

# Distributionally robust parameter identification of a time-delay dynamical system with stochastic measurements

Zhaohua Gong<sup>a</sup>, Chongyang Liu<sup>\*,a,b</sup>, Kok Lay Teo<sup>b,c</sup>, Jie Sun<sup>b</sup>

<sup>a</sup>*School of Mathematics and Information Science, Shandong Institute of Business and Technology, Yantai 264005, China*

<sup>b</sup>*School of Electrical Engineering, Computing, and Mathematical Sciences, Curtin University, Perth 6845, Australia*

<sup>c</sup>*Coordinated Innovation Center for Computable Modeling in Management Science, Tianjin University of Finance and Economics, Tianjin 300222, China*

---

## Abstract

In this paper, we consider a parameter identification problem involving a time-delay dynamical system, in which the measured data are stochastic variable. However, the probability distribution of this stochastic variable is not available and the only information we have is its first moment. This problem is formulated as a distributionally robust parameter identification problem governed by a time-delay dynamical system. Using duality theory of linear optimization in a probability space, the distributionally robust parameter identification problem, which is a bi-level optimization problem, is transformed into a single-level optimization problem with a semi-infinite constraint. By applying problem transformation and smoothing techniques, the semi-infinite constraint is approximated by a smooth constraint and the convergence of the smooth approximation method is established. Then, the gradients of the cost and constraint functions with respect to time-delay and parameters are derived. On this basis, a gradient-based optimization method for solving the transformed problem is developed. Finally, we present an example, arising in practical fermentation process, to illustrate the applicability of the proposed method.

*Key words:* Parameter identification; distributionally robust optimization; duality theory; time-delay system; semi-infinite optimization

---

## 1. Introduction

The study of the distributionally robust programming (i.e., without dynamical system constraints) can be traced back to the work for the newsboy problem reported in [1], where only the first-order and second-order moments are

---

\*Corresponding author.

*Email address:* liu\_chongyang@yahoo.com (Chongyang Liu)

available for the uncertain parameters involved. Distributionally robust programming problem with the first-order and second-order moments information is transformed into a series of deterministic quadratic programming problems in [2]. Distributionally robust optimization under moment uncertainty with application to data-driven problems is investigated in [3]. By using duality theory [4], a distributionally robust programming problem is transformed into a tractable semi-definite programming problem. A general framework is proposed for the modelling of distributionally robust programming in [5], where tractable conditions are derived for the problem. An entropy approximation method is developed for a class of distributionally robust programming problems in [6]. The convergence analysis is studied for distributionally robust optimization and equilibrium problem in [7]. However, for all these studies, the distributionally robust problems do not include dynamical systems. Recently, the work in [8] is a study of the distributionally robust control of linear time-invariant discrete-time systems. A distributionally robust control problem of linear discrete-time system arising in signal processing is studied in [9]. However, for the distributionally robust optimization problem involving continuous-time dynamical system, it is still in its infancy. The main difficulties are: (i) continuous-time dynamical system constraint cannot be expressed as a finite number of equality constraints as in the case of discrete-time system; and (ii) the inner-level objective is expressed in the form of mathematical expectation of a non-convex functional due to the presence of the continuous-time dynamical system.

An optimal parameter identification problem is a type of optimal control problem in which some parameters need to be chosen subject to a given dynamic system. This paper is concerned with parameter identification for nonlinear time-delay systems, in which the time-delay and some other parameters are needed to be identified. Parameter identification for time-delay systems has attracted considerable research interest over the past years; see, for example [10, 11]. A gradient-based optimization approach is proposed in [12, 13], where the estimate of the unknown model parameters and time-delay is achieved through minimizing the deviation between predicted and measured system output. A parameter identification scheme for a time-delay system arising from the study of microbial batch fermentation is developed in [14]. However, the presence of noise in the measurement data is not taken into account in the above parameter identification problems, i.e., the output measurements used in those cost functions are assumed to be exact. In fact, it is impossible to measure the output data accurately in a real system. As a result, computational methods for parameter estimation with noisy measurements are proposed in [15, 16]. However, in those parameter estimation problems, it is assumed that the distributions of output measurements are exactly given.

In a real practical system, the probability distribution of the output measurement data is usually not available. Only partial statistical knowledge is known. This motivates the study of this paper. We consider a distributionally robust parameter identification problem governed by a time-delay dynamical system, in which the optimization objective is divided into two levels: (i) the inner-level is to maximize the deviation between measured and predicted system outputs

with respect to the probability distribution of the measured output; and (ii) the outer-level is to minimize with respect to the time-delay and parameters. By applying the duality theory, the distributionally robust parameter identification problem is transformed into a single-level optimization problem with a semi-infinite constraint. By problem transformation and smoothing techniques [17], the semi-infinite constraint is approximated by a smooth constraint and the convergence of the smooth approximation is established. Furthermore, a gradient-based optimization method for solving the transformed problem is developed. Numerical example illustrates the applicability of the proposed method.

The rest of this paper is organised as follows. The problem formulation is described in Section 2. The transformation procedure, which includes duality theory and smoothing technique, is given in Section 3. A gradient-based computational procedure is developed in Section 4. A numerical example is given in Section 5. Finally, Section 6 concludes the paper.

## 2. Problem formulation

Consider the following time-delay dynamical system

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{x}(t - \tau), \mathbf{q}), & t \in (0, T], \\ \mathbf{x}(t) = \boldsymbol{\phi}(t, \mathbf{q}), & t \leq 0, \end{cases} \quad (1)$$

where  $\mathbf{x}(t) \in \mathbb{R}^n$  is the state vector;  $\tau \in \mathbb{R}$  is a time-delay;  $\mathbf{q} \in \mathbb{R}^r$  is the unknown parameter vector;  $T$  is a given terminal time;  $\mathbf{f} : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^r \rightarrow \mathbb{R}^n$  is a given continuously differentiable function; and  $\boldsymbol{\phi} : \mathbb{R} \times \mathbb{R}^r \rightarrow \mathbb{R}^n$  is a given history function. For system (1), we assume that the output function  $\mathbf{y}(t) \in \mathbb{R}^p$  is given by the following equation:

$$\mathbf{y}(t) = \mathbf{g}(\mathbf{x}(t)), \quad t \in [0, T], \quad (2)$$

where  $\mathbf{g} : \mathbb{R}^n \rightarrow \mathbb{R}^p$  is a given continuously differentiable function.

In system (1), time-delay  $\tau$  and parameter vector  $\mathbf{q}$  are unknown and need to be estimated. Suppose that

$$\tau \in \mathcal{T} := [0, \hat{\tau}],$$

where  $\hat{\tau} > 0$  is a given upper bound. Furthermore, we assume that

$$\mathbf{q} \in \mathcal{Q} := \{\mathbf{q} : c_j \leq q_j \leq d_j, j = 1, 2, \dots, r\},$$

where  $c_j$  and  $d_j$  are given constants such that  $c_j < d_j$ . Any pair  $(\tau, \mathbf{q}) \in \mathcal{T} \times \mathcal{Q}$  is called an admissible delay-parameter pair.

Let  $\mathbf{x}(\cdot | \tau, \mathbf{q})$  denote the solution of time-delay system (1) corresponding to each  $(\tau, \mathbf{q}) \in \mathcal{T} \times \mathcal{Q}$ . Furthermore, let  $\mathbf{y}(\cdot | \tau, \mathbf{q})$  denote the corresponding output obtained by substituting  $\mathbf{x}(\cdot | \tau, \mathbf{q})$  into (2). Our goal is to estimate the unknown pair  $(\tau, \mathbf{q}) \in \mathcal{T} \times \mathcal{Q}$  such that the deviation between the system output  $\mathbf{y}(\cdot | \tau, \mathbf{q})$  and the measured output is minimized at a set of sample times. In a real

practical situation, the measured output, which is obtained through experiment at a set of sample times, can not be exact since experiment data always contain some noises. Thus, in this paper, the measured output vector  $\tilde{\mathbf{z}}$  is regarded as a stochastic variable with its first-order moment given by

$$\mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{z}}] = \boldsymbol{\mu}, \quad (3)$$

where  $\mathbb{E}_{\mathbb{P}}$  denotes the expectation under probability distribution of  $\mathbb{P}$ ;  $\tilde{\mathbf{z}} := ((\tilde{\mathbf{z}}^1)^\top, \dots, (\tilde{\mathbf{z}}^l)^\top)^\top = (\tilde{z}_1^1, \dots, \tilde{z}_p^1, \dots, \tilde{z}_1^l, \dots, \tilde{z}_p^l)^\top \in \mathbb{R}^{lp}$ ;  $\boldsymbol{\mu} := (\mu_1^1, \dots, \mu_p^1, \dots, \mu_1^l, \dots, \mu_p^l)^\top \in \mathbb{R}^{lp}$ ; and  $\tilde{\mathbf{z}}^k$  is the output vector at sample time  $t_k$ ,  $k = 1, 2, \dots, l$ , such that  $0 \leq t_1 < \dots < t_l \leq T$ . Let  $\Omega$  be the support set of  $\tilde{\mathbf{z}}$  defined by

$$\Omega := \{\tilde{\mathbf{z}} \in \mathbb{R}^{lp} : a_h^k \leq \tilde{z}_h^k \leq b_h^k, k = 1, 2, \dots, l, h = 1, 2, \dots, p\}, \quad (4)$$

where  $a_h^k$  and  $b_h^k$  are given constants such that  $a_h^k < b_h^k$ . Furthermore, let  $\mathcal{F}$  be the ambiguity set of probability measures of  $\tilde{\mathbf{z}}$  defined by

$$\mathcal{F} := \{\mathbb{P} : \mathbb{P}(\tilde{\mathbf{z}} \in \Omega) = 1, \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{z}}] = \boldsymbol{\mu}\}. \quad (5)$$

We use the following least-squares error function to measure the deviation between the system output and measurement data:

$$J(\tau, \mathbf{q}, \tilde{\mathbf{z}}) = \sum_{k=1}^l \|\mathbf{y}(t_k | \tau, \mathbf{q}) - \tilde{\mathbf{z}}^k\|^2, \quad (6)$$

where  $\|\cdot\|$  denotes the Euclidean norm. However,  $\tilde{\mathbf{z}}$  is a stochastic variable for which its probability distribution is not known exactly. Note that we only know that the probability distribution  $\mathbb{P}$  of the stochastic variable  $\tilde{\mathbf{z}}$  is in the ambiguity set  $\mathcal{F}$ . Thus, we consider the worst-case scenario of  $J(\tau, \mathbf{q}, \tilde{\mathbf{z}})$

$$\max_{\mathbb{P} \in \mathcal{F}} \mathbb{E}_{\mathbb{P}}[J(\tau, \mathbf{q}, \tilde{\mathbf{z}})]. \quad (7)$$

Then, our distributionally robust parameter identification problem can be stated as follows.

**Problem (A).** Given the time-delay system (1), choose an admissible delay-parameter pair  $(\tau, \mathbf{q}) \in \mathcal{T} \times \mathcal{Q}$  such that (7) is minimized.

### 3. Problem transformation

Problem (A) is a bi-level optimization problem involving a nonlinear time-delay system. In this section, we shall transform it into a deterministic single-level optimization problem.

#### 3.1. Single-level optimization problem

The inner-level optimization of Problem (A) can be explicitly written as

$$\begin{aligned} \text{(IA)} \quad & \max_{\mathbb{P}} \mathbb{E}_{\mathbb{P}}[J(\tau, \mathbf{q}, \tilde{\mathbf{z}})] \\ & \text{s.t.} \quad \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{z}}] = \boldsymbol{\mu}, \\ & \quad \mathbb{P}(\tilde{\mathbf{z}} \in \Omega) = 1. \end{aligned}$$

Note that (IA) is an optimization problem with respect to a probability measure  $\mathbb{P} \in \mathcal{F}$  on support set  $\Omega$ . For convenience of discussion, we only consider discrete or absolutely continuous stochastic variable, i.e., the expectation can be expressed by Lebesgue integral. Then, (IA) can be written as

$$\begin{aligned} \text{(IA')} \quad & \max_{\mathbb{P}} \int_{\Omega} J(\tau, \mathbf{q}, \tilde{\mathbf{z}}) d\mathbb{P} \\ & \text{s.t.} \quad \int_{\Omega} \tilde{\mathbf{z}} d\mathbb{P} = \boldsymbol{\mu}, \\ & \int_{\Omega} \mathbf{1}_{(\tilde{\mathbf{z}} \in \Omega)} d\mathbb{P} = 1. \end{aligned}$$

Using the duality theory of linear optimization in probability spaces [18], the dual problem of (IA') is

$$\begin{aligned} \text{(DIA)} \quad & \min_{\boldsymbol{\beta} \in \mathbb{R}^{lp}, \gamma \in \mathbb{R}} \gamma + \boldsymbol{\mu}^{\top} \boldsymbol{\beta} \\ & \text{s.t.} \quad \mathbf{z}^{\top} \boldsymbol{\beta} + \gamma \geq J(\tau, \mathbf{q}, \tilde{\mathbf{z}}), \quad \forall \mathbf{z} \in \Omega, \end{aligned} \quad (8)$$

where  $\boldsymbol{\beta} = (\beta_1^1, \dots, \beta_p^1, \dots, \beta_1^l, \dots, \beta_p^l)^{\top} \in \mathbb{R}^{lp}$  and  $\gamma \in \mathbb{R}$  are dual variables; and  $J(\tau, \mathbf{q}, \tilde{\mathbf{z}})$  is as defined in (6).

In (DIA), the constraint (8) is a semi-infinite constraint, and

$$\begin{aligned} (8) \iff & \sum_{k=1}^l \sum_{h=1}^p (z_h^k)^2 - \sum_{k=1}^l \sum_{h=1}^p (2y_h(t_k|\tau, \mathbf{q}) + \beta_h^k) z_h^k - \gamma + \sum_{k=1}^l \sum_{h=1}^p (y_h(t_k|\tau, \mathbf{q}))^2 \\ & \leq 0, \quad \forall z_h^k \in [a_h^k, b_h^k], \quad k = 1, 2, \dots, l, \quad h = 1, 2, \dots, p \\ \iff & \max_{\mathbf{z} \in \Omega} \left\{ \sum_{k=1}^l \sum_{h=1}^p \left[ (z_h^k)^2 - (2y_h(t_k|\tau, \mathbf{q}) + \beta_h^k) z_h^k + (y_h(t_k|\tau, \mathbf{q}))^2 \right] \right\} - \gamma \\ & \leq 0. \end{aligned} \quad (9)$$

Due to the separability of the variables  $z_h^k$ ,  $k = 1, 2, \dots, l, h = 1, 2, \dots, p$ , the semi-infinite constraint (9) can be equivalently written as

$$\begin{aligned} & \sum_{k=1}^l \sum_{h=1}^p \max_{z_h^k \in [a_h^k, b_h^k]} \left\{ (z_h^k)^2 - (2y_h(t_k|\tau, \mathbf{q}) + \beta_h^k) z_h^k + \right. \\ & \quad \left. + (y_h(t_k|\tau, \mathbf{q}))^2 \right\} - \gamma \leq 0 \\ \iff & \sum_{k=1}^l \sum_{h=1}^p \max \left\{ (a_h^k)^2 - (2y_h(t_k|\tau, \mathbf{q}) + \beta_h^k) a_h^k + (y_h(t_k|\tau, \mathbf{q}))^2, \right. \\ & \quad \left. (b_h^k)^2 - (2y_h(t_k|\tau, \mathbf{q}) + \beta_h^k) b_h^k + (y_h(t_k|\tau, \mathbf{q}))^2 \right\} - \gamma \leq 0. \end{aligned} \quad (10)$$

It is easy to show that

$$\max\{\kappa, \nu\} = \frac{\kappa + \nu}{2} + \frac{|\kappa - \nu|}{2}. \quad (11)$$

Hence,

$$(10) \iff \frac{1}{2} \sum_{k=1}^l \sum_{h=1}^p \left\{ (a_h^k)^2 + (b_h^k)^2 - (2y_h(t_k|\tau, \mathbf{q}) + \beta_h^k)(a_h^k + b_h^k) + 2(y_h(t_k|\tau, \mathbf{q}))^2 + |(a_h^k - b_h^k)(a_h^k + b_h^k - 2y(t_k|\tau, \mathbf{q}) - \beta_h^k)| \right\} - \gamma \leq 0. \quad (12)$$

Now, we denote the constraint (12) by  $u(\tau, \mathbf{q}, \boldsymbol{\beta}, \gamma)$ :

$$u(\tau, \mathbf{q}, \boldsymbol{\beta}, \gamma) := \frac{1}{2} \sum_{k=1}^l \sum_{h=1}^p \left\{ w(y_h(t_k|\tau, \mathbf{q}), \beta_h^k) + |v(y_h(t_k|\tau, \mathbf{q}), \beta_h^k)| \right\} - \gamma \leq 0, \quad (13)$$

where

$$\begin{aligned} w(y_h(t_k|\tau, \mathbf{q}), \beta_h^k) &:= (a_h^k)^2 + (b_h^k)^2 - (2y_h(t_k|\tau, \mathbf{q}) + \beta_h^k)(a_h^k + b_h^k) \\ &\quad + 2(y_h(t_k|\tau, \mathbf{q}))^2, \\ v(y_h(t_k|\tau, \mathbf{q}), \beta_h^k) &:= (a_h^k - b_h^k)(a_h^k + b_h^k - 2y(t_k|\tau, \mathbf{q}) - \beta_h^k). \end{aligned}$$

Furthermore, let

$$\tilde{J}(\boldsymbol{\beta}, \gamma) := \gamma + \boldsymbol{\mu}^\top \boldsymbol{\beta}. \quad (14)$$

Then, the distributionally robust parameter identification Problem (A) can be transformed into the following single-level optimization problem.

**Problem (B).** Given the time-delay system (1), choose a  $(\tau, \mathbf{q}, \boldsymbol{\beta}, \gamma) \in \mathcal{T} \times \mathcal{Q} \times \mathbb{R}^{lp} \times \mathbb{R}$  such that the cost function (14) is minimized subject to the constraint (13).

### 3.2. Smoothing technique

Note that the second term in the constraint (13) is non-smooth in  $\tau, \mathbf{q}$  and  $\boldsymbol{\beta}$ . Any gradient-based optimization algorithm would have difficulty in solving Problem (B) directly. To overcome this difficulty, we will transform the constraint (13) by the smoothing technique.

To begin with, we introduce a smooth function:

$$\psi_\epsilon(\varpi) = \sqrt{\varpi^2 + \epsilon}, \quad (15)$$

where  $\epsilon > 0$  is an adjustable parameter. It can be easily shown that the function  $\psi_\epsilon$  satisfies the following properties.

**Lemma 1.** For any  $\varpi \in \mathbb{R}$  and  $\epsilon > 0$ ,  $\psi_\epsilon(\varpi)$  has the following properties:

- $\lim_{\epsilon \rightarrow 0^+} \psi_\epsilon(\varpi) = |\varpi|$ ;
- $\psi_\epsilon(\varpi) > 0$ ;
- $\psi'_\epsilon(\varpi) < 1$ ;
- $0 < \psi_\epsilon(\varpi) - |\varpi| \leq \sqrt{\epsilon}$ .

*Proof.* The proof is similar to that given for Lemma 2 in [19].  $\square$

Based on Lemma 1, the non-smooth constraint (13) can be approximated by the following smooth constraint:

$$\begin{aligned} u_\epsilon(\tau, \mathbf{q}, \boldsymbol{\beta}, \gamma) \\ := \frac{1}{2} \sum_{k=1}^l \sum_{h=1}^p \left\{ w(y_h(t_k|\tau, \mathbf{q}), \beta_h^k) + \psi_\epsilon(v(y_h(t_k|\tau, \mathbf{q}), \beta_h^k)) \right\} - \gamma \leq 0. \end{aligned} \quad (16)$$

Thus, Problem (B) is transformed into a smooth optimization problem as follows.

**Problem (C).** Given the time-delay system (1), choose a  $(\tau, \mathbf{q}, \boldsymbol{\beta}, \gamma) \in \mathcal{T} \times \mathcal{Q} \times \mathbb{R}^{lp} \times \mathbb{R}$  such that the cost function (14) is minimized subject to the constraint (16).

Let

$$\begin{aligned} \mathcal{F} &:= \{(\tau, \mathbf{q}, \boldsymbol{\beta}, \gamma) \in \mathcal{T} \times \mathcal{Q} \times \mathbb{R}^{lp} \times \mathbb{R} : u(\tau, \mathbf{q}, \boldsymbol{\beta}, \gamma) \leq 0\}, \\ \mathcal{F}_\epsilon &:= \{(\tau, \mathbf{q}, \boldsymbol{\beta}, \gamma) \in \mathcal{T} \times \mathcal{Q} \times \mathbb{R}^{lp} \times \mathbb{R} : u_\epsilon(\tau, \mathbf{q}, \boldsymbol{\beta}, \gamma) \leq 0\}. \end{aligned}$$

Furthermore, we assume that  $\mathcal{F}$  satisfies the following conditions:

(H1).  $\text{int}(\mathcal{F}) \neq \emptyset$ , where  $\text{int}$  denotes the interior.

(H2). For the optimal solution  $(\tau^*, \mathbf{q}^*, \boldsymbol{\beta}^*, \gamma^*)$  of Problem (B), there exists a  $(\bar{\tau}, \bar{\mathbf{q}}, \bar{\boldsymbol{\beta}}, \bar{\gamma})$  such that  $\iota(\bar{\tau}, \bar{\mathbf{q}}, \bar{\boldsymbol{\beta}}, \bar{\gamma}) + (1 - \iota)(\tau^*, \mathbf{q}^*, \boldsymbol{\beta}^*, \gamma^*) \in \text{int}(\mathcal{F})$  holds for each  $\iota \in (0, 1]$ .

Under assumptions (H1) and (H2), we obtain the following convergence theorems.

**Theorem 1.** For each  $\epsilon > 0$ , the optimal solution of Problem (C) is feasible for Problem (B).

*Proof.* For any  $\epsilon > 0$ , suppose that  $(\tau_\epsilon^*, \mathbf{q}_\epsilon^*, \boldsymbol{\beta}_\epsilon^*, \gamma_\epsilon^*) \in \mathcal{F}_\epsilon$  is the optimal solution of Problem (C). Then,  $u_\epsilon(\tau_\epsilon^*, \mathbf{q}_\epsilon^*, \boldsymbol{\beta}_\epsilon^*, \gamma_\epsilon^*) \leq 0$  holds. By Lemma 1, we can obtain that  $u(\tau_\epsilon^*, \mathbf{q}_\epsilon^*, \boldsymbol{\beta}_\epsilon^*, \gamma_\epsilon^*) \leq u_\epsilon(\tau_\epsilon^*, \mathbf{q}_\epsilon^*, \boldsymbol{\beta}_\epsilon^*, \gamma_\epsilon^*) \leq 0$ . Thus,  $(\tau_\epsilon^*, \mathbf{q}_\epsilon^*, \boldsymbol{\beta}_\epsilon^*, \gamma_\epsilon^*) \in \mathcal{F}$ .  $\square$

**Theorem 2.** Suppose that  $(\tau^*, \mathbf{q}^*, \boldsymbol{\beta}^*, \gamma^*) \in \mathcal{F}$  is an optimal solution of Problem (B). For  $\epsilon > 0$ , let  $(\tau_\epsilon^*, \mathbf{q}_\epsilon^*, \boldsymbol{\beta}_\epsilon^*, \gamma_\epsilon^*) \in \mathcal{F}_\epsilon$  denote an optimal solution of Problem (C). If  $\epsilon \rightarrow 0^+$ , then  $(\tau_\epsilon^*, \mathbf{q}_\epsilon^*, \boldsymbol{\beta}_\epsilon^*, \gamma_\epsilon^*) \rightarrow (\tau^*, \mathbf{q}^*, \boldsymbol{\beta}^*, \gamma^*)$ .

*Proof.* Under assumptions (H1) and (H2), since  $(\tau^*, \mathbf{q}^*, \boldsymbol{\beta}^*, \gamma^*) \in \mathcal{F}$  is an optimal solution of Problem (B), there exists a  $(\bar{\tau}, \bar{\mathbf{q}}, \bar{\boldsymbol{\beta}}, \bar{\gamma}) \in \text{int}(\mathcal{F})$  such that for all  $\iota \in (0, 1]$ ,

$$(\tau_\iota, \mathbf{q}_\iota, \boldsymbol{\beta}_\iota, \gamma_\iota) := \iota(\bar{\tau}, \bar{\mathbf{q}}, \bar{\boldsymbol{\beta}}, \bar{\gamma}) + (1 - \iota)(\tau^*, \mathbf{q}^*, \boldsymbol{\beta}^*, \gamma^*) \in \text{int}(\mathcal{F}). \quad (17)$$

Furthermore, for any  $\varepsilon_1 > 0$ , there exists a  $\iota_1 \in (0, 1]$  such that for all  $\iota \in (0, \iota_1)$ , we have

$$\tilde{J}(\boldsymbol{\beta}^*, \gamma^*) \leq \tilde{J}(\boldsymbol{\beta}_\iota, \gamma_\iota) \leq \tilde{J}(\boldsymbol{\beta}^*, \gamma^*) + \varepsilon_1. \quad (18)$$

Set  $\iota_2 = \iota_1/2$ , then by (17), we have  $(\tau_{\iota_2}, \mathbf{q}_{\iota_2}, \boldsymbol{\beta}_{\iota_2}, \gamma_{\iota_2}) \in \text{int}(\mathcal{F})$ . So there exists an  $\varepsilon_2 > 0$  such that  $u(\tau_{\iota_2}, \mathbf{q}_{\iota_2}, \boldsymbol{\beta}_{\iota_2}, \gamma_{\iota_2}) < -\varepsilon_2$ . By Lemma 1,

$$\lim_{\epsilon \rightarrow 0^+} u_\epsilon(\tau_{\iota_2}, \mathbf{q}_{\iota_2}, \boldsymbol{\beta}_{\iota_2}, \gamma_{\iota_2}) = u(\tau_{\iota_2}, \mathbf{q}_{\iota_2}, \boldsymbol{\beta}_{\iota_2}, \gamma_{\iota_2}).$$

Hence, for any  $\varepsilon_3 > 0$ , there exists an  $\epsilon_0 > 0$  such that for each  $0 < \epsilon < \epsilon_0$

$$u_\epsilon(\tau_{\iota_2}, \mathbf{q}_{\iota_2}, \boldsymbol{\beta}_{\iota_2}, \gamma_{\iota_2}) < u(\tau_{\iota_2}, \mathbf{q}_{\iota_2}, \boldsymbol{\beta}_{\iota_2}, \gamma_{\iota_2}) + \varepsilon_3.$$

Set  $\varepsilon_3 = \varepsilon_2$ , then we have  $u_\epsilon(\tau_{\iota_2}, \mathbf{q}_{\iota_2}, \boldsymbol{\beta}_{\iota_2}, \gamma_{\iota_2}) < 0$ . Since  $(\tau_\epsilon^*, \mathbf{q}_\epsilon^*, \boldsymbol{\beta}_\epsilon^*, \gamma_\epsilon^*)$  is the optimal solution of Problem (C), we have

$$\tilde{J}(\boldsymbol{\beta}_\epsilon^*, \gamma_\epsilon^*) \leq \tilde{J}(\boldsymbol{\beta}_{\iota_2}, \gamma_{\iota_2}). \quad (19)$$

By Theorem 1,  $(\tau_\epsilon^*, \mathbf{q}_\epsilon^*, \boldsymbol{\beta}_\epsilon^*, \gamma_\epsilon^*)$  is also feasible for Problem (B), so

$$\tilde{J}(\boldsymbol{\beta}^*, \gamma^*) \leq \tilde{J}(\boldsymbol{\beta}_\epsilon^*, \gamma_\epsilon^*). \quad (20)$$

Combined with (18), (19) and (20), we have

$$\tilde{J}(\boldsymbol{\beta}^*, \gamma^*) \leq \tilde{J}(\boldsymbol{\beta}_\epsilon^*, \gamma_\epsilon^*) \leq \tilde{J}(\boldsymbol{\beta}^*, \gamma^*) + \varepsilon_1.$$

Consequently, letting  $\epsilon \rightarrow 0^+$  and noting that  $\varepsilon_1 > 0$  is arbitrary, the proof is complete.  $\square$

Theorems 1 and 2 guarantee that we can obtain the optimal solution of Problem (B) by solving Problem (C).

#### 4. Gradient formulae and algorithm

Problem (C) is a dynamic optimization problem. It is well known that gradient-based optimization methods are very effective in solving the dynamic optimization problems [20]. However, the gradient formulae of the cost function (14) and the constraint function (16) with respect to the decision variables  $\tau, \mathbf{q}, \boldsymbol{\beta}$  and  $\gamma$  are needed. In this section, we will investigate the required gradient computation.

##### 4.1. Gradient formulae

It is not straightforward to obtain the gradient formulae of the constraint function  $u_\epsilon(\tau, \mathbf{q}, \boldsymbol{\beta}, \gamma)$  with respect to time-delay  $\tau$  and system parameter  $\mathbf{q}$  because it is not an explicit function of  $\tau$  and  $\mathbf{q}$ . In what follows, we shall propose a method for computing the gradients of  $u_\epsilon(\tau, \mathbf{q}, \boldsymbol{\beta}, \gamma)$  with respect to  $\tau$  and  $\mathbf{q}$ .



Define

$$\begin{aligned}\tilde{\mathbf{f}}(t|\tau, \mathbf{q}) &= f(\mathbf{x}(t), \mathbf{x}(t-\tau), \mathbf{q}), \\ \frac{\partial \tilde{\mathbf{f}}(t|\tau, \mathbf{q})}{\partial \mathbf{x}} &= \frac{\partial \mathbf{f}(\mathbf{x}(t), \mathbf{x}(t-\tau), \mathbf{q})}{\partial \mathbf{x}}, \\ \frac{\partial \tilde{\mathbf{f}}(t|\tau, \mathbf{q})}{\partial \tilde{\mathbf{x}}} &= \frac{\partial \mathbf{f}(\mathbf{x}(t), \mathbf{x}(t-\tau), \mathbf{q})}{\partial \mathbf{x}(t-\tau)}, \\ \frac{\partial \tilde{\mathbf{f}}(t|\tau, \mathbf{q})}{\partial q_i} &= \frac{\partial \mathbf{f}(\mathbf{x}(t), \mathbf{x}(t-\tau), \mathbf{q})}{\partial q_i}.\end{aligned}$$

From [12], it is known that the state vector  $\mathbf{x}(\cdot|\tau, \mathbf{q})$  of time-delay system (15) is differentiable with respect to  $\tau$  and  $\mathbf{q}$ . For brevity, we will omit  $\tau$  and  $\mathbf{q}$  in  $\mathbf{x}(\cdot|\tau, \mathbf{q})$  in the sequel. Moreover, define the state variations with respect to  $\tau$  and  $\mathbf{q}$  as follows:

$$\frac{\partial \mathbf{x}(t)}{\partial \tau} = \boldsymbol{\rho}(t), \quad t \in (-\infty, T], \quad (21)$$

$$\frac{\partial \mathbf{x}(t)}{\partial q_i} = \boldsymbol{\sigma}_i(t), \quad i = 1, 2, \dots, r, \quad t \in (-\infty, T]. \quad (22)$$

Then, it can be shown that

$$\frac{\partial \mathbf{x}(t-\tau)}{\partial \tau} = \begin{cases} -\frac{\partial \phi(t-\tau, \mathbf{q})}{\partial t}, & \text{if } t < \tau, \\ \boldsymbol{\rho}(t-\tau) - \tilde{\mathbf{f}}(t-\tau|\tau, \mathbf{q}), & \text{if } t \geq \tau. \end{cases} \quad (23)$$

Consequently,

$$\begin{aligned}& \frac{\partial \mathbf{x}(t-\tau|\tau, \mathbf{q})}{\partial \tau} \\ &= (\boldsymbol{\rho}(t-\tau) - \tilde{\mathbf{f}}(t-\tau|\tau, \mathbf{q}))\chi_{[\tau, +\infty)}(t) - \frac{\partial \phi(t-\tau, \mathbf{q})}{\partial t}\chi_{(-\infty, \tau)}(t),\end{aligned} \quad (24)$$

and

$$\frac{\partial \mathbf{x}(t-\tau)}{\partial q_i} = \boldsymbol{\sigma}_i(t-\tau), \quad (25)$$

where, for a given interval  $I$ ,

$$\chi_I(t) = \begin{cases} 1, & \text{if } t \in I, \\ 0, & \text{otherwise.} \end{cases} \quad (26)$$

Now, consider the following impulsive auxiliary system:

$$\begin{cases} \dot{\boldsymbol{\lambda}}(t) = -\left(\frac{\partial \tilde{\mathbf{f}}(t|\tau, \mathbf{q})}{\partial \mathbf{x}}\right)^\top \boldsymbol{\lambda}(t) - \left(\frac{\partial \tilde{\mathbf{f}}(t+\tau|\tau, \mathbf{q})}{\partial \tilde{\mathbf{x}}}\right)^\top \boldsymbol{\lambda}(t+\tau), & t \in [0, T] \\ \boldsymbol{\lambda}(t_k^-) = \frac{1}{2} \sum_{h=1}^p \left\{ \frac{\partial \mathbf{g}_h(\mathbf{x}(t_k))}{\partial \mathbf{x}(t_k)} \right\}^\top \left\{ \frac{\partial w(y_h(t_k|\tau, \mathbf{q}))}{\partial y_h(t_k)} + \frac{\partial \psi_\epsilon(v(y_h(t_k|\tau, \mathbf{q})))}{\partial y_h(t_k)} \right\}^\top \\ \quad + \boldsymbol{\lambda}(t_k^+), \quad k = 1, 2, \dots, l, \\ \boldsymbol{\lambda}(t) = \mathbf{0}, \quad t \geq t_l. \end{cases} \quad (27)$$

**Theorem 3.** For each  $(\tau, \mathbf{q}, \boldsymbol{\beta}, \gamma) \in \mathcal{T} \times \mathcal{Q} \times \mathbb{R}^{lp} \times \mathbb{R}$ , the gradient of  $u_\epsilon(\tau, \mathbf{q}, \boldsymbol{\beta}, \gamma)$  with respect to time-delay  $\tau$  is given by

$$\begin{aligned} \frac{\partial u_\epsilon(\tau, \mathbf{q}, \boldsymbol{\beta}, \gamma)}{\partial \tau} &= - \int_0^T \boldsymbol{\lambda}^\top(t) \frac{\partial \tilde{\mathbf{f}}(t|\tau, \mathbf{q})}{\partial \tilde{\mathbf{x}}} \left\{ \tilde{\mathbf{f}}(t-\tau|\tau, \mathbf{q}) \chi_{[\tau, +\infty)}(t) \right. \\ &\quad \left. - \frac{\partial \phi(t-\tau, \mathbf{q})}{\partial t} \chi_{(-\infty, \tau)}(t) \right\} dt. \end{aligned} \quad (28)$$

*Proof.* Let  $\boldsymbol{\nu} : [0, +\infty) \rightarrow \mathbb{R}^n$  be an arbitrary absolutely continuous function. Now, the constraint function  $u_\epsilon(\tau, \mathbf{q}, \boldsymbol{\beta}, \gamma)$  in the constraint (16) can be expressed as

$$\begin{aligned} u_\epsilon(\tau, \mathbf{q}, \boldsymbol{\beta}, \gamma) &= \frac{1}{2} \sum_{k=1}^l \sum_{h=1}^p \left\{ w(y_h(t_k|\tau, \mathbf{q}), \beta_h^k) + \psi_\epsilon(v(y_h(t_k|\tau, \mathbf{q}), \beta_h^k)) \right\} - \gamma \\ &\quad + \sum_{k=1}^l \int_{t_{k-1}}^{t_k} \boldsymbol{\nu}^\top(t) (\tilde{\mathbf{f}}(t|\tau, \mathbf{q}) - \dot{\mathbf{x}}(t)) dt. \end{aligned} \quad (29)$$

By applying integral by parts, the last term in the right side of (29) becomes

$$\begin{aligned} &\sum_{k=1}^l \int_{t_{k-1}}^{t_k} \boldsymbol{\nu}^\top(t) (\tilde{\mathbf{f}}(t|\tau, \mathbf{q}) - \dot{\mathbf{x}}(t)) dt \\ &= \sum_{k=1}^l \int_{t_{k-1}}^{t_k} \boldsymbol{\nu}^\top(t) \tilde{\mathbf{f}}(t|\tau, \mathbf{q}) dt - \sum_{k=1}^l \int_{t_{k-1}}^{t_k} \boldsymbol{\nu}^\top(t) d\mathbf{x}(t) \\ &= -\boldsymbol{\nu}^\top(t_l^-) \mathbf{x}(t_l) + \boldsymbol{\nu}^\top(t_0^+) \phi(0, \mathbf{q}) + \sum_{k=1}^{l-1} (\boldsymbol{\nu}^\top(t_k^+) - \boldsymbol{\nu}^\top(t_k^-)) \mathbf{x}(t_k) \\ &\quad + \int_0^T \boldsymbol{\nu}^\top(t) \tilde{\mathbf{f}}(t|\tau, \mathbf{q}) dt + \int_0^T \dot{\boldsymbol{\nu}}^\top(t) \mathbf{x}(t) dt. \end{aligned} \quad (30)$$

Substituting (30) into (29) yields

$$\begin{aligned}
u_\epsilon(\tau, \mathbf{q}, \boldsymbol{\beta}, \gamma) &= \frac{1}{2} \sum_{k=1}^l \sum_{h=1}^p \left\{ w(y_h(t_k|\tau, \mathbf{q}), \beta_h^k) + \psi_\epsilon(v(y_h(t_k|\tau, \mathbf{q}), \beta_h^k)) \right\} - \gamma \\
&\quad - \boldsymbol{\nu}^\top(t_l^-) \mathbf{x}(t_l) + \boldsymbol{\nu}^\top(t_0^+) \boldsymbol{\phi}(0, \mathbf{q}) + \sum_{k=1}^{l-1} (\boldsymbol{\nu}^\top(t_k^+) - \boldsymbol{\nu}^\top(t_k^-)) \mathbf{x}(t_k) \\
&\quad + \int_0^T \boldsymbol{\nu}^\top(t) \tilde{\mathbf{f}}(t|\tau, \mathbf{q}) dt + \int_0^T \dot{\boldsymbol{\nu}}^\top(t) \mathbf{x}(t) dt. \tag{31}
\end{aligned}$$

By (21) and (24), and differentiating (31) with respect to the time-delay  $\tau$ , we obtain

$$\begin{aligned}
&\frac{\partial u_\epsilon(\tau, \mathbf{q}, \boldsymbol{\beta}, \gamma)}{\partial \tau} \\
&= \frac{1}{2} \sum_{k=1}^l \sum_{h=1}^p \left\{ \frac{\partial w(y_h(t_k|\tau, \mathbf{q}))}{\partial y_h(t_k)} + \frac{\partial \psi_\epsilon(v(y_h(t_k|\tau, \mathbf{q})))}{\partial y_h(t_k)} \right\} \frac{\partial \mathbf{g}_h(\mathbf{x}(t_k))}{\partial \mathbf{x}(t_k)} \boldsymbol{\rho}(t_k) \\
&\quad - \boldsymbol{\nu}^\top(t_l^-) \boldsymbol{\rho}(t_l) + \sum_{k=1}^{l-1} (\boldsymbol{\nu}^\top(t_k^+) - \boldsymbol{\nu}^\top(t_k^-)) \boldsymbol{\rho}(t_k) \\
&\quad + \int_0^T \left\{ \boldsymbol{\nu}^\top(t) \frac{\partial \tilde{\mathbf{f}}(t|\tau, \mathbf{q})}{\partial \mathbf{x}} + \dot{\boldsymbol{\nu}}^\top(t) \right\} \boldsymbol{\rho}(t) dt \\
&\quad + \int_0^T \boldsymbol{\nu}^\top(t) \frac{\partial \tilde{\mathbf{f}}(t|\tau, \mathbf{q})}{\partial \tilde{\mathbf{x}}} \boldsymbol{\rho}(t - \tau) \chi_{[\tau, +\infty)}(t) dt \\
&\quad - \int_0^T \boldsymbol{\nu}^\top(t) \frac{\partial \tilde{\mathbf{f}}(t|\tau, \mathbf{q})}{\partial \tilde{\mathbf{x}}} \left\{ \tilde{\mathbf{f}}(t - \tau|\tau, \mathbf{q}) \chi_{[\tau, +\infty)}(t) - \frac{\partial \boldsymbol{\phi}(t - \tau, \mathbf{q})}{\partial t} \chi_{(-\infty, \tau)}(t) \right\} dt \\
&= \sum_{k=1}^{l-1} \left\{ \frac{1}{2} \sum_{h=1}^p \left\{ \frac{\partial w(y_h(t_k|\tau, \mathbf{q}))}{\partial y_h(t_k)} + \frac{\partial \psi_\epsilon(v(y_h(t_k|\tau, \mathbf{q})))}{\partial y_h(t_k)} \right\} \frac{\partial \mathbf{g}_h(\mathbf{x}(t_k))}{\partial \mathbf{x}(t_k)} \right. \\
&\quad \left. + \boldsymbol{\nu}^\top(t_k^+) - \boldsymbol{\nu}^\top(t_k^-) \right\} \boldsymbol{\rho}(t_k) \\
&\quad + \left\{ \frac{1}{2} \sum_{h=1}^p \left\{ \frac{\partial w(y_h(t_l|\tau, \mathbf{q}))}{\partial y_h(t_l)} + \frac{\partial \psi_\epsilon(v(y_h(t_l|\tau, \mathbf{q})))}{\partial y_h(t_l)} \right\} \frac{\partial \mathbf{g}_h(\mathbf{x}(t_l))}{\partial \mathbf{x}(t_l)} - \boldsymbol{\nu}^\top(t_l^-) \right\} \boldsymbol{\rho}(t_l) \\
&\quad + \int_0^T \left\{ \boldsymbol{\nu}^\top(t) \frac{\partial \tilde{\mathbf{f}}(t|\tau, \mathbf{q})}{\partial \mathbf{x}} + \dot{\boldsymbol{\nu}}^\top(t) \right\} \boldsymbol{\rho}(t) dt \\
&\quad + \int_0^{T-\tau} \boldsymbol{\nu}^\top(t + \tau) \frac{\partial \tilde{\mathbf{f}}(t + \tau|\tau, \mathbf{q})}{\partial \tilde{\mathbf{x}}} \boldsymbol{\rho}(t) \chi_{[0, +\infty)}(t) dt \\
&\quad - \int_0^T \boldsymbol{\nu}^\top(t) \frac{\partial \tilde{\mathbf{f}}(t|\tau, \mathbf{q})}{\partial \tilde{\mathbf{x}}} \left\{ \tilde{\mathbf{f}}(t - \tau|\tau, \mathbf{q}) \chi_{[\tau, +\infty)}(t) - \frac{\partial \boldsymbol{\phi}(t - \tau, \mathbf{q})}{\partial t} \chi_{(-\infty, \tau)}(t) \right\} dt. \tag{32}
\end{aligned}$$

Since  $\boldsymbol{\nu} : [0, +\infty) \rightarrow \mathbb{R}^n$  is an arbitrary absolutely continuous function, we

choose  $\boldsymbol{\nu}(\cdot) = \boldsymbol{\lambda}(\cdot)$ , where  $\boldsymbol{\lambda}(\cdot)$  satisfies impulsive auxiliary system (27). Then, we obtain (28) and the proof is complete.  $\square$

**Theorem 4.** For each  $(\tau, \mathbf{q}, \boldsymbol{\beta}, \gamma) \in \mathcal{T} \times \mathcal{Q} \times \mathbb{R}^{lp} \times \mathbb{R}$ , the gradients of constraint function  $u_\epsilon(\tau, \mathbf{q}, \boldsymbol{\beta}, \gamma)$  with respect to parameter  $q_i, i = 1, 2, \dots, r$ , are given by

$$\begin{aligned} & \frac{\partial u_\epsilon(\tau, \mathbf{q}, \boldsymbol{\beta}, \gamma)}{\partial q_i} \\ &= \boldsymbol{\nu}^\top(t_0^+) \frac{\partial \phi(0, \mathbf{q})}{\partial q_i} + \int_{-\tau}^0 \boldsymbol{\nu}^\top(t + \tau) \frac{\partial \tilde{\mathbf{f}}(t + \tau | \tau, \mathbf{q})}{\partial \tilde{\mathbf{x}}} \frac{\partial \phi(0, \mathbf{q})}{\partial q_i} dt \\ & \quad + \int_0^T \boldsymbol{\nu}^\top(t) \frac{\partial \tilde{\mathbf{f}}(t | \tau, \mathbf{q})}{\partial q_i} dt. \end{aligned} \quad (33)$$

*Proof.* Let  $\boldsymbol{\nu}(\cdot)$  be as defined in Theorem 3. From (31), we obtain

$$\begin{aligned} u_\epsilon(\tau, \mathbf{q}, \boldsymbol{\beta}, \gamma) &= \frac{1}{2} \sum_{k=1}^l \sum_{h=1}^p \left\{ w(y_h(t_k | \tau, \mathbf{q}), \beta_h^k) + \psi_\epsilon(v(y_h(t_k | \tau, \mathbf{q}), \beta_h^k)) \right\} - \gamma \\ & \quad - \boldsymbol{\nu}^\top(t_l^-) \mathbf{x}(t_l) + \boldsymbol{\nu}^\top(t_0^+) \phi(0, \mathbf{q}) + \sum_{k=1}^{l-1} (\boldsymbol{\nu}^\top(t_k^+) - \boldsymbol{\nu}^\top(t_k^-)) \mathbf{x}(t_k) \\ & \quad + \int_0^T \boldsymbol{\nu}^\top(t) \tilde{\mathbf{f}}(t | \tau, \mathbf{q}) dt + \int_0^T \dot{\boldsymbol{\nu}}^\top(t) \mathbf{x}(t) dt. \end{aligned} \quad (34)$$

By (22) and (25), it follows from differentiating (34) with respect to the system parameter  $q_i$  that

$$\begin{aligned} & \frac{\partial u_\epsilon(\tau, \mathbf{q}, \boldsymbol{\beta}, \gamma)}{\partial q_i} \\ &= \frac{1}{2} \sum_{k=1}^l \sum_{h=1}^p \left\{ \frac{\partial w(y_h(t_k | \tau, \mathbf{q}))}{\partial y_h(t_k)} + \frac{\partial \psi_\epsilon(v(y_h(t_k | \tau, \mathbf{q})))}{\partial y_h(t_k)} \right\} \frac{\partial \mathbf{g}_h(\mathbf{x}(t_k))}{\partial \mathbf{x}(t_k)} \boldsymbol{\sigma}_i(t_k) \\ & \quad - \boldsymbol{\nu}^\top(t_l^-) \boldsymbol{\sigma}_i(t_l) + \boldsymbol{\nu}^\top(t_0^+) \boldsymbol{\sigma}_i(0) + \sum_{k=1}^{l-1} (\boldsymbol{\nu}^\top(t_k^+) - \boldsymbol{\nu}^\top(t_k^-)) \boldsymbol{\sigma}_i(t_k) \\ & \quad + \int_0^T \left\{ \boldsymbol{\nu}^\top(t) \frac{\partial \tilde{\mathbf{f}}(t | \tau, \mathbf{q})}{\partial \mathbf{x}} + \dot{\boldsymbol{\nu}}^\top(t) \right\} \boldsymbol{\sigma}_i(t) dt + \int_0^T \boldsymbol{\nu}^\top(t) \frac{\partial \tilde{\mathbf{f}}(t | \tau, \mathbf{q})}{\partial \tilde{\mathbf{x}}} \boldsymbol{\sigma}_i(t - \tau) dt \\ & \quad + \int_0^T \boldsymbol{\nu}^\top(t) \frac{\partial \tilde{\mathbf{f}}(t | \tau, \mathbf{q})}{\partial q_i} dt \end{aligned}$$

$$\begin{aligned}
&= \sum_{k=1}^{l-1} \left\{ \frac{1}{2} \sum_{h=1}^p \left\{ \frac{\partial w(y_h(t_k|\tau, \mathbf{q}))}{\partial y_h(t_k)} + \frac{\partial \psi_\epsilon(v((y_h(t_k|\tau, \mathbf{q}))))}{\partial y_h(t_k)} \right\} \frac{\partial \mathbf{g}_h(\mathbf{x}(t_k))}{\partial \mathbf{x}(t_k)} \right. \\
&\quad \left. + \boldsymbol{\nu}^\top(t_k^+) - \boldsymbol{\nu}^\top(t_k^-) \right\} \boldsymbol{\sigma}_i(t_k) + \boldsymbol{\nu}^\top(t_0^+) \frac{\partial \phi(0, \mathbf{q})}{\partial q_i} \\
&\quad + \left\{ \frac{1}{2} \sum_{h=1}^p \left\{ \frac{\partial w(y_h(t_l|\tau, \mathbf{q}))}{\partial y_h(t_l)} + \frac{\partial \psi_\epsilon(v((y_h(t_l|\tau, \mathbf{q}))))}{\partial y_h(t_l)} \right\} \frac{\partial \mathbf{g}_h(\mathbf{x}(t_l))}{\partial \mathbf{x}(t_l)} - \boldsymbol{\nu}^\top(t_l^-) \right\} \boldsymbol{\sigma}_i(t_l) \\
&\quad + \int_0^T \left\{ \boldsymbol{\nu}^\top(t) \frac{\partial \tilde{\mathbf{f}}(t|\tau, \mathbf{q})}{\partial \mathbf{x}} + \dot{\boldsymbol{\nu}}^\top(t) \right\} \boldsymbol{\sigma}_i(t) dt + \int_0^{T-\tau} \boldsymbol{\nu}^\top(t+\tau) \frac{\partial \tilde{\mathbf{f}}(t+\tau|\tau, \mathbf{q})}{\partial \tilde{\mathbf{x}}} \boldsymbol{\sigma}_i(t) dt \\
&\quad + \int_{-\tau}^0 \boldsymbol{\nu}^\top(t+\tau) \frac{\partial \tilde{\mathbf{f}}(t+\tau|\tau, \mathbf{q})}{\partial \tilde{\mathbf{x}}} \frac{\partial \phi(0, \mathbf{q})}{\partial q_i} dt + \int_0^T \boldsymbol{\nu}^\top(t) \frac{\partial \tilde{\mathbf{f}}(t|\tau, \mathbf{q})}{\partial q_i} dt.
\end{aligned}$$

Choose  $\boldsymbol{\nu}(\cdot) = \boldsymbol{\lambda}(\cdot)$ , where  $\boldsymbol{\lambda}(\cdot)$  satisfies impulsive auxiliary system (27). Then, the proof is complete.  $\square$

Note that the gradients of  $u_\epsilon(\tau, \mathbf{q}, \boldsymbol{\beta}, \gamma)$  and  $\tilde{J}(\beta, \gamma)$  with respect to  $\boldsymbol{\beta}$  and  $\gamma$  are easy to obtain.

#### 4.2. Algorithm for solving Problem (C)

Based on Theorems 3 and 4 and the gradients of  $u_\epsilon(\tau, \mathbf{q}, \boldsymbol{\beta}, \gamma)$  and  $\tilde{J}(\beta, \gamma)$  with respect to  $\boldsymbol{\beta}$  and  $\gamma$ , we propose the following Algorithm A for computing the gradients of  $u_\epsilon(\tau, \mathbf{q}, \boldsymbol{\beta}, \gamma)$  and  $\tilde{J}(\beta, \gamma)$  with respect to  $\tau$  and  $q_i, i = 1, 2, \dots, r$ ,  $\boldsymbol{\beta}$  and  $\gamma$ .

##### Algorithm A

**Step 1.** Solve the time-delay system (1) forward in time to obtain the state  $\mathbf{x}(\cdot|\tau, \mathbf{q})$  and the output  $\mathbf{y}(\cdot|\tau, \mathbf{q})$ .

**Step 2.** Solve the impulsive auxiliary system (27) backward in time to obtain  $\boldsymbol{\lambda}(\cdot|\tau, \mathbf{q})$ .

**Step 3.** Compute the gradients  $\frac{\partial u_\epsilon(\tau, \mathbf{q}, \boldsymbol{\beta}, \gamma)}{\partial \tau}$  and  $\frac{\partial u_\epsilon(\tau, \mathbf{q}, \boldsymbol{\beta}, \gamma)}{\partial q_i}, i = 1, 2, \dots, r$ .

**Step 4.** Compute the gradients of  $u_\epsilon(\tau, \mathbf{q}, \boldsymbol{\beta}, \gamma)$  and  $\tilde{J}(\beta, \gamma)$  with respect to  $\boldsymbol{\beta}$  and  $\gamma$ .

We now develop a solution algorithm incorporating Algorithm A to solve Problem (C) as follows.

##### Algorithm B

**Step 1.** Set initial smoothing parameter  $\epsilon_0 (\epsilon_0 > 0)$ , parameter for checking the convergence criterion  $\varepsilon_0 (\varepsilon_0 > 0)$ , decreasing coefficient  $\delta (0 < \delta < 1)$ , initial pair  $(\tau_{\epsilon_0}^0, \mathbf{q}_{\epsilon_0}^0, \boldsymbol{\beta}_{\epsilon_0}^0, \gamma_{\epsilon_0}^0) \in \mathcal{T} \times \mathcal{Q} \times \mathbb{R}^{lp} \times \mathbb{R}$  and the iteration number  $s = 0$ .

**Step 2.** Compute the cost  $\tilde{J}(\boldsymbol{\beta}_{\epsilon_s}^s, \gamma_{\epsilon_s}^s)$  and compute the gradients of  $\tilde{J}(\boldsymbol{\beta}_{\epsilon_s}^s, \gamma_{\epsilon_s}^s)$  and  $u_{\epsilon_s}(\tau_{\epsilon_s}^s, \mathbf{q}_{\epsilon_s}^s, \boldsymbol{\beta}_{\epsilon_s}^s, \gamma_{\epsilon_s}^s)$  with respect to  $\tau_{\epsilon_s}^s, \mathbf{q}_{\epsilon_s}^s, \boldsymbol{\beta}_{\epsilon_s}^s$  and  $\gamma_{\epsilon_s}^s$  using Algorithm A.

**Step 3.** Solve Problem (C) using gradient-based optimization (e.g., sequential quadratic programming [21]) to obtain a new pair  $(\tau_{\epsilon_{s+1}}^{s+1}, \mathbf{q}_{\epsilon_{s+1}}^{s+1}, \boldsymbol{\beta}_{\epsilon_{s+1}}^{s+1}, \gamma_{\epsilon_{s+1}}^{s+1})$ . If  $|\tilde{J}(\boldsymbol{\beta}_{\epsilon_{s+1}}^{s+1}, \gamma_{\epsilon_{s+1}}^{s+1}) - \tilde{J}(\boldsymbol{\beta}_{\epsilon_s}^s, \gamma_{\epsilon_s}^s)| \leq \varepsilon_0$ , then stop and  $(\tau_{\epsilon_{s+1}}^{s+1}, \mathbf{q}_{\epsilon_{s+1}}^{s+1}, \boldsymbol{\beta}_{\epsilon_{s+1}}^{s+1}, \gamma_{\epsilon_{s+1}}^{s+1})$  is an approximate optimal solution of Problem (B). Otherwise, go to Step 4.

**Step 4.** Set  $s = s + 1, \epsilon_{s+1} = \delta \epsilon_s$ , then go to Step 2.

**Remark 1.** *Theorems 1 and 2 guarantee the convergence of the proposed Algorithm B.*

## 5. Numerical example

In this section, we consider a batch fermentation process of converting glycerol to 1,3-propanediol (1,3-PD) by *Klebsiella pneumoniae* [14, 22]. In batch fermentation, a quantity of glycerol and biomass is put in the reactor and stirred uniformly. During the fermentation process, the concentration of glycerol decreases gradually and tends to zero finally. Although the uptake of nutrient by cells is an essential instantaneous process, cells have to undergo growth process before they produce products. Thus, a time-delay should be taken into account in modelling of the batch process.

Based on previous work [14, 22], the batch process can be described by the following delay-differential equations:

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \dot{x}_3(t) \\ \dot{x}_4(t) \\ \dot{x}_5(t) \end{bmatrix} = \begin{bmatrix} q_1(\mathbf{x}(t))x_1(t-\tau) \\ -q_2(\mathbf{x}(t))x_1(t-\tau) \\ q_3(\mathbf{x}(t))x_1(t-\tau) \\ q_4(\mathbf{x}(t))x_1(t-\tau) \\ q_5(\mathbf{x}(t))x_1(t-\tau) \end{bmatrix}, \quad t \in (0, T], \quad (35)$$

where  $x_1(t), x_2(t), x_3(t), x_4(t)$  and  $x_5(t)$  are the concentrations of biomass, glycerol, 1,3-PD, acetate and ethanol in the reactor at time  $t$ , respectively;  $\mathbf{x}(t) = (x_1(t), x_2(t), x_3(t), x_4(t), x_5(t))^\top$ ;  $\tau$  is an unknown time-delay; and  $T = 6.0\text{h}$  is the terminal time of batch culture. Furthermore, the functions  $q_1(\mathbf{x}(t)), q_2(\mathbf{x}(t)), q_3(\mathbf{x}(t)), q_4(\mathbf{x}(t))$  and  $q_5(\mathbf{x}(t))$  are given as follows:

$$q_1(\mathbf{x}(t)) = \frac{\mu_m x_2(t)}{k_1 + x_2(t)} \prod_{\ell=2}^5 \left(1 - \frac{x_\ell(t)}{x_\ell^*}\right), \quad (36)$$

$$q_\ell(\mathbf{x}(t)) = m_\ell + q_1(\mathbf{x}(t))Y_\ell, \quad \ell = 2, 3, 4, 5. \quad (37)$$

In (36),  $x_\ell^*, \ell = 2, 3, 4, 5$ , are the critical concentrations for cells growth, where  $x_2^* = 2039 \text{ mmolL}^{-1}$ ,  $x_3^* = 939.5 \text{ mmolL}^{-1}$ ,  $x_4^* = 1026 \text{ mmolL}^{-1}$  and  $x_5^* = 360.9 \text{ mmolL}^{-1}$ . In the batch fermentation, at sample time  $t_k$ , there are some measured data denoted by  $\tilde{z}^k, k = 1, 2, \dots, 7$ . Due to measurement errors, in this paper, we assume that the measured data  $\tilde{z}^k$  are stochastic, whose expectations are the experimental data  $z^k$  given in [14]. Let  $\tilde{\mathbf{z}} = ((\tilde{z}^1)^\top, \dots, (\tilde{z}^7)^\top)^\top = (\tilde{z}_1^1, \dots, \tilde{z}_5^1, \dots, \tilde{z}_1^7, \dots, \tilde{z}_5^7)^\top \in \mathbb{R}^{35}$ . The lower and upper bounds of  $\tilde{z}^k$  are defined by  $a_h^k = z_h^k - \delta_0 z_h^k$  and  $b_h^k = z_h^k + \delta_0 z_h^k$ , respectively, where  $\delta_0 > 0$  is a given parameter. The support set and the ambiguity set are as defined in (4) and (5) with  $\boldsymbol{\mu}$  being the experimental data in [14].

Then the distributionally robust identification model of the batch culture is: given the system (35), choose  $(\tau, \mathbf{q}) \in \mathcal{T} \times \mathcal{Q}$  such that

$$\max_{\mathbb{P} \in \mathcal{F}} \mathbb{E}_{\mathbb{P}}[J(\tau, \mathbf{q}, \tilde{\mathbf{z}})] = \sum_{k=1}^l \|\mathbf{x}(t_k | \tau, \mathbf{q}) - \tilde{z}^k\|^2 \quad (38)$$

Table 1: The lower and upper bounds of kinetic parameters and optimal kinetic parameters.

	$\mu_m$	$k_1$	$m_2$	$m_3$	$m_4$	$m_5$	$Y_2$	$Y_3$	$Y_4$	$Y_5$
$\mathbf{c}$ [23]	0.1	0.01	0.01	-10	-10	-10	0.001	1	1	1
$\mathbf{d}$ [23]	1	100	10	10	10	10	10000	100	100	100
$\mathbf{q}^*$	0.95	77.54	9.30	2.00	-2.93	1.02	123.46	94.17	34.85	11.95

is minimized, where  $\mathbf{q} := (\mu_m, k_1, m_2, m_3, m_4, m_5, Y_2, Y_3, Y_4, Y_5)^\top$ ,  $\mathcal{T} := [0, 1]$ ,  $\mathcal{Q} := \prod_{\ell=1}^{10} [c_\ell, d_\ell]$  ( $c_\ell$  and  $d_\ell$  are listed in Table 1) and  $\mathbf{x}(\cdot|\tau, \mathbf{q})$  is the solution of the system (35) for  $(\tau, \mathbf{q}) \in \mathcal{T} \times \mathcal{Q}$ . In particular, the history function is obtained by applying cubic spline interpolation [24] experimental data before point zero. Furthermore, in Algorithm B, the initial smoothing parameter  $\epsilon_0$ , parameter for checking the convergence criterion  $\epsilon_0$ , decreasing coefficient  $\delta$ , the parameter  $\delta_0$  and initial pair  $(\tau_{\epsilon_0}^0, \mathbf{q}_{\epsilon_0}^0)$  are  $10^{-3}$ ,  $10^{-6}$ , 0.5, 0.05, (0.05, 1, 74.58, 9.68, 2.08, -2.44, 1.05, 0, 95.06, 33.07, 11.6), respectively. Moreover, the initial dual vector  $(\beta_{\epsilon_0}^0, \gamma_{\epsilon_0}^0)$  was generated randomly.

We wrote a Matlab 7.10.0 (The Mathworks Inc.) program to implement Algorithms A and B in Section 4, where the intrinsic delay-differential equations (DDEs) solver DDE23 which solved DDEs with explicit Runge-Kutta triples [25] is used to solve state and auxiliary state systems (1) and (27). By running our program, we obtain that  $\tau^* = 0.278\text{h}$ , and  $\mathbf{q}^*$  is also listed in Table 1. Under the obtained optimal time-delay  $\tau^*$  and parameters  $\mathbf{q}^*$ , the concentration changes in biomass, glycerol, 1,3-PD, acetate and ethanol with respect to fermentation time are shown in Figure 1. From Figure 1, we can see that the simulation results reasonably fit the experimental data with the worst-case probability distribution. For comparison, we also plot the concentration changes in biomass, glycerol, 1,3-PD, acetate and ethanol with respect to fermentation time using the obtained optimal time-delay and parameters [14] in Figure 2. Note that, unlike parameter identification problem in [14], the distributions of experimental data in our parameter identification problem are assumed to be with only first-order moment information.

## 6. Conclusions

This paper studies the distributionally robust parameter identification of a time-delay system with stochastic measurements. This problem arises in the modelling of batch culture of 1,3-PD. We first proposed the distributionally robust identification model involving a time-delay system. Then by problem transformation and smooth techniques, we constructed a gradient-based computational method to solve the distributionally robust parameter identification problem. Finally, we verified the effectiveness of the proposed approach via an example of 1,3-PD batch production. Note that, in this paper, only one time-delay is involved in the system dynamics. However, the results in this paper can

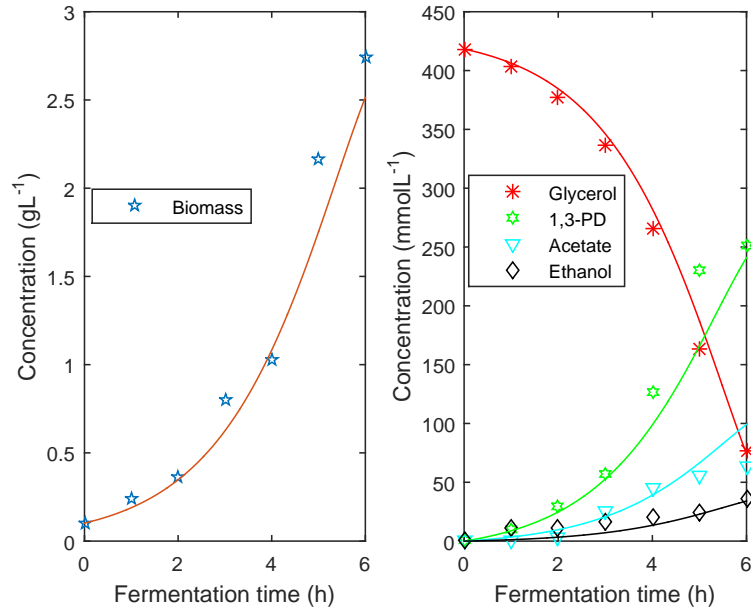


Figure 1: Concentration changes of biomass, glycerol, 1,3-PD, acetate and ethanol with respect to fermentation time in this paper.

be easily extended to system with multiple time-delays. In addition, we assume that we only know the expectation information of the output measurement. If the information of higher moments is known, then Monte Carlo sampling [26] method may be an interesting tool to deal with such problems. We will study this situation in our future research work.

### Acknowledgements

This work was supported by the Natural Science Foundation of China (No. 11771008), the Natural Science Foundation of Shandong Province, China (Nos. ZR2017MA005 and ZR2015AL010), and the Australian Research Council (Nos. DP140100289 and DP160102819).

### References

- [1] H. Scarf, K. Arrow, S. Karlin, A min-max solution of an inventory problem, *Studies in the Mathematical Theory of Inventory and Production*, 1958, 10: 201–209.
- [2] I. Popescu, Robust mean-covariance solutions for stochastic optimization, *Operations Research*, 2007, 55: 98–112.



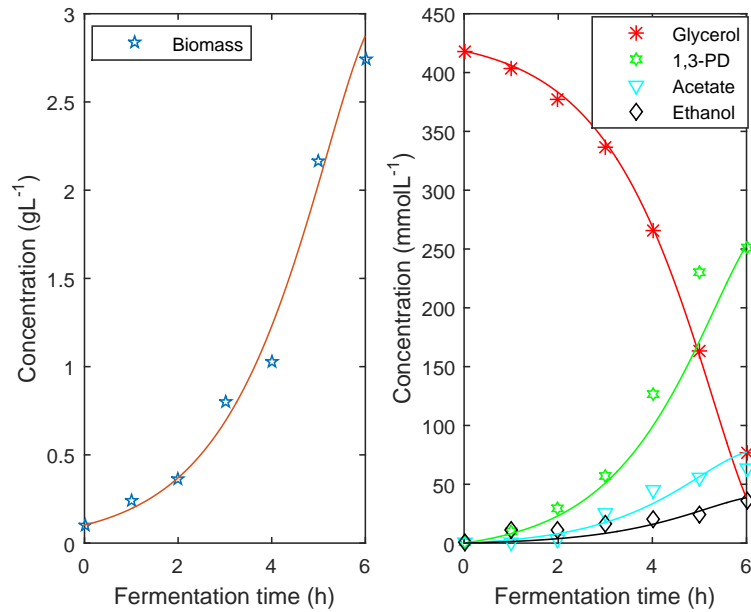


Figure 2: Concentration changes of biomass, glycerol, 1,3-PD, acetate and ethanol with respect to fermentation time in [14].

- [3] E. Delage, Y. Ye, Distributionally robust optimization under moment uncertainty with application to data-driven problems, *Operations Research*, 2010, 58: 596–612.
- [4] A. Shapiro, On duality theory of conic linear problems, *Semi-Infinite Programming: Recent Advances* (Goberna M. A., Lopez M. A. Eds.), Kluwer Academic Publishers, Dordrecht, 2001, 135–165.
- [5] W. Wiesemann, D. Kuhn, M. Sim, Distributionally robust convex optimization, *Operations Research*, 2014, 62: 1358–1376.
- [6] Y. Liu, H. Xu, Entropic approximation for mathematical programs with robust equilibrium constraints, *SIAM Journal of Optimization*, 2014, 24: 933–958.
- [7] H. Sun, H. Xu, Convergence analysis for distributionally robust optimization and equilibrium problems, *Mathematics of Operations Research*, 2016, 41: 377–401.
- [8] B. Van Parys, D. Kuhn, P. Goulart, M. Morari, Distributionally robust control of constrained stochastic systems, *IEEE Transactions on Automatic Control*, 2016, 61: 430–442.

- [9] B. Li, Y. Rong, J. Sun, K. L. Teo, A distributionally robust linear receiver design for multi-access space-time block coded MIMO systems, *IEEE Transactions on Wireless Communications*, 2017, 16: 464–474.
- [10] L. Belkoura, J. P. Richard, M. Fliess, Parameters estimation of systems with delayed and structured entries, *Automatica*, 2009, 45: 1117–1125.
- [11] G. Zheng, J. P. Barbot, D. Boutat, Identification of the delay parameter for nonlinear time-delay systems with unknown inputs, *Automatica*, 2013, 49: 1755–1760.
- [12] Q. Chai, R. Loxton, K. L. Teo, C. Yang, A unified parameter identification method for nonlinear time-delay systems, *Journal of Industrial and Management Optimization*, 2013, 9: 471–486.
- [13] Q. Chai, R. Loxton, K. L. Teo, C. Yang, Time-delay estimation for nonlinear systems with piecewise-constant input, *Applied Mathematics and Computation*, 2013, 219: 9543–9560.
- [14] C. Liu, Modelling and parameter identification for a nonlinear time-delay system in microbial batch fermentation, *Applied Mathematical Modelling*, 2013, 37: 6899–6908.
- [15] Q. Lin, R. Loxton, C. Xu, K. L. Teo, Parameter estimation for nonlinear time-delay systems with noisy output measurements, *Automatica*, 2015, 60: 48–56.
- [16] C. Liu, Z. Gong, K. L. Teo, Robust parameter estimation for nonlinear multistage time-delay systems with noisy measurement data, *Applied Mathematical Modelling*, 2018, 53: 353–368.
- [17] W. Liu, C. Wang, A smoothing Levenberg-Marquardt method for generalized semi-infinite programming, *Computational and Applied Mathematics*, 2013, 32: 89–105.
- [18] E. J. Anderson, P. Nash, *Linear Programming in Infinite-Dimensional Spaces: Theory and Applications*, Wiley, Chichester, United Kingdom, 1987.
- [19] X. Wu, K. Zhang, Three-dimensional trajectory design for horizontal well based on optimal switching algorithms, *ISA Transactions*, 2015, 58: 348–356.
- [20] K. L. Teo, C. J. Goh, K. H. Wong, *A Unified Computational Approach to Optimal Control Problems*, Longman Scientific & Technical, Essex, 1991.
- [21] J. Nocedal, S. J. Wright, *Numerical Optimization*, Springer, New York, 2006.
- [22] Y. Yu, Optimal control of a nonlinear time-delay system in batch fermentation process, *Mathematical Problems in Engineering*, 2014, 2014: 1–7.

- [23] C. Gao, Z. Wang, E. Feng, Z. Xiu, Parameter identification and optimization of process for bio-dissimilation of glycerol to 1,3- propanediol in batch culture, *Journal of Dalian University of Technology*, 2006, 46: 771–774.
- [24] J. Stoer, R. Bulirsch, *Introduction to Numerical Analysis*, Springer-Verlag, New York, 1980
- [25] L. F. Shampine, S. Thompson, Solving DDEs in Matlab, *Applied Numerical Mathematics*, 2001, 37: 441–458.
- [26] A. Shapiro, Monte Carlo sampling methods, *Handbooks in Operations Research and Management Science*, 2003, 10: 353–425.