ROBUST PARAMETER ESTIMATION FOR CONSTRAINED TIME-DELAY SYSTEMS WITH INEXACT MEASUREMENTS

CHONGYANG LIU1, MEIJIA HAN2, ZHAOHUA GONG1∗ AND KOK LAY TEO3,4

1. School of Mathematics and Information Science
   Shandong Technology and Business University
   Yantai 264005, China
2. School of Computer Science and Technology
   Shandong Technology and Business University
   Yantai 264005, China
3. School of Electrical Engineering, Computing, and Mathematical Sciences
   Curtin University
   Perth 6845, Australia
4. Coordinated Innovation Center for Computable Modeling in Management Science
   Tianjin University of Finance and Economics
   Tianjin 300222, China

(Communicated by Changjun Yu)

ABSTRACT. In this paper, we consider estimation problems involving constrained nonlinear systems with the 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 measurements is minimized subject to the characteristic time constraints specifying the restrictions. We first present the classical estimation formulation, where the expectation of the error function is regarded as the cost function. Then, in order to obtain robust estimates against the noises in measurements, we propose a robust estimation formulation, in which the cost function is the variance of the error function and an additional constraint indicates an allowable sacrifice from the optimal expectation value of the classical estimation problem. For these two estimation problems, we derive the gradients of the corresponding cost and constraint functions with respect to time-delays and system parameters by solving some auxiliary time-delay systems backward in time. On this basis, we develop gradient-based optimization algorithms to determine the optimal time-delays and system parameters. Finally, we consider two example problems, including a parameter estimation problem in microbial batch fermentation process, to illustrate the effectiveness and applicability of our proposed algorithms.

1. Introduction. Parameter estimation deals with the problem of building mathematical models of dynamic systems based on observed measurements. Even though considerable efforts have been undertaken on various aspects of parameter estimation, there are still many open problems due to the specific structure of the underlying models. In particular, parameter estimation of continuous-time systems with time-delays has received particular attention during the past decades [20, 21].

2010 Mathematics Subject Classification. Primary: 49M37; Secondary: 39B72.
Key words and phrases. Nonlinear time-delay system, parameter estimation, characteristic time constraint, gradient computation, numerical optimization.
∗ Corresponding author: Zhaohua Gong.
A parameter estimation problem is usually formulated as an optimization problem, in which a cost function measures the discrepancy between the predicted output and observed system output measurements at a set of sample times. Time-delays are commonly encountered in various engineering systems, such as chemical processes, mechanical systems, network control systems and economic systems [4]. For estimation problem involving linear systems with a single delay and no other unknown parameters, many computational methods, such as exact least squares algorithm [6], steepest descent algorithm [5] and genetic algorithm [19], have been developed. In contrast, estimation problems for nonlinear systems with multiple delays have yet to receive significant attention in the literature. A finite-dimension approximation scheme is considered for such problems in [1]. Recently, a gradient-based algorithm for estimating time-delays in a nonlinear system is developed in [15]. This algorithm is further extended to some more general nonlinear systems with unknown time-delays and unknown parameters in [3, 14]. However, the observed system output measurements are assumed to be obtained exactly in the above research work. This, of course, is an idealistic assumption since the observed system output from the real plant can never be obtained with perfect precision. In practice, there exist noises in the system output measurements. The optimal estimates should be able to withstand the noises. Namely, it needs to be robust against the noises in measurements. More recently, considering noises in the system output measurements, robust parameter estimation for nonlinear time-delay systems, in which the cost function is the weighted sum of mean and variance of a least-squares error between actual and predicted system output, is investigated in [7]. However, it does not provide quantitative information showing the sacrifice of the obtained expectation value from the optimal expectation value of the classical estimation problem. For this, a novel robust parameter estimation formulation, in which the cost function is the variance of an error function and an additional constraint specifies the sacrifice from the optimal expectation value of classical estimation problem, is proposed in [12]. Although the above achieved results are interesting, no constraint is involved in these robust parameter estimation problems. In fact, many estimation problems arising in engineering are subject to constraints; see, for example [2, 17, 30].

In this paper, we consider nonlinear systems with unknown time-delays and unknown system parameters and subject to characteristic time constraints. This type of constraints, which takes the canonical constraint [24] as its special case, often arises in many real-world applications [11, 16, 26]. Time-delays and system parameters are to be estimated robustly such that a least-squares error function between the system output and a set of noisy measurements is minimized. As with [12], we present the classical and robust parameter estimation problems. For these two estimation problems, we transform them into equivalent nonlinear optimization problems. Furthermore, we derive the gradients of the corresponding cost and constraint functions with respect to time-delays and system parameters by solving auxiliary time-delay systems backward in time. On this basis, gradient-based optimization algorithms are developed to solve these two estimation problems. Note that our new algorithms are different from the reported method in [12], which involves solving a large number of auxiliary systems (there is one auxiliary system for each of time-delays and system parameters). Finally, two example problems, including a real-world practical example involving batch fermentation, are considered to test the performance of our new approaches.
2. Problem statement. Consider the following nonlinear time-delay system:
\[
\begin{align*}
\dot{x}(t) &= f(x(t), x(t - \alpha_1), \ldots, x(t - \alpha_m), \zeta), \quad t \in (0, T], \\
x(t) &= \phi(t, \zeta), \quad t \leq 0,
\end{align*}
\]
where \( x(t) \in \mathbb{R}^n \) is the state vector; \( \alpha_i, i = 1, 2, \ldots, m, \) are time-delays; \( \zeta \in \mathbb{R}^p \) is the system parameter vector; \( T \) is the terminal time; and \( f : \mathbb{R}^{(m+1)n} \times \mathbb{R}^p \to \mathbb{R}^n \) and \( \phi : \mathbb{R} \times \mathbb{R}^p \to \mathbb{R}^n \) are given continuously differentiable functions.

For system (1), we assume that the output \( y(t) \in \mathbb{R}^q \) is given by
\[
y(t) = g(x(t), \zeta), \quad t \geq 0,
\]
where \( g : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^q \) is a given continuously differentiable function.

In system (1), both time-delays \( \alpha_i, i = 1, 2, \ldots, m, \) and system parameter vector \( \zeta \) are unknown and need to be estimated. Let \( a_i \) and \( b_i \) denote the lower and upper bounds of the \( i \)th time-delay. Then,
\[
a_i \leq \alpha_i \leq b_i, \quad i = 1, 2, \ldots, m.
\]
Any \( \alpha = (\alpha_1, \ldots, \alpha_m)^T \in \mathbb{R}^m \) with components satisfying (3) is called an admissible time-delay vector for system (1). Let \( D \) denote the set of all admissible time-delay vectors. Furthermore, let \( c_j \) and \( d_j \) denote the lower and upper bounds of the \( j \)th system parameter in vector \( \zeta \). Then,
\[
c_j \leq \zeta_j \leq d_j, \quad j = 1, 2, \ldots, p.
\]
Any vector \( \zeta \in \mathbb{R}^p \) with components satisfying (4) is called an admissible system parameter vector for system (1). Let \( Z \) denote the set of all admissible parameter vectors. Any pair \( (\alpha, \zeta) \in D \times Z \) is called an admissible delay-parameter pair.

For a given \( (\alpha, \zeta) \in D \times Z \), let \( x(\cdot | \alpha, \zeta) \) be the solution of system (1) on \(( -\infty, T] \).

We suppose that, without loss of generality, system (1) is subject to the following inequality constraints:
\[
\begin{align*}
h_k(\alpha, \zeta) &= \Phi_k(x(t_1), x(t_2), \ldots, x(t_r)) \\
&\quad + \int_0^T \mathcal{L}_k(x(t), x(t - \alpha_1), \ldots, x(t - \alpha_m), \zeta)dt \geq 0, \quad k = 1, 2, \ldots, N,
\end{align*}
\]
where \( \Phi_k : \mathbb{R}^m \to \mathbb{R} \) and \( \mathcal{L}_k : \mathbb{R}^{(m+1)n} \times \mathbb{R}^p \to \mathbb{R} \) are given continuously differential functions; and \( t_l, l = 1, 2, \ldots, r, \) are given sample times such that
\[
0 = t_0 < t_1 < t_2 < \cdots < t_r < T = t_{r+1}.
\]

Constraints in (5) are known as characteristic time constraints with wide real-world applications [11, 16, 26]. In particular, such constraints take the canonical constraints in [24] as special cases.

Let \( y(\cdot | \alpha, \zeta) \) be the corresponding output obtained by substituting \( x(\cdot | \alpha, \zeta) \) and \( \zeta \) into (2), and let \( \hat{y} \) be the measurement at sample time \( t = t_1 \). Obviously, these measurements are generally imprecise due to measurement noises. Thus, in this paper, we assume that the measurements \( \hat{y}^l, l = 1, 2, \ldots, r, \) are random vectors, where the corresponding mean vector (of dimension \( qr \)) and covariance matrix (of dimension \( qr \times qr \)) can be obtained. Our goal is to estimate the unknown time-delays and system parameters subject to constraints in (5) by comparing the predicted outputs with the actual measurements at the sample times. To measure estimation
accuracy, we use the following least-squares error function:

\[ J(\alpha, \zeta) = \sum_{l=1}^{r} \| y(\tau_l|\alpha, \zeta) - \hat{y}^l \|^2, \]  

(7)

where \( \| \cdot \| \) denotes the Euclidean norm.

The least-squares error function (7) is a function containing random vectors \( \hat{y}^l \), \( l = 1, 2, \ldots, r \). Thus, the classical estimation problem, which minimizes the expectation of \( J(\alpha, \zeta) \) and is subject to constraints in (5), is stated as follows.

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

\[ G_1(\alpha, \zeta) = E\{J(\alpha, \zeta)\} \]

is minimized subject to characteristic time constraints in (5), where \( E\{\cdot\} \) denotes the expectation.

Problem (P) is, in essence, a dynamic optimization problem with time-delays and system parameters as the decision variables subject to characteristic time constraints. Nevertheless, it is assumed that the exact distributions of measurements can be obtained. In fact, the distributions of measurements are not known exactly and there is uncertainty in these distributions. Thus, a robust delay-parameter pair against the uncertainty is required. As a result, we choose the variance of the least-squares error function (7) as the cost function, and consider the following constraint specifying the sacrifice from the optimal expectation value of Problem (P):

\[ E\{J(\alpha, \zeta)\} \leq (1 + \beta)E\{J(\alpha^*, \zeta^*)\}, \]  

(8)

where \( (\alpha^*, \zeta^*) \) is the optimal delay-parameter estimate of Problem (P); and \( \beta \geq 0 \) is a given constant. Accordingly, we propose our robust estimation problem as follows.

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

\[ G_2(\alpha, \zeta) = \text{Var}\{J(\alpha, \zeta)\} \]

is minimized subject to characteristic time constraints in (5) and an additional constraint (8), where \( \text{Var}\{\cdot\} \) denotes the variance.

Problem (Q) is a dynamic optimization problem subject to characteristic time constraints in (5) and an additional constraint (8). It is different from the existing estimation formulations in [7, 12], in which the weighted sum of expectation and variance of the least-squares error function is taken as the cost function in [7] and no characteristic time constraint is considered in [12]. Obviously, those existing solution methods are not applicable to solve Problem (Q). To solve Problems (P) and (Q), we may rewrite the least-squares error function \( J(\alpha, \zeta) \) as

\[ J(\alpha, \zeta) = \sum_{l=1}^{r} y(\tau_l|\alpha, \zeta)^\top y(\tau_l|\alpha, \zeta) - 2 \sum_{l=1}^{r} y(\tau_l)^\top \hat{y}^l + \sum_{l=1}^{r} [\hat{y}^l]^\top \hat{y}^l. \]

By using similar arguments as in [12], we obtain

\[ E\{J(\alpha, \zeta)\} \]

\[ = \sum_{l=1}^{r} y(\tau_l|\alpha, \zeta)^\top y(\tau_l|\alpha, \zeta) - 2 \sum_{l=1}^{r} y(\tau_l)^\top E\{\hat{y}^l\} + \sum_{l=1}^{r} E\{[\hat{y}^l]^\top \hat{y}^l\}, \]  

(9)
and
\[
\text{Var}\{J(\alpha, \zeta)\} = 4 \sum_{i=1}^{r} \sum_{\ell=1}^{r} y(\tau_i|\alpha, \zeta)^\top \Theta^{i,\ell} y(\tau_\ell|\alpha, \zeta) - 4 \sum_{i=1}^{r} \sum_{\ell=1}^{r} y(\tau_i|\alpha, \zeta)^\top \Delta^{i,\ell} 1^q
\]
\[
+ \text{Var}\left\{\sum_{i=1}^{r} \left[\hat{y}_i^\top \hat{y}_i\right]\right\},
\]
(10)
where $1^q$ is a column vector of all ones in $\mathbb{R}^q$; $\Theta_{i,\ell} = \left[\theta_{i,\ell}\right]$ and $\Delta_{i,\ell} = \left[\delta_{i,\ell}\right]$ are, respectively, $q \times q$ matrices whose elements are defined by
\[
\theta_{i,\ell} = \text{Cov}\{\hat{y}_i, \hat{y}_\ell\}
\]
and $\delta_{i,\ell} = \text{Cov}\{\hat{y}_i, [\hat{y}_\ell]^2\}$;
(11)
and Cov\{·, ·\} denotes the covariance. Note that the last terms in the right hand-side of (9) and (10) are independent of $\alpha$ and $\zeta$. Thus, Problems (P) and (Q) are, respectively, equivalent to the following problems.

**Problem (P).** Given system (1), choose an admissible delay-parameter pair $(\alpha, \zeta) \in D \times Z$ such that the cost function
\[
\tilde{G}_1(\alpha, \zeta) = \sum_{i=1}^{r} \left[y(\tau_i|\alpha, \zeta)^\top y(\tau_i|\alpha, \zeta) - 2y(\tau_i|\alpha, \zeta)^\top \mathbb{E}\{\hat{y}_i\}\right]
\]
(12)
is minimized subject to characteristic time constraints in (5).

**Problem (Q).** Given system (1), choose an admissible delay-parameter pair $(\alpha, \zeta) \in D \times Z$ such that the cost function
\[
\tilde{G}_2(\alpha, \zeta) = 4 \sum_{i=1}^{r} \sum_{\ell=1}^{r} \left[y(\tau_i|\alpha, \zeta)^\top \Theta^{i,\ell} y(\tau_\ell|\alpha, \zeta) - y(\tau_i|\alpha, \zeta)^\top \Delta^{i,\ell} 1^q\right]
\]
(13)
is minimized subject to characteristic time constraints in (5) and an additional constraint (8).

Problems (P) and (Q) can be viewed as constrained mathematical programming problems. It is well known that gradient-based optimization is very effective to solve mathematical programming problems [18]. To do so, the gradients of the cost and constraint functions with respect to time-delays and system parameters are required. However, since the cost and constraint functions are implicit instead of explicit functions of time-delays and system parameters, these gradients cannot be obtained directly. In addition, recall that the gradients of constraint function (8) can be computed by the gradients of cost function (12). Thus, in next section, we will present these required gradient computations of cost functions (12) and (13) as well as characteristic time constraint functions in (5) with respect to time-delays and system parameters.

3. Gradient computation.

3.1. Gradient formulas. In this subsection, we will derive the required gradient formulas in solving Problems (P) and (Q).
Define
\begin{align}
\hat{f}(t|\alpha, \zeta) &= f(x(t), x(t - \alpha_1), \ldots, x(t - \alpha_m), \zeta), \quad (14) \\
\hat{\mathcal{L}}_k(t|\alpha, \zeta) &= \mathcal{L}_k(x(t), x(t - \alpha_1), \ldots, x(t - \alpha_m), \zeta), \quad (15)
\end{align}
\varphi(t|\alpha, \zeta) = \begin{cases} \\
\partial \phi(t|\zeta) \frac{\partial}{\partial t}, & \text{if } t \leq 0, \\
\hat{f}(t|\alpha, \zeta), & \text{if } t \in (0, T].
\end{cases} \quad (16)

Clearly, for almost all \( t \in (-\infty, T] \), we have \( \dot{x}(t|\alpha, \zeta) = \varphi(t|\alpha, \zeta) \). Moreover, define
\begin{align}
\frac{\partial \hat{f}(t|\alpha, \zeta)}{\partial \tilde{x}^t} &= \frac{\partial f(x(t), x(t - \alpha_1), \ldots, x(t - \alpha_m), \zeta)}{\partial x(t - \alpha_\ell)}, \\
\frac{\partial \hat{\mathcal{L}}_k(t|\alpha, \zeta)}{\partial \tilde{x}^t} &= \frac{\partial \mathcal{L}_k(x(t), x(t - \alpha_1), \ldots, x(t - \alpha_m), \zeta)}{\partial x(t - \alpha_\ell)}.
\end{align}

Note that \( \alpha_0 = 0 \) and \( \partial \tilde{x}^0 \) denotes the differentiation with respect to \( x(t) \).

Now, consider the first auxiliary time-delay system defined as follows:
\begin{equation}
\dot{\lambda}(t) = -\sum_{\ell=0}^{m} \left[ \frac{\partial \hat{f}(t + \alpha_\ell|\alpha, \zeta)}{\partial \tilde{x}^t} \right] \lambda(t + \alpha_\ell), \quad t \in (\tau_{i-1}, \tau_i), \quad (17)
\end{equation}
with the intermediate jump conditions
\begin{equation}
\lambda(\tau_i -) = \lambda(\tau_i +) + \psi^i(\alpha, \zeta), \quad l = 1, 2, \ldots, r, \quad (18)
\end{equation}
and the terminal condition
\begin{equation}
\lambda(t) = 0, \quad t > \tau_r, \quad (19)
\end{equation}
where \( \lambda(\tau_i -) \) and \( \lambda(\tau_i +) \) are, respectively, the left and right limits of \( \lambda(t) \) from \( \tau_i \); and
\begin{equation}
\psi^i(\alpha, \zeta) = 2 \left[ \frac{\partial g(x(\tau_i|\alpha, \zeta), \zeta)}{\partial x} \right]^\top \left[ g(x(\tau_i|\alpha, \zeta), \zeta) - E\{\hat{y}^i}\right].
\end{equation}

Let \( \lambda(\cdot|\alpha, \zeta) \) denote the solution of auxiliary system (17)-(19) corresponding to each \( (\alpha, \zeta) \in \mathcal{D} \times \mathcal{Z} \). We now express the gradient of cost function \( \tilde{G}_1(\alpha, \zeta) \) with respect to \( \alpha \) in terms of \( \lambda(\cdot|\alpha, \zeta) \).

**Theorem 3.1.** Let \( (\alpha, \zeta) \in \mathcal{D} \times \mathcal{Z} \) and \( i \in \{1, 2, \ldots, m\} \). Then,
\begin{equation}
\frac{\partial \tilde{G}_1(\alpha, \zeta)}{\partial \alpha_i} = - \int_0^{\tau_r} \lambda(t|\alpha, \zeta)^\top \frac{\partial \hat{f}(t|\alpha, \zeta)}{\partial \tilde{x}^t} \varphi(t - \alpha_i|\alpha, \zeta) dt, \quad (20)
\end{equation}
where \( \varphi(\cdot) \) is as defined in (16).

**Proof.** Let \( \omega^l: [\tau_{i-1}, \tau_i] \to \mathbb{R}^n, \ l = 1, 2, \ldots, r, \) be set of arbitrary absolutely continuous functions. Multiplying system (1) and integrating over intervals \([\tau_{i-1}, \tau_i], \ l = 1, 2, \ldots, r, \) give
\begin{equation}
\sum_{l=1}^r \int_{\tau_{i-1}}^{\tau_i} \omega^l(t)^\top \dot{x}(t) dt = \sum_{l=1}^r \int_{\tau_{i-1}}^{\tau_i} \omega^l(t)^\top \hat{f}(t) dt, \quad (21)
\end{equation}
where we have omitted the arguments \( \alpha \) and \( \zeta \) in \( x(\cdot | \alpha, \zeta) \) and \( \tilde{f}(\cdot | \alpha, \zeta) \) for simplicity. Applying integration by parts to the left-hand-side term of (21) yields

\[
\omega^r(\tau_r-)^\top x(\tau_r) - \omega^l(0+)^\top \phi(0, \zeta) - \sum_{l=1}^{r-1} \left[ \omega^{l+1}(\tau_l+) - \omega^l(\tau_l-) \right]^\top x(\tau_l) = \sum_{l=1}^{r} \int_{\tau_{l-1}}^{\tau_l} \left\{ \omega^l(t)^\top x(t) + \omega^l(t)^\top \tilde{f}(t) \right\} dt.
\] (22)

Furthermore, define the state variation with respect to \( \alpha_i \) as

\[
\Xi^i(t) = \frac{\partial x(t|\alpha, \zeta)}{\partial \alpha_i}, \quad t \in (-\infty, T].
\] (23)

Then for \( \ell \in \{1, 2, \ldots, m\}, \)

\[
\frac{\partial x(t - \alpha_{\ell}|\alpha, \zeta)}{\partial \alpha_{\ell}} = \Xi^i(t - \alpha_{\ell}) - \delta_{i\ell} \varphi(t - \alpha_{\ell}), \quad t \in (-\infty, T],
\] (24)

where we have omitted the arguments \( \alpha \) and \( \zeta \) in \( \varphi(\cdot | \alpha, \zeta) \); and \( \delta_{i\ell} \) denotes the Kronecker delta function. Thus, differentiating (22) with respect to \( \alpha_i \) and applying (23) and (24) yield

\[
\omega^r(\tau_r-)^\top \Xi^i(\tau_r) - \sum_{l=1}^{r-1} \left[ \omega^{l+1}(\tau_l+) - \omega^l(\tau_l-) \right]^\top \Xi^i(\tau_l) = \sum_{l=1}^{r} \int_{\tau_{l-1}}^{\tau_l} \left\{ \omega^l(t)^\top \Xi^i(t) + \omega^l(t)^\top \frac{\partial \tilde{f}(t)}{\partial \tilde{\theta}^i} \varphi(t - \alpha_i) \right\} dt,
\]

where we have omitted the arguments \( \alpha \) and \( \zeta \) in \( \partial \tilde{f}(\cdot | \alpha, \zeta)/\partial \tilde{\theta}^i \). Since \( \omega^l, l = 1, 2, \ldots, r \), are arbitrary functions, we can choose

\[
\omega^l(t) = \begin{cases} \lambda(\tau_{l-1}+), & \text{if } t = \tau_{l-1}, \\ \lambda(t), & \text{if } t \in (\tau_{l-1}, \tau_l), \\ \lambda(\tau_l-), & \text{if } t = \tau_l, 
\end{cases}
\]

where the arguments \( \alpha \) and \( \zeta \) have been omitted. Moreover, by differentiating \( \tilde{G}_1(\alpha, \zeta) \) with respect to \( \alpha_i \) and using (18) and (19), we obtain

\[
\frac{\partial \tilde{G}_1(\alpha, \zeta)}{\partial \alpha_i} = \int_{0}^{\tau_r} \lambda(t)^\top \tilde{\xi}^i(t) dt - \int_{0}^{\tau_r} \lambda(t)^\top \frac{\partial \tilde{f}(t)}{\partial \tilde{\theta}^i} \varphi(t - \alpha_i) dt + \sum_{\ell=1}^{m} \int_{0}^{\tau_r} \lambda(t)^\top \frac{\partial \tilde{f}(t)}{\partial \tilde{\theta}^i} \Xi^i(t - \alpha_{\ell}) dt.
\] (25)

Perform a change of variable in the last integral term of (25) to give

\[
\sum_{\ell=0}^{m} \int_{0}^{\tau_r} \lambda(t)^\top \frac{\partial \tilde{f}(t)}{\partial \tilde{\theta}^i} \Xi^i(t - \alpha_{\ell}) dt = \sum_{\ell=0}^{m} \int_{-\alpha_{\ell}}^{\tau_r-\alpha_{\ell}} \lambda(t + \alpha_{\ell})^\top \frac{\partial \tilde{f}(t + \alpha_{\ell})}{\partial \tilde{\theta}^i} \Xi^i(t) dt.
\] (26)

Recall that \( x(t) = \phi(t, \zeta), \ t \leq 0 \). Thus, \( \Xi^i(t) = 0, \ t \leq 0 \). Furthermore, in view of (19), equation (26) can be simplified as

\[
\sum_{\ell=0}^{m} \int_{0}^{\tau_r} \lambda(t)^\top \frac{\partial \tilde{f}(t)}{\partial \tilde{\theta}^i} \Xi^i(t - \alpha_{\ell}) dt = \sum_{\ell=0}^{m} \int_{0}^{\tau_r} \lambda(t + \alpha_{\ell})^\top \frac{\partial \tilde{f}(t + \alpha_{\ell})}{\partial \tilde{\theta}^i} \Xi^i(t) dt.
\] (27)

Substituting (27) into (25) and applying (17) complete the proof. \( \square \)
The following theorem gives the gradient of cost function \( \tilde{G}_1(\alpha, \zeta) \) with respect to \( \zeta \).

**Theorem 3.2.** Let \((\alpha, \zeta) \in \mathcal{D} \times \mathcal{Z}\) and \(j \in \{1, 2, \ldots, p\}\). Then,

\[
\frac{\partial \tilde{G}_1(\alpha, \zeta)}{\partial \zeta_j} = \lambda(0|\alpha, \zeta)^\top \frac{\partial \phi(0, \zeta)}{\partial \zeta_j} + \int_0^{T_r} \lambda(t|\alpha, \zeta)^\top \frac{\partial \tilde{f}(t|\alpha, \zeta)}{\partial \zeta_j} dt + \sum_{\ell=1}^m \int_{-\alpha+}^0 \lambda(t+\alpha\ell|\alpha, \zeta)\frac{\partial \phi(\ell, \zeta)}{\partial \ell} \frac{\partial \tilde{f}(t)}{\partial \ell} dt. \tag{28}
\]

**Proof.** Let \(\omega^j(\cdot), l = 1, 2, \ldots, r\), be as defined in the proof of Theorem 3.1. Recall from (22) that

\[
\omega^r(\tau_r-)^\top x(\tau_r) - \omega^r(0+)^\top \phi(0, \zeta) - \sum_{l=1}^{r-1} \left[ \omega^{l+1}(\tau_l+) - \omega^l(\tau_l-) \right]^\top x(\tau_l)
\]

\[
= \sum_{l=1}^r \int_{\tau_{l-1}}^{\tau_l} \left\{ \omega^l(t)^\top x(t) + \omega^l(t)^\top \tilde{f}(t) \right\} dt,
\]

where, as in the proof of Theorem 3.1, we have omitted the arguments \(\alpha\) and \(\zeta\) for clarity.

Let the state variation with respect to \(\zeta_j\) be

\[
\Pi^j(t) = \frac{\partial x(t|\alpha, \zeta)}{\partial \zeta_j}, \quad t \in (-\infty, T].
\]

Then for \(\ell \in \{1, 2, \ldots, m\}\),

\[
\frac{\partial x(t-\alpha\ell|\alpha, \zeta)}{\partial \zeta_j} = \Pi^j(t-\alpha\ell), \quad t \in (-\infty, T].
\]

Differentiating (29) with respect to \(\zeta_j\) yields

\[
\omega^r(\tau_r-)^\top \Pi^r(\tau_r) - \omega^r(0+)^\top \phi(0, \zeta) - \sum_{l=1}^{r-1} \left[ \omega^{l+1}(\tau_l+) - \omega^l(\tau_l-) \right]^\top \Pi^l(\tau_l)
\]

\[
= \sum_{l=1}^r \int_{\tau_{l-1}}^{\tau_l} \left\{ \omega^l(t)^\top \Pi^l(t) + \omega^l(t)^\top \sum_{\ell=0}^m \frac{\partial \tilde{f}(t)}{\partial x^\ell} \Pi^l(t-\alpha\ell) + \omega^l(t)^\top \frac{\partial \tilde{f}(t)}{\partial \ell} \right\} dt,
\]

where, as in the proof of Theorem 3.1, we have omitted the arguments \(\alpha\) and \(\zeta\) in \(\partial \tilde{f}(\cdot|\alpha, \zeta)/\partial x^\ell\) and \(\partial \tilde{f}(\cdot|\alpha, \zeta)/\partial \ell\). Furthermore, we choose

\[
\omega^l(t) = \begin{cases} 
\lambda(\tau_{l-1}+), & \text{if } t = \tau_{l-1}, \\
\lambda(t), & \text{if } t \in (\tau_{l-1}, \tau_l), \\
\lambda(\tau_l-), & \text{if } t = \tau_l,
\end{cases}
\]

where, as in the proof of Theorem 3.1, we have omitted the arguments \(\alpha\) and \(\zeta\) in \(\lambda(\cdot|\alpha, \zeta)\). Then, differentiating \(\tilde{G}_1(\alpha, \zeta)\) with respect to \(\zeta_j\) and using (18) and (19) give

\[
\frac{\partial \tilde{G}_1(\alpha, \zeta)}{\partial \zeta_j} = \omega^l(0+)^\top \frac{\partial \phi(0, \zeta)}{\partial \zeta_j} + \int_0^{T_r} \lambda(t)^\top \Pi^j(t) dt + \int_0^{T_r} \lambda(t)^\top \frac{\partial \tilde{f}(t)}{\partial \zeta_j} dt + \sum_{\ell=0}^m \int_0^{T_r} \lambda(t)^\top \frac{\partial \tilde{f}(t)}{\partial x^\ell} \Pi^j(t-\alpha\ell) dt. \tag{30}
\]
Performing a change of variable in the last integral of the equation yields
\[
\int_0^\tau \lambda(t)^\top \sum_{\ell=0}^m \frac{\partial \tilde{f}(t)}{\partial \bar{x}^\ell} \Pi^j(t - \alpha_{\ell}) dt = \sum_{\ell=0}^m \int_{-\alpha_\ell}^{\tau - \alpha_\ell} \lambda(t + \alpha_{\ell})^\top \frac{\partial \tilde{f}(t + \alpha_{\ell})}{\partial \bar{x}^\ell} \Pi^j(t) dt. \tag{31}
\]
Recall that \( x(t) = \phi(t, \zeta), t \leq 0 \). Thus, \( \Pi^j(t) = \partial \phi(t, \zeta) / \partial \zeta_j, t \leq 0 \). In view of (19), (31) can be simplified as
\[
\int_0^\tau \lambda(t)^\top \sum_{\ell=0}^m \frac{\partial \tilde{f}(t)}{\partial \bar{x}^\ell} \Pi^j(t - \alpha_{\ell}) dt = \sum_{\ell=0}^m \int_0^\tau \lambda(t + \alpha_{\ell})^\top \frac{\partial \tilde{f}(t + \alpha_{\ell})}{\partial \bar{x}^\ell} \Pi^j(t) dt
+ \sum_{\ell=0}^m \int_{-\alpha_\ell}^0 \lambda(t + \alpha_{\ell})^\top \frac{\partial \tilde{f}(t + \alpha_{\ell})}{\partial \bar{x}^\ell} \frac{\partial \phi(t, \zeta)}{\partial \zeta_j} dt. \tag{32}
\]
Substituting (32) into (30) and then applying (17) complete the proof. \( \square \)

To investigate the gradients of cost function (13) with respect to time-delays and system parameters, we consider the second auxiliary time-delay system defined as follows:
\[
\dot{\lambda}(t) = -\sum_{\ell=0}^m \left[ \frac{\partial \tilde{f}(t + \alpha_{\ell} | \alpha, \zeta)}{\partial \bar{x}^\ell} \right]^\top \lambda(t + \alpha_{\ell}), \quad t \in (\tau_{l-1}, \tau_l), \tag{33}
\]
with the intermediate jump conditions
\[
\lambda(\tau_{l-1}) = \lambda(\tau_{l+}) + \sigma^l(\alpha, \zeta), \quad l = 1, 2, \ldots, r, \tag{34}
\]
and the terminal condition
\[
\dot{\lambda}(t) = 0, \quad t > \tau_r, \tag{35}
\]
where
\[
\sigma^l(\alpha, \zeta) = 4 \left[ \frac{\partial g(x(\tau_l | \alpha, \zeta), \zeta)}{\partial x} \right]^\top \sum_{q=1}^r \left[ 2 \Theta^{l,q} g(x(\tau_l | \alpha, \zeta), \zeta) - \Delta^{l,q} \right].
\]

Let \( \tilde{\lambda}(|\alpha, \zeta) \) denote the solution of auxiliary system (33)-(35) corresponding to each \((\alpha, \zeta) \in D \times Z\). We present the gradients of cost function \( G_2(\alpha, \zeta) \) with respect to \( \alpha \) and \( \zeta \) in the following theorem.

**Theorem 3.3.** Let \( \alpha, \zeta \in D \times Z \). Then,
\[
\frac{\partial G_2(\alpha, \zeta)}{\partial \alpha_i} = -\int_0^\tau \tilde{\lambda}(t | \alpha, \zeta)^\top \frac{\partial \tilde{f}(t | \alpha, \zeta)}{\partial \bar{x}^\ell} \varphi(t - \alpha_{\ell} | \alpha, \zeta) dt, \quad i = 1, 2, \ldots, m, \tag{36}
\]
and
\[
\frac{\partial G_2(\alpha, \zeta)}{\partial \zeta_j} = \tilde{\lambda}(0 | \alpha, \zeta)^\top \frac{\partial \phi(0, \zeta)}{\partial \zeta_j} + \sum_{\ell=1}^m \int_{-\alpha_\ell}^{\tau - \alpha_\ell} \tilde{\lambda}(t + \alpha_{\ell} | \alpha, \zeta)^\top \frac{\partial \tilde{f}(t + \alpha_{\ell} | \alpha, \zeta)}{\partial \bar{x}^\ell} \frac{\partial \phi(t, \zeta)}{\partial \zeta_j} dt
+ \int_0^\tau \tilde{\lambda}(t | \alpha, \zeta)^\top \frac{\partial \tilde{f}(t | \alpha, \zeta)}{\partial \zeta_j} dt, \quad j = 1, 2, \ldots, p. \tag{37}
\]

**Proof.** The proof is similar to that given for Theorems 3.1 and 3.2. \( \square \)
To investigate the gradients of characteristic time constraints in (5), for each $k \in \{1, 2, \ldots, N\}$, we consider the third auxiliary time-delay system:

$$
\dot{x}^k(t) = -\sum_{l=0}^{m} \left\{ \left[ \frac{\partial \tilde{L}_k(t + \alpha_l | \alpha, \zeta)}{\partial x^l} \right]^T - \left[ \frac{\partial f(t + \alpha_l | \alpha, \zeta)}{\partial x^l} \right]^T \dot{x}^k(t + \alpha_l) \right\},
$$

with the intermediate jump conditions

$$
\tilde{x}^k_i(t) = \tilde{x}^k_j(t) + \frac{\partial \tilde{\Phi}_k(x(\tau_1), x(\tau_2), \ldots, x(\tau_r))}{\partial x(\tau_i)}, \quad l = 1, 2, \ldots, r,
$$

and the terminal conditions

$$
\tilde{x}^k_i(T) = \tilde{x}^k_j(T), \\
\tilde{x}^k_i(t) = 0, \quad t \geq T.
$$

Let $\tilde{x}^k_i(\alpha, \zeta)$ be the solution of auxiliary system (38)-(41) corresponding to each $(\alpha, \zeta) \in D \times Z$. The following theorem gives the gradients of constraints $h^k_i(\alpha, \zeta)$, $k = 1, 2, \ldots, N$, with respect to $\alpha$ and $\zeta$.

**Theorem 3.4.** Let $(\alpha, \zeta) \in D \times Z$ and $k \in \{1, 2, \ldots, N\}$. Then,

$$
\frac{\partial h^k_i(\alpha, \zeta)}{\partial \alpha_l} = -\int_0^T \left\{ \frac{\partial \tilde{L}_k(t | \alpha, \zeta)}{\partial x^l} + \tilde{x}^k_i(t | \alpha, \zeta) \frac{\partial \tilde{f}(t | \alpha, \zeta)}{\partial x^l} \right\} \varphi(t - \alpha_l | \alpha, \zeta) dt,
$$

$$
\quad i = 1, 2, \ldots, m,
$$

and

$$
\frac{\partial h^k_i(\alpha, \zeta)}{\partial \zeta_j} = \tilde{x}^k_i(0) \frac{\partial \phi(0 | \alpha, \zeta)}{\partial \zeta_j} + \sum_{l=1}^{m} \int_{-\alpha_l}^{0} \left\{ \frac{\partial \tilde{L}_k(t + \alpha_l | \alpha, \zeta)}{\partial x^l} \right\} \left\{ \frac{\partial \tilde{\Phi}_k(x(\tau_1), x(\tau_2), \ldots, x(\tau_r))}{\partial x(\tau_i)} \right\} dt + \int_0^T \left\{ \frac{\partial \tilde{f}(t + \alpha_l | \alpha, \zeta)}{\partial x^l} \right\} \left\{ \frac{\partial \tilde{\Phi}_k(x(\tau_1), x(\tau_2), \ldots, x(\tau_r))}{\partial x(\tau_i)} \right\} \varphi(t - \alpha_l | \alpha, \zeta) dt,
$$

$$
\quad j = 1, 2, \ldots, p.
$$

**Proof.** The proof is similar to that given for Theorems 3.1 and 3.2.

3.2. A computational procedure. Based on Theorems 3.1-3.4, we now propose the following computational procedure to compute the cost functions (12) and (13), characteristic time constraints in (5) and the gradients (20), (28), (36), (37), (42) and (43) corresponding to an admissible delay-parameter pair $(\alpha, \zeta)$.

**Step 1.** Solve the time-delay system (1) from $t = 0$ to $t = T$ to obtain $x(\cdot | \alpha, \zeta)$.

**Step 2.** Using $x(\cdot | \alpha, \zeta)$, compute $\tilde{G}_1(\alpha, \zeta)$ and $h^k(\alpha, \zeta)$, $k = 1, 2, \ldots, N$ (or $\tilde{G}_2(\alpha, \zeta)$ and $h^k(\alpha, \zeta)$, $k = 1, 2, \ldots, N$).

**Step 3.** Using $x(\cdot | \alpha, \zeta)$, solve auxiliary time-delay systems (17)-(19) and (38)-(40) (or (17)-(19), (33)-(35) and (38)-(40)) from $t = T$ to $t = 0$ to obtain $\lambda(\cdot | \alpha, \zeta)$ and $\tilde{\lambda}(\cdot | \alpha, \zeta)$ (or $\lambda(\cdot | \alpha, \zeta)$, $\tilde{\lambda}(\cdot | \alpha, \zeta)$).

**Step 4.** Using $x(\cdot | \alpha, \zeta)$, $\lambda(\cdot | \alpha, \zeta)$ and $\tilde{\lambda}(\cdot | \alpha, \zeta)$ (or $x(\cdot | \alpha, \zeta)$, $\lambda(\cdot | \alpha, \zeta)$, $\tilde{\lambda}(\cdot | \alpha, \zeta)$), compute $\frac{\partial \tilde{G}_1(\alpha, \zeta)}{\partial \alpha}$, $\frac{\partial \tilde{G}_1(\alpha, \zeta)}{\partial \zeta}$, $\frac{\partial h^k_i(\alpha, \zeta)}{\partial \alpha}$ and $\frac{\partial h^k_i(\alpha, \zeta)}{\partial \zeta}$, $k = 1, 2, \ldots, N$ (or $\frac{\partial \tilde{G}_2(\alpha, \zeta)}{\partial \alpha}$, $\frac{\partial \tilde{G}_2(\alpha, \zeta)}{\partial \zeta}$, $\frac{\partial h^k_i(\alpha, \zeta)}{\partial \alpha}$ and $\frac{\partial h^k_i(\alpha, \zeta)}{\partial \zeta}$, $k = 1, 2, \ldots, N$) by (20), (28), (42) and (43) (or (20), (28), (36), (37), (42) and (43)).
This computation procedure can be readily incorporated into a gradient-based optimization method (e.g., sequential quadratic programming [18]) to solve Problems \((\tilde{P})\) and \((\tilde{Q})\) as standard mathematical programming problems. In next section, we will use this approach to solve two numerical examples.

4. Numerical examples. We consider two example problems. To solve these problems, we wrote a Fortran program that combines the gradient computation procedure in Section 3 with the optimization software NLPQLP [22]. This program uses the 6th order Runge-Kutta method to solve the state and auxiliary systems. Lagrange interpolation [23] is used when Runge-Kutta method requires the value of the original state or the auxiliary state at an intermediate time between two adjacent knot points.

4.1. Example 1. Consider the following nonlinear systems with two delays:

\[
\begin{align*}
\dot{x}_1(t) &= -0.7x_1(t - \alpha_1) - \zeta_1x_2(t) + 2x_2(t - \alpha_2) + 0.1 \tanh(x_1(t)), \quad t \in (0, 1.5), \\
\dot{x}_2(t) &= \zeta_2x_1(t) - 6.7x_2(t) - \sin(x_2(t - \alpha_2)),
\end{align*}
\]

with the initial conditions

\[
x_1(t) = 6, \quad x_2(t) = t^2 + 2, \quad t \leq 0,
\]

where \(\alpha = (\alpha_1, \alpha_2)^T\) and \(\zeta = (\zeta_1, \zeta_2)^T\) are unknown time-delay and system parameter vectors that need to be identified. Here, we assume that \(\alpha_1, \alpha_2 \in [0.01, 2]\) and \(\zeta_1, \zeta_2 \in [0.1, 5]\).

The system output is given by

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

Thus, the least-squares error function is

\[J(\alpha, \zeta) = \sum_{l=1}^{10} [y(\tau_l) - \hat{y}_l]^2,\]

where \(\tau_l = 0.15l - 0.075\) are the sample times; and \(\hat{y}_l\) is a random variable representing the output measurement at the \(l\)th sample time. In addition, we assume that the system (44)-(45) is subject to the following terminal state constraint

\[x_1(1.5) \geq 2.46.\]

Note that this constraint is equivalent to the following constraint

\[3.64 + \int_0^{1.5} \{-0.7x_1(t - \alpha_1) - \zeta_1x_2(t) + 2x_2(t - \alpha_2) + 0.1 \tanh(x_1(t))\} dt \geq 0.\]

Based on the discussion in Section 2, the classical estimation problem is equivalent to the following problem: given time-delay system (44)-(45), choose \(\alpha_1, \alpha_2 \in [0.01, 2]\) and \(\zeta_1, \zeta_2 \in [0.1, 5]\) such that

\[G_1(\alpha, \zeta) = \sum_{l=1}^{10} [y^2(\tau_l) - 2y(\tau_l)E\{\hat{y}_l\}]\]

is minimized subject to constraint (47). Accordingly,

\[G_1(\alpha, \zeta) = \sum_{l=1}^{10} \left[ y^2(\tau_l) - 2y(\tau_l)E\{\hat{y}_l\} + E\{[\hat{y}_l]^2\} \right].\]
Furthermore, the robust estimation problem can be reformulated as follows: given system \((44)-(45)\), choose \(\alpha_1, \alpha_2 \in [0.01, 2]\) and \(\zeta_1, \zeta_2 \in [0.1, 5]\) such that

\[
\tilde{G}_2(\alpha, \zeta) = 4 \sum_{i=1}^{10} \sum_{\ell=1}^{10} \left[ y^2(\tau_i)y(\tau_\ell)\theta^{i,\ell} - y(\tau_i)\delta^{i,\ell} \right]
\]

is minimized subject to constraints \((47)\) and

\[
G_1(\alpha, \zeta) \leq (1 + \beta)G_1(\alpha^*, \zeta^*),
\]

where \(\theta^{i,\ell} = \text{Cov}\{\hat{y}_i, \hat{y}_\ell\}\); \(\delta^{i,\ell} = \text{Cov}\{\hat{y}_i, [\hat{y}_\ell]^2\}\); and \(\alpha^*\) and \(\zeta^*\) are the optimal estimates of time-delay and system parameter for the classical estimation problem.

To generate the reference output trajectory, we take \(\alpha = (0.6, 0.8)^T\) and \(\zeta = (3.5, 0.7)^T\) as the nominal time-delay and system parameter vectors. Furthermore, we randomly perturbed the reference trajectory using independent normal random distributions. These data are shown in Figure 1. To obtain the required expectations and covariances, suppose the noisy measurement data are generated by

\[
\hat{y}_l = \gamma_l + \sum_{\ell=1}^{l} \beta_\ell, \quad l \in \{1, 2, \ldots, 10\},
\]

where \(\gamma_l\) follows gamma distribution with parameters 5 and \(\bar{y}_l/5\); \(\hat{y}^l\) is the \(l\)th sample data in Figure 1; and \(\beta_\ell = \bar{\beta}_\ell - \text{E}\{\bar{\beta}_\ell\}\) with \(\bar{\beta}_\ell\) following beta distributions of parameters 2 and 5. By using \((48)\), we generate 100 random data at each sample time as the observed data. In addition, the required expectations in \(\tilde{G}_1\) and covariances in \(\tilde{G}_2\) are computed by statistic method.

By implementing our program, we first solved the classical estimation problem and obtained the optimal time-delay and parameter estimates. It is important to note that, in the computation process, only two auxiliary time-delay systems are solved to obtain the required gradients. In contrast, it must solve four auxiliary time-delay systems when one implements the algorithm developed in [12]. For notation simplicity, this optimal parameter estimate is denoted as for \(\beta = 0\). Then, we solved the robust parameter estimation problems for \(\beta = 0.001, 0.002, \ldots, 0.195\). In particular, for each \(\beta\), the number of solving auxiliary time-delay systems is also less.
Table 1. Optimal time-delay and parameter estimates for Example 1.

| β  | α_1^* | α_2^* | ζ_1^* | ζ_2^* |
|----|-------|-------|-------|-------|
| 0  | 0.434658 | 0.713674 | 3.068431 | 0.690517 |
| 0.001 | 0.414390 | 0.764191 | 2.986242 | 0.809873 |
| 0.004 | 0.408673 | 0.790692 | 2.930212 | 0.845580 |
| 0.007 | 0.396182 | 0.804993 | 2.830212 | 0.898728 |
| 0.01 | 0.365811 | 0.819147 | 2.805867 | 0.914341 |
| 0.04 | 0.319229 | 0.839143 | 2.761121 | 0.939468 |
| 0.07 | 0.306182 | 0.896250 | 2.748475 | 0.943645 |
| 0.1  | 0.283349 | 0.942031 | 2.536738 | 0.975285 |
| 0.17 | 0.253673 | 0.951192 | 2.245683 | 1.077120 |
| 0.178| 0.249934 | 0.954031 | 2.231923 | 1.073528 |

Figure 2. Output trajectories corresponding to the optimal time-delay parameter estimates in Table 1 for Example 1.

than that of solving auxiliary systems in implementing algorithm developed in [12]. The optimal parameter estimates for β = 0.001, 0.004, 0.007, 0.01, 0.04, 0.07, 0.1, 0.17 and 0.178 are listed in Table 1. Note that the optimal parameter estimates for β > 0.178 are the same as the one for β = 0.178. This implies the maximal sacrifice for the optimal expectation of the classical parameter estimation problem is 17.8%. The output trajectories corresponding to the optimal time-delay 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 trajectory (black line) as the values of β decrease. Moreover, the green line in Figure 2 denotes x_1(t) ≡ 2.46. It can be seen that, from Figure 2, constraint (47) is satisfied in all these output trajectories and is active for β = 0.

To investigate solution robustness, we generated 100,000 realization of the output data ˆy_l, l = 1, 2, . . . , 10. For each realization, we compute the least-squares error corresponding to the optimal time-delay and parameter estimates for β = 0.001, 0.002, . . . , 0.178. Figure 3 shows the mean and variance of the least-squares error for 100,000 realizations. From Figure 3, we see that the dashed red curve (error variance) is much steeper than the solid blue curve (error mean) near β = 0. For
example, when $\beta$ changes from $\beta = 0$ to $\beta = 0.001$, the variance decreases from 667.1314 to 600.2678 with a slight increase of mean from 42.4660 to 42.8224. This clearly shows the solution robustness increases with negligible sacrifice of the error mean. Furthermore, we investigate the robustness of solution with respect to uncertainty in the distributions of measurements. Thus, instead of (48), we suppose that the distributions of measurements are given by

$$\tilde{y}^l = \bar{\gamma}^l + \sum_{\ell=1}^{l} \beta^\ell, \quad l = 1, 2, \ldots, 10,$$

where $\bar{\gamma}^l$ follow gamma distributions with parameters 5 and $(\tilde{y}^l + \epsilon)/5$; $\epsilon$ is a small parameter ensuring that the distributions of $\bar{\gamma}^l$ are slightly different from the distributions of $\gamma^l$; and $\beta^l$ follow beta distributions as in (48). For each $\epsilon = 0, 0.1, \ldots, 1$, we generate 100,000 realizations of the output data according to (49) and calculate the means of the least-squares error according to the optimal time-delay and parameter estimates for $\beta = 0, 0.01, 0.178$. These variation curves with respect to $\epsilon$ are illustrated in Figure 4. Note that, as expected, the mean according to the optimal time-delay and parameter estimates for $\beta = 0$ varies more notable than the ones for $\beta > 0$. This clearly demonstrates the advantage of our robust parameter estimation problem.

4.2. Example 2. We now consider the batch fermentation process for converting glycerol to 1,3-propanediol (1,3-PD) using the microorganism *Klebsiella pneumoniae*. In this process, a quantity of biomass and glycerol is added to the reactor only once and stirred uniformly to ferment under given conditions. In particular, since nutrient metabolism does not immediately lead to the production of new biomass, time-delay exists in the fermentation process [27].

The mass balance equations for biomass, substrate and 1,3-PD are given by

$$\begin{align*}
\dot{x}_1(t) &= \mu(x(t))x_1(t - \alpha), \\
\dot{x}_2(t) &= -q_2(x(t))x_1(t - \alpha), \quad t \in (0, 7.0], \\
\dot{x}_3(t) &= q_3(x(t))x_1(t - \alpha),
\end{align*}$$

where $\bar{\gamma}^l$ follow gamma distributions with parameters 5 and $(\tilde{y}^l + \epsilon)/5$; $\epsilon$ is a small parameter ensuring that the distributions of $\bar{\gamma}^l$ are slightly different from the distributions of $\gamma^l$; and $\beta^l$ follow beta distributions as in (48). For each $\epsilon = 0, 0.1, \ldots, 1$, we generate 100,000 realizations of the output data according to (49) and calculate the means of the least-squares error according to the optimal time-delay and parameter estimates for $\beta = 0, 0.01, 0.178$. These variation curves with respect to $\epsilon$ are illustrated in Figure 4. Note that, as expected, the mean according to the optimal time-delay and parameter estimates for $\beta = 0$ varies more notable than the ones for $\beta > 0$. This clearly demonstrates the advantage of our robust parameter estimation problem.

4.2. Example 2. We now consider the batch fermentation process for converting glycerol to 1,3-propanediol (1,3-PD) using the microorganism *Klebsiella pneumoniae*. In this process, a quantity of biomass and glycerol is added to the reactor only once and stirred uniformly to ferment under given conditions. In particular, since nutrient metabolism does not immediately lead to the production of new biomass, time-delay exists in the fermentation process [27].

The mass balance equations for biomass, substrate and 1,3-PD are given by

$$\begin{align*}
\dot{x}_1(t) &= \mu(x(t))x_1(t - \alpha), \\
\dot{x}_2(t) &= -q_2(x(t))x_1(t - \alpha), \quad t \in (0, 7.0], \\
\dot{x}_3(t) &= q_3(x(t))x_1(t - \alpha),
\end{align*}$$

where $\bar{\gamma}^l$ follow gamma distributions with parameters 5 and $(\tilde{y}^l + \epsilon)/5$; $\epsilon$ is a small parameter ensuring that the distributions of $\bar{\gamma}^l$ are slightly different from the distributions of $\gamma^l$; and $\beta^l$ follow beta distributions as in (48). For each $\epsilon = 0, 0.1, \ldots, 1$, we generate 100,000 realizations of the output data according to (49) and calculate the means of the least-squares error according to the optimal time-delay and parameter estimates for $\beta = 0, 0.01, 0.178$. These variation curves with respect to $\epsilon$ are illustrated in Figure 4. Note that, as expected, the mean according to the optimal time-delay and parameter estimates for $\beta = 0$ varies more notable than the ones for $\beta > 0$. This clearly demonstrates the advantage of our robust parameter estimation problem.

4.2. Example 2. We now consider the batch fermentation process for converting glycerol to 1,3-propanediol (1,3-PD) using the microorganism *Klebsiella pneumoniae*. In this process, a quantity of biomass and glycerol is added to the reactor only once and stirred uniformly to ferment under given conditions. In particular, since nutrient metabolism does not immediately lead to the production of new biomass, time-delay exists in the fermentation process [27].

The mass balance equations for biomass, substrate and 1,3-PD are given by

$$\begin{align*}
\dot{x}_1(t) &= \mu(x(t))x_1(t - \alpha), \\
\dot{x}_2(t) &= -q_2(x(t))x_1(t - \alpha), \quad t \in (0, 7.0], \\
\dot{x}_3(t) &= q_3(x(t))x_1(t - \alpha),
\end{align*}$$

where $\bar{\gamma}^l$ follow gamma distributions with parameters 5 and $(\tilde{y}^l + \epsilon)/5$; $\epsilon$ is a small parameter ensuring that the distributions of $\bar{\gamma}^l$ are slightly different from the distributions of $\gamma^l$; and $\beta^l$ follow beta distributions as in (48). For each $\epsilon = 0, 0.1, \ldots, 1$, we generate 100,000 realizations of the output data according to (49) and calculate the means of the least-squares error according to the optimal time-delay and parameter estimates for $\beta = 0, 0.01, 0.178$. These variation curves with respect to $\epsilon$ are illustrated in Figure 4. Note that, as expected, the mean according to the optimal time-delay and parameter estimates for $\beta = 0$ varies more notable than the ones for $\beta > 0$. This clearly demonstrates the advantage of our robust parameter estimation problem.
where $t$ denotes the process time; $x_1(t)$, $x_2(t)$ and $x_3(t)$ are, respectively, the concentrations of biomass, glycerol and 1,3-PD; $\alpha$ is a time-delay argument; $x(t) = (x_1(t), x_2(t), x_3(t))^T$; $\mu(x(t))$ is the cell growth rate; $q_2(x(t))$ is the substrate consumption rate; and $q_3(x(t))$ is the product formation rate of 1,3-PD. The functions $\mu(x(t))$, $q_2(x(t))$ and $q_3(x(t))$ are

\begin{align*}
\mu(x(t)) &= \Delta_1 x_2(t) x_2(t) + k_1 \left(1 - \frac{x_2(t)}{x^*_2}\right) \left(1 - \frac{x_3(t)}{x^*_3}\right), \\
q_2(x(t)) &= m_2 + \frac{\mu(x(t))}{Y_2}, \\
q_3(x(t)) &= -m_3 + Y_3 \mu(x(t)),
\end{align*}

(51)

(52)

(53)

where $\Delta_1$, $k_1$, $m_2$, $Y_2$, $m_3$ and $Y_3$ are model parameters; and $x^*_2 = 203.9\text{mmolL}^{-1}$ and $x^*_3 = 93.95\text{mmolL}^{-1}$ are the maximal residual concentrations of substrate and 1,3-PD [28].

During the fermentation process, the concentrations of biomass, glycerol and 1,3-PD must be restricted to biological meaningful ranges since the cells will cease to grow outside these ranges. Thus, all the states must satisfy

\begin{equation}
x(t) \in [x_1^*, x_1^*] \times [x_2^*, x_2^*] \times [x_3^*, x_3^*], \quad t \in [0, 7.0],
\end{equation}

(54)

where $x_1^* = 0.01\text{gL}^{-1}$, $x_2^* = 10\text{mmolL}^{-1}$, $x_3^* = 0\text{mmolL}^{-1}$ are the lower thresholds for cell growth of biomass, glycerol and 1,3-PD, respectively; and $x_1^* = 7$, $x_2^*$ and $x_3^*$ (as used in the formula for $\mu(x(t))$) are the corresponding upper thresholds. Note that constraint (54) has been considered in the literature; see, for example [9, 10, 13, 25, 29].

For the batch fermentation process, there are some real experimental data conducted in China [28]. The data consist of 7 sample times and the biomass, glycerol and 1,3-PD concentrations at each of these sample times. We denote the concentrations of biomass, glycerol and 1,3-PD measured at the sample time $\tau_l$ by $\bar{x}_1^l$, $\bar{x}_2^l$ and $\bar{x}_3^l$, $l \in \{1, 2, \ldots, 7\}$, respectively. The initial function for system (50) was constructed by applying Lagrange interpolation [23] to these experimental data before

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure4}
\caption{Error mean variations with respect to $\epsilon$ for 100,000 output realization according to (49) for Example 1.}
\end{figure}
Table 2. Bounds, optimal time-delay and parameter estimates for Example 2.

| Parameters | Lower bounds | Upper bounds | Optimal values |
|------------|--------------|--------------|----------------|
| \( \beta \) = 0 | 0.0010 | 0.0863 | 0.0743 | 0.0700 | 0.0648 |
| \( \beta \) = 0.01 | 0.1000 | 2.0000 | 1.5346 | 1.5366 | 1.6385 | 1.9040 |
| \( k_1 \) | 10.000 | 500.00 | 287.91 | 289.79 | 331.63 | 442.56 |
| \( m_2 \) | 1.0000 | 20.000 | 5.2545 | 7.5042 | 10.759 | 15.629 |
| \( Y_2 \) | 0.0001 | 2.0000 | 0.0083 | 0.0092 | 1.0325 | 0.0124 |
| \( m_3 \) | 1.0000 | 20.000 | 13.598 | 15.787 | 16.587 | 17.001 |
| \( Y_3 \) | 10.000 | 200.00 | 131.91 | 136.49 | 138.26 | 139.31 |

the zero time point. We assume that the system output is given by

\[ y(t) = \omega x(t), \]

where \( \omega \) is a \( 3 \times 3 \) diagonal matrix defined by

\[ \omega = \begin{bmatrix} 1.620485 & 0 & 0 \\ 0 & 0.011955 & 0 \\ 0 & 0 & 0.017585 \end{bmatrix}. \] (55)

Let \( \zeta := (\Delta_1, k_1, m_2, Y_2, m_3, Y_3) \) be the model parameter vector. We assume that the bounds of \( \alpha \) and \( \zeta \) are as listed in Table 2. Now, the least-squares error function is defined by

\[ J(\alpha, \zeta) = \sum_{l=1}^{7} \sum_{\ell=1}^{3} |y(\tau_l|\alpha, \zeta) - \hat{y}_l|, \] (56)

where \( \hat{y}_1, \hat{y}_2, \hat{y}_3 \) are, respectively, random variables representing the output measurement of biomass, glycerol and 1,3-PD concentrations at the \( l \)th sample time.

Based on the discussion in Section 2, the classical parameter estimation problem is: given the system (50), choose \( \alpha \) and \( \zeta \) to minimize the cost function

\[ \hat{G}_1(\alpha, \zeta) = \sum_{l=1}^{7} \sum_{\ell=1}^{3} \left\{ y(\tau_l|\alpha, \zeta)^2 - 2y(\tau_l|\alpha, \zeta)E\{\hat{y}_l\} \right\}, \] (57)

subject to the bounds of \( \alpha \) and \( \zeta \) in Table 2 and constraint (54). Then,

\[ G_1(\alpha, \zeta) = \sum_{l=1}^{7} \sum_{\ell=1}^{3} \left\{ y(\tau_l|\alpha, \zeta)^2 - 2y(\tau_l|\alpha, \zeta)E\{\hat{y}_l\} + E\{(\hat{y}_l)^2\} \right\}. \]

As a result, the robust parameter estimation problem is: given system (50), choose \( \alpha \) and \( \zeta \) to minimize the cost function

\[ \tilde{G}_2(\alpha, \zeta) = 4 \sum_{l=1}^{7} \sum_{\ell=1}^{3} \left\{ y(\tau_l|\alpha, \zeta)^\top \Theta^{l,\ell} y(\tau_l|\alpha, \zeta) - y(\tau_l|\alpha, \zeta)^\top \Delta^{l,\ell} \right\} \] (58)

subject to the constraints (54),

\[ G_1(\alpha, \zeta) \leq (1 + \beta)G_1(\hat{\alpha}^*, \hat{\zeta}^*), \]

and the bounds of \( \alpha \) and \( \zeta \) in Table 2, where \( \Theta^{l,\ell} \) and \( \Delta^{l,\ell} \) as defined in (11); and \( (\hat{\alpha}^*, \hat{\zeta}^*) \) is the optimal solution of the parameter estimation problem.
Now, suppose that the noisy measurement outputs take the following form:

\[ \hat{y}^l = \sigma_l + \rho_l, \quad l = 1, 2, \ldots, 7, \]  

(59)

where \( \sigma_l \) and \( \rho_l \), \( l = 1, 2, \ldots, 7 \) are independent random vectors. Moreover, we assume that \( \sigma_l \) follows a gamma distribution with parameters 10 and \( \bar{y}^l/10 \), where \( \bar{y}^l = \omega \bar{x}^l \) is the output sample at the \( l \)th sample point. We also assume that \( \rho_l \) follows a standard normal distribution. By using (59), we generate 500 random data at each sample time as the observed data. For the observed data, the mean and covariance matrices in (57) and (58) are computed by statistic method. In particular, the constraint (54) is a continuous state inequality constraint. Using the approximation method \([8, 24]\), we approximate (54) by the following constraint:

\[ \eta + \sum_{k=1}^{6} \int_{0}^{7.0} \pi_\rho(\varpi_\zeta(x(s))) ds \geq 0, \]  

(60)

where \( \eta > 0 \) and \( \varrho > 0 \) are adjustable parameters;

\[ \varpi_\zeta(x(t)) = x^*_t - x(t), \]

\[ \varpi_{\zeta+3}(x(t)) = x_t(t) - x_{s+1}, \quad t = 1, 2, 3; \]

and

\[ \pi_\rho(\nu) = \begin{cases} 
\nu, & \text{if } \nu < -\varrho, \\
-(\frac{\nu - \varrho}{4\varrho}), & \text{if } -\varrho < \nu < \varrho, \\
0, & \text{otherwise.}
\end{cases} \]

It can be shown, as in [24], that for each \( \varrho > 0 \), there exists a corresponding \( \eta(\varrho) > 0 \) such that whenever \( 0 < \eta < \eta(\varrho) \), constraint (60) implies constraint (54).

Using our program, we first solve the parameter estimation problem. The obtained optimal time-delay and parameters are also listed in TABLE 2. For notation simplicity, these optimal time-delay and parameters are denoted as \( \beta = 0 \). We then solve the robust parameter estimation for \( \beta = 0.001, 0.002, \ldots, 0.3 \). Note that, in our program, at most three auxiliary time-delay systems are solved for each \( \beta \) to obtain the required gradients comparing with solving seven auxiliary time-delay systems when one implements the algorithm developed in [12]. In addition, the obtained optimal time-delays and parameters for \( \beta > 0.3 \) are same as the ones for \( \beta = 0.3 \). The obtained optimal time-delays and parameters for \( \beta = 0, 0.001, 0.01, 0.3 \) are also listed in TABLE 2. The state trajectories corresponding to the optimal time-delays and parameters in TABLE 2 are shown in FIGURE 5. In particular, the blue line in FIGURE 5 denotes lower critical concentration \( x_{s+2} = 10 \) and the constraint \( \varpi_\zeta(x(t)) \) is active at the terminal time of state trajectory for \( \beta = 0.3 \).

For each \( l = 1, 2, \ldots, 7 \), we generate 100,000 realizations of the noisy measurement output vectors \( \hat{y}^l \) to investigate the solution robustness. For each realization, we compute the least-squares error (56) corresponding to the optimal parameters for \( \beta = 0, 0.001, 0.002, \ldots, 0.3 \). The mean and variance of the least-squares errors for 100,000 output realizations are shown in FIGURE 6. As expected, the solution robustness increases clearly with negligible cost to the error mean. Moreover, we investigate how the optimal time-delays and parameters perform when the actual distribution differs from the assumed distribution (59). Thus, we now suppose that the output distribution is given by

\[ \hat{y}^l = \bar{\eta} + \rho_l, \quad l = 1, \ldots, 7, \]  

(61)
Figure 5. State trajectories corresponding to the optimal parameters in Table 2 for Example 2.

where $\rho_l$ is a random vector as defined in (59) and $\sigma_l$ follows a gamma distribution with parameters 10 and $(\bar{y}_l^d + \epsilon)/10$ with a small parameter $\epsilon$. The parameter $\epsilon$ ensures that $\sigma_l$ differs from $\sigma_l$ in (59). For each $\epsilon = 0, 0.1, \ldots, 1.0$, we generated 100,000 realizations of the output data according to (61) and calculated the mean
of the least-squares errors according to the optimal time-delays and parameters obtained for $\beta = 0, 0.001, 0.01, 0.3$. Figure 7 shows these error means vary with respect to the parameter $\epsilon$. As expected, the bigger of the value of $\beta$ takes, the more slowly of the error mean changes, which indicate the advantage of our robust parameter estimation formulation facing uncertainty in the output distribution.

5. Conclusions. In this paper, we investigated the parameter estimation problem for constrained nonlinear time-delay systems with noisy measurements. This problem class arises in the mathematical modelling of the batch fermentation process for 1,3-propanediol production (see Example 2). We first proposed robust parameter estimation formulation based on the classical parameter estimation formulation. We then derived the gradients of the cost and constraint functions with respect to
time-delays and system parameters. On this basis, we developed gradient-based optimization algorithms to numerically solve the classical and robust parameter estimation problems. The effectiveness of this algorithm was verified using two highly nonlinear numerical examples. In closing, it is worth noting that our methods proposed here are more efficient than the existing method developed in [12] for problems with optimization parameters outnumbering constraints.

Acknowledgments. This research was supported by the Natural Science Foundation of China (grant 11771008), the Australian Research Council (grant DP190103361) and the Natural Science Foundation of Shandong Province, China (grants ZR2017MA005 and ZR2019MA031).

REFERENCES

[1] H. T. Banks, J. A. Burns and E. M. Cliff, Parameter estimation and identification for systems with delay, SIAM J. Control Optim., 19 (1981), 791–828.
[2] Y. Bard, Comparison of gradient methods for the solution of nonlinear parameter estimation problems, SIAM J. Numer. Anal., 7 (1970), 157–186.
[3] Q. Q. Chai, R. Loxton, K. L. Teo and C. H. Yang, A unified parameter identification method for nonlinear time-delay systems, J. Ind. Manag. Optim., 9 (2013), 471–486.
[4] D. Debeljkovic, Time-Delay Systems, InTech, 2011.
[5] S. Diop, I. Kolmanovsky, P. E. Moraal and M. V. Nieuwstadt, Preserving stability/performance when facing an unknown time-delay, Control Eng. Pract., 9 (2001), 1319–1325.
[6] P. J. Gawthrop and M. T. Nihtilä, Identification of time-delays using a polynomial identification method, Syst. Control Lett., 5 (1985), 267–271.
[7] Q. Lin, R. Loxton, C. Xu and K. L. Teo, Parameter estimation for nonlinear time-delay systems with noisy output measurements, Automatica J. IFAC, 60 (2015), 48–56.
[8] Q. Lin, R. Loxton and K. L. Teo, The control parameterization method for nonlinear optimal control: A survey, J. Ind. Manag. Optim., 10 (2014), 275–309.
[9] C. Y. Liu and Z. H. Gong, Optimal Control of Switched Systems Arising in Fermentation Processes, Springer Optimization and Its Applications, 97. Springer, Heidelberg, Tsinghua University Press, Beijing, 2014.
[10] C. Y. Liu, Z. H. Gong, E. Feng and H. C. Yin, Modelling and optimal control for nonlinear multistage dynamical system of microbial fed-batch culture, J. Ind. Manag. Optim., 5 (2009), 835–850.
[11] C. Y. Liu, Z. H. Gong, K. L. Teo, J. Sun and L. Caccetta, Robust multi-objective optimal switching control arising in 1,3-propanediol microbial fed-batch process, Nonlinear Anal. Hybrid Syst., 25 (2017), 1–20.
[12] C. Y. Liu, Z. H. Gong and K. L. Teo, Robust parameter estimation for nonlinear multistage time-delay systems with noisy measurement data, Appl. Math. Model., 53 (2018), 353–368.
[13] C. Y. Liu, Z. H. Gong, H. W. J. Lee and K. L. Teo, Robust bi-objective optimal control of 1,3-propanediol microbial batch production process, J. Process Contr., 78 (2019), 170–182.
[14] C. Y. Liu, R. Loxton and K. L. Teo, Optimal parameter selection for nonlinear multistage systems with time-delays, Comput. Optim. Appl., 59 (2014), 285–306.
[15] R. Loxton, K. L. Teo and V. Rehbock, An optimization approach to state-delay identification, IEEE Trans. Aut. Control, 55 (2010), 2113–2119.
[16] R. B. Martin, Optimal control drug scheduling of cancer chemotherapy, Automatica J. IFAC, 28 (1992), 1113–1123.
[17] P. Mendes and D. Kell, Non-linear optimization of biochemical pathways: Applications to metabolic engineering and parameter estimation, Bioinformatics, 14 (1998), 869–883.
[18] J. Nocedal and S. J. Wright, Numerical Optimization, Springer Series in Operations Research and Financial Engineering, Springer, New York, 2006.
[19] F. Pan, R. C. Han and D. M. Feng, An identification method of time-varying delay based on genetic algorithm, in Proceedings of the Second International Conference on Machine Learning and Cybernetics, (2003), 781–783.
[20] X. M. Ren, A. B. Rad, P. T. Chan and W. L. Lo, Online identification of continuous-time systems with unknown time-delay, IEEE Trans. Aut. Control, 50 (2005), 1418–1422.
[21] J. P. Richard, *Time-delay systems: An overview of some recent advances and open problems*, *Automatica J. IFAC*, 39 (2003), 1667–1694.

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

[23] J. Stoer and R. Bulirsch, *Introduction to Numerical Analysis*, Springer, New York-Heidelberg, 1980.

[24] K. L. Teo, C. J. Goh and K. H. Wong, *A Unified Computational Approach to Optimal Control Problems*, Pitman Monographs and Surveys in Pure and Applied Mathematics, 55, Longman Scientific & Technical, Harlow, copublished in the United States with John Wiley & Sons, Inc., New York, 1991.

[25] L. Wang, Q. Lin, R. Loxton, K. L. Teo and G. Cheng, *Optimal 1,3-propanediol production: Exploring the trade-off between process yield and feeding rate variation*, *J. Process Contr.*, 32 (2015), 1–9.

[26] L. Y. Wang, W. H. Gui, K. L. Teo, R. Loxton and C. H. Yang, *Optimal control problems arising in the zinc sulphate electrolyte purification process*, *J. Glob. Optim.*, 54 (2012), 307–323.

[27] Z. L. Xiu, B.-H. Song, L.-H. Sun and A.-P. Zeng, *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 (2002), 101–109.

[28] Z. L. Xiu, A. P. Zeng and L. J. An, *Mathematical modelling of kinetics and research on multiplicity of glycerol bioconversion to 1,3-propanediol*, *J. Dalian Univ. Tech.*, 40 (2000), 428–433.

[29] J. L. Yuan, X. Zhang, X. Zhu, E. Feng, H. C. Yin and Z. L. Xiu, *Pathway identification using parallel optimization for a nonlinear hybrid system in batch culture*, *Nonlinear Anal. Hybrid Syst.*, 15 (2015), 112–131.

[30] J. L. Yuan, X. Zhu, X. Zhang, H. C. Yin, E. Feng and Z. L. Xiu, *Robust identification of enzymatic nonlinear dynamical systems for 1,3-propanediol transport mechanisms in microbial batch culture*, *Appl. Math. Comput.*, 232 (2014), 150–163.

Received January 2019; revised April 2019.

E-mail address: liu_chongyang@yahoo.com
E-mail address: meijia_han@sina.com
E-mail address: zhaohuagong@163.com