **21** (2024), 2526–2540. http://doi.org/10.1109/TDSC.2023.3309778

8. S. Das, P. Dey, D. Chatterjee, Almost sure detection of the presence of malicious components in cyber-physical systems, *Automatica*, **167** (2024), 111789. http://doi.org/10.1016/j.automatica.2024.111789

9. C. Fioravanti, S. Panzieri, G. Oliva, Negativizability: a useful property for distributed state estimation and control in cyber-physical systems, *Automatica*, **157** (2023), 111240. http://doi.org/10.1016/j.automatica.2023.111240

10. H. N. AlEisa, F. Alrowais, R. Allafi, N. S. Almalki, R. Faqih, R. Marzouk, et al., Transforming transportation: safe and secure vehicular communication and anomaly detection with intelligent cyber-physical system and deep learning, *IEEE Trans. Consum. Electron.*, **70** (2024), 1736–1746. http://doi.org/10.1109/TCE.2023.3325827

11. L. Khoshnevisan, X. Liu, Resilient neural network-based control of nonlinear heterogeneous multi-agent systems: a cyber-physical system approach, *Nonlinear Dyn.*, **111** (2023), 19171–19185. http://doi.org/10.1007/s11071-023-08840-w

12. L. Chen, Y. Li, S. Tong, Robust adaptive control for nonlinear cyber-physical systems with FDI attacks via attack estimation, *Int. J. Robust Nonlinear Control*, **33** (2023), 9299–9316. http://doi.org/10.1002/rnc.6851

13. G. Routray, R. M. Hegde, Sparsity-driven loudspeaker gain optimization for sound field reconstruction with spherical microphone array, *Digital Signal Process.*, **154** (2024), 104688. http://doi.org/10.1016/j.dsp.2024.104688

14. K. Zhou, Y. Wang, B. Qiao, J. Liu, M. Liu, Z. Yang, et al., Non-convex sparse regularization via convex optimization for blade tip timing, *Mech. Syst. Signal Process.*, **222** (2025), 111764. http://doi.org/10.1016/j.ymssp.2024.111764

15. T. Zhang, F. Peng, X. Tang, R. Yan, R. Deng, S. Zhao, A sparse knowledge embedded configuration optimization method for robotic machining system toward improving machining quality, *Rob. Comput-Integr. Manuf.*, **90** (2024), 102818. http://doi.org/10.1016/j.rcim.2024.102818

16. J. Yuan, C. Wu, K. L. Teo, J. Xie, S. Wang, Computational method for feedback perimeter control of multiregion urban traffic networks with state-dependent delays, *Transp. Res. Part C: Emerging Technol.*, **153** (2023), 104231. https://doi.org/10.1016/j.trc.2023.104231

17. J. Yuan, C. Wu, K. L. Teo, S. Zhao, L. Meng, Perimeter control with state-dependent delays: optimal control model and computational method, *IEEE Trans. Intell. Transp. Syst.*, **23** (2022), 20614–20627. https://doi.org/10.1109/TITS.2022.3179729

18. C. Zhao, Z. Luo, N. Xiu, Some advances in theory and algorithms for sparse optimization, *Oper. Res. Trans.*, **24** (2020), 1–24.

19. S. Dai, Variable selection in convex quantile regression: L_1-norm or L_0-norm regularization?, *Eur. J. Oper. Res.*, **305** (2023), 338–355. http://doi.org/10.1016/j.ejor.2022.05.041

20. B. Hong, H. Qian, Z. Wang, Iterative hard thresholding algorithm-based detector for compressed OFDM-IM systems, *IEEE Commun. Lett.*, **26** (2022), 2205–2209. http://doi.org/10.1109/LCOMM.2022.3187451

21. H. Liu, T. Wang, Z. Liu, Some modified fast iterative shrinkage thresholding algorithms with a new adaptive non-monotone stepsize strategy for nonsmooth and convex minimization problems, *Comput. Optim. Appl.*, **83** (2022), 651–691. http://doi.org/10.1007/s10589-022-00396-6

22. B. Liu, K. Gong, L. Zhang, Convergence analysis of the augmented Lagrangian method for l_p-norm cone optimization problems with $p \geq 2$, *Numer. Algorithms*, **2024** (2024). http://doi.org/10.1007/s11075-024-01912-x

23. D. Han, D. Sun, L. Zhang, Linear rate convergence of the alternating direction method of multipliers for convex composite programming, *Math. Oper. Res.*, **43** (2018), 622–637. https://doi.org/10.1287/moor.2017.0875

24. E. Hernández, P. Merino, Sparse optimal control of timoshenko's beam using a locking-free finite element approximation, *Optim. Control Appl. Methods*, **45** (2024), 1007–1029. https://doi.org/10.1002/oca.3085

25. K. Ito, T. Ikeda, K. Kashima, Sparse optimal stochastic control, *Automatica*, **125** (2021), 109438. http://doi.org/10.1016/j.automatica.2020.109438

26. F. Lejarza, M. Baldea, Economic model predictive control for robust optimal operation of sparse storage networks, *Automatica*, **125** (2021), 109346. http://doi.org/10.1016/j.automatica.2020.109346

27. M. Liu, H. Zhu, F. Zhang, J. Wang, C. Zhou, Y. Lv, Model-based sparse optimal control of the hydrogen sulfide synthesis process for acidic wastewater sulfidation, *J. Water Process Eng.*, **65** (2024), 105836. http://doi.org/10.1016/j.jwpe.2024.105836

28. J. Yuan, C. Wu, Z. Liu, S. Zhao, C. Yu, K. L. Teo, et al., Koopman modeling for optimal control of the perimeter of multi-region urban traffic networks, *Appl. Math. Modell.*, **138** (2025), 115742. https://doi.org/10.1016/j.apm.2024.115742

29. Q. Li, Y. Bai, C. Yu, Y. Yuan, A new piecewise quadratic approximation approach for L_0 norm minimization problem, *Sci. China Math.* **62** (2019), 185–204. http://doi.org/10.1007/s11425-017-9315-9

30. A. Modi, M. K. S. Faradonbeh, A. Tewari, G. Michailidis, Joint learning of linear time-invariant dynamical systems, *Automatica*, **164** (2024), 111635. http://doi.org/10.1016/j.automatica.2024.111635

31. K. L. Teo, B. Li, C. Yu, V. Rehbock, *Applied and Computational Optimal Control: a Control Parametrization Approach*, Springer, Cham, 2021. http://doi.org/10.1007/978-3-030-69913-0

32. X. Wu, K. Zhang, M. Chen, A gradient-based algorithm for non-smooth constrained optimization problems governed by discrete-time nonlinear equations with application to long-term hydrothermal optimal scheduling control, *J. Comput. Appl. Math.*, **412** (2022), 114335. https://doi.org/10.1016/j.cam.2022.114335

AIMS Press