Robust parameter estimation for nonlinear multistage time-delay systems with noisy measurement data

Chongyang Liu[a,b,*], Zhaohua Gong[a], Kok Lay Teo[b]

[a] *School of Mathematics and Information Science, Shandong Institute of Business and Technology, Yantai 264005, China*
[b] *Department of Mathematics and Statistics, Curtin University, Perth 6102, Australia*

Abstract

In this paper, we consider estimation problems involving a class of nonlinear systems characterized by two non-standard attributes: (i) such systems evolve over multiple stages; and (ii) the dynamics in each stage involve unknown time-delays and unknown system parameters. These unknown quantities are to be estimated such that a least-squares error function between the system output and a set of noisy measurement data from a real plant is minimized. We first present the classical parameter estimation formulation, where the expectation of the error function is regarded as the cost function. However, in practice, there exists uncertainty in the distribution of the measurement data. The optimal parameter estimate should be able to withstand this uncertainty. Accordingly, we propose a new parameter estimation formulation, in which the cost function is the variance of the error function and the constraint indicates an allowable sacrifice from the optimal expectation value of the classical parameter estimation problem. For these two estimation problems, we show that the gradients of their cost functions and the constraint function with respect to the time-delays and system parameters can be computed through solving a set of auxiliary time-delay systems in conjunction with the governing multistage time-delay system, simultaneously. On this basis, we develop gradient-based optimization algorithms to determine the unknown time-delays and system parameters. Finally, we consider two example problems to illustrate the effectiveness and applicability of our proposed algorithms.

Keywords: Multistage system; time-delay system; parameter estimation; robust parameter estimation; nonlinear optimization

1. Introduction

A dynamic system that evolves over multiple stages is referred to as a multistage system. The system is called a multistage time-delay system if time-delays

*Corresponding author
Email address: `liu_chongyang@yahoo.com` (Chongyang Liu)

appear in any of the stages of the multistage system. Both multistage system and multistage time-delay system are encountered in various real-world applications, such as population models [1], production-inventory systems [2], and fermentation processes [3, 4].

In solving a real-world problem, the first step is to construct a dynamic system to model the behaviour of the real plant. Clearly, some of the parameters, which may include time-delays, are required to be estimated. This process is known as parameter estimation, which is a key step in the construction of reliable dynamic models in science and engineering problems [5]. A parameter estimation problem is usually formulated as an optimization problem, in which a cost function quantifying the difference between the output of the constructed dynamic model and the measured output data from a real plant is minimized. There are many results, both in theory and algorithms, available in the literature for parameter estimation problems involving nonlinear systems with a single stage and no time-delays; see, for example [6, 7, 8]. However, for nonlinear multistage systems, they contain both discrete and continuous characteristics [9], and hence are much harder to handle. A gradient-based algorithm to compute the optimal parameters for a class of multistage systems without path constraints is developed in [10], and it is extended to a class of multistage systems with path constraints in [11]. A hybrid stochastic-deterministic method is proposed to optimize the parameters for a class of multistage systems in [12]. More recently, parameter estimation problems involving nonlinear multistage systems, which arise in the study of fermentation processes, are investigated in [13, 14, 15]. However, all the results mentioned above do not take into account the effect of time-delays that exist in the multistage systems.

Time-delays are commonly encountered in various engineering systems, such as chemical processes, mechanical systems, network control systems and economic systems [16]. It is well known that time-delays can degrade system performance [17]. Therefore, the effect of time-delays must not be ignored during the construction of the system model. As a result, parameter estimation for time-delay systems has attracted a considerable interest amongst researches over the past decades; see, for example [18, 19, 20, 21]. Many computation methods, such as finite-dimensional approximation scheme [22], steepest-descent algorithm [23], modified least-squares technique [24] and particle swarm optimization [25], have been developed. Recently, a new algorithm for estimating unknown time-delays in a nonlinear dynamic system is developed in [26]. This algorithm is extended to cater for nonlinear systems that contain unknown system parameters as well as unknown time-delays in [27]. However, these parameter estimation algorithms are only designed for time-delay systems with a single stage. More recently, a gradient-based optimization algorithm is developed to solve parameter estimation problem involving nonlinear multistage systems with unknown time-delays as well as unknown system parameters in [28]. However, the measured output data from the real plant are assumed to be exact in [28]. This is, of course, an idealistic assumption, as the measurement of the output from the real plant can never be obtained with a perfect precision. In practice, there exists uncertainty in the distribution of the measurement data. The optimal parameter estimate should be able to withstand the uncertainty, i.e., it is robust against the uncertainty. As a result, by minimizing the weighted sum of mean and variance of a least-squares error between actual and predicted system output, parameter estimation for nonlinear time-delay systems with noisy output measurements is investigated in [29, 30]. However, there are two limitations in [29, 30]: (i) it does not provide quantitative information showing the sacrifice of the obtained expectation value from the optimal expectation value of the classical parameter estimation problem; and (ii) the parameter estimation method is only applicable to time-delay systems with a single stage.

In this paper, we propose a new formulation of the parameter estimation which is free of these drawbacks. To be more specific, we consider a general nonlinear multistage system with multiple time-delays and multiple system parameters. Robust time-delays and system parameters are to be estimated such that a least-squares error function between the system output and a set of noisy measurement data from the real plant is minimized. It is to be achieved in two stages. We first present the classical parameter estimation formulation, in which the cost function is the expectation of the error function. Then, based on the solution obtained, we propose a new formulation for finding robust time-delays and system parameters, where the cost function is the variance of the error function and the constraint explicitly specifies the allowable sacrifice in the expectation value of the error function. For these two estimation problems, we transform them into equivalent nonlinear optimization problems. Furthermore, we show that the gradients of the corresponding cost and constraint functions with respect to time-delays and system parameters can be computed through solving the original multistage time-delay system and a set of auxiliary systems forward in time, simultaneously. On this basis, gradient-based optimization methods are developed to solve these two estimation problems. Note that, unlike the method reported in [29, 30], our new methods are applicable to a much larger array of problems, not restricted to single-stage parameter estimation problems. Finally, two example problems are considered to test the performance of our new approaches.

The rest of the paper is organized as follows. Section 2 gives two parameter estimation problems. Section 3 presents the equivalent nonlinear optimization problems. The computational algorithms for the equivalent optimization problems are provided in Section 4. Numerical examples are discussed in Section 5. Finally, Section 6 provides some concluding remarks.

2. Problem formulation

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

$$\begin{cases} \dot{x}(t) = f^i(t, x(t), x(t-\alpha_1), \dots, x(t-\alpha_m), \zeta), t \in (t_{i-1}, t_i), i = 1, 2, \dots, N, & \text{(1a)} \\ x(t_i+) = x(t_i-), \quad i = 0, 1, \dots, N, & \text{(1b)} \\ x(t) = \phi(t, \zeta), \quad t \leq 0, & \text{(1c)} \end{cases}$$

where $x(t) \in R^n$ is the state vector; $\zeta \in R^v$ is the parameter vector; α_j, $j = 1, 2, \ldots, m$, are time-delays; t_i, $i = 1, 2, \ldots, N$, 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 $f^i : R \times R^n \times R^{nm} \times R^v \to R^n$ and $\phi : R \times R^v \to R^n$ are given functions. Here, the switching times in (1a)-(1b) are assumed to be pre-assigned such that

$$0 = t_0 < t_1 < \cdots < t_N = T,$$

where T is the terminal time. For system (1), it 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).

The system output $y(t) \in R^q$ of system (1) is given by

$$y(t) = g(t, x(t), \zeta), \quad t \geq 0, \tag{2}$$

where $g : R \times R^n \times R^v \to R^q$ is a given function.

In system (1), both time-delays α_j, $j = 1, 2, \ldots, m$, and the parameter vector ζ are unknown, requiring to be estimated. Now, define

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

where a_j and b_j are the lower and upper bounds of the jth time-delay. Any vector $\alpha \in \mathcal{D}$ is called an admissible time-delay vector. Furthermore, define

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

where c_k and d_k are the lower and upper bounds of the kth system parameter in ζ. Any vector $\zeta \in \mathcal{Z}$ is called an admissible parameter vector. Accordingly, any pair $(\alpha, \zeta) \in \mathcal{D} \times \mathcal{Z}$ is called an admissible delay-parameter pair.

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

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

Assumption 2. *There exists a positive real number $L > 0$ such that for all $t \in [t_{i-1}, t_i]$, $i = 1, 2, \ldots, N$, $x^j \in R^n$, $j = 0, 1, \ldots, m$, and $\zeta \in \mathcal{Z}$,*

$$\|f^i(t, x^0, x^1, \ldots, x^m, \zeta)\| \leq L(1 + \|x^0\| + \|x^1\| + \cdots + \|x^m\|),$$

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

Assumptions 1 and 2 ensure that system (1) admits a unique continuous solution corresponding to each delay-parameter pair $(\alpha, \zeta) \in \mathcal{D} \times \mathcal{Z}$ [31]. We denote this solution by $x(\cdot|\alpha, \zeta)$.

Let $y(\cdot|\alpha, \zeta)$ denote the system output obtained by substituting $x(\cdot|\alpha, \zeta)$ and ζ into (2). Our aim is to estimate the unknown time-delays and system parameters by comparing the system output with some measurement data of the real plant at a set of sample times

$$0 < \tau_1 < \tau_2 < \cdots < \tau_p < T.$$

Let $\hat{y}^l$ denote the measurement data at time $t = \tau_l$. Obviously, these measurement data are generally imprecise due to measurement noise. Here, we assume that the measurement data, $\hat{y}^l$, $l = 1, 2, \ldots, p$, are random vectors, where the corresponding mean vector (of dimension pq) and covariance matrix (of dimension $pq \times pq$) can be obtained. The following least-squares error function is used to measure the difference between the system output and the measurement data:

$$J(\alpha, \zeta) = \sum_{l=1}^{p} \|y(\tau_l|\alpha, \zeta) - \hat{y}^l\|^2. \tag{3}$$

Note that $J(\cdot, \cdot)$ is a function containing random vectors $\hat{y}^l$, $l = 1, 2, \ldots, p$. Now, we present the classical parameter estimation problem, which is denoted by Problem A, given below.

Problem A. *Given system* (1)*, choose an admissible delay-parameter pair* $(\alpha, \zeta) \in \mathcal{D} \times \mathcal{Z}$ *such that the cost function*

$$G_1(\alpha, \zeta) = \mathrm{E}\{J(\alpha, \zeta)\}$$

is minimized, where $\mathrm{E}\{\cdot\}$ *denotes the expectation.*

Problem A is a dynamic optimization problem in which the time-delays and system parameters in system (1) are to be optimized. In case that the distribution of the measurement data is known exactly, the optimal delay-parameter pair can be obtained by solving Problem A. However, in practice, we do not know the distribution of the measurement data exactly. There exists uncertainty in the distribution of measurement data. Therefore, it is important to find the optimal delay-parameter pair which is robust against the uncertainty. To this end, we take the variance of the least-squares error function (3), which measures how far the random error functions are spread out from their expectation value, as the cost function. Moreover, we impose the the following constraint to ensure that the expectation value is within the allowable sacrifice from the optimal expectation value of Problem A:

$$\mathrm{E}\{J(\alpha, \zeta)\} \leq (1 + \beta)\mathrm{E}(\tilde{\alpha}^*, \tilde{\zeta}^*), \tag{4}$$

where $(\tilde{\alpha}^*, \tilde{\zeta}^*)$ is the optimal delay-parameter pair of Problem A; and $\beta > 0$ is a weighting factor specifying the allowable sacrifice of the expectation value from the optimal expectation value of Problem A. Thus, our parameter estimation problem, which takes robustness into account, can be stated as follows.

Problem B. *Given system* (1)*, choose an admissible delay-parameter pair* $(\alpha, \zeta) \in \mathcal{D} \times \mathcal{Z}$ *such that the variance of the least-squares error function*

$$G_2(\alpha, \zeta) = \mathrm{Var}\{J(\alpha, \zeta)\}$$

is minimized subject to constraint (4), where $\mathrm{Var}\{\cdot\}$ *denotes variance.*

Problem B is a constrained optimization problem involving a nonlinear multistage system with time-delays and system parameters. In particular, constraint (4) quantitatively specifies the discrepancy between the delay-parameter pair and the optimal delay-parameter pair of Problem A. This is quite different from the parameter estimation formulation in [29, 30], in which the cost function is the weighted sum of mean and variance of the least-squares error function. Furthermore, the optimization algorithms reported in [27, 29, 30] are only designed for parameter estimation problems involving time-delay system with a single stage. Therefore, new approaches are needed to solve Problem A and Problem B. For this, we need to transform Problem A and Problem B into equivalent optimization problems.

3. Problem transformation

In what follows, we shall omit the arguments α and ζ in the output $y(\cdot|\alpha,\zeta)$ for brevity. Then, the least-squares error function (3) can be written as follows.

$$\begin{aligned} J(\alpha,\zeta) &= \sum_{l=1}^{p}(y(\tau_l)-\hat{y}^l)^\top(y(\tau_l)-\hat{y}^l) \\ &= \sum_{l=1}^{p} y(\tau_l)^\top y(\tau_l) - 2\sum_{l=1}^{p} y(\tau_l)^\top \hat{y}^l + \sum_{l=1}^{p}(\hat{y}^l)^\top \hat{y}^l. \end{aligned}$$

Thus,

$$\begin{aligned} G_1(\alpha,\zeta) &= \mathrm{E}\{J(\alpha,\zeta)\} \\ &= \sum_{l=1}^{p} y(\tau_l)^\top y(\tau_l) - 2\sum_{l=1}^{p} y(\tau_l)^\top \mathrm{E}\{\hat{y}^l\} + \sum_{l=1}^{p}\mathrm{E}\{(\hat{y}^l)^\top \hat{y}^l\}, \end{aligned} \tag{5}$$

and

$$\begin{aligned} G_2(\alpha,\zeta) = \mathrm{Var}\{J(\alpha,\zeta)\} &= 4\mathrm{Var}\left\{\sum_{l=1}^{p} y(\tau_l)^\top \hat{y}^l\right\} + \mathrm{Var}\left\{\sum_{l=1}^{p}(\hat{y}^l)^\top \hat{y}^l\right\} \\ &\quad - 4\mathrm{Cov}\left\{\sum_{l=1}^{p} y(\tau_l)^\top \hat{y}^l, \sum_{l=1}^{p}(\hat{y}^l)^\top \hat{y}^l\right\}, \end{aligned} \tag{6}$$

where $\mathrm{Cov}\{\cdot,\cdot\}$ denotes covariance. Note that

$$\begin{aligned} \mathrm{Var}\left\{\sum_{l=1}^{p} y(\tau_l)^\top \hat{y}^l\right\} &= \sum_{l=1}^{p}\sum_{r=1}^{p}\mathrm{Cov}\{y(\tau_l)^\top \hat{y}^l, y(\tau_r)^\top \hat{y}^r\} \\ &= \sum_{l=1}^{p}\sum_{r=1}^{p} y(\tau_l)^\top \Gamma^{l,r} y(\tau_r), \end{aligned} \tag{7}$$

where $\Gamma^{l,r} = [\gamma_{\imath\jmath}^{l,r}]$ is a $q \times q$ matrix whose $(\imath, \jmath)$th element is defined as

$$\gamma_{\imath\jmath}^{l,r} = \mathrm{Cov}\{\hat{y}_\imath^l, \hat{y}_\jmath^r\}. \tag{8}$$

Moreover,

$$\begin{aligned}\mathrm{Cov}\Bigg\{\sum_{l=1}^{p} y(\tau_l)^\top \hat{y}^l, \sum_{l=1}^{p} (\hat{y}^l)^\top \hat{y}^l\Bigg\} &= \sum_{l=1}^{p}\sum_{r=1}^{p} y(\tau_l)^\top \mathrm{Cov}\{\hat{y}^l, (\hat{y}^r)^\top \hat{y}^r\} \\ &= \sum_{l=1}^{p}\sum_{r=1}^{p} y(\tau_l)^\top \Lambda^{l,r} 1^q, \end{aligned}\tag{9}$$

where 1^q is a column vector of all ones in R^q; and $\Lambda^{l,r} = [\lambda_{\imath\jmath}^{l,r}]$ is a $q \times q$ matrix whose $(\imath, \jmath)$th element is defined as

$$\lambda_{\imath\jmath}^{l,r} = \mathrm{Cov}\{\hat{y}_\imath^l, (\hat{y}_\jmath^r)^2\}. \tag{10}$$

Substituting (7) and (9) into (6) gives

$$\begin{aligned} G_2(\alpha, \zeta) = 4\sum_{l=1}^{p}\sum_{r=1}^{p} y(\tau_l)^\top \Gamma^{l,r} y(\tau_r) - 4\sum_{l=1}^{p}\sum_{r=1}^{p} y(\tau_l)^\top \Lambda^{l,r} 1^q \\ + \mathrm{Var}\Bigg\{\sum_{l=1}^{p} (\hat{y}^l)^\top \hat{y}^l\Bigg\}. \end{aligned}\tag{11}$$

Since the last terms on the right-hand side of (5) and (11) are independent of decision vectors α and ζ, Problem A and Problem B are, respectively, equivalent to the following problems:

Problem C. *Given system* (1), *choose an admissible delay-parameter pair* $(\alpha, \zeta) \in \mathcal{D} \times \mathcal{Z}$ *to minimize*

$$H_1(\alpha, \zeta) = \sum_{l=1}^{p} y(\tau_l)^\top y(\tau_l) - 2\sum_{l=1}^{p} y(\tau_l)^\top \mathrm{E}\{\hat{y}^l\}, \tag{12}$$

where $y(\cdot) = y(\cdot|\alpha, \zeta)$.

Problem D. *Given system* (1), *choose an admissible delay-parameter pair* $(\alpha, \zeta) \in \mathcal{D} \times \mathcal{Z}$ *to minimize*

$$H_2(\alpha, \zeta) = 4\sum_{l=1}^{p}\sum_{r=1}^{p} y(\tau_l)^\top \Gamma^{l,r} y(\tau_r) - 4\sum_{l=1}^{p}\sum_{r=1}^{p} y(\tau_l)^\top \Lambda^{l,r} 1^q \tag{13}$$

subject to the constraint

$$G_1(\alpha, \zeta) \leq (1 + \beta) G_1(\tilde{\alpha}^*, \tilde{\zeta}^*), \tag{14}$$

where $y(\cdot) = y(\cdot|\alpha, \zeta)$*;* $\beta > 0$*; and* $(\tilde{\alpha}^*, \tilde{\zeta}^*)$ *is the optimal solution of* Problem C.

Note that constraint (14) is to specify the discrepancy between the optimal delay-parameter pair of Problem D and that of Problem C.

4. Computational algorithms

Problems C and D can be regarded as nonlinear dynamic optimization problems with decision vectors α and ζ. It is well known that gradient-based optimization methods, e.g., sequential quadratic programming (SQP) algorithm [32], are effective methods for solving nonlinear dynamic optimization problems [33]. However, such methods require the gradients of the cost function (and the gradients of the constraint functions, if applicable). But since the cost functions in Problem C and Problem D, and the constraint function in Problem D are implicit (rather than explicit) functions of the decision vectors α and ζ, it is not obvious how to determine their gradients. In this section, we will determine these required gradients and develop computational algorithms to solve Problem C and Problem D.

Let

$$\tilde{f}^i(t|\alpha,\zeta) = f^i(t, x(t), x(t-\alpha_1), \ldots, x(t-\alpha_m), \zeta), \tag{15}$$

$$\chi(t|\alpha,\zeta) = \begin{cases} \dfrac{\partial \phi(t,\zeta)}{\partial t}, & \text{if } t \leq 0, \\ \tilde{f}^i(t|\alpha,\zeta), & \text{if } t \in (t_{i-1}, t_i] \text{ for some } i \in \{1,2,\ldots,N\}. \end{cases} \tag{16}$$

Clearly, for almost all $t \in R$, we have $\dot{x}(t|\alpha,\zeta) = \chi(t|\alpha,\zeta)$. In the following, we will use $\partial \tilde{x}^j$ to denote partial differentiation with respect to $x(t-\alpha_j)$.

For each $j = 1,2,\ldots,m$ and a given $(\alpha,\zeta) \in \mathcal{D} \times \mathcal{Z}$, we consider the following auxiliary time-delay system:

$$\begin{aligned} \dot{\Xi}^j(t) = {} & \frac{\partial \tilde{f}^i(t|\alpha,\zeta)}{\partial x} \Xi^j(t) + \sum_{j=1}^{m} \frac{\partial \tilde{f}^i(t|\alpha,\zeta)}{\partial \tilde{x}^j} \Xi^j(t-\alpha_j) \\ & - \frac{\partial \tilde{f}^i(t|\alpha,\zeta)}{\partial \tilde{x}^j} \chi(t-\alpha_j|\alpha,\zeta), \quad t \in (t_{i-1}, t_i), \quad i = 1,2,\ldots,N, \end{aligned} \tag{17a}$$

with the intermediate conditions

$$\Xi^j(t_i+) = \Xi^j(t_i-), \quad i = 0,1,\ldots,N, \tag{17b}$$

and the initial condition

$$\Xi^j(t) = 0, \quad t \leq 0. \tag{17c}$$

Let $\Xi^j(\cdot|\alpha,\zeta)$ be the unique continuous solution of auxiliary system (17) corresponding to each delay-parameter pair $(\alpha,\zeta) \in \mathcal{D} \times \mathcal{Z}$. We have the following result.

Theorem 1. *For each* $(\alpha,\zeta) \in \mathcal{D} \times \mathcal{Z}$,

$$\frac{\partial x(t|\alpha,\zeta)}{\partial \alpha_j} = \Xi^j(t|\alpha,\zeta), \quad t \in (-\infty, T], \quad j = 1,2,\ldots,m. \tag{18}$$

Proof. Since (α, ζ) is fixed throughout this proof, we write $x(\cdot)$ instead of $x(\cdot|\alpha, \zeta)$ for brevity.

First, note that

$$\frac{\partial x(t)}{\partial \alpha_j} = \frac{\partial}{\partial \alpha_j}\Big\{\phi(t, \zeta)\Big\} = 0, \quad t \leq 0. \tag{19}$$

Thus, $\partial x(\cdot|\alpha, \zeta)/\partial \alpha_j$ satisfies the initial condition (17c).

Now, by (1),

$$x(t) = \phi(0, \zeta) + \sum_{\imath=1}^{i-1} \int_{t_{\imath-1}}^{t_\imath} \tilde{f}^\imath(s|\alpha, \zeta) ds + \int_{t_{i-1}}^{t} \tilde{f}^i(s|\alpha, \zeta) ds,$$
$$t \in (t_{i-1}, t_i), \quad i = 1, 2, \ldots, N, \tag{20}$$

where $\tilde{f}^i(\cdot|\alpha, \zeta)$ is as defined in (15). It can be shown that for each fixed $t \in (-\infty, T]$, $x(t|\alpha, \zeta)$ is a continuously differentiable function of α_j, $j = 1, \ldots, m$ (see [28]). Hence, by using Leibniz's rule to differentiate (20) with respect to α_j, we obtain

$$\begin{aligned}\frac{\partial x(t)}{\partial \alpha_j} = &\sum_{\imath=1}^{i-1} \int_{t_{\imath-1}}^{t_\imath} \Bigg\{ \frac{\partial \tilde{f}^\imath(s|\alpha, \zeta)}{\partial x} \frac{\partial x(s)}{\partial \alpha_j} + \sum_{\jmath=1}^{m} \frac{\partial \tilde{f}^\imath(s|\alpha, \zeta)}{\partial \tilde{x}^\jmath} \frac{\partial x(s - \alpha_\jmath)}{\partial \alpha_j} \\ &- \frac{\partial \tilde{f}^\imath(s|\alpha, \zeta)}{\partial \tilde{x}^j} \chi(s - \alpha_j|\alpha, \zeta) \Bigg\} ds + \int_{t_{i-1}}^{t} \Bigg\{ \frac{\partial \tilde{f}^i(s|\alpha, \zeta)}{\partial x} \frac{\partial x(s)}{\partial \alpha_j} \\ &+ \sum_{\jmath=1}^{m} \frac{\partial \tilde{f}^i(s|\alpha, \zeta)}{\partial \tilde{x}^\jmath} \frac{\partial x(s - \alpha_\jmath)}{\partial \alpha_j} - \frac{\partial \tilde{f}^i(s|\alpha, \zeta)}{\partial \tilde{x}^j} \chi(s - \alpha_j|\alpha, \zeta) \Bigg\} ds,\end{aligned}$$

where $\chi(\cdot|\alpha, \zeta)$ is as defined in (16). Differentiating the above equation with respect to time yields

$$\begin{aligned}\frac{d}{dt}\left\{\frac{\partial x(t)}{\partial \alpha_j}\right\} = &\frac{\partial \tilde{f}^i(t|\alpha, \zeta)}{\partial x} \frac{\partial x(t)}{\partial \alpha_j} + \sum_{\jmath=1}^{m} \frac{\partial \tilde{f}^i(t|\alpha, \zeta)}{\partial \tilde{x}^\jmath} \frac{\partial x(t - \alpha_\jmath)}{\partial \alpha_j} \\ &- \frac{\partial \tilde{f}^i(t|\alpha, \zeta)}{\partial \tilde{x}^j} \chi(t - \alpha_j|\alpha, \zeta), t \in (t_{i-1}, t_i), i = 1, 2, \ldots, N.\end{aligned} \tag{21}$$

Furthermore, by (1b), it is obvious that

$$\frac{\partial x(t_i+)}{\partial \alpha_j} = \frac{\partial x(t_i-)}{\partial \alpha_j}, \quad i = 0, 1, \ldots, N. \tag{22}$$

Equations (19),(21), and (22) show that $\partial x(\cdot|\alpha, \zeta)/\partial \alpha_j$ is the solution of auxiliary system (17), as required. $\square$

On the basis of Theorem 1, we obtain the following result which gives the gradients of H_1 and H_2 with respect to α.

Theorem 2. *Let $(\alpha,\zeta) \in \mathcal{D} \times \mathcal{Z}$. Then, for $j = 1, 2, \ldots, m$,*

$$\frac{\partial H_1(\alpha,\zeta)}{\partial \alpha_j} = 2\sum_{l=1}^{p} \varphi^l(\alpha,\zeta)^\top \frac{\partial g(\tau_l, x(\tau_l|\alpha,\zeta),\zeta)}{\partial x} \Xi^j(\tau_l|\alpha,\zeta), \tag{23}$$

$$\frac{\partial H_2(\alpha,\zeta)}{\partial \alpha_j} = 4\sum_{l=1}^{p} \psi^l(\alpha,\zeta)^\top \frac{\partial g(\tau_l, x(\tau_l|\alpha,\zeta),\zeta)}{\partial x} \Xi^j(\tau_l|\alpha,\zeta), \tag{24}$$

where

$$\varphi^l(\alpha,\zeta) = g(\tau_l, x(\tau_l|\alpha,\zeta),\zeta)) - \mathrm{E}\{\hat{y}^l\}; \tag{25}$$

and

$$\psi^l(\alpha,\zeta) = \sum_{r=1}^{p} \left\{ 2\Gamma^{l,r} g(\tau_r, x(\tau_r|\alpha,\zeta),\zeta) - \Lambda^{l,r} 1^q \right\}. \tag{26}$$

Proof. Differentiate H_1 and H_2 with respect to α_j and then applying (18) and the symmetry of $\Gamma^{l,r}$. □

From Theorem 2, we can see that the gradients of H_1 and H_2 with respect to time-delays can be computed by solving a set of auxiliary systems (17) together with system (1), simultaneously. Next, we will investigate the gradients of H_1 and H_2 with respect to system parameters. To this end, for each $k = 1, 2, \ldots, v$ and a given $(\alpha,\zeta) \in \mathcal{D} \times \mathcal{Z}$, we consider the following auxiliary time-delay system:

$$\begin{aligned}\dot{\Psi}^k(t) = {} & \frac{\partial \tilde{f}^i(t|\alpha,\zeta)}{\partial x} \Psi^k(t) + \sum_{\jmath=1}^{m} \frac{\partial \tilde{f}^i(t|\alpha,\zeta)}{\partial \tilde{x}^\jmath} \Psi^k(t-\alpha_\jmath) \\ & + \frac{\partial \tilde{f}^i(t|\alpha,\zeta)}{\partial \zeta_k}, \quad t \in (t_{i-1}, t_i), \quad i = 1, 2, \ldots, N, \end{aligned} \tag{27a}$$

with the intermediate conditions

$$\Psi^k(t_i+) = \Psi^k(t_i-), \quad i = 0, 1, \ldots, N, \tag{27b}$$

and the initial condition

$$\Psi^k(t) = \frac{\partial \phi(t,\zeta)}{\partial \zeta_k}, \quad t \le 0. \tag{27c}$$

Let $\Psi^k(\cdot|\alpha,\zeta)$ be the solution of system (27) corresponding to each delay-parameter pair $(\alpha,\zeta) \in \mathcal{D} \times \mathcal{Z}$. Then, we have the following result.

Theorem 3. *For each $(\alpha,\zeta) \in \mathcal{D} \times \mathcal{Z}$,*

$$\frac{\partial x(t|\alpha,\zeta)}{\partial \zeta_k} = \Psi^k(t|\alpha,\zeta), \quad t \in (-\infty, T], \quad k = 1, 2, \ldots, v.$$

Proof. The proof is similar to that given for Theorem 1. □

On the basis of Theorem 3, we present the gradients of H_1 and H_2 with respect to ζ in the following theorem.

Theorem 4. *Let $(\alpha, \zeta) \in \mathcal{D} \times \mathcal{Z}$. Then, for $k = 1, 2, \ldots, v$,*

$$\frac{\partial H_1(\alpha, \zeta)}{\partial \zeta_k} = 2 \sum_{l=1}^{p} \varphi^l(\alpha, \zeta)^\top \left\{ \frac{\partial g(\tau_l, x(\tau_l), \zeta)}{\partial x} \Psi^k(\tau_l) + \frac{\partial g(\tau_l, x(\tau_l), \zeta)}{\partial \zeta_k} \right\}, \quad (28)$$

$$\frac{\partial H_2(\alpha, \zeta)}{\partial \zeta_k} = 4 \sum_{l=1}^{p} \psi^l(\alpha, \zeta)^\top \left\{ \frac{\partial g(\tau_l, x(\tau_l), \zeta)}{\partial x} \Psi^k(\tau_l) + \frac{\partial g(\tau_l, x(\tau_l), \zeta)}{\partial \zeta_k} \right\}, \quad (29)$$

where $x(\cdot) = x(\cdot|\alpha, \zeta)$; $\Psi^k(\cdot) = \Psi^k(\cdot|\alpha, \zeta)$; and $\varphi^l(\alpha, \zeta)$ and $\psi^l(\alpha, \zeta)$ are as defined in (25) *and* (26), *respectively.*

Proof. The proof is similar to that given for Theorem 2. □

Based on Theorems 2 and 4, we now present the following algorithm for solving Problem C:

Algorithm 1. Step 1. Choose an initial delay-parameter pair $(\alpha, \zeta) \in \mathcal{D} \times \mathcal{Z}$.
Step 2. Compute $H_1(\alpha, \zeta)$ according to (12).
Step 3. Compute $\partial H_1(\alpha, \zeta)/\partial \alpha_j$, $j = 1, 2, \ldots, m$, and $\partial H_1(\alpha, \zeta)/\partial \zeta_k$, $k = 1, 2, \ldots, v$, using (23) and (28), respectively.
Step 4. If (α, ζ) is optimal, then $(\alpha, \zeta) \to (\tilde{\alpha}^*, \tilde{\zeta}^*)$ and stop. Otherwise, use the gradient information obtained in Step 3 to compute a descent direction.
Step 5. Perform a line search along this direction to obtain a new pair $(\alpha', \zeta') \in \mathcal{D} \times \mathcal{Z}$.
Step 6. Set $(\alpha', \zeta') \to (\alpha, \zeta)$ and return to Step 2.

Remark 1. *Theorems 2 and 4 present the gradients of expectation and variance of the error function (3) with respect to each delay-parameter pair. Thus, if we additionally compute $H_2(\alpha, \zeta)$, $\partial H_2(\alpha, \zeta)/\partial \alpha_j$, $j = 1, 2, \ldots, m$, and $\partial H_2(\alpha, \zeta)/\partial \zeta_k$, $k = 1, 2, \ldots, v$, in Steps 2 and 3, then Algorithm 1 can be used to solve the parameter estimation problem, in which the cost function is the weighted sum of expectation and variance of the error function and dynamic system is a multistage time-delay system. This extension of Algorithm 1 is called Extended Algorithm 1 in the sequel. In particular, if such parameter estimation problem involves only single-stage time-delay systems as the one in [29, 30], then the Extended Algorithm 1 is reduced to the algorithm reported in [29, 30].*

Based on the obtained optimal delay-parameter pair $(\tilde{\alpha}^*, \tilde{\zeta}^*)$ in Algorithm 1, we present the following algorithm for solving Problem D:

Algorithm 2. Step 1. Set a weighting factor $\beta > 0$ and choose an initial delay-parameter pair $(\alpha, \zeta) \in \mathcal{D} \times \mathcal{Z}$.
Step 2. Compute $H_2(\alpha, \zeta)$ and the constraint function according to (13) and (14).
Step 3. Compute $\partial H_1(\alpha, \zeta)/\partial \alpha_j$, $\partial H_2(\alpha, \zeta)/\partial \alpha_j$, $j = 1, 2, \ldots, m$, and $\partial H_1(\alpha, \zeta)/\partial \zeta_k$, $\partial H_2(\alpha, \zeta)/\partial \zeta_k$, $k = 1, 2, \ldots, v$, using (23), (24), (28) and (29), respectively.
Step 4. If (α, ζ) is optimal, then stop. Otherwise, use the gradient information obtained in Step 3 to compute a descent direction.
Step 5. Perform a line search along this direction to obtain a new pair $(\alpha'', \zeta'') \in$

$\mathcal{D} \times \mathcal{Z}$.
Step 6. Set $(\alpha'', \zeta'') \to (\alpha, \zeta)$ and return to Step 2.

Remark 2. *In Algorithm 1 and Algorithm 2, the gradients are computed in a unified manner in the sense that the multistage time-delay system (1), the auxiliary systems (17) and (27) are solved simultaneously forward in time. This makes the algorithms very convenient to be implemented.*

Remark 3. *In Algorithm 1 and Algorithm 2, Steps 4 and 5 can be implemented using a standard nonlinear programming method such as SQP or the conjugate gradient method [32].*

5. Numerical examples

In this section, we solve two numerical examples using our proposed computational algorithms in Section 4. Example 1 is a parameter estimation problem for a single-stage time-delay system first considered in [30]. Example 2 is a time-delay estimation problem for a multistage time-delay system. To solve these problems, we wrote a Fortran program that implemented Algorithm 1, Extended Algorithm 1 and Algorithm 2 with the optimization software NLPQLP [34], which is a Fortran implementation of SQP algorithm. This program invokes the differential equation software LSODA [35] to solve the expanded multistage time-delay systems. Lagrange interpolation [36] is used when LSODA requires the value of the state at an intermediate time between two adjacent knot points.

5.1. Example 1

Consider a continuously-stirred tank reactor in [30]. The reaction dynamics are described by the following single-stage time-delay system:

$$\begin{cases} \dot{x}_1(t) = \zeta_1 x_1(t) - (1 - x_1(t)) \exp\left[\dfrac{20x_2(t)}{x_2(t)+20}\right] + x_1(t-\alpha), \\ \dot{x}_2(t) = \zeta_2 x_2(t) - (1 - x_1(t)) \exp\left[\dfrac{20x_2(t)}{x_2(t)+20}\right] + x_2(t-\alpha), \end{cases} \tag{30a}$$

with initial conditions

$$x_1(t) = 1, \quad x_2(t) = 1, 2, \quad t \le 0, \tag{30b}$$

where x_1 is the dimensionless concentration of reactant; x_2 is the dimensionless temperature of the reactor; and α, ζ_1, and ζ_2 are unknown model parameters to be estimated.

As in [30], the system output is given by

$$y(t) = 10x_2(t), \quad t \ge 0.$$

The least-squares error function is

$$J(\alpha, \zeta_1, \zeta_2) = \sum_{l=1}^{20} \left\{ y(\tau_l) - \hat{y}^l \right\}^2 = \sum_{l=1}^{20} \left\{ 10x_2(\tau_l) - \hat{y}^l \right\}^2, \tag{31}$$

where $\hat{y}^l$ is a random variable representing the output measurement at the lth sample time; and τ_l is the lth sample time given by

$$\tau_l = \frac{1}{2}l, \quad l = 1, 2, \ldots, 20.$$

Based on the discussion in Section 3, the classical parameter estimation problem can be reformulated as follows: given system (30), choose α, ζ_1, and ζ_2 to minimize

$$H_1(\alpha, \zeta_1, \zeta_2) = \sum_{l=1}^{20} \Big\{ y(\tau_l)^2 - 2y(\tau_l)\mathrm{E}\{\hat{y}^l\} \Big\}. \tag{32}$$

From (5), we note that

$$G_1(\alpha, \zeta_1, \zeta_2) = \sum_{l=1}^{20} \Big\{ y(\tau_l)^2 - 2y(\tau_l)\mathrm{E}\{\hat{y}^l\} + \mathrm{E}\{(\hat{y}^l)^2\} \Big\}. \tag{33}$$

Accordingly, our parameter estimation problem can be reformulated as follows: given system (30), choose α, ζ_1, and ζ_2 such that

$$H_2(\alpha, \zeta_1, \zeta_2) = 4 \sum_{l=1}^{20} \sum_{r=1}^{20} \Big\{ y(\tau_l)y(\tau_r)\Gamma^{l,r} - y(\tau_l)\Lambda^{l,r} \Big\} \tag{34}$$

is minimized subject to the constraint

$$G_1(\alpha, \zeta_1, \zeta_2) \leq (1+\beta)G_1(\tilde{\alpha}^*, \tilde{\zeta}_1^*, \tilde{\zeta}_2^*),$$

where $\Gamma^{l,r} = \mathrm{Cov}\{\hat{y}^l, \hat{y}^r\}$, $\Lambda^{l,r} = \mathrm{Cov}\{\hat{y}^l, (\hat{y}^r)^2\}$, $\beta > 0$, and $\tilde{\alpha}^*$, $\tilde{\zeta}_1^*$, and $\tilde{\zeta}_2^*$ are the optimal parameter estimates of the classical parameter estimation problem.

As in [30], consider the output trajectory of system (30) corresponding to the nominal parameter estimates

$$(\alpha, \zeta_1, \zeta_2) = (2, -2, -\frac{5}{2}).$$

This output trajectory is known as reference trajectory. To generate the output data for numerical experiments, we randomly perturbed the reference trajectory using independent normal random variables. The sample points obtained are shown in Figure 1. Moreover, the noisy measurement outputs take the following form:

$$\hat{y}^l = \sigma_l + \sum_{\ell=1}^{l} \rho_\ell, \quad l = 1, 2, \ldots, 20, \tag{35}$$

where σ_l and ρ_ℓ are independent random variables. Here, σ_l follows a gamma distribution with parameters 10 and $\bar{y}^l/10$, where $\bar{y}^l$ is the lth sample point in Figure 1, and $\rho_\ell = \bar{\rho}_\ell - \mathrm{E}\{\bar{\rho}_\ell\}$, where $\bar{\rho}_\ell$ follows a beta distribution with parameters 2 and 5. By using (35), we generate 100 random data at each sample

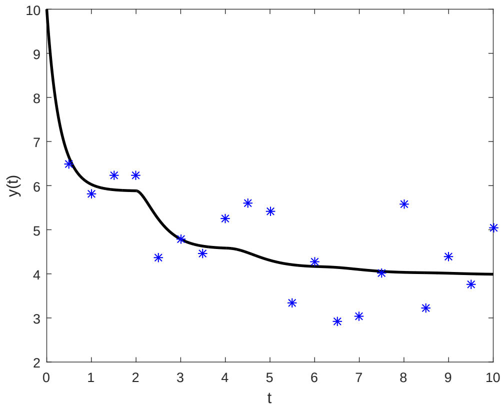


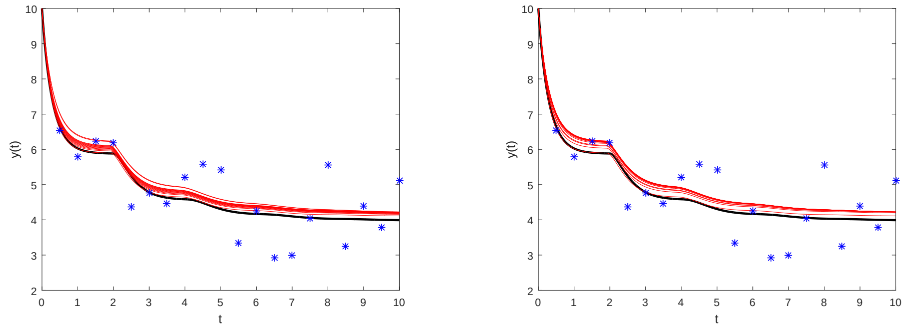
Figure 1: Reference output trajectory (black line) and corresponding perturbed sample points (blue stars) for Example 1.

time as the observed data. For the observed data, the mean and covariance matrices in (32)-(34) are computed by statistic method.

Using our Fortran program, we first solve the classical parameter estimation problem and obtain the optimal parameter estimates. For notation simplicity, these optimal parameter estimates are denoted as for $\beta = 0$. Then, we solve our parameter estimation problem for $\beta = 0.001, 0.002, \ldots, 1.0$. The optimal parameter estimates for $\beta = 0,\ 0.001,\ \ldots,\ 0.009,\ 0.01,\ \ldots,\ 0.03$ are listed in Table 1. Note that the optimal parameter estimates for $\beta > 0.03$ are same as the ones for $\beta = 0.03$. This indicates that the maximal allowable sacrifice of the expectation value is 3% of the optimal expectation value of the classical parameter estimation problem. For comparison, we also consider the parameter estimation formulation in [29, 30], that is, given system (30), choose α, ζ_1, and ζ_2 to minimize

$$(1-\beta_1)H_1(\alpha,\zeta_1,\zeta_2) + \beta_1 H_2(\alpha,\zeta_1,\zeta_2), \tag{36}$$

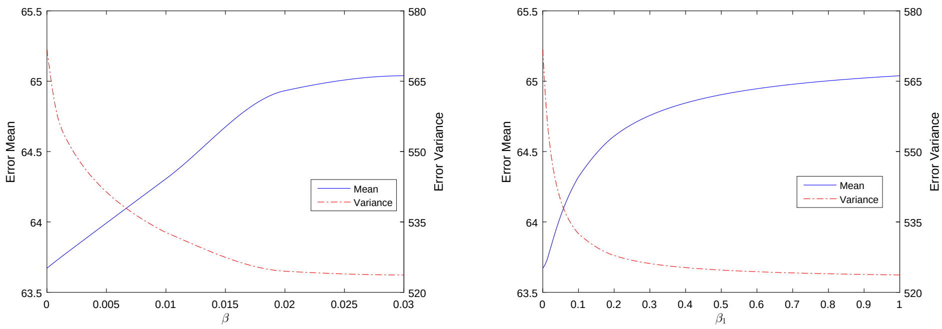
where $\beta_1 \in [0, 1]$ is a weighting factor. This estimation problem is solved for $\beta_1 = 0, 0.01, \ldots, 1.0$ using our Fortran program, in which the Extended Algorithm 1 is involved and it is reduced to the algorithm report in [29, 30]. The optimal parameter estimates for $\beta_1 = 0, 0.04, 0.08, 0.1, \ldots, 1.0$ are also listed in Table 1. From Table 1, we can see that the optimal parameter estimates for $\beta = 0$ and $\beta = 0.03$ of the classical and our estimation formulations are, respectively, same as the ones for $\beta_1 = 0$ and $\beta_1 = 1.0$ of the estimation formulation in [29, 30]. The output trajectories corresponding to the optimal parameter estimates in Table 1 are shown in Figure 2. From Figure 2, we can see that these output trajectories (red lines) converge to the reference output trajectory (black line) as the values of β and β_1 decease. Note that, due to the uncertain nature of the output measurements, the optimal output trajectory for $\beta = \beta_1 = 0$ is not identical with the reference output trajectory.



(a) Optimal parameter estimates (Algorithm 1 or 2).

(b) Optimal parameter estimates (Algorithm in [29, 30]).

Figure 2: Output trajectories corresponding to the optimal parameter estimates in Table 1.



(a) Optimal parameter estimates (Algorithm 1 or 2).

(b) Optimal parameter estimates (Algorithm in [29, 30]).

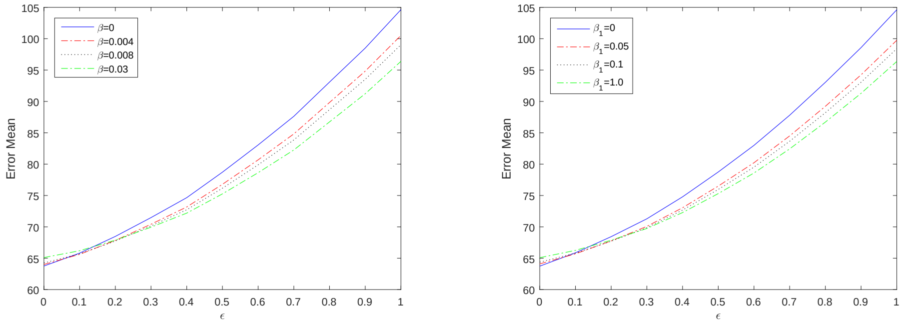
Figure 3: Means and variances of the least-squares errors for 100,000 output realizations generated according to (35).

Table 1: Optimal parameter estimates for Example 1.

Optimal parameter estimates (Algorithm 1 or 2)				Optimal parameter estimates (Algorithm in [29, 30])			
β	α^*	ζ_1^*	ζ_2^*	β_1	α^*	ζ_1^*	ζ_2^*
0	1.9245	-2.1298	-2.5687	0	1.9245	-2.1298	-2.5687
0.001	1.9189	-2.0895	-2.5232	0.04	1.9100	-2.0227	-2.4570
0.002	1.9160	-2.0678	-2.5010	0.08	1.9019	-1.9638	-2.4016
0.003	1.9135	-2.0493	-2.4826	0.1	1.8990	-1.9443	-2.3834
0.004	1.9113	-2.0324	-2.4662	0.2	1.8907	-1.8893	-2.3323
0.005	1.9092	-2.0165	-2.4511	0.3	1.8866	-1.8638	-2.3086
0.006	1.9071	-2.0015	-2.4368	0.4	1.8841	-1.8491	-2.2949
0.007	1.9051	-1.9871	-2.4233	0.5	1.8825	-1.8396	-2.2860
0.008	1.9032	-1.9733	-2.4103	0.6	1.8813	-1.8329	-2.2798
0.009	1.9013	-1.9599	-2.3979	0.7	1.8805	-1.8280	-2.2751
0.01	1.8994	-1.9469	-2.3858	0.8	1.8798	-1.8242	-2.2715
0.02	1.8817	-1.8351	-2.2818	0.9	1.8793	-1.8211	-2.2687
0.03	1.8788	-1.8187	-2.2664	1.0	1.8788	-1.8187	-2.2664

To investigate solution robustness, we generate 100,000 realizations of the noisy measurement outputs $\hat{y}^l$, $l = 1, 2, \ldots, 20$. For each realization, we compute the least-squares error (31) corresponding to the optimal parameter estimates for $\beta = 0$, 0.001, ..., 0.03 of the classical and our estimation formulations. For comparison, we also compute the least-squares error (31) corresponding to the optimal parameter estimates for $\beta_1 = 0$, 0.01, ..., 1.0 of the estimation formulation in [29, 30]. Figure 3 shows the means and variances of the least-squares errors for 100000 output realizations. From Figure 3, we can see that the dashed red curves (error variances) are much steeper than the solid blue curves (error means) near $\beta = \beta_1 = 0$. This shows the advantages of both our estimation formulation and the estimation formulation in [29, 30]: the solution robustness increases significantly at the expense of a negligible cost to the error mean. It is important to note that the value of β in our estimation formulation provides useful quantitative information on the allowable sacrifice of the expectation value from the optimal expectation value of the classical parameter estimation problem. On the other hand, the estimation formulation in [29, 30] does not provide such important information.

The obtained optimal parameter estimates by our program are based on the output distribution specified by (35). We now investigate how these optimal parameter estimates perform when the actual distribution differs from the assumed distribution. Thus, instead of (35), we now suppose that the output



(a) Optimal parameter estimates (Algorithm 1 or 2).

(b) Optimal parameter estimates (Algorithm in [29, 30]).

Figure 4: Mean variations of the least-squares error for 100,000 output realizations with respect to ϵ.

distribution is given by

$$\hat{y}^l = \tilde{\sigma}_l + \sum_{\ell=1}^{l} \rho_\ell, \quad l = 1, 2, \ldots, 20, \tag{37}$$

where $\tilde{\sigma}_l$ and ρ_l are independent random variables. Here, ρ_ℓ is as defined in (35) and $\tilde{\sigma}_l$ follows a gamma distribution with parameters 10 and $(\bar{y}^l + \epsilon)/10$, where ϵ is a small parameter. The parameter ϵ ensures that $\tilde{\sigma}_l$ differs from σ_l in (35). For each $\epsilon = 0, 0.1, \ldots, 1.0$, we generated 100000 realizations of the output data according to (37) and calculated the means of the least-squares error according to the optimal parameter estimates for $\beta = 0, 0.004, 0.008, 0.03$ of the classical and our estimation formulations. For comparison, we also calculated the means of the least-squares error according to the optimal parameter estimates for $\beta_1 = 0, 0.05, 0.1, 1.0$ of the estimation formulation in [29, 30]. These mean changes with respect to ϵ are all plotted in Figure 4. Note that, as expected, the error means of both the estimation formulations corresponding to the optimal parameter estimates for $\beta > 0$ and $\beta_1 > 0$ vary more slowly than the ones corresponding to the optimal parameter estimates for $\beta = \beta_1 = 0$ do. This shows that both our estimation formulation and the estimation formulation in [29, 30] are robust facing uncertainty in the output distribution. In particular, compared with the estimation formulation in [29, 30], our parameter estimation formulation provides explicit information on the allowable sacrifice of the expectation value from the optimal expectation value of the classical parameter estimation problem.

5.2. Example 2

Consider the following multistage time-delay system with 2 stages and 2 time-delays:

$$\begin{cases} \dot{x}_1(t) = -0.7x_1(t-\alpha_1) - 3.5x_2(t) + 2x_2(t-\alpha_2) \\ \qquad\quad +0.1\tanh(x_1(t)), \\ \dot{x}_2(t) = 0.7x_1(t) - 6.7x_2(t) - \sin(x_2(t-\alpha_2)), \end{cases} \quad t \in (0, 1.0), \tag{38a}$$

$$\begin{cases} \dot{x}_1(t) = -4x_1(t-\alpha_1) + 0.5x_2(t) + 0.2\sin(x_2(t)) \\ \qquad\quad +t^2 + 8, \\ \dot{x}_2(t) = 4.7x_1(t) - 5.6x_2(t) + 0.5\sin(x_1(t-\alpha_2)), \end{cases} \quad t \in (1.0, 1.5), \tag{38b}$$

with initial conditions

$$x_1(t) = 6, \quad x_2(t) = t^2 + 2, \quad t \leq 0, \tag{38c}$$

where α_1 and α_2 are unknown time-delays that need to be estimated. We assume that both α_1 and α_2 lie within the interval $[0.01, 2]$.

The system output is given by

$$y(t) = x_1(t), \quad t \geq 0.$$

Thus, the least-squares error function is

$$J(\alpha_1, \alpha_2) = \sum_{l=1}^{p} \left\| x_1(\tau_l) - \hat{y}^l \right\|^2, \tag{39}$$

where p is the number of sample times; τ_l is the lth sample time; and $\hat{y}^l$ is a random variable representing the output measurement at the lth sample time. We choose $p = 10$ and

$$\tau_l = 0.15l - 0.075, \quad l = 1, 2, \ldots, 10.$$

Based on the discussion in Section 3, the classical parameter estimation problem can be reformulated as follows: given multistage system (38), choose α_1 and α_2 to minimize

$$H_1(\alpha_1, \alpha_2) = \sum_{l=1}^{10} \Big\{ y(\tau_l)^2 - 2y(\tau_l)\mathrm{E}\{\hat{y}^l\} \Big\}, \tag{40}$$

subject to bounds $\alpha_1, \alpha_2 \in [0.01, 2]$. Accordingly,

$$G_1(\alpha_1, \alpha_2) = \sum_{l=1}^{10} \Big\{ y(\tau_l)^2 - 2y(\tau_l)\mathrm{E}\{\hat{y}^l\} + \mathrm{E}\{(\hat{y}^l)^2\} \Big\}. \tag{41}$$

Furthermore, our estimation problem can be reformulated as follows: given system (38), choose α_1 and α_2 such that

$$H_2(\alpha_1, \alpha_2) = 4\sum_{l=1}^{10}\sum_{r=1}^{10} \Big\{ y(\tau_l)y(\tau_r)\Gamma^{l,r} - y(\tau_l)\Lambda^{l,r} \Big\} \tag{42}$$

is minimized subject to

$$G_1(\alpha_1, \alpha_2) \leq (1+\beta) G_1(\tilde{\alpha}_1^*, \tilde{\alpha}_2^*),$$

and $\alpha_1, \alpha_2 \in [0.01, 2]$, where $\Gamma^{l,r} = \text{Cov}\{\hat{y}^l, \hat{y}^r\}$, $\Lambda^{l,r} = \text{Cov}\{\hat{y}^l, (\hat{y}^r)^2\}$, $\beta > 0$, and $\tilde{\alpha}_1^*$ and $\tilde{\alpha}_2^*$ are the optimal solutions of the classical parameter estimation problem.

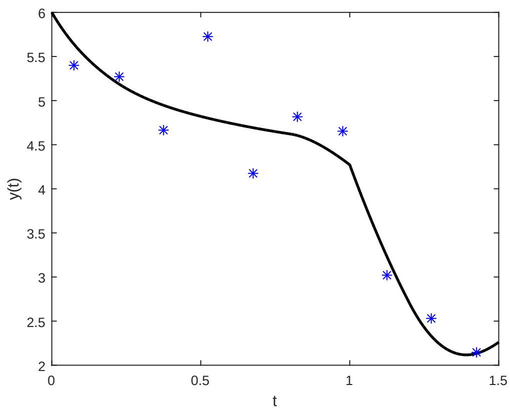


Figure 5: Reference output trajectory (black line) and corresponding perturbed sample points (blue stars) for Example 2.

To generate the reference trajectory, we simulate the multistage system (38) with $\alpha_1 = 0.2$ and $\alpha_2 = 0.8$. Moreover, the output data at sample times are obtained by adding independent normal variables to the values on the reference trajectory. The sample output data at the sample times are shown as in Figure 5. Now, suppose that the noisy measurement outputs take the following form:

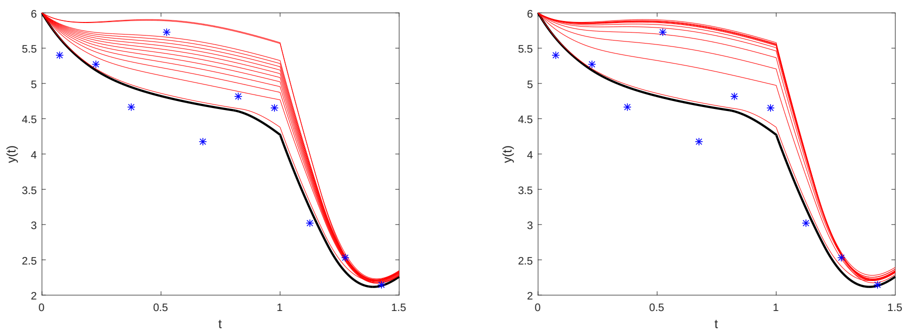
$$\hat{y}^l = \sigma_l + \sum_{\ell=1}^{l} \rho_\ell, \quad l = 1, 2, \ldots, 10, \tag{43}$$

where σ_l and ρ_ℓ are independent random variables. In addition, we assume that σ_l follows a gamma distribution with parameters 6 and $\bar{y}^l/6$, where $\bar{y}^l$ is the lth sample point in Figure 5. We also assume that $\rho_\ell = \bar{\rho}_\ell - \text{E}\{\bar{\rho}_\ell\}$, where $\bar{\rho}_\ell$ follows a beta distribution with parameters 2 and 5. By using (43), we generate 200 random data at each sample time as the observed data. For the observed data, the mean and covariance matrices in (40)-(42) are computed by statistic method.

We first solve the classical parameter estimation problem using our Fortran program and obtain the optimal time-delay estimates. For notation simplicity, we also denote the optimal time-delay estimates as for $\beta = 0$. Then, we solve our parameter estimation problem for $\beta = 0.01, 0.02, \ldots, 1.0$. The optimal time-delay estimates for $\beta = 0, 0.01, \ldots, 0.09, 0.1, \ldots, 0.3$ are listed in Table 2. Note that the optimal time-delay estimates for $\beta > 0.3$ are same as the ones for $\beta = 0.3$. This indicates that the maximal allowable sacrifice of the expectation value is 30% of the optimal expectation value of the classical parameter estimation problem. Like the estimation formulation in [29, 30], we also consider the parameter estimation formulation, in which the cost function is the weighted sum of expectation and variance of the error function (39), that is, given multistage system (38), choose α_1 and α_2 to minimize

$$(1 - \beta_1)H_1(\alpha_1, \alpha_2) + \beta_1 H_2(\alpha_1, \alpha_2) \tag{44}$$

subject to constraints $\alpha_1, \alpha_2 \in [0.01, 2]$, where $\beta_1 \in [0, 1]$ is a weighting factor. This estimation problem is solved for $\beta_1 = 0.01, \ldots, 1.0$ using our Fortran program, where the Extended Algorithm 1 is involved. Note that, however, the algorithm reported in [29, 30] cannot be used to solve this estimation problem involving multistage time-delay system (38). The optimal time-delay estimates for $\beta_1 = 0, 0.04, 0.07, 0.1, \ldots, 1.0$ using this estimation formulation are also listed in Table 2. From Table 2, we can see that the optimal time-delay estimates for $\beta = 0$ and $\beta = 0.3$ using the classical and our estimation formulations are, respectively, same as the ones for $\beta_1 = 0$ and $\beta_1 = 1.0$ using this estimation formulation. The output trajectories corresponding to the optimal time-delay estimates in Table 2 are shown in Figure 6. From Figure 6, we can see that the output trajectories (red lines) converge to the reference output trajectory (black line) as the values of β and β_1 decease.



(a) Optimal time-delay estimates (Algorithm 1 or 2).

(b) Optimal time-delay estimates (Extended Algorithm 1).

Figure 6: Output trajectories corresponding to the optimal time-delay estimates in Table 2.

To investigate solution robustness, we generate 100,000 realizations of the noisy measurement outputs $\hat{y}^l$, $l = 1, 2, \ldots, 10$. For each realization, we compute the least-squares error (39) corresponding to the optimal time-delay estimates for $\beta = 0, 0.01, \ldots, 0.3$ and $\beta_1 = 0, 0.01, \ldots, 1.0$. Figure 7 shows the means and variances of the least-squares error for 100,000 output realizations. From Figure 7, we can see that the dashed red curves (error variances) are much

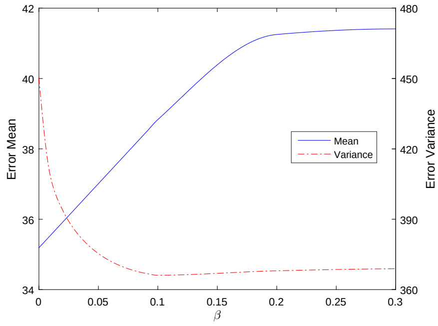
Table 2: Optimal time-delay estimates for Example 2.

Optimal time-delay estimates (Algorithm 1 or 2)			Optimal time-delay estimates (Extended Algorithm 1)		
β	α_1^*	α_2^*	β_1	α_1^*	α_2^*
0	0.1863	0.8368	0	0.1863	0.8368
0.01	0.1866	1.0874	0.04	0.1775	1.2512
0.02	0.1828	1.1777	0.07	0.1466	1.3305
0.03	0.1800	1.2388	0.1	0.1681	1.4004
0.04	0.1777	1.2867	0.2	0.1660	1.4891
0.05	0.1758	1.3267	0.3	0.1619	1.5378
0.06	0.1742	1.3613	0.4	0.1666	1.5606
0.07	0.1728	1.3920	0.5	0.1675	1.5748
0.08	0.1716	1.4196	0.6	0.1666	1.5801
0.09	0.1705	1.4447	0.7	0.1659	1.5800
0.1	0.1697	1.4679	0.8	0.1674	1.5867
0.2	0.1709	1.5893	0.9	0.1705	1.5855
0.3	0.1651	1.5949	1.0	0.1651	1.5949

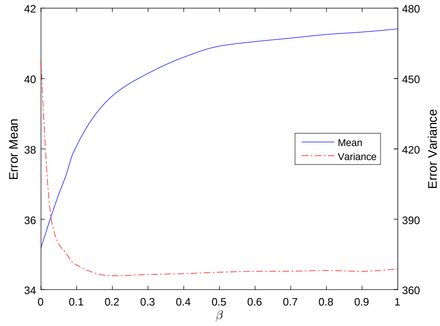
steeper than the solid blue curves (error means) near $\beta = \beta_1 = 0$. This shows the solutions of both our estimation formulation and the estimation formulation using cost function (44) are robust. Note that an additional benefit of our estimation formulation is that it explicitly specifies the allowable sacrifice from the optimal expectation value of the classical parameter estimation problem. Now, suppose that the output distribution is given by

$$\hat{y}^l = \tilde{\sigma}_l + \sum_{\ell=1}^{l} \rho_\ell, \quad l = 1, 2, \ldots, 10, \tag{45}$$

where $\tilde{\sigma}_l$ and ρ_l are independent random variables. Here, ρ_ℓ is as defined in (43) and $\tilde{\sigma}_l$ follows a gamma distribution with parameters 6 and $(\bar{y}^l + \epsilon)/6$, where ϵ is a small parameter. The parameter ϵ ensures that $\tilde{\sigma}_l$ differs from σ_l in (43). For each $\epsilon = 0, 0.1, \ldots, 1$, we generated 100000 realizations of the output data according to (45) and calculated the means of the least-squares errors according to the optimal time-delay estimates for $\beta = 0, 0.05, 0.09, 0.3$ of the classical and our estimation formulations. For comparison, we also calculated the means of the least-squares errors according to the optimal time-delay estimates for $\beta_1 = 0, 0.07, 0.2, 1.0$ of the estimation formulation using cost function (44). These mean changes with respect to the parameter ϵ are plotted in Figure 8. From Figure 8, we see that the bigger the values of β and β_1 are, the slower the changes of the error means. This indicates that the robustness of the optimal time-delay estimates increases as the values of β and β_1 increase when facing distribution uncertainty.



(a) Optimal time-delay estimates (Algorithm 1 or 2).



(b) Optimal time-delay estimates (Extended Algorithm 1).

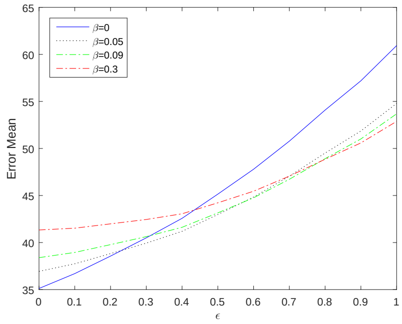
Figure 7: Means and variances of the least-squares errors for 100,000 output realizations generated according to (43).

6. Conclusions

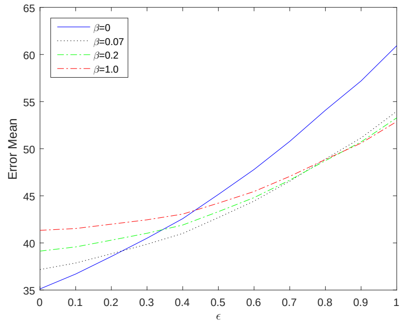
This paper has studied the parameter estimation problem in which the goal is to choose the robust optimal estimates for unknown time-delays and system parameter in multistage time-delay systems. This parameter estimation problem is an extension of the one formulated in [29,30] with two important differences: (i) it involves multistage time-delay system, not restricted to single-stage time-delay system; and (ii) it explicitly specifies the allowable sacrifice in the expectation value from the optimal expectation value of the classical parameter estimation problem. Numerical examples in Section 5 show that our parameter estimation approach is capable of solving parameter estimation problems with multiple stages and multiple time-delays, and, compared with the classical parameter estimation, it is able to withstand the uncertainty in the distribution of measurement data. Nevertheless, one of the limitations of the new parameter estimation method proposed in this paper is that it relies on the statistical distribution of the noisy measurement output. Future work will focus on extending this method to the case where the statistical distribution is unavailable and to consider the optimal parameter estimates which are robust with respect to statistical distribution.

Acknowledgments

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



(a) Optimal time-delay estimates (Algorithm 1 or 2).



(b) Optimal time-delay estimates (Extended Algorithm 1).

Figure 8: Mean variations of the least-squares error for 100,000 output realizations with respect to ϵ.

References

[1] A. Mazanov, A multistage population model, J. Theor. Biol., **39** (1973), 581–587.

[2] S.K. Goyal, A. Gunasekaran, Multi-stage production-inventory systems, Eur. J. Oper. Res., **46** (1990), 1–20.

[3] F. Deindoerfer, A. Humphrey, Design of multistage systems for simple fermentation processes, Ind. Eng. Chem., **51** (1959), 809–812.

[4] C. Liu, Sensitivity analysis and parameter identification for a nonlinear time-delay system in microbial fed-batch process, Appl. Math. Model., **38** (2014), 1449–1463.

[5] A. van den Bos, Parameter Estimation for Scientists and Engineers, John Wiley & Sons, New Jersey, 2007.

[6] Y. Bard, Nonlinear Parameter Estimation, Academic Press, New York, 1974.

[7] K. Schittkowski, Parameter estimation in differential equations, In: Recent Trends in Optimization Theory and Applications (R.P. Agarwal ed.), World Scientific Publishing Co., Singapore, 1995.

[8] H. Bock, E. Kostina, O. Kostyukova, Covariance matrices for parameter estimtes of constrained parameter estimation problems, SIAM J. Matrix Anal. Appl., **29** (2007), 626–642.

[9] F. Clark, R. Vinter, Optimal multiprocesses, SIAM J. Control Optim., **27** (1989), 1072–1091.

[10] V.S. Vassiliadis, R.W.H. Sargen, C.C. Pantelides, Solution of a class of multistage dynamic optimization problems: 1. Problem without path constraints, Ind. Eng. Chem. Res., **33** (1994), 2111–2122.

[11] V.S. Vassiliadis, R.W.H. Sargen, C.C. Pantelides, Solution of a class of multistage dynamic optimization problems: 2. Problem with path constraints, Ind. Eng. Chem. Res., **33** (1994), 2123–2133.

[12] E. Balsa-Canto, V.S. Vassiliadis, J.R. Banga, Dynamic optimization of single- and multi-stage systems using a hybrid stochastic-deterministic method, Ind. Eng. Chem. Res., **44** (2005), 1514–1523.

[13] Z. Gong, A multistage system of microbial fed-batch fermentation and its parameter identification, Math. Comput. Simul., **80** (2010), 1903–1910.

[14] Z. Jiang, J. Yuan, E. Feng, Robust identification and its properties of nonlinear bilevel multistage dynamic system, Appl. Math. Comput., **219** (2013), 6979–6985.

[15] H. Yin, J. Yuan, X. Zhu, J. Xie, E. Feng, Z. Xiu, Modeling and parameter identification for a nonlinear multi-stage system for dha regulon in batch culture, Appl. Math. Model., **40** (2016), 468–484.

[16] D. Debeljković, Time-Delay Systems, InTech, Rijeka, 2011.

[17] J.P. Richard, Time-delay systems: An overview of some recent advances and open problems, Automatica, **39** (2003), 1667–1694.

[18] C. Carlemalm, S. Halvarsson, T. Wigren, B. Wahlberg, Algorithms for time delay estimation using a low complexity exhaustive search, IEEE T. Automat. Contr., **44** (1999), 1031–1037.

[19] W. Horbelt, J. Timmer, H.U. Voss, Parameter estimation in nonlinear delayed feedback systems from noisy data, Phys. Lett. A., **299** (2002), 513–521.

[20] L. Belkoura, J.P. Richard, M. Fliess, Parameter estimation of systems with delayed and stuctured entries, Automatica, **45** (2009), 1117–1125.

[21] G. Zheng, J.P. Barbot, D. Boutat, Identification of the delay parameter for nonlinear time-delay systems with unknown inputs, Automatica, **49** (2013), 1755–1760.

[22] H.T. Banks, J.A. Burns, E.M. Cliff, Parameter estimation and identification for systems with delays, SIAM J. Control Optim., **19** (1981), 791–828.

[23] X. Kong, V. Solo, Stochastic averaging analysis of a steepest-descent-type adaptive time-delay estimation algorithm, Math. Control Signals. Syst., **7** (1994), 121–147.

[24] X. Ren, A. Rad, P. Chan, W. Lo, On line identification of continuous-time systems with unknown time delay, IEEE T. Automat. Contr., **50** (2005), 1418–1694.

[25] Y. Tang, X. Guan, Parameter estimation for time-delay chaotic system by particle swarm optimization, Chaos Soliton. Fract., **40** (2009), 1391–1398.

[26] R. Loxton, K.L. Teo,V. Rehbock, An optimization approach to state-delay identification, IEEE T. Automat. Control, **55** (2010), 2113–2119.

[27] Q. Chai, R. Loxton, K.L. Teo, C. Yang, A unified parameter identification method for nonlinear time-delay systems, J. Ind. Manag. Optim., **9** (2013), 471–486.

[28] C. Liu, R. Loxton, K.L. Teo, Optimal parameter selection for nonlinear multistage systems with time-delays, Comput. Optim. Appl., **59** (2014), 285–306.

[29] Q. Lin, R. Loxton, C. Xu, K.L. Teo, State-delay estimation for nonlinear systems using inexact output data, The 33rd CCC, Nanjing, China, July 28–30, 2014.

[30] Q. Lin, R. Loxton, C. Xu, K.L. Teo, Parameter estimation for nonlinear time-delay systems with noisy output measurements, Automatica, **60** (2015), 48–56.

[31] N.U. Ahmed, Dynamic Systems and Control with Applications, World Scientific, Singapore, 2006.

[32] J. Nocedal, S.J. Wright, Numerical Optimization, Springer, New York, 2006.

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

[34] K. Schittkowski, A Fortran Implementation of a Sequential Quadratic Programming Algorithm with Distributed and Non-monotone Line Search—User's Guide, University of Bayeuth, Bayreuth, 2007.

[35] A.C. Hindmarsh, Large ordinary differential equation systems and software, IEEE Control Syst. Mag., **2** (1982), 24–30.

[36] J. Stoer, R. Bulirsch, Introduction to Numerical Analysis, Springer-Verlag, New York, 1980.