

(will be inserted by the editor)

Optimal parameter selection for nonlinear multistage systems with time-delays

Chongyang Liu · Ryan Loxton · Kok Lay Teo

Received: date / Accepted: date

Abstract In this paper, we consider a novel dynamic optimization problem for nonlinear multistage systems with time-delays. Such systems evolve over multiple stages, with the dynamics in each stage depending on both the current state of the system and the state at delayed times. The optimization problem involves choosing the values of the time-delays, as well as the values of additional parameters that influence the system dynamics, to minimize a given cost functional. We first show that the partial derivatives of the system state with respect to the time-delays and system parameters can be computed by solving a set of auxiliary dynamic systems in conjunction with the governing multistage system. On this basis, a gradient-based optimization algorithm is proposed to determine the optimal values of the delays and system parameters. Finally, two example problems, one of which involves parameter identification for a realistic fed-batch fermentation process, are solved to demonstrate the algorithm's effectiveness.

Keywords Multistage system · Time-delay system · Nonlinear optimization · Parameter identification

1 Introduction

A *multistage system* refers to a dynamic system that evolves over multiple stages or subsystems [1]. If any of the stages in the multistage system involves time-delays, then the system is known as a *delayed multistage system*. Both conventional and delayed multistage systems arise in many real-world applications such as fermentation processes [2], population models [3], manufacturing systems [4] and image tubes [5].

An *optimal parameter selection problem* is a type of optimal control problem in which a finite set of parameters need to be chosen optimally subject to a given dynamic system. Many common dynamic optimization problems can be formulated as optimal parameter selection problems, such as parameter identification problems [6–8] and controller parameter design problems [9,10]. There are many useful theories and algorithms for solving optimal parameter selection problems for nonlinear systems with

C. Liu (✉)

School of Mathematics and Information Science, Shandong Institute of Business and Technology, Yantai, China

E-mail: liu_chongyang@yahoo.com

R. Loxton (✉)

Department of Mathematics and Statistics, Curtin University, Perth, Australia

Institute of Cyber-Systems and Control, Zhejiang University, Hangzhou, China

E-mail: r.loxton@curtin.edu.au

K.L. Teo

Department of Mathematics and Statistics, Curtin University, Perth, Australia

E-mail: k.l.teo@curtin.edu.au

a single stage and no time-delays; see, for example, [11–13]. However, nonlinear multistage systems, which have both discrete and continuous characteristics [14], are significantly more difficult to handle. Some theoretical issues for a class of dynamic optimization problems involving multistage systems were investigated in [15]. Nevertheless, an implementable algorithm is not described in [15]. An algorithm for solving a class of multistage dynamic optimization problems without path constraints was introduced in [16], and then extended in [17] to handle path constraints on the state variables. Recently, studies of the multistage parameter identification problem for fermentation processes have appeared in [18,19]. Although the aforementioned results are certainly valid and interesting, they ignore the effect of time-delays in the nonlinear multistage system.

Time-delays are frequently encountered in real-life systems. The existence of time-delays often causes undesirable system transient response, or even instability [20]. The estimation of unknown time-delays and possibly other unknown system parameters from a given set of experimental data is one of the key problems in the study of time-delay systems [21]. For estimation problems involving linear systems with a single input delay and no other unknown parameters, many computational methods, such as the exact least squares algorithm [22], the steepest descent algorithm [23] and genetic algorithms [24], have been developed. In contrast, estimation problems for nonlinear systems with multiple delays have yet to receive significant attention in the literature. Motivated by examples from aerodynamics and biochemistry, such problems were studied in [25]. More recently, a new algorithm for estimating unknown time-delays in a nonlinear dynamic system was developed in [26]. This algorithm was extended in [27,28] to cater for more general nonlinear systems with unknown system parameters in addition to unknown delays. However, these parameter estimation algorithms are only designed for delay systems with a single stage. The parameter estimation problem for delayed multistage systems has yet to be tackled in the literature.

In this paper, we consider a general nonlinear multistage system with time-delays and system parameters. Each stage in this system is described by a set of nonlinear delay-differential equations in which the time-delays and system parameters are control variables. The dynamic optimization problem involves choosing values for the time-delays and system parameters to minimize a given cost functional. This cost functional depends on the state values at a finite number of *characteristic time points* [12]. As we will show, this problem class incorporates the parameter estimation problem as a special case. Our main contribution in this paper is the derivation of the partial derivatives of the system state with respect to the time-delays and system parameters. This derivation is based on the corresponding approach for single-stage systems pioneered in [26] and extended in [27]. After deriving the gradient of the state, the gradient of the cost functional can be calculated accordingly. Combined with standard numerical optimization techniques [29], such as sequential quadratic programming (SQP), a gradient-based optimization algorithm is then developed to determine the optimal time-delays and system parameters. A major advantage of our new algorithm is that it is applicable to a much larger array of problems than the existing algorithms in [26,27], which are restricted to parameter identification problems. Finally, two example problems, including a real-world practical example involving fed-batch fermentation, are considered to test the performance of our new algorithm.

2 Problem formulation

Consider the following delayed multistage system with N stages and m time-delays:

$$\begin{cases} \dot{x}(t) = f^i(t, x(t), \tilde{x}(t), \zeta), & t \in (t_{i-1}, t_i), & i = 1, 2, \dots, N, & (1a) \\ x(t_i+) = x(t_i-), & i = 0, 1, \dots, N, & & (1b) \\ x(t) = \phi(t), & t \leq 0, & & (1c) \end{cases}$$

where $x(t) := (x_1(t), x_2(t), \dots, x_n(t))^\top \in R^n$ is the *state vector*; $\tilde{x}(t) := (x(t - \alpha_1)^\top, x(t - \alpha_2)^\top, \dots, x(t - \alpha_m)^\top)^\top \in R^{nm}$ is the *delayed state vector*; $\alpha_j, j = 1, 2, \dots, m$, are *time-delays*; $\zeta := (\zeta_1, \zeta_2, \dots, \zeta_v)^\top \in$

R^v is a vector of *system parameters*; t_i , $i = 1, 2, \dots, N$, are given *switching times*; $x(t_i+)$ is the state immediately after the switching time t_i ; $x(t_i-)$ is the state immediately before the switching time t_i ; and $f^i : R \times R^n \times R^{nm} \times R^v \rightarrow R^n$ and $\phi : R \rightarrow R^n$ are given functions.

We assume that the switching times in (1a) and (1b) are pre-assigned so that

$$0 = t_0 < t_1 < \dots < t_N = T, \quad (2)$$

where T is the terminal time. Thus, the multistage system (1) begins in stage 1 at time $t = 0$, then switches to stage 2 at time $t = t_1$, and so on. We also assume that there are no state jumps at the switching times; see condition (1b).

Now, define

$$\mathcal{D} := \{(\alpha_1, \alpha_2, \dots, \alpha_m)^\top \in R^m : a_j \leq \alpha_j \leq b_j, j = 1, 2, \dots, m\}, \quad (3)$$

where a_j and b_j are given constants such that $0 \leq a_j < b_j$. Clearly, \mathcal{D} is a compact and convex subset of R^m . Any vector $\alpha \in \mathcal{D}$ is called a *feasible time-delay vector*.

Furthermore, define

$$\mathcal{Z} := \{(\zeta_1, \zeta_2, \dots, \zeta_v)^\top \in R^v : c_k \leq \zeta_k \leq d_k, k = 1, 2, \dots, v\}, \quad (4)$$

where c_k and d_k are given constants such that $c_k < d_k$. Obviously, the set \mathcal{Z} is a compact and convex subset of R^v . Any vector $\zeta \in \mathcal{Z}$ is called a *feasible parameter vector*. Accordingly, any pair $(\alpha, \zeta) \in \mathcal{D} \times \mathcal{Z}$ is called a *feasible delay-parameter pair*.

Remark 1 Note that the time-delays and system parameters in (1a) could be different for each subsystem. For example, stage 1 could only involve ζ_1 and $x(t - \alpha_1)$, stage 2 could only involve ζ_2 and $x(t - \alpha_2)$, and so on.

We assume throughout this paper that the following conditions are satisfied.

Assumption 1. *The functions f^i , $i = 1, 2, \dots, N$, are continuously differentiable. Moreover, the function ϕ is twice continuously differentiable.*

Assumption 2. *There exists a positive real number $L_1 > 0$ such that for each $i = 1, 2, \dots, N$,*

$$|f^i(t, x, \tilde{x}, \zeta)| \leq L_1(1 + |x| + |\tilde{x}|), \quad (t, x, \tilde{x}, \zeta) \in [t_{i-1}, t_i] \times R^n \times R^{nm} \times \mathcal{Z},$$

where $|\cdot|$ denotes the Euclidean norm.

Assumptions 1 and 2 ensure that the multistage system (1) admits a unique solution corresponding to each delay-parameter pair $(\alpha, \zeta) \in \mathcal{D} \times \mathcal{Z}$ [30]. We denote this solution by $x(\cdot|\alpha, \zeta)$. Our aim is to select an optimal delay-parameter pair that minimizes the following cost functional:

$$J(\alpha, \zeta) = \Phi(x(\tau_1|\alpha, \zeta), \dots, x(\tau_d|\alpha, \zeta)), \quad (5)$$

where $\Phi : R^{dn} \rightarrow R$ is a given function and τ_l , $l = 1, 2, \dots, d$, are given characteristic time points satisfying $0 < \tau_1 < \tau_2 < \dots < \tau_d < T$.

Now, our problem can be formulated as the following optimal parameter selection problem.

Problem (P). *Find $(\alpha, \zeta) \in \mathcal{D} \times \mathcal{Z}$ such that the cost functional (5) is minimized.*

Problem (P) is a dynamic optimization problem in which the time-delays in the multistage system (1) are decision variables to be optimized. Conventional methods for optimizing time-delay systems—see, for example, [8, 21, 31, 32]—are only applicable when the delays are fixed and known, not decision variables. Therefore, such methods cannot be used to solve Problem (P). Although several promising techniques for optimizing time-delays have recently been developed (see [26–28, 33]), these techniques are based on the assumption that the system dynamics do not change—an assumption that is clearly invalid for multistage systems, which operate by cycling through a series of different subsystems. Moreover, the aforementioned techniques are only applicable to time-delay estimation problems, which are a special case of Problem (P). As a result, a new method must be developed to solve Problem (P).

3 Solving Problem (P)

In essence, Problem (P) is a nonlinear programming problem with decision vectors α and ζ . It is well known that gradient-based optimization methods are very effective at solving nonlinear programming problems [29]. However, such methods require the gradient of the cost function (and the gradients of the constraints, if applicable). But since the cost functional in Problem (P) is an implicit (rather than explicit) function of the decision vectors α and ζ , it is not obvious how to determine its gradient. In this section, we will develop a variational method (also called sensitivity method) to determine the gradient of the cost functional in Problem (P). The key idea is to introduce a set of auxiliary systems and then express the gradient of the cost functional in terms of the solution of these auxiliary systems. On this basis, a numerical optimization method can be developed for solving Problem (P).

The solution of system (1) is usually viewed as a function of time, with α and ζ fixed vectors. We denote this function by $x(\cdot|\alpha, \zeta)$. In this section, we will also consider the function $x(t|\cdot, \cdot) : \mathcal{D} \times \mathcal{Z} \rightarrow R^n$, in which t is fixed and α and ζ are the variables. Thus, $x(t|\cdot, \cdot)$ returns the state value at time t for the given delay-parameter pair.

3.1 Gradient computation: Partial derivatives with respect to time-delays

We assume throughout this subsection that $p \in \{1, 2, \dots, m\}$ and $(\alpha, \zeta) \in \mathcal{D} \times \mathcal{Z}$ are arbitrary but fixed. For simplicity, we write $x(t)$ instead of $x(t|\alpha, \zeta)$, and $x^\epsilon(t)$ instead of $x(t|\alpha + \epsilon e^p, \zeta)$, where e^p denotes the p th unit basis vector in R^m .

Define

$$\Xi := [a_p - \alpha_p, b_p - \alpha_p].$$

Note that $0 \in \Xi$ and thus $\Xi \neq \emptyset$. Clearly,

$$\epsilon \in \Xi \iff \alpha + \epsilon e^p \in \mathcal{D}.$$

For each $\epsilon \in \Xi$, define the following functions:

$$\varphi^\epsilon(t) := x^\epsilon(t) - x(t), \quad t \leq T, \quad (6)$$

$$\theta^{\epsilon,j}(t) := x^\epsilon(t - \alpha_j - \epsilon \delta_{pj}) - x(t - \alpha_j), \quad t \leq T, \quad j = 1, 2, \dots, m, \quad (7)$$

where δ_{pj} denotes the Kronecker delta function. Furthermore, define

$$\chi^\epsilon(t) := \begin{cases} \dot{\phi}(t), & \text{if } t \leq 0, \\ f^i(t, x^\epsilon(t), \tilde{x}^\epsilon(t), \zeta), & \text{if } t \in (t_{i-1}, t_i] \text{ for some } i \in \{1, 2, \dots, N\}, \end{cases} \quad (8)$$

where

$$\tilde{x}^\epsilon(t) := (x^\epsilon(t - \alpha_1 - \epsilon \delta_{p1})^\top, \dots, x^\epsilon(t - \alpha_m - \epsilon \delta_{pm})^\top)^\top.$$

In addition, let

$$\theta^\epsilon(t) := ((\theta^{\epsilon,1}(t))^\top, (\theta^{\epsilon,2}(t))^\top, \dots, (\theta^{\epsilon,m}(t))^\top)^\top \in R^{nm}, \quad t \leq T.$$

Obviously,

$$\theta^{\epsilon,j}(t) = \varphi^\epsilon(t - \alpha_j), \quad t \leq T, \quad j \neq p, \quad (9)$$

and

$$\varphi^\epsilon(t) = 0, \quad t \leq 0. \quad (10)$$

Furthermore, for almost all $t \in (-\infty, T]$,

$$\dot{x}^\epsilon(t) = \chi^\epsilon(t). \quad (11)$$

In the next three lemmas, we present some important properties of the functions defined in (6), (7) and (8). The proofs of these lemmas, which rely on Assumptions 1 and 2, are similar to the proofs of the corresponding lemmas in [26].

Lemma 1 *There exists a positive real number $L_2 > 0$ such that for all $\epsilon \in \Xi$,*

$$|x^\epsilon(t)| \leq L_2, \quad |\chi^\epsilon(t)| \leq L_2, \quad t \in [-\bar{b}, T], \quad (12)$$

where $\bar{b} := \max_{j \in \{1, 2, \dots, m\}} \{b_j\}$.

Lemma 2 *There exists a positive real number $L_3 > 0$ such that for all $\epsilon \in \Xi$,*

$$|\varphi^\epsilon(t)| \leq L_3|\epsilon|, \quad \max_{j \in \{1, 2, \dots, m\}} |\theta^{\epsilon, j}(t)| \leq L_3|\epsilon|, \quad t \in [0, T]. \quad (13)$$

Lemma 3 *For almost all $t \in [0, T]$,*

$$\lim_{\epsilon \rightarrow 0} \frac{\theta^{\epsilon, p}(t) - \varphi^\epsilon(t - \alpha_p)}{\epsilon} = -\chi^0(t - \alpha_p). \quad (14)$$

We now present our main result, which gives gradient formulae for the state with respect to the time-delays. Note that in the following, we use $\partial \tilde{x}^j$ to denote partial differentiation with respect to $x(t - \alpha_j)$.

Theorem 1 *Let $t \in (0, T]$ be a fixed time point. Then $x(t|\cdot, \cdot)$ is differentiable with respect to α_p on $\mathcal{D} \times \mathcal{Z}$. Furthermore, for each $(\alpha, \zeta) \in \mathcal{D} \times \mathcal{Z}$,*

$$\frac{\partial x(t|\alpha, \zeta)}{\partial \alpha_p} = \Lambda_p(t|\alpha, \zeta), \quad p = 1, 2, \dots, m, \quad (15)$$

where $\Lambda_p(\cdot)$ is the solution of the auxiliary multistage system

$$\begin{aligned} \dot{\Lambda}_p(s) &= \frac{\partial f^i(s, x(s), \tilde{x}(s), \zeta)}{\partial x} \Lambda_p(s) + \sum_{j=1}^m \frac{\partial f^i(s, x(s), \tilde{x}(s), \zeta)}{\partial \tilde{x}^j} \Lambda_p(s - \alpha_j) \\ &\quad - \frac{\partial f^i(s, x(s), \tilde{x}(s), \zeta)}{\partial \tilde{x}^p} \chi^0(s - \alpha_p), \quad s \in (t_{i-1}, t_i), \quad i = 1, 2, \dots, N, \end{aligned} \quad (16)$$

with the intermediate conditions

$$\Lambda_p(t_i+) = \Lambda_p(t_i-), \quad i = 0, 1, \dots, N, \quad (17)$$

and the initial condition

$$\Lambda_p(s) = 0, \quad s \leq 0. \quad (18)$$

Proof Let $p \in \{1, 2, \dots, m\}$ and $(\alpha, \zeta) \in \mathcal{D} \times \mathcal{Z}$ be arbitrary but fixed. For each $i = 1, 2, \dots, N$ and $\epsilon \in \Xi$, define the following functions:

$$\bar{f}^{i, \epsilon}(s, \eta) := f^i(s, x(s) + \eta \varphi^\epsilon(s), \tilde{x}(s) + \eta \theta^\epsilon(s), \zeta), \quad (s, \eta) \in [t_{i-1}, t_i] \times [0, 1], \quad (19)$$

$$\Delta_1^{i, \epsilon}(s) := \int_0^1 \left\{ \frac{\partial \bar{f}^{i, \epsilon}(s, \eta)}{\partial x} - \frac{\partial \bar{f}^{i, \epsilon}(s, 0)}{\partial x} \right\} \varphi^\epsilon(s) d\eta, \quad s \in [t_{i-1}, t_i], \quad (20)$$

and

$$\Delta_2^{i, \epsilon}(s) := \sum_{j=1}^m \int_0^1 \left\{ \frac{\partial \bar{f}^{i, \epsilon}(s, \eta)}{\partial \tilde{x}^j} - \frac{\partial \bar{f}^{i, \epsilon}(s, 0)}{\partial \tilde{x}^j} \right\} \theta^{\epsilon, j}(s) d\eta, \quad s \in [t_{i-1}, t_i]. \quad (21)$$

By Assumption 1 and Lemma 1, there exist constants $M_1 > 0$ and $M_2 > 0$ such that for all $\epsilon \in \Xi$,

$$\left| \frac{\partial \bar{f}^{i, \epsilon}(s, 0)}{\partial x} \right| \leq M_1, \quad s \in [t_{i-1}, t_i], \quad i = 1, 2, \dots, N,$$

and

$$\left| \frac{\partial \bar{f}^{i,\epsilon}(s,0)}{\partial \tilde{x}^j} \right| \leq M_2, \quad s \in [t_{i-1}, t_i], \quad i = 1, 2, \dots, N, \quad j = 1, 2, \dots, m,$$

where $|\cdot|$ denotes the natural matrix norm on $R^{n \times n}$. Moreover, by Lemma 2, the following limits exist uniformly with respect to $\eta \in [0, 1]$ and $s \in [0, T]$:

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \{x(s) + \eta \varphi^\epsilon(s)\} &= x(s), \\ \lim_{\epsilon \rightarrow 0} \{\tilde{x}(s) + \eta \theta^\epsilon(s)\} &= \tilde{x}(s). \end{aligned}$$

Thus, it follows from Assumption 1 that for each $\delta > 0$, there exists a corresponding $\epsilon' > 0$ such that for all $\epsilon \in \Xi$ satisfying $|\epsilon| < \epsilon'$,

$$\left| \frac{\partial \bar{f}^{i,\epsilon}(s,\eta)}{\partial x} - \frac{\partial \bar{f}^{i,\epsilon}(s,0)}{\partial x} \right| < \delta, \quad (s,\eta) \in [t_{i-1}, t_i] \times [0, 1], \quad i = 1, 2, \dots, N, \quad (22)$$

and

$$\left| \frac{\partial \bar{f}^{i,\epsilon}(s,\eta)}{\partial \tilde{x}^j} - \frac{\partial \bar{f}^{i,\epsilon}(s,0)}{\partial \tilde{x}^j} \right| < \delta, \quad (s,\eta) \in [t_{i-1}, t_i] \times [0, 1], \quad i = 1, 2, \dots, N, \quad j = 1, 2, \dots, m. \quad (23)$$

Inequalities (22) and (23), together with Lemma 2, imply that for all $\epsilon \in \Xi$ with $|\epsilon| < \epsilon'$,

$$|\Delta_1^{i,\epsilon}(s)| \leq \delta L_3 |\epsilon|, \quad |\Delta_2^{i,\epsilon}(s)| \leq \delta m L_3 |\epsilon|, \quad i = 1, 2, \dots, N. \quad (24)$$

Now, define another function $\rho : \Xi \setminus \{0\} \rightarrow R$ by

$$\rho(\epsilon) := \int_0^T |\epsilon^{-1} \theta^{\epsilon,p}(s) - \epsilon^{-1} \varphi^\epsilon(s - \alpha_p) + \chi^0(s - \alpha_p)| ds.$$

It follows from Lemmas 1 and 2 and (10) that for all $\epsilon \in \Xi \setminus \{0\}$,

$$|\epsilon^{-1} \theta^{\epsilon,p}(s) - \epsilon^{-1} \varphi^\epsilon(s - \alpha_p) + \chi^0(s - \alpha_p)| \leq 2L_3 + L_2, \quad s \in [0, T].$$

Thus, by Lemma 3 and Lebesgue's Dominated Convergence Theorem [11], we conclude that

$$\lim_{\epsilon \rightarrow 0} \rho(\epsilon) = 0. \quad (25)$$

Now, let $\delta > 0$ be arbitrary but fixed. Then by (25), there exists a corresponding $\epsilon'' > 0$ such that $\rho(\epsilon) < \delta$ whenever $0 < |\epsilon| < \epsilon''$. Let $\epsilon \in \Xi$ be such that

$$0 < |\epsilon| < \min\{\epsilon', \epsilon''\}. \quad (26)$$

By the chain rule, we obtain

$$\frac{\partial \bar{f}^{i,\epsilon}(s,\eta)}{\partial \eta} = \frac{\partial \bar{f}^{i,\epsilon}(s,\eta)}{\partial x} \varphi^\epsilon(s) + \sum_{j=1}^m \frac{\partial \bar{f}^{i,\epsilon}(s,\eta)}{\partial \tilde{x}^j} \theta^{\epsilon,j}(s), \quad s \in [t_{i-1}, t_i], \quad i = 1, 2, \dots, N. \quad (27)$$

Recall that $t \in (0, T]$ is a fixed time point. Then $t \in (t_{\varsigma-1}, t_\varsigma]$ for some fixed integer $\varsigma \in \{1, 2, \dots, N\}$. By the fundamental theorem of calculus,

$$\begin{aligned} \varphi^\epsilon(t) &= x^\epsilon(t) - x(t) \\ &= \sum_{i=1}^{\varsigma} \int_{t_{i-1}}^{\min\{t, t_i\}} \{\bar{f}^{i,\epsilon}(s,1) - \bar{f}^{i,\epsilon}(s,0)\} ds \\ &= \sum_{i=1}^{\varsigma} \int_{t_{i-1}}^{\min\{t, t_i\}} \left(\int_0^1 \frac{\partial \bar{f}^{i,\epsilon}(s,\eta)}{\partial \eta} d\eta \right) ds. \end{aligned}$$

Thus, by (27),

$$\varphi^\epsilon(t) = \sum_{i=1}^{\varsigma} \int_{t_{i-1}}^{\min\{t, t_i\}} \left(\int_0^1 \frac{\partial \bar{f}^{i, \epsilon}(s, \eta)}{\partial x} \varphi^\epsilon(s) d\eta + \sum_{j=1}^m \int_0^1 \frac{\partial \bar{f}^{i, \epsilon}(s, \eta)}{\partial \tilde{x}^j} \theta^{\epsilon, j}(s) d\eta \right) ds. \quad (28)$$

Note that

$$\int_0^1 \frac{\partial \bar{f}^{i, \epsilon}(s, \eta)}{\partial x} \varphi^\epsilon(s) d\eta = \Delta_1^{i, \epsilon}(s) + \frac{\partial \bar{f}^{i, \epsilon}(s, 0)}{\partial x} \varphi^\epsilon(s), \quad (29)$$

and

$$\sum_{j=1}^m \int_0^1 \frac{\partial \bar{f}^{i, \epsilon}(s, \eta)}{\partial \tilde{x}^j} \theta^{\epsilon, j}(s) d\eta = \Delta_2^{i, \epsilon}(s) + \sum_{j=1}^m \frac{\partial \bar{f}^{i, \epsilon}(s, 0)}{\partial \tilde{x}^j} \theta^{\epsilon, j}(s). \quad (30)$$

Combining (9) and (30) yields

$$\begin{aligned} \sum_{j=1}^m \int_0^1 \frac{\partial \bar{f}^{i, \epsilon}(s, \eta)}{\partial \tilde{x}^j} \theta^{\epsilon, j}(s) d\eta &= \Delta_2^{i, \epsilon}(s) + \sum_{j=1}^m \frac{\partial \bar{f}^{i, \epsilon}(s, 0)}{\partial \tilde{x}^j} \varphi^\epsilon(s - \alpha_j) \\ &\quad + \frac{\partial \bar{f}^{i, \epsilon}(s, 0)}{\partial \tilde{x}^p} \theta^{\epsilon, p}(s) - \frac{\partial \bar{f}^{i, \epsilon}(s, 0)}{\partial \tilde{x}^p} \varphi^\epsilon(s - \alpha_p). \end{aligned} \quad (31)$$

Next, integrating the auxiliary system (16) with the intermediate conditions (17) and the initial condition (18) yields

$$A_p(t) = \sum_{i=1}^{\varsigma} \int_{t_{i-1}}^{\min\{t, t_i\}} \left\{ \frac{\partial \bar{f}^{i, \epsilon}(s, 0)}{\partial x} A_p(s) + \sum_{j=1}^m \frac{\partial \bar{f}^{i, \epsilon}(s, 0)}{\partial \tilde{x}^j} A_p(s - \alpha_j) - \frac{\partial \bar{f}^{i, \epsilon}(s, 0)}{\partial \tilde{x}^p} \chi^0(s - \alpha_p) \right\} ds. \quad (32)$$

Multiplying (28) by ϵ^{-1} , subtracting (32), simplifying using (29) and (31), and taking the norm of both sides and then applying (24), we obtain

$$\begin{aligned} |\epsilon^{-1} \varphi^\epsilon(t) - A_p(t)| &\leq (m+1) \delta L_3 T + M_2 \rho(\epsilon) + \int_0^t M_1 |\epsilon^{-1} \varphi^\epsilon(s) - A_p(s)| ds \\ &\quad + \sum_{j=1}^m \int_0^t M_2 |\epsilon^{-1} \varphi^\epsilon(s - \alpha_j) - A_p(s - \alpha_j)| ds. \end{aligned} \quad (33)$$

The last integral term on the right-hand side of (33) can be simplified using (10) and (18) as follows:

$$\begin{aligned} \sum_{j=1}^m \int_0^t M_2 |\epsilon^{-1} \varphi^\epsilon(s - \alpha_j) - A_p(s - \alpha_j)| ds &= \sum_{j=1}^m \int_{-\alpha_j}^{t - \alpha_j} M_2 |\epsilon^{-1} \varphi^\epsilon(s) - A_p(s)| ds \\ &\leq \int_0^t m M_2 |\epsilon^{-1} \varphi^\epsilon(s) - A_p(s)| ds. \end{aligned}$$

Hence, (33) becomes

$$|\epsilon^{-1} \varphi^\epsilon(t) - A_p(t)| \leq (m+1) \delta L_3 T + M_2 \rho(\epsilon) + \int_0^t M_3 |\epsilon^{-1} \varphi^\epsilon(s) - A_p(s)| ds,$$

where $M_3 := M_1 + m M_2$.

Recall that $\rho(\epsilon) < \delta$ whenever $0 < |\epsilon| < \epsilon''$. Thus, by (26),

$$|\epsilon^{-1} \varphi^\epsilon(t) - A_p(t)| \leq (m+1) \delta L_3 T + M_2 \delta + \int_0^t M_3 |\epsilon^{-1} \varphi^\epsilon(s) - A_p(s)| ds.$$

By the Gronwall-Bellman Lemma [11], it follows that

$$|\epsilon^{-1}\varphi^\epsilon(t) - A_p(t)| \leq (m\delta L_3 T + \delta L_3 T + M_2\delta) \exp(M_3 T), \quad (34)$$

whenever $0 < |\epsilon| < \min\{\epsilon', \epsilon''\}$. Since δ is arbitrary, this shows that $\epsilon^{-1}\varphi^\epsilon(t) \rightarrow A_p(t)$ as $\epsilon \rightarrow 0$, as required. \square

3.2 Gradient computation: Partial derivatives with respect to system parameters

In this subsection, the gradients of $x(\cdot|\alpha, \zeta)$ with respect to the system parameters are investigated. Let $q \in \{1, 2, \dots, v\}$ and $(\alpha, \zeta) \in \mathcal{D} \times \mathcal{Z}$ be arbitrary but fixed. Also, for notational simplicity, we write $x(t)$ instead of $x(t|\alpha, \zeta)$, and $x^\epsilon(t)$ instead of $x(t|\alpha, \zeta + \epsilon e^q)$, where e^q denotes the q th unit basis vector in R^v .

Define

$$\Theta := [c_q - \zeta_q, d_q - \zeta_q].$$

Then, $0 \in \Theta$ and

$$\epsilon \in \Theta \iff \zeta + \epsilon e^q \in \Theta.$$

Furthermore, for each $\epsilon \in \Theta$, define

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

and

$$\theta^{\epsilon,j}(t) := x^\epsilon(t - \alpha_j) - x(t - \alpha_j), \quad t \leq T, \quad j = 1, 2, \dots, m.$$

Let

$$\theta^\epsilon(t) := ((\theta^{\epsilon,1}(t))^\top, (\theta^{\epsilon,2}(t))^\top, \dots, (\theta^{\epsilon,m}(t))^\top)^\top \in R^{nm}, \quad t \leq T. \quad (35)$$

Clearly,

$$\theta^{\epsilon,j}(t) = \varphi^\epsilon(t - \alpha_j), \quad t \leq T, \quad j = 1, 2, \dots, m, \quad (36)$$

$$\varphi^\epsilon(t) = 0, \quad t \leq 0. \quad (37)$$

The following lemmas are similar to Lemmas 1 and 2 in Subsection 3.1.

Lemma 4 *There exists a positive real number $L_4 > 0$ such that for all $\epsilon \in \Theta$,*

$$|x^\epsilon(t)| \leq L_4, \quad t \in [-\bar{b}, T]. \quad (38)$$

Lemma 5 *There exists a positive real number $L_5 > 0$ such that for all $\epsilon \in \Theta$,*

$$|\varphi^\epsilon(t)| \leq L_5|\epsilon|, \quad \max_{j \in \{1, 2, \dots, m\}} |\theta^{\epsilon,j}(t)| \leq L_5|\epsilon|, \quad t \in [0, T]. \quad (39)$$

The partial derivatives of the system state with respect to the system parameters are given in the following theorem.

Theorem 2 *Let $t \in (0, T]$ be a fixed time point. Then $x(t|\cdot, \cdot)$ is differentiable with respect to ζ_q on $\mathcal{D} \times \mathcal{Z}$. Moreover, for each $(\alpha, \zeta) \in \mathcal{D} \times \mathcal{Z}$,*

$$\frac{\partial x(t|\alpha, \zeta)}{\partial \zeta_q} = \Gamma_q(t|\alpha, \zeta), \quad q = 1, 2, \dots, v, \quad (40)$$

where $\Gamma_q(\cdot)$ is the solution of the following auxiliary multistage system

$$\begin{aligned} \dot{\Gamma}_q(s) &= \frac{\partial f^i(s, x(s), \tilde{x}(s), \zeta)}{\partial x} \Gamma_q(s) + \sum_{j=1}^m \frac{\partial f^i(s, x(s), \tilde{x}(s), \zeta)}{\partial \tilde{x}^j} \Gamma_q(s - \alpha_j) \\ &\quad + \frac{\partial f^i(s, x(s), \tilde{x}(s), \zeta)}{\partial \zeta_q}, \quad s \in (t_{i-1}, t_i), \quad i = 1, 2, \dots, N, \end{aligned} \quad (41)$$

with the intermediate conditions

$$\Gamma_q(t_i+) = \Gamma_q(t_i-), \quad i = 0, 1, \dots, N, \quad (42)$$

and the initial condition

$$\Gamma_q(s) = 0, \quad s \leq 0. \quad (43)$$

Proof Let $q \in \{1, 2, \dots, v\}$ and $(\alpha, \zeta) \in \mathcal{D} \times \mathcal{Z}$ be arbitrary but fixed. For each $i = 1, 2, \dots, N$ and $\epsilon \in \Theta$, define the following functions:

$$\bar{f}^{i,\epsilon}(s, \eta) := f^i(s, x(s) + \eta\varphi^\epsilon(s), \tilde{x}(s) + \eta\theta^\epsilon(s), \zeta + \eta\epsilon e^q), \quad (s, \eta) \in [t_{i-1}, t_i] \times [0, 1], \quad (44)$$

$$\Delta_1^{i,\epsilon}(s) := \int_0^1 \left\{ \frac{\partial \bar{f}^{i,\epsilon}(s, \eta)}{\partial x} - \frac{\partial \bar{f}^{i,\epsilon}(s, 0)}{\partial x} \right\} \varphi^\epsilon(s) d\eta, \quad s \in [t_{i-1}, t_i], \quad (45)$$

$$\Delta_2^{i,\epsilon}(s) := \sum_{j=1}^m \int_0^1 \left\{ \frac{\partial \bar{f}^{i,\epsilon}(s, \eta)}{\partial \tilde{x}^j} - \frac{\partial \bar{f}^{i,\epsilon}(s, 0)}{\partial \tilde{x}^j} \right\} \theta^{\epsilon,j}(s) d\eta, \quad s \in [t_{i-1}, t_i], \quad (46)$$

and

$$\Delta_3^{i,\epsilon}(s) := \int_0^1 \epsilon \left\{ \frac{\partial \bar{f}^{i,\epsilon}(s, \eta)}{\partial \zeta_q} - \frac{\partial \bar{f}^{i,\epsilon}(s, 0)}{\partial \zeta_q} \right\} d\eta, \quad s \in [t_{i-1}, t_i]. \quad (47)$$

By Assumption 1 and Lemma 4, there exist constants $M_4 > 0$ and $M_5 > 0$ such that

$$\left| \frac{\partial \bar{f}^{i,\epsilon}(s, 0)}{\partial x} \right| \leq M_4, \quad s \in [t_{i-1}, t_i], \quad i = 1, 2, \dots, N,$$

and

$$\left| \frac{\partial \bar{f}^{i,\epsilon}(s, 0)}{\partial \tilde{x}^j} \right| \leq M_5, \quad s \in [t_{i-1}, t_i], \quad i = 1, 2, \dots, N, \quad j = 1, 2, \dots, m,$$

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

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \{x(s) + \eta\varphi^\epsilon(s)\} &= x(s), \\ \lim_{\epsilon \rightarrow 0} \{\tilde{x}(s) + \eta\theta^\epsilon(s)\} &= \tilde{x}(s). \end{aligned}$$

Hence, Assumption 1 implies that for each $\delta > 0$, there exists an $\epsilon' > 0$ such that for all ϵ satisfying $|\epsilon| < \epsilon'$,

$$\left| \frac{\partial \bar{f}^{i,\epsilon}(s, \eta)}{\partial x} - \frac{\partial \bar{f}^{i,\epsilon}(s, 0)}{\partial x} \right| < \delta, \quad (s, \eta) \in [t_{i-1}, t_i] \times [0, 1], \quad i = 1, 2, \dots, N,$$

$$\left| \frac{\partial \bar{f}^{i,\epsilon}(s, \eta)}{\partial \tilde{x}^j} - \frac{\partial \bar{f}^{i,\epsilon}(s, 0)}{\partial \tilde{x}^j} \right| < \delta, \quad (s, \eta) \in [t_{i-1}, t_i] \times [0, 1], \quad i = 1, 2, \dots, N, \quad j = 1, 2, \dots, m,$$

and

$$\left| \frac{\partial \bar{f}^{i,\epsilon}(s,\eta)}{\partial \zeta_q} - \frac{\partial \bar{f}^{i,\epsilon}(s,0)}{\partial \zeta_q} \right| < \delta, \quad (s,\eta) \in [t_{i-1}, t_i] \times [0, 1], \quad i = 1, 2, \dots, N.$$

Thus, it follows from Lemma 5 that

$$|\Delta_1^{i,\epsilon}(s)| \leq L_5 \delta |\epsilon|, \quad |\Delta_2^{i,\epsilon}(s)| \leq mL_5 \delta |\epsilon|, \quad |\Delta_3^{i,\epsilon}(s)| \leq \delta |\epsilon|, \quad i = 1, 2, \dots, N. \quad (48)$$

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

$$\begin{aligned} \frac{\partial \bar{f}^{i,\epsilon}(s,\eta)}{\partial \eta} &= \frac{\partial \bar{f}^{i,\epsilon}(s,\eta)}{\partial x} \varphi^\epsilon(s) + \sum_{j=1}^m \frac{\partial \bar{f}^{i,\epsilon}(s,\eta)}{\partial \tilde{x}^j} \theta^{\epsilon,j}(s) + \epsilon \frac{\partial \bar{f}^{i,\epsilon}(s,\eta)}{\partial \zeta_q}, \\ s &\in [t_{i-1}, t_i], \quad i = 1, 2, \dots, N. \end{aligned} \quad (49)$$

Recall that $t \in (0, T]$ is a fixed time point. Then $t \in (t_{\varsigma-1}, t_\varsigma]$ for some fixed integer $\varsigma \in \{1, 2, \dots, N\}$. By the fundamental theorem of calculus,

$$\begin{aligned} \varphi^\epsilon(t) &= x^\epsilon(t) - x(t) \\ &= \sum_{i=1}^{\varsigma} \int_{t_{i-1}}^{\min\{t, t_i\}} \{ \bar{f}^{i,\epsilon}(s, 1) - \bar{f}^{i,\epsilon}(s, 0) \} ds \\ &= \sum_{i=1}^{\varsigma} \int_{t_{i-1}}^{\min\{t, t_i\}} \left(\int_0^1 \frac{\bar{f}^{i,\epsilon}(s,\eta)}{\partial \eta} d\eta \right) ds. \end{aligned}$$

Thus, by (49),

$$\begin{aligned} \varphi^\epsilon(t) &= \sum_{i=1}^{\varsigma} \int_{t_{i-1}}^{\min\{t, t_i\}} \left(\int_0^1 \frac{\partial \bar{f}^{i,\epsilon}(s,\eta)}{\partial x} \varphi^\epsilon(s) d\eta \right) ds \\ &\quad + \sum_{i=1}^{\varsigma} \int_{t_{i-1}}^{\min\{t, t_i\}} \left(\sum_{j=1}^m \int_0^1 \frac{\partial \bar{f}^{i,\epsilon}(s,\eta)}{\partial \tilde{x}^j} \theta^{\epsilon,j}(s) d\eta \right) ds \\ &\quad + \sum_{i=1}^{\varsigma} \int_{t_{i-1}}^{\min\{t, t_i\}} \left(\int_0^1 \epsilon \frac{\partial \bar{f}^{i,\epsilon}(s,\eta)}{\partial \zeta_q} d\eta \right) ds. \end{aligned} \quad (50)$$

Note that

$$\int_0^1 \frac{\partial \bar{f}^{i,\epsilon}(s,\eta)}{\partial x} \varphi^\epsilon(s) d\eta = \Delta_1^{i,\epsilon}(s) + \frac{\partial \bar{f}^{i,\epsilon}(s,0)}{\partial x} \varphi^\epsilon(s), \quad (51)$$

$$\sum_{j=1}^m \int_0^1 \frac{\partial \bar{f}^{i,\epsilon}(s,\eta)}{\partial \tilde{x}^j} \theta^{\epsilon,j}(s) d\eta = \Delta_2^{i,\epsilon}(s) + \sum_{j=1}^m \frac{\partial \bar{f}^{i,\epsilon}(s,0)}{\partial \tilde{x}^j} \theta^{\epsilon,j}(s), \quad (52)$$

and

$$\int_0^1 \epsilon \frac{\partial \bar{f}^{i,\epsilon}(s,\eta)}{\partial \zeta_q} d\eta = \Delta_3^{i,\epsilon}(s) + \epsilon \frac{\partial \bar{f}^{i,\epsilon}(s,0)}{\partial \zeta_q}. \quad (53)$$

Substituting (51)-(53) into (50) gives

$$\begin{aligned} \varphi^\epsilon(t) &= \sum_{i=1}^{\varsigma} \int_{t_{i-1}}^{\min\{t, t_i\}} \left(\Delta_1^{i,\epsilon}(s) + \frac{\partial \bar{f}^{i,\epsilon}(s,0)}{\partial x} \varphi^\epsilon(s) \right) ds \\ &\quad + \sum_{i=1}^{\varsigma} \int_{t_{i-1}}^{\min\{t, t_i\}} \left(\Delta_2^{i,\epsilon}(s) + \sum_{j=1}^m \frac{\partial \bar{f}^{i,\epsilon}(s,0)}{\partial \tilde{x}^j} \theta^{\epsilon,j}(s) \right) ds \\ &\quad + \sum_{i=1}^{\varsigma} \int_{t_{i-1}}^{\min\{t, t_i\}} \left(\Delta_3^{i,\epsilon}(s) + \epsilon \frac{\partial \bar{f}^{i,\epsilon}(s,0)}{\partial \zeta_q} \right) ds. \end{aligned} \quad (54)$$

Next, integrating the auxiliary system (41) with the intermediate conditions (42) and the initial condition (43) gives

$$\Gamma_q(t) = \sum_{i=1}^s \int_{t_{i-1}}^{\min\{t, t_i\}} \left(\frac{\partial \bar{f}^{i, \epsilon}(s, 0)}{\partial x} \Gamma_q(s) + \sum_{j=1}^m \frac{\partial \bar{f}^{i, \epsilon}(s, 0)}{\partial \tilde{x}^j} \Gamma_q(s - \alpha_j) + \frac{\partial \bar{f}^{i, \epsilon}(s, 0)}{\partial \zeta_q} \right) ds. \quad (55)$$

Multiplying (54) by ϵ^{-1} , subtracting (55), simplifying using (36), taking the norm of both sides and finally applying (48) yields

$$\begin{aligned} |\epsilon^{-1} \varphi^\epsilon(t) - \Gamma_q(t)| &\leq (mL_5 + L_5 + 1)\delta T + M_4 \int_0^t |\epsilon^{-1} \varphi^\epsilon(s) - \Gamma_q(s)| ds \\ &\quad + \sum_{j=1}^m \int_0^t M_5 |\epsilon^{-1} \varphi^\epsilon(s - \alpha_j) - \Gamma_q(s - \alpha_j)| ds. \end{aligned} \quad (56)$$

The last integral term on the right-hand side of (56) can be simplified using (37) and (43) as follows:

$$\begin{aligned} \sum_{j=1}^m \int_0^t M_5 |\epsilon^{-1} \varphi^\epsilon(s - \alpha_j) - \Gamma_q(s - \alpha_j)| ds &= \sum_{j=1}^m \int_{-\alpha_j}^{t - \alpha_j} M_5 |\epsilon^{-1} \varphi^\epsilon(s) - \Gamma_q(s)| ds \\ &\leq \int_0^t m M_5 |\epsilon^{-1} \varphi^\epsilon(s) - \Gamma_q(s)| ds. \end{aligned}$$

Thus, (56) becomes

$$|\epsilon^{-1} \varphi^\epsilon(t) - \Gamma_q(t)| \leq (mL_5 + L_5 + 1)\delta T + M_6 \int_0^t |\epsilon^{-1} \varphi^\epsilon(s) - \Gamma_q(s)| ds,$$

where $M_6 := M_4 + m M_5$. By the Gronwall-Bellman Lemma [11], it follows that

$$|\epsilon^{-1} \varphi^\epsilon(t) - \Gamma_q(t)| \leq (mL_5 T \delta + L_5 T \delta + T \delta) \exp(M_6 T),$$

which holds whenever $0 < |\epsilon| < \epsilon'$. Since δ is arbitrarily chosen, we conclude that $\epsilon^{-1} \varphi^\epsilon(t) \rightarrow \Gamma_q(t)$ as $\epsilon \rightarrow 0$. \square

3.3 A computational procedure

Based on Theorems 1 and 2, the partial derivatives of the cost functional in Problem (P) can be expressed as follows:

$$\frac{\partial J(\alpha, \zeta)}{\partial \alpha_p} = \sum_{l=1}^d \frac{\partial \Phi(x(\tau_1 | \alpha, \zeta), \dots, x(\tau_d | \alpha, \zeta))}{\partial x(\tau_l)} A_p(\tau_l | \alpha, \zeta), \quad p = 1, 2, \dots, m, \quad (57)$$

$$\frac{\partial J(\alpha, \zeta)}{\partial \zeta_q} = \sum_{l=1}^d \frac{\partial \Phi(x(\tau_1 | \alpha, \zeta), \dots, x(\tau_d | \alpha, \zeta))}{\partial x(\tau_l)} \Gamma_q(\tau_l | \alpha, \zeta), \quad q = 1, 2, \dots, v. \quad (58)$$

As a result, the cost function $J(\alpha, \zeta)$ and the gradients (57) and (58) corresponding to a feasible delay-parameter pair $(\alpha, \zeta) \in \mathcal{D} \times \mathcal{Z}$ can be computed using the following computational procedure.

Step 1. Obtain $x(\cdot | \alpha, \zeta)$, $A_p(\cdot | \alpha, \zeta)$, $p = 1, 2, \dots, m$, and $\Gamma_q(\cdot | \alpha, \zeta)$, $q = 1, 2, \dots, v$, by solving the expanded multistage system consisting of the original system (1) and the auxiliary systems (16)-(18) and (41)-(43).

Step 2. Use $x(\tau_l|\alpha, \zeta)$, $l = 1, 2, \dots, d$, to compute the cost function $J(\alpha, \zeta)$ by equation (5).

Step 3. Use $A_p(\tau_l|\alpha, \zeta)$ and $\Gamma_q(\tau_l|\alpha, \zeta)$ to compute $\frac{\partial J(\alpha, \zeta)}{\partial \alpha_p}$, $p = 1, 2, \dots, m$, and $\frac{\partial J(\alpha, \zeta)}{\partial \zeta_q}$, $q = 1, 2, \dots, v$, by (57) and (58).

This computational procedure can be readily incorporated into a standard gradient-based optimization method, such as the sequential quadratic programming (SQP) method [29]. Thus, by implementing this computational procedure, we can treat Problem (P) as a standard mathematical programming problem and solve it accordingly using existing techniques. This yields the optimal time-delays and the optimal system parameters that minimize the cost functional.

4 Numerical examples

To test the performance of the optimization method developed in the previous section, we consider two examples. Example 4.1 is a time-delay estimation problem for a nonlinear multistage system first considered in [32]. Example 4.2 is a parameter identification problem for a realistic fed-batch fermentation process.

4.1 Time-delay estimation for a 3-stage system

We consider the following two-dimensional delayed multistage system with 3 stages and 1 time-delay [32]:

$$\begin{cases} \dot{x}_1(t) = 2x_1(t)x_2(t) + x_2(t - \alpha), \\ \dot{x}_2(t) = 3x_1(t) + 4x_2(t - \alpha), \end{cases} \text{ if } t \in (0, 0.15), \quad (59)$$

$$\begin{cases} \dot{x}_1(t) = -2x_1(t)x_2(t) + \sin(x_2(t - \alpha)), \\ \dot{x}_2(t) = x_1(t)x_2(t) + x_1(t - \alpha)x_2(t - \alpha), \end{cases} \text{ if } t \in (0.15, 0.5211), \quad (60)$$

$$\begin{cases} \dot{x}_1(t) = t^2 - 2x_1(t) + 3x_2(t - \alpha), \\ \dot{x}_2(t) = -x_2(t) + x_1(t - \alpha)x_2(t - \alpha), \end{cases} \text{ if } t \in (0.5211, 1), \quad (61)$$

with initial conditions

$$x_1(t) = t - 1, \quad x_2(t) = t^2 + 1, \quad t \leq 0. \quad (62)$$

The switching times $t_1 = 0.15$ and $t_2 = 0.5211$ in (59)-(61) are the optimal switching times determined in [32] under the assumption that α is fixed and known. Here, we assume that α is unknown and our goal is to estimate its value using given output data. We assume that α lies within the interval $[0.01, 1]$. We also assume that x_2 is the only state that can be measured. To generate the observed data, we simulate the multistage system (59)-(62) with $\alpha = 0.5$. This yields the observed data $y^l = x_2(\tau_l|0.5)$, where $\tau_l = 0.1l - 0.05$, $l = 1, \dots, 10$, are fixed sample times. Thus, the optimal time-delay estimation problem is: choose $\alpha \in [0.01, 1]$ to minimize

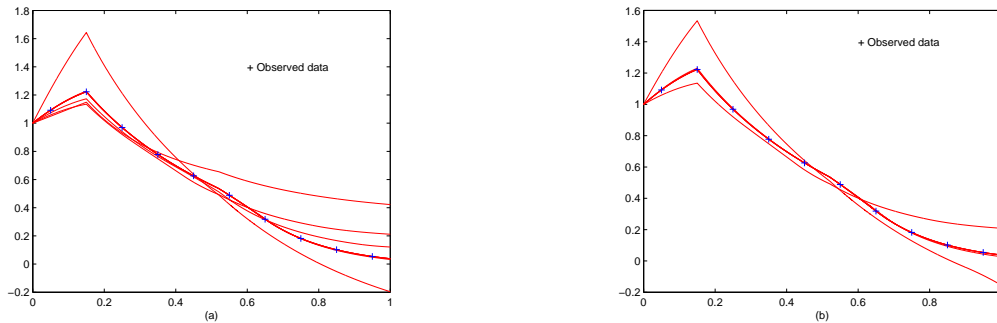
$$J(\alpha) = \sum_{l=1}^{10} |x_2(\tau_l|\alpha) - y^l|^2 \quad (63)$$

subject to the multistage system (59)-(61) and the initial conditions (62).

Clearly, the optimal solution for this problem is $\alpha = 0.5$. Our goal is to recover this optimal solution by using the computational method in Section 3 to minimize (63). To do this, we wrote a Visual C++ program that combines the gradient computation algorithm described in Section 3 with the optimization code NLPQLP [34]. This program uses the fourth-order Runge-Kutta integration scheme to solve the state and auxiliary systems simultaneously. Computational results for initial guesses of $\alpha = 0.01$ and $\alpha = 0.9$ are summarized in Table 1 and Fig. 1. Note that our program only needed at most eight iterations to converge from either initial guess to the optimal solution $\alpha = 0.5$.

Table 1 Convergence of the time-delay estimate in Example 4.1. The initial guesses are $\alpha = 0.01$ and $\alpha = 0.9$ and i denotes the iteration number.

i	α^i	$J(\alpha^i)$	i	α^i	$J(\alpha^i)$
0	0.01000	4.68160×10^{-1}	0	0.90000	1.86337×10^{-1}
1	0.98510	3.29317×10^{-1}	1	0.27700	7.47820×10^{-2}
2	0.43911	5.16643×10^{-3}	2	0.45364	3.00165×10^{-3}
3	0.47823	6.71816×10^{-4}	3	0.51571	3.52834×10^{-4}
4	0.49580	2.54658×10^{-5}	4	0.49953	3.09494×10^{-7}
5	0.49944	4.46961×10^{-7}	5	0.49999	7.28361×10^{-11}
6	0.49998	5.26670×10^{-10}	6	0.50000	1.48039×10^{-18}
7	0.49999	1.25571×10^{-14}			
8	0.50000	9.08580×10^{-19}			

**Fig. 1** Numerical convergence of the observed state in Example 4.1: The red lines show the state x_2 at each iteration; the blue crosses show the observed data (which corresponds to $\alpha = 0.5$). (a) Initial guess $\alpha = 0.01$. (b) Initial guess $\alpha = 0.9$.

4.2 Parameter identification for a fed-batch fermentation process

We now consider the fed-batch fermentation process for converting glycerol to 1,3-propanediol (1,3-PD) using the microorganism *Klebsiella pneumoniae*. This process consists of two modes: batch mode and feeding mode. The fed-batch process begins with batch mode, in which no substrate is added to the fermentor. After some time in batch mode, the biomass begins to grow exponentially. Once the exponential growth phase is completed and the growth rate stabilizes, the feeding mode begins. In the feeding mode, the substrates of glycerol and alkali are added continuously to the fermentor, thus helping to maintain a suitable environment for cell growth. At the end of the feed, another batch process begins, and so on. The fed-batch process switches back and forth between the batch and feeding modes throughout the time horizon [35].

In batch mode, the mass balance equations for the biomass, substrate and 1,3-PD are given by

$$\begin{cases} \dot{x}_1(t) = \mu(x_2(t))x_1(t - \alpha), \\ \dot{x}_2(t) = -q_2(x_2(t))x_1(t - \alpha), \\ \dot{x}_3(t) = q_3(x_2(t))x_1(t - \alpha), \end{cases} \quad (64)$$

where $x_1(t)$, $x_2(t)$, $x_3(t)$ are the extracellular concentrations of biomass, glycerol and 1,3-PD in the fermentor at time t , respectively. Furthermore, α is a time-delay; $\mu(x_2(t))$ is the specific growth rate of cells; $q_2(x_2(t))$ is the specific consumption rate of substrate; and $q_3(x_2(t))$ is the specific formation rate of 1,3-PD. The delay in (64) arises because nutrient metabolization does not immediately lead to the

production of new biomass [36]. The functions $\mu(x_2(t))$, $q_2(x_2(t))$ and $q_3(x_2(t))$ are given by

$$\begin{aligned}\mu(x_2(t)) &= \frac{\Delta_1 x_2(t)}{x_2(t) + k_1}, \\ q_2(x_2(t)) &= m_2 + Y_2 \mu(x_2(t)) + \frac{\Delta_2 x_2(t)}{x_2(t) + k_2}, \\ q_3(x_2(t)) &= -m_3 + Y_3 \mu(x_2(t)) + \frac{\Delta_3 x_2(t)}{x_2(t) + k_3},\end{aligned}$$

where $\Delta_1, k_1, m_2, Y_2, \Delta_2, k_2, m_3, Y_3, \Delta_3, k_3$ are model parameters whose biological meanings are explained in [37].

During the feeding mode, the mass balance equations for the biomass, substrate and 1,3-PD are given by

$$\begin{cases} \dot{x}_1(t) = \mu(x_2(t))x_1(t - \alpha) - D(t)x_1(t), \\ \dot{x}_2(t) = D(t)((1+r)^{-1}c_{s0} - x_2(t)) - q_2(x_2(t))x_1(t - \alpha), \\ \dot{x}_3(t) = q_3(x_2(t))x_1(t - \alpha) - D(t)x_3(t), \end{cases} \quad (65)$$

where r is the velocity ratio of adding alkali to glycerol; c_{s0} is the concentration of the initial feed of glycerol; and $D(t)$ is the dilution rate at time t .

Let $t_i, i = 1, 2, \dots, N$, be the times at which the fermentation process switches between the batch and feeding modes. Then the dilution rate in the interval $[t_{i-1}, t_i]$ is given by

$$D(t) = \frac{(1+r)v_i}{V(t)}, \quad (66)$$

$$V(t) = V_0 + \sum_{j=1}^{i-1} (1+r)(t_j - t_{j-1})v_j + (1+r)(t - t_{i-1})v_i. \quad (67)$$

In (66) and (67), $v_i \geq 0$ is the constant feeding rate of glycerol during the interval $[t_{i-1}, t_i], i = 1, 2, \dots, N$, and V_0 is the initial volume of culture fluid in the fermentor.

Let $\zeta := (\Delta_1, k_1, m_2, Y_2, \Delta_2, k_2, m_3, Y_3, \Delta_3, k_3)^\top$ be the model parameter vector. We assume that the time-delay α lies within the interval $[0.01, 5]$. We also assume that the components of ζ satisfy the lower and upper bounds given in Table 2. Our goal is to optimally choose α and ζ to fit some real output data from an experiment conducted in China [38]. The data set consists of 12 sample times and the biomass, glycerol and 1,3-PD concentrations at each of these times. We denote the concentrations of biomass, glycerol and 1,3-PD measured at the sample time τ_l by y_1^l, y_2^l and $y_3^l, l \in \{1, 2, \dots, 12\}$, respectively. Then, the cost functional in this example is the following weighted least squares error function:

$$J(\alpha, \zeta) = \sum_{\ell=1}^3 \sum_{l=1}^{12} w_\ell^l |x(\tau_l | \alpha, \zeta) - y_\ell^l|^2 \quad (68)$$

where $w_\ell^l := (y_\ell^l)^{-2}$ is a weight used to normalize the contribution of each term. The parameter identification problem is: given the multistage system with subsystems (64) and (65), choose α and ζ to minimize the cost functional (68) subject to $\alpha \in [0.01, 5]$ and the model parameter bounds in Table 2.

The initial functions for this parameter identification problem were constructed by applying cubic spline interpolation [39] to the experimental data before the zero time point [38]. The other parameters are $r = 0.75, c_{s0} = 10762 \text{mmolL}^{-1}, N = 1355, V_0 = 5\text{L}$ and $T = 24.16\text{h}$. The first batch mode lasts for 5.33 hours. After this, the fed-batch process proceeds in phases, where each phase consists of multiple batch and feeding modes. There are 9 phases in total (Phs.I-IX). In each one of Phs.I-IX, the same feeding rates, namely, $0.2103 \text{mLs}^{-1}, 0.2103 \text{mLs}^{-1}, 0.1992 \text{mLs}^{-1}, 0.2103 \text{mLs}^{-1}, 0.2103 \text{mLs}^{-1}, 0.2214 \text{mLs}^{-1}, 0.2548 \text{mLs}^{-1}, 0.2548 \text{mLs}^{-1}$, and 0.2548mLs^{-1} for the feeding processes were applied. The durations of

Table 2 Optimal values, initial guesses, and lower and upper bounds for the model parameters in Example 4.2.

Parameter	Δ_1	k_1	m_2	Y_2	Δ_2
Optimal value	0.8037	0.4856	0.2977	144.912	7.8367
Initial guess	1.4746	0.5148	0.9167	275.33	19.991
Lower bound	0.438	0.14	0.2977	64.10	5.945
Upper bound	2.19	0.7	3.988	320.512	29.727
Parameter	k_2	m_3	Y_3	Δ_3	k_3
Optimal value	9.4632	12.2557	80.8439	20.2757	38.75
Initial guess	12.615	8.1909	114.79	26.002	24.071
Lower bound	4.367	2.451	33.84	8.865	7.75
Upper bound	21.847	12.257	169.225	44.32	38.75

the feeding processes in Pfs.I-IX were 5, 7, 8, 7, 6, 4, 3, 2, and 1 seconds in each 100 seconds, leaving 95, 93, 92, 93, 94, 96, 97, 98, and 99 seconds for batch processes, respectively.

By extending the Visual C++ program used in Example 4.1, we obtained the optimal time-delay $\alpha^* = 0.1568$ h and the optimal model parameters listed in Table 2. The entire optimization process took 38 iterations. The corresponding optimal cost functional value is 0.51486. Here, the initial guess of the time-delay is 0.31559h and the initial guesses of the model parameters are listed in Table 2. For comparison, we also used the Particle Swarm Optimization (PSO) algorithm developed in [35] to solve this example. The parameters in the PSO algorithm are: the number of particles in the swarm is 200, the maximal number of iterations is 300, and all other parameters are the same as those in [35]. Using PSO, we obtained an optimal cost functional value of 0.61387. The time taken by PSO is 15 minutes, which is 10 times longer than our new method. Furthermore, under the obtained optimal time-delay and model parameters, we plotted the concentration changes of the biomass, glycerol and 1,3-PD with respect to the fermentation time. We also showed the numerical convergence of the state trajectories in Fig. 2. From Fig. 2, we can see that the obtained optimal state trajectories are good fits for the experimental data.

5 Conclusions

In this paper, we investigated the optimal parameter selection problem for nonlinear multistage systems with time-delays. This problem class arises in the mathematical modelling of the fed-batch fermentation process for 1,3-propanediol production (see Example 4.2). We first derived the gradients of the system state with respect to the time-delays and system parameters. We then developed a gradient-based optimization algorithm to find the optimal time-delays and system parameters. To the best of our knowledge, this is the first algorithm in the literature for simultaneously optimizing time-delays and system parameters in a nonlinear multistage system. The effectiveness of this algorithm was verified using two highly nonlinear numerical examples.

Acknowledgements The first author is supported by the Natural Science Foundation for the Youth of China (Grant no. 11201267 and Grant no. 11001153), the Tian Yuan Special Funds of the National Natural Science Foundation of China (Grant no. 11126077) and the Shandong Province Natural Science Foundation of China (Grant no. ZR2010AQ016 and Grant no. ZR2011AL003). The second author is supported by the Natural Science Foundation of China (Grant no. 11350110208) and the Australian Research Council (Discovery Grant DP110100083). The third author is supported by the Australian Research Council (Discovery Grant DP110100083).

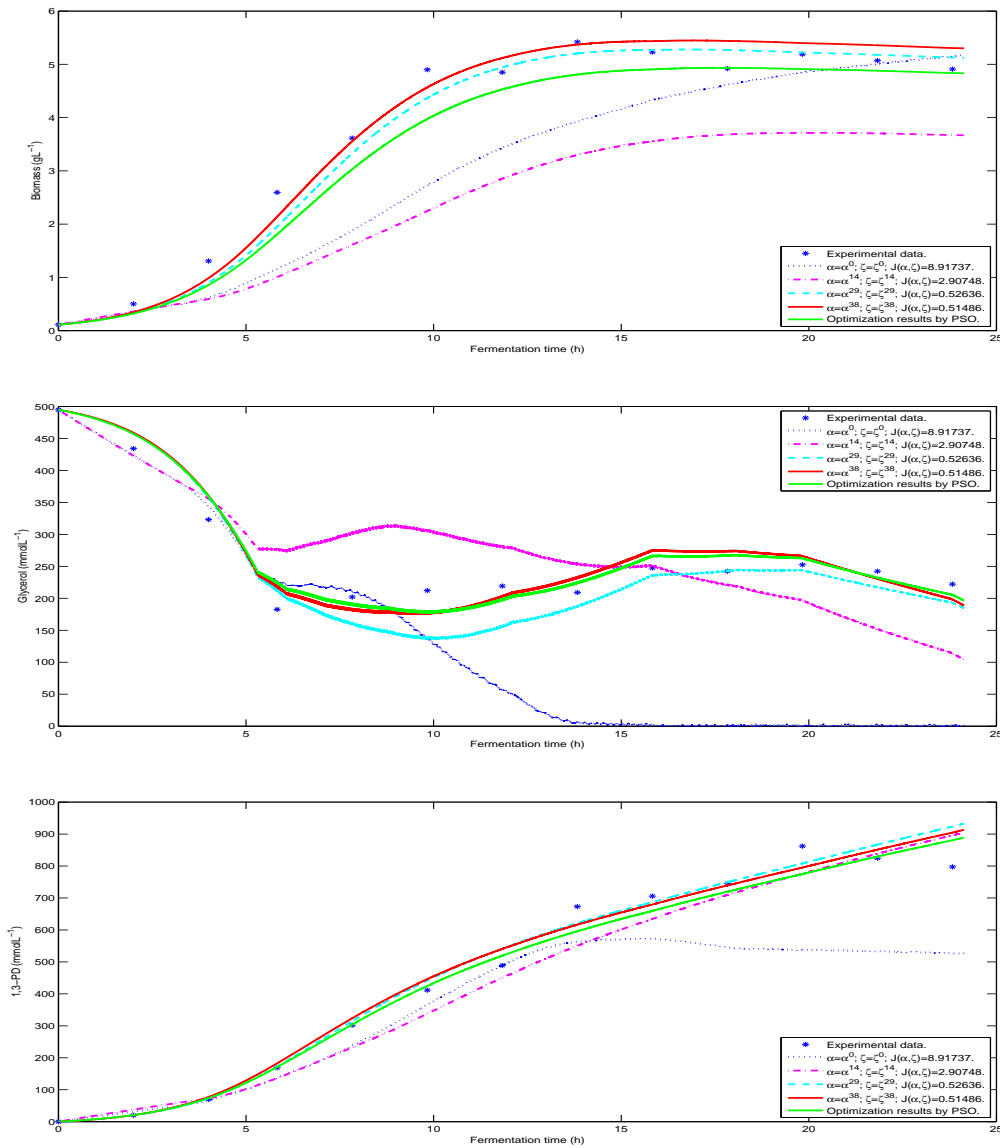


Fig. 2 Convergence of the optimal state trajectories in Example 4.2. The read lines show the final trajectories from our new algorithm; the green lines show the optimal trajectories from PSO.

References

1. Shi, J.J., Zhou, S.Y.: Quality control and improvement for multistage systems: A survey. *IIE Trans.* **41**, 744–753 (2009)
2. Deindoerfer, F.H., Humphrey, A.E.: Design of multistage systems for simple fermentation processes. *Ind. Eng. Chem.* **51**, 809–812 (1959)
3. Mazanov, A.: A multistage population model. *J. Theor. Biol.* **39**, 581–587 (1973)
4. Colledani, M., Matta, A., Tolio T.: Analysis of the production variability in multi-stage manufacturing systems. *CIRP Ann. – Manuf. Techn.* **59**, 449–452 (2010)
5. Flam, R.P.: Time delay in multistage image tubes. *P. IEEE* **57**, 2152–2153 (1969)

6. Livk, I., Pohar, C., Ilievski, D.: Estimation of batch precipitation kinetics by a simplified differential method. *AIChE J.* **45**, 1593–1596 (1999)
7. Li, T.S., Livk, I., Ilievski, D.: Influence of the estimation procedure on the accuracy and precision of aluminum trihydroxide crystallization kinetics from dynamic data. *Ind. Eng. Chem. Res.* **40**, 5005–5013 (2001)
8. Wang, L.Y., Gui, W.H., Teo, K.L., Loxton, R., Yang, C.H.: Time delayed optimal control problems with multiple characteristic time points: Computation and industrial applications. *J. Ind. Manag. Optim.* **5**, 705–718 (2009)
9. Wu, C.Z., Teo, K.L., Rehbock, V.: Optimal control of piecewise affine systems with piecewise affine state feedback. *J. Ind. Manag. Optim.* **5**, 737–747 (2009)
10. Feng, Z.G., Teo, K.L.: Optimal feedback control for stochastic impulsive linear systems subject to Poisson processes. In: *Optimization and Optimal Control: Theory and Applications*, Springer, 241–258 (2010)
11. Teo, K.L., Goh, C.J., Wong, K.H.: *A Unified Computational Approach to Optimal Control Problems*. Long Scientific and Technical, Essex (1991)
12. Loxton, R., Teo, K.L., Rehbock, V.: Optimal control problems with multiple characteristic time points in the objective and constraints. *Automatica* **44**, 2923–2929 (2008)
13. Feng, Z.G., Yiu, K.F.C., Teo, K.L.: A smoothing approach for the optimal parameter selection problem with continuous inequality constraint. *Optim. Method. Softw.* **28**(4), 689–705 (2013)
14. Pantelides, C.C., Barton, P.I.: Equation-oriented dynamic simulation: current status and future perspectives. *Comput. Chem. Eng.* **17S**, S268–S285 (1993)
15. Morison, K.R., Sargent, R.W.H.: Optimization of multistage processes described by differential-algebraic equations. *Lect. Notes Math.* **1230**, 86–102 (1986)
16. Vassiliadis, V.S., Sargen, R.W.H., Pantelides, C.C.: Solution of a class of multistage dynamic optimization problems: 1. Problem without path constraints. *Ind. Eng. Chem. Res.* **33**, 2111–2122 (1994)
17. Vassiliadis, V.S., Sargen, R.W.H., Pantelides, C.C.: Solution of a class of multistage dynamic optimization problems: 2. Problem with path constraints. *Ind. Eng. Chem. Res.* **33**, 2123–2133 (1994)
18. Shen, L.J., Wang, Y., Feng, E.M., Xiu, Z.L.: Bilevel parameters identification for the multi-stage nonlinear impulsive system in microorganisms fed-batch cultures. *Nonlinear Anal. – Real.* **9**, 1068–1077 (2008)
19. Jiang, Z.G., Yuan, J.L., Feng, E.M.: Robust identification and its properties of nonlinear bilevel multi-stage dynamic system. *Appl. Math. Comput.* **219**, 6979–6985 (2013)
20. Debeljković, D.: *Time-Delay Systems*. InTech, Rijeka (2011)
21. Richard, J.P.: Time-delay systems: An overview of some recent advances and open problems. *Automatica* **29**, 1667–1694 (2003)
22. Gawthrop, P.J., Nihtilä, M.T.: Identification of time-delays using a polynomial identification method. *Syst. Control Lett.* **5**, 267–271 (1985)
23. Diop, S., Kolmanovsky, I., Moraal, P.E., Nieuwstadt, M.V.: Preserving stability/performance when facing an unknown time-delay. *Control Eng. Pract.* **9**, 1319–1325 (2001)
24. Pan, F., Han, R.C., Feng, D.M.: An identification method of time-varying delay based on genetic algorithm. In: *Proceedings of the Second International Conference on Machine Learning and Cybernetics*, pp. 781–783 (2003)
25. Banks, H.T., Burns, J.A., Cliff, E.M.: Parameter estimation and identification for systems with delay. *SIAM J. Control Optim.* **19**, 791–828 (1981)
26. Loxton, R., Teo, K.L., Rehbock, V.: An optimization approach to state-delay identification. *IEEE Trans. Autom. Control* **55**, 2113–2119 (2010)
27. Chai, Q., Loxton, R., Teo, K.L., Yang, C.: A unified parameter identification method for nonlinear time-delay systems. *J. Ind. Manag. Optim.* **9**, 471–486 (2013)
28. Chai, Q., Loxton, R., Teo, K.L., Yang, C.: A class of optimal state-delay control problems. *Nonlinear Anal. – Real.* **14**, 1536–1550 (2013)
29. Nocedal, J., Wright, S.J.: *Numerical Optimization*. Springer-Verlag, New York (1999)
30. Ahmed, N.U.: *Dynamic Systems and Control with Applications*. World Scientific, Singapore (2006)
31. Wong, K.H., Jennings, L.S., Benyah, F.: The control parametrization enhancing transform for constrained time-delayed optimal control problems. *ANZIAM J.* **43**, E154–E185 (2002)
32. Wu, C.Z., Teo, K.L., Li, R., Zhao, Y.: Optimal control of switched systems with time delay. *Appl. Math. Lett.* **19**, 1062–1067 (2006)
33. Chai, Q., Loxton, R., Teo, K.L., Yang, C.H.: Time-delay estimation for nonlinear systems with piecewise-constant input. *Appl. Math. Comput.* **219**, 9543–9560 (2013)
34. Schittkowski, K.: *NLPQLP; A Fortran Implementation of a Sequential Quadratic Programming Algorithm with Distributed and Non-Monotone Line Search*. University of Bayreuth, Bayreuth (2007)
35. Liu, C.Y., Gong, Z.H., Feng, E.M., Yin, H.C.: Modelling and optimal control for nonlinear multistage dynamical system of microbial fed-batch culture. *J. Ind. Manag. Optim.* **5**, 835–850 (2009)
36. Xiu, Z.L., Song, B.H., Sun, L.H., Zeng, A.P.: Theoretical analysis of effects of metabolic overflow and time delay on the performance and dynamic behavior of a two-stage fermentation process. *Biochem. Eng. J.* **11**, 101–109 (2002)
37. Liu, C.Y., Gong, Z.H., Shen, B.Y., Feng, E.M.: Modelling and optimal control for a fed-batch fermentation process. *Appl. Math. Model.* **37**, 695–706 (2013)
38. Mu, Y., Zhang, D.J., Teng, H., Wang, W., Xiu, Z.L.: Microbial production of 1,3-Propanediol by *Klebsiella pneumoniae* using crude glycerol from biodiesel preparation. *Biotechnol. Lett.* **28**, 1755–1759 (2006)
39. Stoer, J., Bulirsch, R.: *Introduction to Numerical Analysis*. Springer-Verlag, New York (1980)