

Parameter Estimation for Nonlinear Time-delay Systems with Noisy Output Measurements [★]

Qun Lin ^a, Ryan Loxton ^{a,b}, Chao Xu ^b, Kok Lay Teo ^{a,b}

^a*Department of Mathematics and Statistics, Curtin University, Perth, Australia*

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

Abstract

This paper considers the problem of using noisy output data to estimate unknown time-delays and unknown system parameters in a general nonlinear time-delay system. We formulate the problem as a dynamic optimization problem in which the unknown quantities are decision variables to be chosen optimally, with the cost function penalizing the mean and variance of the least-squares error between actual and predicted system output. Since the time-delays and system parameters influence the cost function implicitly through the governing time-delay system, the cost function's gradient—which is required to solve the problem using gradient-based optimization techniques—cannot be computed analytically using standard differentiation rules. We instead develop two computational methods for evaluating this gradient: one involves solving an auxiliary time-delay system forward in time; the other involves solving an auxiliary time-advance system backward in time. On this basis, we propose an efficient optimization algorithm for determining optimal estimates for the time-delays and system parameters. We conclude the paper by examining the performance of this algorithm on a dynamic model of a continuously-stirred tank reactor.

Key words: Time-delay, Nonlinear system, Parameter estimation, Dynamic optimization, Nonlinear optimization

1 Introduction

Developing a mathematical model is a two-step process: first, the general structure of the model is derived based on fundamental physical principles; then, the model is matched to a particular system of interest by tuning various model parameters. This second step, known as *parameter estimation* or *parameter identification*, usually involves comparing the system output predicted by the model with the real system output measured during an experiment (or a series of experiments).

This paper is concerned with parameter estimation for nonlinear time-delay systems. We consider a general dynamic model consisting of nonlinear delay-differential

equations with multiple time-delays and multiple system parameters, each of which is unknown and needs to be estimated appropriately. The problem that we investigate—called the *parameter estimation problem*—is to determine optimal estimates for the time-delays and system parameters so that the dynamic model best fits the real system under consideration. Such problems are commonly referred to as *inverse problems*.

Parameter estimation for time-delay systems has attracted considerable research interest over the past two decades (see Tuch, Feuer & Palmor (1994); Lunel (2001); Orlov, Belkoura, Richard & Dambrine (2002, 2003); Drakunov, Perruquetti, Richard & Belkoura (2006); Belkoura, Richard & Fliess (2009); Park, Han & Kwon (2013); Zheng, Barbot & Boutat (2013)). Popular approaches for solving the parameter estimation problem include swarm intelligence algorithms such as particle swarm optimization (Tang & Guan (2009); Gao, Qi, Yin & Xiao (2010)), or finite-dimensional approximation schemes for the original infinite-dimensional time-delay model (Banks, Rehm & Sutton (2010)). Recently, a new gradient-based optimization approach has been proposed by Loxton, Teo & Rehbock (2010) and Chai, Loxton, Teo & Yang (2013a, 2013b). In this approach, the parameter estimates are chosen as the so-

[★] This paper was not presented at any IFAC meeting. The authors acknowledge support from the National Natural Science Foundation of China (Grant Numbers 61104048 and 11350110208) and the State Key Laboratory for Industrial Control Technology at Zhejiang University, China (Open Research Project ICT1301). Corresponding author C. Xu. Tel. +86-571-87952233. Fax +86-571-87952279.

Email addresses: q.lin@curtin.edu.au (Qun Lin), r.loxton@curtin.edu.au (Ryan Loxton), cxu@zju.edu.cn (Chao Xu), k.l.teo@curtin.edu.au (Kok Lay Teo).

lution of a dynamic optimization problem in which the cost function penalizes the deviation between predicted and measured system output. Special dynamic optimization techniques can then be deployed to solve this problem and obtain accurate estimates for the model parameters. This approach was introduced by Loxton et al. (2010) for nonlinear time-delay systems in which each nonlinear term contains a single delay and no other model parameters, and then extended by Chai et al. (2013a) to more general nonlinear systems with multiple delays and multiple system parameters. In Chai et al. (2013b), the approach was applied to a more difficult parameter estimation problem in which the dynamic system contains both state- and input-delays, and the input function is discontinuous.

The two main advantages of the parameter estimation methods proposed by Loxton et al. (2010) and Chai et al. (2013a, 2013b) are: (i) these methods can readily handle system nonlinearities; and (ii) these methods can simultaneously compute optimal estimates for the time-delays and system parameters in a unified fashion. One limitation, however, is that these methods do not take into account the possibility of noise in the output data. Thus, the output measurements used in the cost function (recall that the cost function penalizes the discrepancy between predicted and measured system output) are assumed to be exact. This is, of course, an idealistic assumption, as it is impossible to guarantee perfect precision when measuring the output of a real system.

The purpose of this paper is to address this limitation. Building on the results in Loxton et al. (2010) and Chai et al. (2013a, 2013b), we will devise a new method for parameter estimation that explicitly takes output measurement noise into account. The main idea is to consider the output data points as random variables, rather than fixed constants. This allows for possible discrepancies between the actual and observed system output due to measurement errors. With the output data as random variables, our parameter estimation problem is formulated as a stochastic dynamic optimization problem in which the aim is to choose the time-delays and system parameters to minimize a weighted sum of the expectation and variance of the least-squares error between actual and predicted system output. We will develop a computational approach for solving this problem based on novel dynamic optimization techniques. The result is a unified parameter estimation method for nonlinear time-delay systems that is fast, versatile, and capable of handling uncertainties in the measured output data.

2 Problem Statement

Consider the following nonlinear time-delay system:

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{f}(\mathbf{x}(t), \mathbf{x}(t - \tau_1), \dots, \mathbf{x}(t - \tau_m), \boldsymbol{\zeta}), \quad t \geq 0, \quad (1) \\ \mathbf{x}(t) &= \boldsymbol{\phi}(t, \boldsymbol{\zeta}), \quad t \leq 0, \quad (2)\end{aligned}$$

where $\mathbf{x}(t) \in \mathbb{R}^n$ is the *state vector*; $\boldsymbol{\zeta} \in \mathbb{R}^r$ is the *parameter vector*; $\tau_i, i = 1, \dots, m$, are *time-delays*; and $\mathbf{f} : \mathbb{R}^{(m+1)n} \times \mathbb{R}^r \rightarrow \mathbb{R}^n$ and $\boldsymbol{\phi} : \mathbb{R} \times \mathbb{R}^r \rightarrow \mathbb{R}^n$ are given continuously differentiable functions.

The output $\mathbf{y}(t) \in \mathbb{R}^q$ of system (1)-(2) is given by the following equation:

$$\mathbf{y}(t) = \mathbf{g}(\mathbf{x}(t), \boldsymbol{\zeta}), \quad t \geq 0, \quad (3)$$

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

Both the time-delays $\tau_i, i = 1, \dots, m$, and the parameter vector $\boldsymbol{\zeta}$ 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

$$a_i \leq \tau_i \leq b_i, \quad i = 1, \dots, m. \quad (4)$$

Any vector $\boldsymbol{\tau} \in \mathbb{R}^m$ with components satisfying (4) is called a *candidate time-delay vector* for system (1)-(3). Let \mathcal{T} denote the set of all candidate time-delay vectors.

Similarly, let c_j and d_j denote the lower and upper bounds of the j th system parameter in $\boldsymbol{\zeta}$. Then

$$c_j \leq \zeta_j \leq d_j, \quad j = 1, \dots, r. \quad (5)$$

Any vector $\boldsymbol{\zeta} \in \mathbb{R}^r$ with components satisfying (5) is called a *candidate parameter vector* for system (1)-(3). Let \mathcal{Z} denote the set of all candidate parameter vectors.

For each candidate pair $(\boldsymbol{\tau}, \boldsymbol{\zeta}) \in \mathcal{T} \times \mathcal{Z}$, let $\mathbf{x}(\cdot | \boldsymbol{\tau}, \boldsymbol{\zeta})$ denote the state trajectory obtained by solving equations (1)-(2) with the components of $\boldsymbol{\tau}$ and $\boldsymbol{\zeta}$ used as the time-delays and system parameters, respectively. Furthermore, let $\mathbf{y}(\cdot | \boldsymbol{\tau}, \boldsymbol{\zeta})$ denote the corresponding output function obtained by substituting $\mathbf{x}(\cdot | \boldsymbol{\tau}, \boldsymbol{\zeta})$ into (3).

Our goal is to estimate the unknown time-delays and system parameters by comparing the predicted system output (obtained by solving the model (1)-(3)) with the actual system output (measured during a series of experiments) at a set of sample times $\{t_k\}_{k=1}^p$, where

$$0 = t_0 < t_1 < t_2 < \dots < t_{p-1} < t_p.$$

Let $\hat{\mathbf{y}}^k$ denote the actual system output at time $t = t_k$. In Loxton et al. (2010) and Chai et al. (2013a, 2013b), we assumed that the output vectors $\hat{\mathbf{y}}^k, k = 1, \dots, p$, can be measured exactly. However, this assumption is unrealistic; due to system noise and measurement errors, the true system output will often differ slightly from the measured output. Thus, in this paper, we view $\hat{\mathbf{y}}^k, k = 1, \dots, p$, as random vectors of known distribution.

We assume that the following matrices can be obtained

from the distribution of $\hat{\mathbf{y}}^k$, $k = 1, \dots, p$:

$$\Xi^{k,l} = \text{Cov}\{\hat{\mathbf{y}}^k, \hat{\mathbf{y}}^l\}, \quad \Upsilon^{k,l} = \text{Cov}\{(\hat{\mathbf{y}}^k)^2, \hat{\mathbf{y}}^l\}, \quad (6)$$

where $(\hat{\mathbf{y}}^k)^2$ denotes the vector obtained by squaring each element of $\hat{\mathbf{y}}^k$. The issue of computing these matrices is discussed in Section 4.

Any $\boldsymbol{\tau} \in \mathcal{T}$ is a candidate for the real time-delay vector. Similarly, any $\boldsymbol{\zeta} \in \mathcal{Z}$ is a candidate for the real parameter vector. To measure estimation accuracy, we use the following least-squares error function:

$$J(\boldsymbol{\tau}, \boldsymbol{\zeta}) = \sum_{k=1}^p \|\mathbf{y}(t_k | \boldsymbol{\tau}, \boldsymbol{\zeta}) - \hat{\mathbf{y}}^k\|^2.$$

Our parameter estimation problem is stated as follows.

Problem P. Choose $\boldsymbol{\tau} \in \mathcal{T}$ and $\boldsymbol{\zeta} \in \mathcal{Z}$ to minimize

$$G(\boldsymbol{\tau}, \boldsymbol{\zeta}) = \gamma \text{E}\{J(\boldsymbol{\tau}, \boldsymbol{\zeta})\} + (1 - \gamma) \text{Var}\{J(\boldsymbol{\tau}, \boldsymbol{\zeta})\},$$

where $\text{E}\{\cdot\}$ denotes expectation, $\text{Var}\{\cdot\}$ denotes variance, and $\gamma \in [0, 1]$ is a given weight.

The aim in Problem P is to minimize *both* the average error and the error variance. The weight γ controls the relative importance between these two objectives. If γ is close to one, then the priority is to minimize average error; if γ is close to zero, then the priority is to minimize error variance. When the output distribution is known exactly, $\gamma = 1$ is the best option for minimizing the expected error. However, as we show in the numerical simulations in Section 6, when there are errors and/or uncertainties in the output distribution, it is essential to choose $\gamma < 1$ to ensure solution robustness.

3 Equivalent Formulation

Omitting the $\boldsymbol{\tau}$ and $\boldsymbol{\zeta}$ arguments in $\mathbf{y}(\cdot | \boldsymbol{\tau}, \boldsymbol{\zeta})$ for simplicity, the least-squares error function $J(\boldsymbol{\tau}, \boldsymbol{\zeta})$ can be written as follows:

$$\begin{aligned} J(\boldsymbol{\tau}, \boldsymbol{\zeta}) &= \sum_{k=1}^p \{\mathbf{y}(t_k)^\top \mathbf{y}(t_k) - 2(\hat{\mathbf{y}}^k)^\top \mathbf{y}(t_k) + (\hat{\mathbf{y}}^k)^\top \hat{\mathbf{y}}^k\} \\ &= \sum_{k=1}^p \mathbf{y}(t_k)^\top \mathbf{y}(t_k) - 2 \sum_{k=1}^p (\hat{\mathbf{y}}^k)^\top \mathbf{y}(t_k) + \sum_{k=1}^p (\hat{\mathbf{y}}^k)^\top \hat{\mathbf{y}}^k. \end{aligned}$$

Hence,

$$\begin{aligned} \text{E}\{J(\boldsymbol{\tau}, \boldsymbol{\zeta})\} &= \sum_{k=1}^p \mathbf{y}(t_k)^\top \mathbf{y}(t_k) \\ &\quad - 2 \sum_{k=1}^p \text{E}\{\hat{\mathbf{y}}^k\}^\top \mathbf{y}(t_k) + \sum_{k=1}^p \text{E}\{(\hat{\mathbf{y}}^k)^\top \hat{\mathbf{y}}^k\} \end{aligned} \quad (7)$$

and

$$\begin{aligned} \text{Var}\{J(\boldsymbol{\tau}, \boldsymbol{\zeta})\} &= 4 \text{Var}\left\{\sum_{k=1}^p (\hat{\mathbf{y}}^k)^\top \mathbf{y}(t_k)\right\} \\ &\quad + \text{Var}\left\{\sum_{k=1}^p (\hat{\mathbf{y}}^k)^\top \hat{\mathbf{y}}^k\right\} \\ &\quad - 4 \text{Cov}\left\{\sum_{k=1}^p (\hat{\mathbf{y}}^k)^\top \hat{\mathbf{y}}^k, \sum_{l=1}^p (\hat{\mathbf{y}}^l)^\top \mathbf{y}(t_l)\right\}. \end{aligned} \quad (8)$$

Note that

$$\begin{aligned} \text{Var}\left\{\sum_{k=1}^p (\hat{\mathbf{y}}^k)^\top \mathbf{y}(t_k)\right\} &= \sum_{k=1}^p \sum_{l=1}^p \text{Cov}\{(\hat{\mathbf{y}}^k)^\top \mathbf{y}(t_k), (\hat{\mathbf{y}}^l)^\top \mathbf{y}(t_l)\} \\ &= \sum_{k=1}^p \sum_{l=1}^p \mathbf{y}(t_k)^\top \Xi^{k,l} \mathbf{y}(t_l), \end{aligned} \quad (9)$$

where $\Xi^{k,l}$ is as defined in (6). Furthermore,

$$\begin{aligned} \text{Cov}\left\{\sum_{k=1}^p (\hat{\mathbf{y}}^k)^\top \hat{\mathbf{y}}^k, \sum_{l=1}^p (\hat{\mathbf{y}}^l)^\top \mathbf{y}(t_l)\right\} &= \sum_{k=1}^p \sum_{l=1}^p \text{Cov}\{(\hat{\mathbf{y}}^k)^\top \hat{\mathbf{y}}^k, (\hat{\mathbf{y}}^l)^\top \mathbf{y}(t_l)\} \\ &= \sum_{k=1}^p \sum_{l=1}^p \mathbf{1}^q \Upsilon^{k,l} \mathbf{y}(t_l), \end{aligned} \quad (10)$$

where $\Upsilon^{k,l}$ is as defined in (6) and $\mathbf{1}^q$ denotes a row vector of ones in \mathbb{R}^q . Substituting (9) and (10) into (8) gives

$$\begin{aligned} \text{Var}\{J(\boldsymbol{\tau}, \boldsymbol{\zeta})\} &= 4 \sum_{k=1}^p \sum_{l=1}^p \mathbf{y}(t_k)^\top \Xi^{k,l} \mathbf{y}(t_l) \\ &\quad + \text{Var}\left\{\sum_{k=1}^p (\hat{\mathbf{y}}^k)^\top \hat{\mathbf{y}}^k\right\} - 4 \sum_{k=1}^p \sum_{l=1}^p \mathbf{1}^q \Upsilon^{k,l} \mathbf{y}(t_l). \end{aligned} \quad (11)$$

Using (7) and (11), we have

$$\begin{aligned} G(\boldsymbol{\tau}, \boldsymbol{\zeta}) &= \gamma \sum_{k=1}^p \mathbf{y}(t_k)^\top \mathbf{y}(t_k) - 2\gamma \sum_{k=1}^p \text{E}\{\hat{\mathbf{y}}^k\}^\top \mathbf{y}(t_k) \\ &\quad + \gamma \sum_{k=1}^p \text{E}\{(\hat{\mathbf{y}}^k)^\top \hat{\mathbf{y}}^k\} + (1 - \gamma) \left[4 \sum_{k=1}^p \sum_{l=1}^p \mathbf{y}(t_k)^\top \Xi^{k,l} \mathbf{y}(t_l) \right. \\ &\quad \left. + \text{Var}\left\{\sum_{k=1}^p (\hat{\mathbf{y}}^k)^\top \hat{\mathbf{y}}^k\right\} - 4 \sum_{k=1}^p \sum_{l=1}^p \mathbf{1}^q \Upsilon^{k,l} \mathbf{y}(t_l) \right]. \end{aligned}$$

Thus, since the third and fifth terms on the right-hand side of this equation are independent of $\boldsymbol{\tau}$ and $\boldsymbol{\zeta}$, Problem P is equivalent to the following problem.

Problem Q. Choose $\boldsymbol{\tau} \in \mathcal{T}$ and $\boldsymbol{\zeta} \in \mathcal{Z}$ to minimize

$$\begin{aligned} H(\boldsymbol{\tau}, \boldsymbol{\zeta}) &= \gamma \sum_{k=1}^p \mathbf{y}(t_k)^\top \mathbf{y}(t_k) - 2\gamma \sum_{k=1}^p \mathbf{E}\{\hat{\mathbf{y}}^k\}^\top \mathbf{y}(t_k) \\ &+ 4(1-\gamma) \sum_{k=1}^p \sum_{l=1}^p [\mathbf{y}(t_k)^\top \boldsymbol{\Xi}^{k,l} \mathbf{y}(t_l) - \mathbf{1}^q \boldsymbol{\Upsilon}^{k,l} \mathbf{y}(t_l)], \end{aligned}$$

where $\mathbf{y}(\cdot) = \mathbf{y}(\cdot | \boldsymbol{\tau}, \boldsymbol{\zeta})$.

4 Computing the Matrices $\boldsymbol{\Xi}^{k,l}$ and $\boldsymbol{\Upsilon}^{k,l}$

In this section, we show how to compute the matrices $\boldsymbol{\Xi}^{k,l}$ and $\boldsymbol{\Upsilon}^{k,l}$ when the random vectors $\hat{\mathbf{y}}^k$, $k = 1, \dots, p$, can be expressed as linear combinations of d independent random variables z_1, \dots, z_d (not necessarily of the same distribution). In this case, there exists matrices $\boldsymbol{\Theta}^k = [\theta_{ij}^k]$, $k = 1, \dots, p$, of dimension $q \times d$ such that

$$\hat{\mathbf{y}}^k = \boldsymbol{\Theta}^k \mathbf{z}, \quad k = 1, \dots, p,$$

where $\mathbf{z} = [z_1, \dots, z_d]^\top \in \mathbb{R}^d$. Clearly,

$$\hat{y}_i^k = \sum_{u=1}^d \theta_{iu}^k z_u, \quad \hat{y}_j^l = \sum_{v=1}^d \theta_{jv}^l z_v.$$

Therefore, since the components of \mathbf{z} are independent,

$$\begin{aligned} [\boldsymbol{\Xi}^{k,l}]_{ij} &= \text{Cov}\{\hat{y}_i^k, \hat{y}_j^l\} \\ &= \sum_{u=1}^d \sum_{v=1}^d \theta_{iu}^k \theta_{jv}^l \text{Cov}\{z_u, z_v\} = \sum_{u=1}^d \theta_{iu}^k \theta_{ju}^l \text{Var}\{z_u\}. \end{aligned}$$

Furthermore,

$$\begin{aligned} [\boldsymbol{\Upsilon}^{k,l}]_{ij} &= \text{Cov}\{(\hat{y}_i^k)^2, \hat{y}_j^l\} \\ &= \sum_{u=1}^d \sum_{v=1}^d \sum_{w=1}^d \theta_{iu}^k \theta_{iv}^k \theta_{jw}^l \text{Cov}\{z_u z_v, z_w\}. \quad (12) \end{aligned}$$

Now, since $\text{Cov}\{z_u z_v, z_w\} = \mathbf{E}\{z_u z_v z_w\} - \mathbf{E}\{z_u z_v\} \mathbf{E}\{z_w\}$, the summand on the right-hand side of (12) can be simplified by considering four cases:

- $u = v = w$: $\text{Cov}\{z_u z_v, z_w\} = \mathbf{E}\{z_u^3\} - \mathbf{E}\{z_u\} \text{Var}\{z_u\} - \mathbf{E}\{z_u\}^3$.
- $u \neq v, u = w$: $\text{Cov}\{z_u z_v, z_w\} = \mathbf{E}\{z_v\} \text{Var}\{z_u\}$.
- $u \neq v, v = w$: $\text{Cov}\{z_u z_v, z_w\} = \mathbf{E}\{z_u\} \text{Var}\{z_v\}$.
- $u \neq w, v \neq w$: $\text{Cov}\{z_u z_v, z_w\} = 0$.

Thus, equation (12) becomes

$$\begin{aligned} [\boldsymbol{\Upsilon}^{k,l}]_{ij} &= \sum_{u=1}^d \sum_{\substack{v=1 \\ v \neq u}}^d \left\{ \theta_{iu}^k \theta_{iv}^k \theta_{ju}^l \mathbf{E}\{z_v\} \text{Var}\{z_u\} \right. \\ &\quad \left. + \theta_{iu}^k \theta_{iv}^k \theta_{jv}^l \mathbf{E}\{z_u\} \text{Var}\{z_v\} \right\} \\ &+ \sum_{u=1}^d (\theta_{iu}^k)^2 \theta_{ju}^l \left\{ \mathbf{E}\{z_u^3\} - \mathbf{E}\{z_u\} \text{Var}\{z_u\} - \mathbf{E}\{z_u\}^3 \right\}. \end{aligned}$$

This equation expresses $\boldsymbol{\Upsilon}^{k,l}$ in terms of the first, second, and third moments of the random variables z_1, \dots, z_d .

5 Gradient Computation

Problem Q is a dynamic optimization problem in which the decision vectors $\boldsymbol{\tau}$ and $\boldsymbol{\zeta}$ must be chosen to minimize the cost function H subject to the time-delay system (1)-(3) and the bound constraints (4) and (5). In principle, this problem can be solved using existing gradient-based optimization algorithms. However, to do so, the partial derivatives of the cost function H are required. These partial derivatives are difficult to obtain because H is not an explicit function of the decision vectors $\boldsymbol{\tau}$ and $\boldsymbol{\zeta}$. In this section, we will develop two numerical methods for computing the partial derivatives of H . The key idea is to view the state vector $\mathbf{x}(t | \boldsymbol{\tau}, \boldsymbol{\zeta})$ as a function of $\boldsymbol{\tau}$ and $\boldsymbol{\zeta}$, as well as time. It is then possible to consider the partial derivatives of $\mathbf{x}(t | \boldsymbol{\tau}, \boldsymbol{\zeta})$ with respect to $\boldsymbol{\tau}$ and $\boldsymbol{\zeta}$. This is similar to the classical sensitivity approach in the field of ordinary differential equations (Khalil (2002)).

Let

$$\boldsymbol{\chi}(t | \boldsymbol{\tau}, \boldsymbol{\zeta}) = \begin{cases} \frac{\partial \phi(t, \boldsymbol{\zeta})}{\partial t}, & \text{if } t \leq 0, \\ \bar{\mathbf{f}}(t | \boldsymbol{\tau}, \boldsymbol{\zeta}), & \text{if } t > 0, \end{cases}$$

where

$$\bar{\mathbf{f}}(t | \boldsymbol{\tau}, \boldsymbol{\zeta}) = \mathbf{f}(\mathbf{x}(t), \mathbf{x}(t - \tau_1), \dots, \mathbf{x}(t - \tau_m), \boldsymbol{\zeta}). \quad (13)$$

Clearly, for almost all time points $t \in \mathbb{R}$, we have $\dot{\mathbf{x}}(t | \boldsymbol{\tau}, \boldsymbol{\zeta}) = \boldsymbol{\chi}(t | \boldsymbol{\tau}, \boldsymbol{\zeta})$. Now, define

$$\frac{\partial \bar{\mathbf{f}}(t | \boldsymbol{\tau}, \boldsymbol{\zeta})}{\partial \zeta_j} = \frac{\partial \mathbf{f}(\mathbf{x}(t), \mathbf{x}(t - \tau_1), \dots, \mathbf{x}(t - \tau_m), \boldsymbol{\zeta})}{\partial \zeta_j}$$

and

$$\frac{\partial \bar{\mathbf{f}}(t | \boldsymbol{\tau}, \boldsymbol{\zeta})}{\partial \tilde{\mathbf{x}}^i} = \frac{\partial \mathbf{f}(\mathbf{x}(t), \mathbf{x}(t - \tau_1), \dots, \mathbf{x}(t - \tau_m), \boldsymbol{\zeta})}{\partial \mathbf{x}(t - \tau_i)},$$

where $\mathbf{x}(\cdot) = \mathbf{x}(\cdot | \boldsymbol{\tau}, \boldsymbol{\zeta})$ and $\tau_0 = 0$ (i.e., $\partial \tilde{\mathbf{x}}^0$ denotes differentiation with respect to \mathbf{x}).

For each $i = 1, \dots, m$, consider the following *sensitivity system*:

$$\begin{aligned} \dot{\Lambda}^i(t) &= \sum_{\varsigma=0}^m \frac{\partial \bar{\mathbf{f}}(t|\boldsymbol{\tau}, \boldsymbol{\zeta})}{\partial \bar{\mathbf{x}}^\varsigma} \Lambda^i(t - \tau_\varsigma) \\ &\quad - \frac{\partial \bar{\mathbf{f}}(t|\boldsymbol{\tau}, \boldsymbol{\zeta})}{\partial \bar{\mathbf{x}}^i} \boldsymbol{\chi}(t - \tau_i|\boldsymbol{\tau}, \boldsymbol{\zeta}), \quad t \geq 0, \quad (14) \\ \Lambda^i(t) &= \mathbf{0}, \quad t \leq 0, \quad (15) \end{aligned}$$

where $(\boldsymbol{\tau}, \boldsymbol{\zeta}) \in \mathcal{T} \times \mathcal{Z}$ is a given pair. Let $\Lambda^i(\cdot|\boldsymbol{\tau}, \boldsymbol{\zeta})$ denote the solution of (14)-(15) corresponding to $(\boldsymbol{\tau}, \boldsymbol{\zeta}) \in \mathcal{T} \times \mathcal{Z}$.

Under mild assumptions on the functions \mathbf{f} and ϕ , it can be shown (see Loxton et al. (2010) and Chai et al. (2013a)) that the state vector $\mathbf{x}(t|\boldsymbol{\tau}, \boldsymbol{\zeta})$ is differentiable with respect to each τ_i . Moreover, it turns out that the value of the corresponding partial derivative is given by the solution of (14)-(15):

$$\frac{\partial \mathbf{x}(t|\boldsymbol{\tau}, \boldsymbol{\zeta})}{\partial \tau_i} = \Lambda^i(t|\boldsymbol{\tau}, \boldsymbol{\zeta}), \quad t \in \mathbb{R}, \quad i = 1, \dots, m. \quad (16)$$

On this basis, we have the following theorem for the partial derivatives of H with respect to the time-delays.

Theorem 1. *For each $i = 1, \dots, m$, the partial derivative of H with respect to τ_i is given by*

$$\frac{\partial H(\boldsymbol{\tau}, \boldsymbol{\zeta})}{\partial \tau_i} = 2 \sum_{k=1}^p \boldsymbol{\sigma}^k(\boldsymbol{\tau}, \boldsymbol{\zeta})^\top \frac{\partial \mathbf{g}(\mathbf{x}(t_k), \boldsymbol{\zeta})}{\partial \mathbf{x}} \Lambda^i(t_k),$$

where $\mathbf{x}(\cdot) = \mathbf{x}(\cdot|\boldsymbol{\tau}, \boldsymbol{\zeta})$, $\Lambda^i(\cdot) = \Lambda^i(\cdot|\boldsymbol{\tau}, \boldsymbol{\zeta})$, and

$$\begin{aligned} \boldsymbol{\sigma}^k(\boldsymbol{\tau}, \boldsymbol{\zeta}) &= \gamma \mathbf{g}(\mathbf{x}(t_k), \boldsymbol{\zeta}) - \gamma \mathbf{E}\{\hat{\mathbf{y}}^k\} \\ &\quad + 2(1 - \gamma) \sum_{l=1}^p \{2\Xi^{k,l} \mathbf{g}(\mathbf{x}(t_l), \boldsymbol{\zeta}) - (\mathbf{1}^q \Upsilon^{l,k})^\top\}. \quad (17) \end{aligned}$$

Proof. Differentiate H with respect to τ_i and then apply equation (16) and the identity $\Xi^{l,k} = (\Xi^{k,l})^\top$. \square

We now consider the partial derivatives of H with respect to the system parameters. For each $j = 1, \dots, r$, define the following *sensitivity system*:

$$\begin{aligned} \dot{\Gamma}^j(t) &= \sum_{\varsigma=0}^m \frac{\partial \bar{\mathbf{f}}(t|\boldsymbol{\tau}, \boldsymbol{\zeta})}{\partial \bar{\mathbf{x}}^\varsigma} \Gamma^j(t - \tau_\varsigma) \\ &\quad + \frac{\partial \bar{\mathbf{f}}(t|\boldsymbol{\tau}, \boldsymbol{\zeta})}{\partial \zeta_j}, \quad t \geq 0, \quad (18) \end{aligned}$$

$$\Gamma^j(t) = \frac{\partial \phi(t, \boldsymbol{\zeta})}{\partial \zeta_j}, \quad t \leq 0, \quad (19)$$

where $(\boldsymbol{\tau}, \boldsymbol{\zeta}) \in \mathcal{T} \times \mathcal{Z}$ is a given pair. Let $\Gamma^j(\cdot|\boldsymbol{\tau}, \boldsymbol{\zeta})$ denote the solution of (18)-(19) corresponding to $(\boldsymbol{\tau}, \boldsymbol{\zeta}) \in \mathcal{T} \times \mathcal{Z}$. Then according to the results in Loxton et al. (2010) and Chai et al. (2013a), the following equation holds:

$$\frac{\partial \mathbf{x}(t|\boldsymbol{\tau}, \boldsymbol{\zeta})}{\partial \zeta_j} = \Gamma^j(t|\boldsymbol{\tau}, \boldsymbol{\zeta}), \quad t \in \mathbb{R}, \quad j = 1, \dots, r.$$

Thus, the solution of (18)-(19) coincides with the partial derivative of the state vector with respect to the j th system parameter. We now present the analogue of Theorem 1 for the partial derivatives of H with respect to the system parameters. The proof is similar to the proof of Theorem 1.

Theorem 2. *For each $j = 1, \dots, r$, the partial derivative of H with respect to ζ_j is given by*

$$\begin{aligned} \frac{\partial H(\boldsymbol{\tau}, \boldsymbol{\zeta})}{\partial \zeta_j} &= 2 \sum_{k=1}^p \boldsymbol{\sigma}^k(\boldsymbol{\tau}, \boldsymbol{\zeta})^\top \left\{ \frac{\partial \mathbf{g}(\mathbf{x}(t_k), \boldsymbol{\zeta})}{\partial \mathbf{x}} \Gamma^j(t_k) \right. \\ &\quad \left. + \frac{\partial \mathbf{g}(\mathbf{x}(t_k), \boldsymbol{\zeta})}{\partial \zeta_j} \right\}, \end{aligned}$$

where $\mathbf{x}(\cdot) = \mathbf{x}(\cdot|\boldsymbol{\tau}, \boldsymbol{\zeta})$, $\Gamma^j(\cdot) = \Gamma^j(\cdot|\boldsymbol{\tau}, \boldsymbol{\zeta})$, and $\boldsymbol{\sigma}^k(\boldsymbol{\tau}, \boldsymbol{\zeta})$ is defined by (17).

Recall from our discussion at the beginning of this section that computing the partial derivatives of H is the key step to solving Problem Q using gradient-based optimization techniques. By virtue of Theorems 1 and 2, these partial derivatives can be obtained by solving the state system (1)-(2) and the auxiliary systems (14)-(15) and (18)-(19). There are m auxiliary systems of the form (14)-(15) and r auxiliary systems of the form (18)-(19), with each auxiliary system of dimension n (the same dimension as the state space). Thus, evaluating the derivative formulae in Theorems 1 and 2 requires solving a large number of differential equations, which can be time-consuming in practice. We now derive an alternative method for computing the partial derivatives of H that requires solving only one auxiliary system.

Consider a new auxiliary system defined as follows:

$$\dot{\lambda}(t) = - \sum_{\varsigma=0}^m \left[\frac{\partial \bar{\mathbf{f}}(t + \tau_\varsigma|\boldsymbol{\tau}, \boldsymbol{\zeta})}{\partial \bar{\mathbf{x}}^\varsigma} \right]^\top \boldsymbol{\lambda}(t + \tau_\varsigma), \quad t \leq t_p, \quad (20)$$

with the intermediate jump conditions

$$\boldsymbol{\lambda}(t_k^-) = \boldsymbol{\lambda}(t_k^+) + 2 \left[\frac{\partial \mathbf{g}(\mathbf{x}(t_k|\boldsymbol{\tau}, \boldsymbol{\zeta}), \boldsymbol{\zeta})}{\partial \mathbf{x}} \right]^\top \boldsymbol{\sigma}^k(\boldsymbol{\tau}, \boldsymbol{\zeta}), \quad (21)$$

$$k = 1, \dots, p,$$

and the terminal condition

$$\boldsymbol{\lambda}(t) = \mathbf{0}, \quad t > t_p, \quad (22)$$

where $(\boldsymbol{\tau}, \boldsymbol{\zeta}) \in \mathcal{T} \times \mathcal{Z}$ is a given pair and $\boldsymbol{\sigma}^k(\boldsymbol{\tau}, \boldsymbol{\zeta})$ is defined by (17).

Note that system (20)-(22) is a ‘‘time-advance’’ system that must be solved backward in time starting from the terminal condition (22). Note also that this system is analogous to the well-known costate system in optimal control (Lin, Loxton & Teo (2014)). Let $\boldsymbol{\lambda}(\cdot|\boldsymbol{\tau}, \boldsymbol{\zeta})$ denote the left-continuous solution of (20)-(22) corresponding to $(\boldsymbol{\tau}, \boldsymbol{\zeta}) \in \mathcal{T} \times \mathcal{Z}$. We now express the partial derivatives of H with respect to the time-delays in terms of $\boldsymbol{\lambda}(\cdot|\boldsymbol{\tau}, \boldsymbol{\zeta})$.

Theorem 3. *For each $i = 1, \dots, m$, the partial derivative of H with respect to τ_i is given by*

$$\frac{\partial H(\boldsymbol{\tau}, \boldsymbol{\zeta})}{\partial \tau_i} = - \int_0^{t_p} \boldsymbol{\lambda}(t)^\top \frac{\partial \bar{\mathbf{f}}(t|\boldsymbol{\tau}, \boldsymbol{\zeta})}{\partial \tilde{\mathbf{x}}^i} \boldsymbol{\chi}(t - \tau_i|\boldsymbol{\tau}, \boldsymbol{\zeta}) dt,$$

where $\boldsymbol{\lambda}(\cdot) = \boldsymbol{\lambda}(\cdot|\boldsymbol{\tau}, \boldsymbol{\zeta})$.

Proof. Let $\mathbf{v}^k : [t_{k-1}, t_k] \rightarrow \mathbb{R}^n$, $k = 1, \dots, p$, be a set of arbitrary absolutely continuous functions. Multiplying equation (1) by \mathbf{v}^k and integrating over the subintervals $[t_{k-1}, t_k]$, $k = 1, \dots, p$, gives

$$\sum_{k=1}^p \int_{t_{k-1}}^{t_k} \mathbf{v}^k(t)^\top \dot{\mathbf{x}}(t) dt = \sum_{k=1}^p \int_{t_{k-1}}^{t_k} \mathbf{v}^k(t)^\top \bar{\mathbf{f}}(t|\boldsymbol{\tau}, \boldsymbol{\zeta}) dt,$$

where $\bar{\mathbf{f}}(\cdot|\boldsymbol{\tau}, \boldsymbol{\zeta})$ is defined by (13).

Thus, using integration-by-parts,

$$\sum_{k=1}^p \left\{ \mathbf{v}^k(t_k)^\top \mathbf{x}(t_k) - \mathbf{v}^k(t_{k-1})^\top \mathbf{x}(t_{k-1}) - \int_{t_{k-1}}^{t_k} \dot{\mathbf{v}}^k(t)^\top \mathbf{x}(t) dt \right\} = \sum_{k=1}^p \int_{t_{k-1}}^{t_k} \mathbf{v}^k(t)^\top \bar{\mathbf{f}}(t|\boldsymbol{\tau}, \boldsymbol{\zeta}) dt.$$

This equation can be rearranged to give

$$\begin{aligned} \mathbf{v}^p(t_p)^\top \mathbf{x}(t_p) + \sum_{k=1}^{p-1} \{ \mathbf{v}^k(t_k)^\top - \mathbf{v}^{k+1}(t_k)^\top \} \mathbf{x}(t_k) \\ - \mathbf{v}^1(0)^\top \boldsymbol{\phi}(0, \boldsymbol{\zeta}) \\ = \sum_{k=1}^p \int_{t_{k-1}}^{t_k} \left\{ \dot{\mathbf{v}}^k(t)^\top \mathbf{x}(t) + \mathbf{v}^k(t)^\top \bar{\mathbf{f}}(t|\boldsymbol{\tau}, \boldsymbol{\zeta}) \right\} dt. \end{aligned}$$

Now, by differentiating this equation with respect to τ_i ,

and then applying equation (16), we obtain

$$\begin{aligned} \mathbf{v}^p(t_p)^\top \boldsymbol{\Lambda}^i(t_p) + \sum_{k=1}^{p-1} \{ \mathbf{v}^k(t_k)^\top - \mathbf{v}^{k+1}(t_k)^\top \} \boldsymbol{\Lambda}^i(t_k) \\ = \sum_{k=1}^p \int_{t_{k-1}}^{t_k} \left\{ \dot{\mathbf{v}}^k(t)^\top \boldsymbol{\Lambda}^i(t) - \mathbf{v}^k(t)^\top \frac{\partial \bar{\mathbf{f}}(t|\boldsymbol{\tau}, \boldsymbol{\zeta})}{\partial \tilde{\mathbf{x}}^i} \boldsymbol{\chi}(t - \tau_i) \right. \\ \left. + \sum_{\varsigma=0}^m \mathbf{v}^k(t)^\top \frac{\partial \bar{\mathbf{f}}(t|\boldsymbol{\tau}, \boldsymbol{\zeta})}{\partial \tilde{\mathbf{x}}^\varsigma} \boldsymbol{\Lambda}^i(t - \tau_\varsigma) \right\} dt, \quad (23) \end{aligned}$$

where $\boldsymbol{\chi}(\cdot) = \boldsymbol{\chi}(\cdot|\boldsymbol{\tau}, \boldsymbol{\zeta})$ and $\boldsymbol{\Lambda}^i(\cdot) = \boldsymbol{\Lambda}^i(\cdot|\boldsymbol{\tau}, \boldsymbol{\zeta})$.

Since $\mathbf{v}^k : [t_{k-1}, t_k] \rightarrow \mathbb{R}^n$, $k = 1, \dots, p$, are arbitrary functions, we can choose

$$\mathbf{v}^k(t) := \begin{cases} \boldsymbol{\lambda}(t_{k-1}^+), & \text{if } t = t_{k-1}, \\ \boldsymbol{\lambda}(t), & \text{if } t \in (t_{k-1}, t_k), \\ \boldsymbol{\lambda}(t_k^-), & \text{if } t = t_k, \end{cases} \quad (24)$$

where $\boldsymbol{\lambda}(\cdot) = \boldsymbol{\lambda}(\cdot|\boldsymbol{\tau}, \boldsymbol{\zeta})$ is the solution of the auxiliary system (20)-(22). Then by using Theorem 1 and equations (21) and (22), equation (23) becomes

$$\begin{aligned} \frac{\partial H(\boldsymbol{\tau}, \boldsymbol{\zeta})}{\partial \tau_i} &= \int_0^{t_p} \dot{\boldsymbol{\lambda}}(t)^\top \boldsymbol{\Lambda}^i(t) dt \\ &+ \sum_{\varsigma=0}^m \int_0^{t_p} \boldsymbol{\lambda}(t)^\top \frac{\partial \bar{\mathbf{f}}(t|\boldsymbol{\tau}, \boldsymbol{\zeta})}{\partial \tilde{\mathbf{x}}^\varsigma} \boldsymbol{\Lambda}^i(t - \tau_\varsigma) dt \\ &- \int_0^{t_p} \boldsymbol{\lambda}(t)^\top \frac{\partial \bar{\mathbf{f}}(t|\boldsymbol{\tau}, \boldsymbol{\zeta})}{\partial \tilde{\mathbf{x}}^i} \boldsymbol{\chi}(t - \tau_i) dt. \end{aligned}$$

Hence, by performing a change of variable in the second integral on the right-hand side,

$$\begin{aligned} \frac{\partial H(\boldsymbol{\tau}, \boldsymbol{\zeta})}{\partial \tau_i} &= \int_0^{t_p} \dot{\boldsymbol{\lambda}}(t)^\top \boldsymbol{\Lambda}^i(t) dt \\ &+ \sum_{\varsigma=0}^m \int_{-\tau_\varsigma}^{t_p - \tau_\varsigma} \boldsymbol{\lambda}(t + \tau_\varsigma)^\top \frac{\partial \bar{\mathbf{f}}(t + \tau_\varsigma|\boldsymbol{\tau}, \boldsymbol{\zeta})}{\partial \tilde{\mathbf{x}}^\varsigma} \boldsymbol{\Lambda}^i(t) dt \\ &- \int_0^{t_p} \boldsymbol{\lambda}(t)^\top \frac{\partial \bar{\mathbf{f}}(t|\boldsymbol{\tau}, \boldsymbol{\zeta})}{\partial \tilde{\mathbf{x}}^i} \boldsymbol{\chi}(t - \tau_i) dt. \end{aligned}$$

In view of (15) and (22), this can be simplified to yield

$$\begin{aligned} \frac{\partial H(\boldsymbol{\tau}, \boldsymbol{\zeta})}{\partial \tau_i} &= \int_0^{t_p} \dot{\boldsymbol{\lambda}}(t)^\top \boldsymbol{\Lambda}^i(t) dt \\ &+ \sum_{\varsigma=0}^m \int_0^{t_p} \boldsymbol{\lambda}(t + \tau_\varsigma)^\top \frac{\partial \bar{\mathbf{f}}(t + \tau_\varsigma|\boldsymbol{\tau}, \boldsymbol{\zeta})}{\partial \tilde{\mathbf{x}}^\varsigma} \boldsymbol{\Lambda}^i(t) dt \\ &- \int_0^{t_p} \boldsymbol{\lambda}(t)^\top \frac{\partial \bar{\mathbf{f}}(t|\boldsymbol{\tau}, \boldsymbol{\zeta})}{\partial \tilde{\mathbf{x}}^i} \boldsymbol{\chi}(t - \tau_i) dt. \end{aligned}$$

Applying (20) completes the proof. \square

The next theorem expresses the partial derivatives of H with respect to the system parameters in terms of the solution of (20)-(22). The proof is similar to the proof of Theorem 3.

Theorem 4. For each $j = 1, \dots, r$, the partial derivative of H with respect to ζ_j is given by

$$\begin{aligned} \frac{\partial H(\boldsymbol{\tau}, \boldsymbol{\zeta})}{\partial \zeta_j} &= 2 \sum_{k=1}^p \boldsymbol{\sigma}^k(\boldsymbol{\tau}, \boldsymbol{\zeta})^\top \frac{\partial \mathbf{g}(\mathbf{x}(t_k), \boldsymbol{\zeta})}{\partial \zeta_j} \\ &+ \boldsymbol{\lambda}(0^+)^\top \frac{\partial \phi(0, \boldsymbol{\zeta})}{\partial \zeta_j} + \int_0^{t_p} \boldsymbol{\lambda}(t)^\top \frac{\partial \bar{\mathbf{f}}(t|\boldsymbol{\tau}, \boldsymbol{\zeta})}{\partial \zeta_j} dt \\ &+ \sum_{\varsigma=1}^m \int_{-\tau_\varsigma}^0 \boldsymbol{\lambda}(t + \tau_\varsigma)^\top \frac{\partial \bar{\mathbf{f}}(t + \tau_\varsigma|\boldsymbol{\tau}, \boldsymbol{\zeta})}{\partial \bar{\mathbf{x}}^\varsigma} \frac{\partial \phi(t, \boldsymbol{\zeta})}{\partial \zeta_j} dt, \end{aligned}$$

where $\boldsymbol{\lambda}(\cdot) = \boldsymbol{\lambda}(\cdot|\boldsymbol{\tau}, \boldsymbol{\zeta})$ and $\boldsymbol{\sigma}^k(\boldsymbol{\tau}, \boldsymbol{\zeta})$ is defined by (17).

By exploiting the derivative formulae in either Theorems 1 and 2 or Theorems 3 and 4, Problem Q can be solved using gradient-based optimization algorithms. Note that, in general, such algorithms are only guaranteed to find locally-optimal points that satisfy the first-order optimality conditions (the well-known KKT conditions). Nevertheless, for Problem Q, it is highly likely that a global solution will be obtained. This is because the number of decision variables in Problem Q is small compared with most nonlinear optimization problems. Moreover, from practical experience, it is usually possible to determine good initial guesses for the optimal parameter estimates.

6 Simulation Results

Consider a continuously-stirred tank reactor in which the reaction $A \rightarrow B$ occurs. The reaction dynamics can be described by the following delay-differential equations (Loxton et al. (2010)):

$$\begin{aligned} \dot{x}_1(t) &= \zeta_1 x_1(t) + (1 - x_1(t)) \exp \left[\frac{20x_2(t)}{x_2(t) + 20} \right] \\ &+ x_1(t - \tau), \end{aligned} \quad (24)$$

$$\begin{aligned} \dot{x}_2(t) &= \zeta_2 x_2(t) + (1 - x_1(t)) \exp \left[\frac{20x_2(t)}{x_2(t) + 20} \right] \\ &+ x_2(t - \tau), \end{aligned} \quad (25)$$

with initial conditions

$$x_1(t) = 1, \quad x_2(t) = 1, \quad t \leq 0, \quad (26)$$

where x_1 is the (dimensionless) concentration of A ; x_2 is the (dimensionless) temperature of the reactor; and τ , ζ_1 , and ζ_2 are unknown model parameters that need to be identified.

The system output is given by

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

Hence, the least-squares error function is

$$J(\tau, \zeta_1, \zeta_2) = \sum_{k=1}^p \{y(t_k) - \hat{y}^k\}^2 = \sum_{k=1}^p \{10x_2(t_k) - \hat{y}^k\}^2,$$

where p is the number of sample times, t_k is the k th sample time, and \hat{y}^k is a random variable representing the output measurement at the k th sample time. We choose $p = 20$ equidistant sample times:

$$t_k = \frac{1}{2}k, \quad k = 1, \dots, 20.$$

The problem is to choose τ , ζ_1 , and ζ_2 to minimize

$$G(\tau, \zeta_1, \zeta_2) = \gamma \mathbb{E}\{J(\tau, \zeta_1, \zeta_2)\} + (1 - \gamma) \text{Var}\{J(\tau, \zeta_1, \zeta_2)\}$$

subject to the delay-differential equations (24)-(25) and the initial conditions (26), where $\gamma \in [0, 1]$ is a given weight. Based on the discussion in Section 3, this problem can be reformulated as follows: choose τ , ζ_1 , and ζ_2 to minimize

$$\begin{aligned} H(\tau, \zeta_1, \zeta_2) &= \gamma \sum_{k=1}^p y(t_k)^2 - 2\gamma \sum_{k=1}^p \mathbb{E}\{\hat{y}^k\} y(t_k) \\ &+ 4(1 - \gamma) \sum_{k=1}^p \sum_{l=1}^p [y(t_k) \Xi^{k,l} y(t_l) - \Upsilon^{k,l} y(t_l)] \end{aligned}$$

subject to the delay-differential equations (24)-(25) and the initial conditions (26), where

$$\Xi^{k,l} = \text{Cov}\{\hat{y}^k, \hat{y}^l\}, \quad \Upsilon^{k,l} = \text{Cov}\{(\hat{y}^k)^2, \hat{y}^l\}.$$

Consider the output trajectory of (24)-(26) corresponding to the nominal parameter estimates

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

This output trajectory is known as the *reference trajectory*. To generate the output data for our numerical experiments, we randomly perturbed the reference trajectory using independent normal random variables. The sample points obtained are shown in Figure 1.

Now, suppose that the noisy output measurements take the following form:

$$\hat{y}^k = \alpha_k + \sum_{j=1}^k \beta_j, \quad k = 1, \dots, 20, \quad (27)$$

where α_k , $k = 1, \dots, 20$, and β_j , $j = 1, \dots, 20$, are inde-

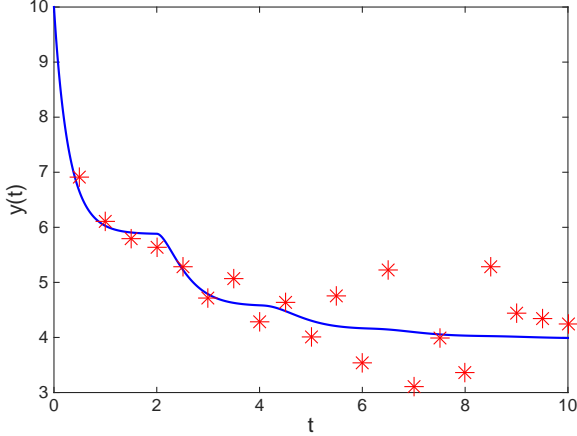


Fig. 1. Reference output trajectory (blue line) and corresponding perturbed sample points (red stars).

pendent random variables.

We assume that α_k follows a gamma distribution with parameters 10 and $10/\bar{y}^k$, where \bar{y}^k is the k th sample point in Figure 1. The mean, variance, and third moment of α_k are determined by the distribution parameters as follows:

$$\mathbb{E}\{\alpha_k\} = \bar{y}^k, \quad \text{Var}\{\alpha_k\} = \frac{1}{10}(\bar{y}^k)^2, \quad \mathbb{E}\{\alpha_k^3\} = \frac{33}{25}(\bar{y}^k)^3.$$

We also assume that $\beta_j = \bar{\beta}_j - \mathbb{E}\{\bar{\beta}_j\}$, where $\bar{\beta}_j$ follows a beta distribution with parameters 2 and 5. The mean, variance, and third moment of $\bar{\beta}_j$ are

$$\mathbb{E}\{\bar{\beta}_j\} = \frac{2}{7}, \quad \text{Var}\{\bar{\beta}_j\} = \frac{5}{196}, \quad \mathbb{E}\{\bar{\beta}_j^3\} = \frac{1}{21}.$$

Thus,

$$\begin{aligned} \mathbb{E}\{\beta_j\} &= 0, \quad \text{Var}\{\beta_j\} = \text{Var}\{\bar{\beta}_j\} = \frac{5}{196}, \\ \mathbb{E}\{\beta_j^3\} &= \mathbb{E}\{\bar{\beta}_j^3\} - 3\mathbb{E}\{\bar{\beta}_j^2\}\mathbb{E}\{\bar{\beta}_j\} + 2\mathbb{E}\{\bar{\beta}_j\}^3 = \frac{5}{2058}. \end{aligned}$$

Based on the distributions of α_k and β_j , we have

$$\mathbb{E}\{\hat{y}^k\} = \mathbb{E}\{\alpha_k\} + \sum_{j=1}^k \mathbb{E}\{\beta_j\} = \bar{y}^k, \quad k = 1, \dots, 20.$$

To solve the estimation problem, we wrote a Fortran program that combines the optimization code NLPQLP (Schittkowski (2007)) with the derivative formulae in Section 5 (the user can choose between the formulae in Theorems 1-2 and the formulae in Theorems 3-4). The state and auxiliary systems are solved using the sixth-order Runge-Kutta method, with the delay/advance values calculated using Hermite interpolation.

Using our Fortran program, we solved the estimation problem for $\gamma = 0, 0.01, 0.02, \dots, 1.0$. Both the forward

γ	τ^*	ζ_1^*	ζ_2^*
0.0	1.50040	-1.67516	-2.09934
0.1	1.50129	-1.67723	-2.10157
0.2	1.50249	-1.67979	-2.10434
0.3	1.50394	-1.68303	-2.10782
0.4	1.50591	-1.68725	-2.11236
0.5	1.50856	-1.69299	-2.11853
0.6	1.51245	-1.70126	-2.12739
0.7	1.51867	-1.71414	-2.14118
0.8	1.52999	-1.73706	-2.16561
0.9	1.55661	-1.78917	-2.22072
1.0	1.65698	-2.01224	-2.45646

Table 1
Optimal parameter estimates for $\gamma = 0, 0.1, \dots, 1.0$.

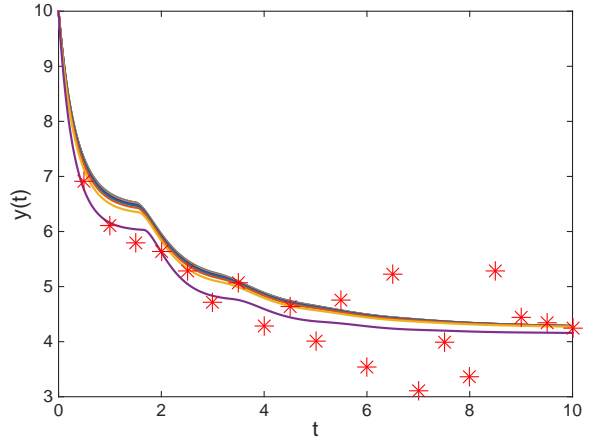


Fig. 2. Output trajectories corresponding to the optimal parameter estimates in Table 1. The red crosses are the sample points $\{(t_k, \bar{y}^k)\}_{k=1}^{20}$.

and the backward gradient computation schemes worked equally well in our computations. The optimal parameter estimates for $\gamma = 0, 0.1, \dots, 1.0$ are given in Table 1 and the corresponding output trajectories are shown in Figure 2. Note that, due to the uncertain nature of the output measurements, the optimal system trajectories do not intersect the sample points.

To investigate solution robustness, we generated 100,000 realizations of the output data \hat{y}^k , $k = 1, \dots, 20$. For each realization, we calculated the least-squares error corresponding to the optimal parameter estimates for $\gamma = 0, 0.01, 0.02, \dots, 1.0$. Figure 3 shows the mean and variance of the least-squares error for the 100,000 output realizations. The tradeoff between error mean and error variance is clearly apparent. Note that the dashed blue curve (error variance) is much steeper than the solid red

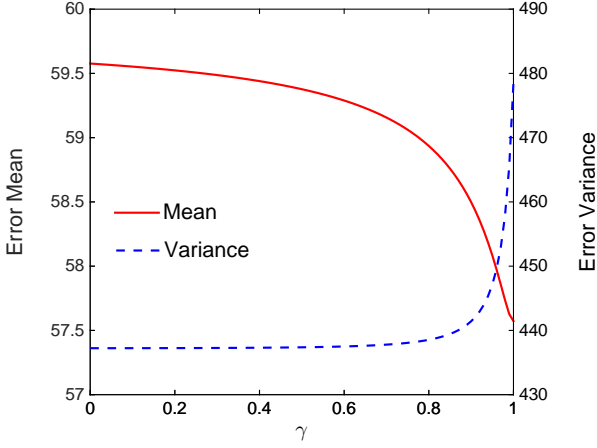


Fig. 3. Mean and variance of the least-squares error for 100,000 output realizations generated according to (27).

γ	Mean Error	Error Variance
0.95	58.08380	447.12879
0.96	57.97462	449.53893
0.97	57.85729	452.91093
0.98	57.73659	457.84531
0.99	57.62714	465.52275
1.00	57.57255	478.54902

Table 2
Mean and variance of the least-squares error for 100,000 realizations generated according to (27).

curve (error mean) near $\gamma = 1$ (see also Table 2). For example, when γ changes from $\gamma = 0.99$ to $\gamma = 1.0$, the variance increases from 465.52275 to 478.54902, but the mean drops only slightly from 57.62714 to 57.57255. This clearly demonstrates the advantage of our new stochastic formulation: imposing a small penalty on the error variance increases solution robustness, with negligible cost to the error mean.

Note that, for $\gamma = 1$, the cost function in Problem P is

$$\begin{aligned}
\mathbb{E}\{J(\tau, \zeta_1, \zeta_2)\} &= \mathbb{E}\left\{\sum_{k=1}^p \{y(t_k) - \hat{y}^k\}^2\right\} \\
&= \sum_{k=1}^p \left\{y(t_k)^2 - 2\mathbb{E}\{\hat{y}^k\}y(t_k) + \mathbb{E}\{(\hat{y}^k)^2\}\right\} \\
&= \sum_{k=1}^p \left\{y(t_k)^2 - 2\mathbb{E}\{\hat{y}^k\}y(t_k) + \text{Var}\{\hat{y}^k\} + \mathbb{E}\{\hat{y}^k\}^2\right\} \\
&= \sum_{k=1}^p \{y(t_k) - \mathbb{E}\{\hat{y}^k\}\}^2 + \sum_{k=1}^p \text{Var}\{\hat{y}^k\},
\end{aligned}$$

where $\mathbb{E}\{\hat{y}^k\} = \bar{y}^k$. Since the last term is constant with respect to τ , ζ_1 , and ζ_2 , Problem P for $\gamma = 1$ is equivalent to the following deterministic estimation problem: choose τ , ζ_1 , and ζ_2 to minimize

$$\sum_{k=1}^p \{y(t_k) - \mathbb{E}\{\hat{y}^k\}\}^2 = \sum_{k=1}^p \{10x_2(t_k) - \mathbb{E}\{\hat{y}^k\}\}^2,$$

subject to the delay-differential equations (24)-(25) and the initial conditions (26). This deterministic formulation, in which the output measurements are fixed values instead of random variables, was proposed in Loxton et al. (2010) and Chai et al. (2013a). Table 2 shows that our new stochastic formulation gives more robust results than the deterministic formulation (which is equivalent to $\gamma = 1$). In addition, a side benefit of our new approach is that the backward gradient computation scheme in Theorems 3-4 requires solving fewer differential equations than the corresponding methods in Loxton et al. (2010) and Chai et al. (2013a).

The optimal parameter estimates obtained by our program are based on the output distribution specified by equation (27). We now investigate how these optimal estimates perform when the actual output distribution differs slightly from the assumed distribution—a common occurrence in practice. Thus, instead of equation (27), we now suppose that the output distribution is given by

$$\hat{y}^k = \eta_k + \sum_{j=1}^k \beta_j, \quad k = 1, \dots, 20, \quad (28)$$

where η_k and β_j are independent random variables. Here, β_j is as defined in (27) and η_k follows a gamma distribution with parameters 10 and $10/(\bar{y}^k + \epsilon)$, where ϵ is a small parameter. The parameter ϵ ensures that η_k differs slightly from α_k in (27). We generated 100,000 realizations of the output data according to (28) with $\epsilon = 0.2$. For each realization, we calculated the least-squares estimation error corresponding to the optimal solutions obtained previously for $\gamma = 0, 0.01, \dots, 1.0$. The mean and variance of the least-squares error are shown in Figure 4. Note that, as expected, the mean error in Figure 4 is higher than the mean error in Figure 3 due to the discrepancy between the actual output distribution (28) and the assumed output distribution (27). In the face of such discrepancies, the benefits of our new stochastic formulation are clearly apparent: since the deviation in the error mean is much smaller than the deviation in the error variance, imposing a small penalty on the error variance hardly changes the error mean, but significantly improves solution robustness. In particular, the solution for $\gamma = 1$ (which is the same as the solution from the deterministic formulation in Loxton et al. (2010) and Chai et al. (2013a)) performs poorly, yielding more error on average with much higher variance.

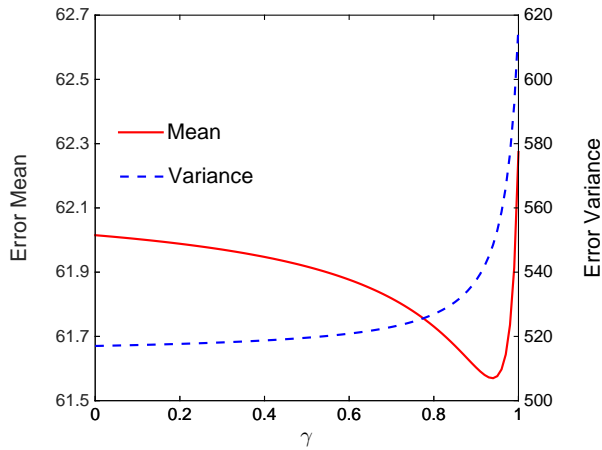


Fig. 4. Mean and variance of the least-squares error for 100,000 output realizations generated according to (28).

7 Conclusion

This paper has considered a dynamic optimization problem in which the goal is to choose optimal estimates for unknown time-delays and system parameters in a time-delay dynamic model. This dynamic optimization problem is similar to those formulated in earlier work (see Loxton et al. (2010) and Chai et al. (2013a, 2013b)), but with one important difference: the output measurements are random variables—not fixed constants—to allow for system noise and measurement inaccuracies. In contrast, the output measurements in the aforementioned references are assumed to be exact—clearly an idealistic assumption. Indeed, identical experiments run at different times are unlikely to yield exactly the same results—there will usually be some random variation due to system disturbances and measurement errors.

The numerical results in Section 6 demonstrate the effectiveness of our new parameter estimation approach. Nevertheless, several interesting research questions remain unanswered. For example, how does the choice of sample times influence estimation accuracy, and what properties should the functions f (system dynamics) and g (output function) satisfy to guarantee high-quality estimates? The second question is related to the so-called *identifiability issue*, which has been studied extensively in the literature (see Lunel (2001), Denis-Vidal, Jauberthie & Joly-Blanchard (2006), Zheng et al. (2013), and the references cited therein). One of the limitations of the new estimation method proposed in this paper is that it relies on the statistical properties of the output noise. Future work will focus on extending this method to the case where such statistical data is unavailable.

References

[1] Banks, H.T., Rehm, K., & Sutton, K. (2010). Inverse problems for nonlinear delay systems. *Methods and Applications of Analysis*, 17(4), 331-356.

[2] Belkoura, L., Richard, J.P., & Fliess, M. (2009). Parameters estimation of systems with delayed and structured entries. *Automatica*, 45(5), 1117-1125.

[3] Chai, Q., Loxton, R., Teo, K.L., & Yang, C. (2013). A unified parameter identification method for nonlinear time-delay systems. *Journal of Industrial and Management Optimization*, 9(2), 471-486.

[4] Chai, Q., Loxton, R., Teo, K.L., & Yang, C. (2013). Time-delay estimation for nonlinear systems with piecewise-constant input. *Applied Mathematics and Computation*, 219(17), 9543-9560.

[5] Denis-Vidal, L., Jauberthie, C., & Joly-Blanchard, G. (2006). Identifiability of a nonlinear delayed-differential aerospace model. *IEEE Transactions on Automatic Control*, 51(1), 154-158.

[6] Drakunov, S.V., Perruquetti, W., Richard, J.P., & Belkoura, L. (2006). Delay identification in time-delay systems using variable structure observers. *Annual Reviews in Control*, 30(2), 143-158.

[7] Gao, F., Qi, Y., Yin, Q., & Xiao, J. (2010). An artificial bee colony algorithm for unknown parameters and time-delays identification of chaotic systems. In *Proceedings of the 5th International Conference on Computer Sciences and Convergence Information Technology*, Seoul, Korea.

[8] Khalil, H.K. (2002). *Nonlinear Systems* (3rd ed.), New Jersey, Prentice Hall.

[9] Lin, Q., Loxton, R., & Teo, K.L. (2014). The control parameterization method for nonlinear optimal control: A survey. *Journal of Industrial and Management Optimization*, 10(1), 275-309.

[10] Loxton, R., Teo, K.L., & Rehbock, V. (2010). An optimization approach to state-delay identification. *IEEE Transactions on Automatic Control*, 55(9), 2113-2119.

[11] Lunel, S.M.V. (2001). Parameter identifiability of differential delay equations. *International Journal of Adaptive Control and Signal Processing*, 15(6), 655-678.

[12] Orlov, Y., Belkoura, L., Richard, J.P., & Dambrine, M. (2002). On identifiability of linear time-delay systems. *IEEE Transactions on Automatic Control*, 47(8), 1319-1324.

[13] Orlov, Y., Belkoura, L., Richard, J.P., & Dambrine, M. (2003). Adaptive identification of linear time-delay systems. *International Journal of Robust and Nonlinear Control*, 13(9), 857-872.

[14] Park, J.H., Han, S., & Kwon, B. (2013). On-line model parameter estimations for time-delay systems. *IEICE Transactions on Information and Systems*, E96-D(8), 1867-1870.

[15] Schittkowski, K. (2007). NLPQLP: A Fortran Implementation of a Sequential Quadratic Programming Algorithm with Distributed and Non-monotone Line Search – User’s Guide. Bayreuth, University of Bayreuth.

[16] Tang, Y. & Guan, X. (2009). Parameter estimation for time-delay chaotic system by particle swarm optimization. *Chaos, Solitons and Fractals*, 40(3), 1391-1398.

[17] Tuch, J., Feuer, A., & Palmor, Z.J. (1994). Time delay estimation in continuous linear time-invariant systems. *IEEE Transactions on Automatic Control*, 39(4), 823-827.

[18] Zheng, G., Barbot, J.P., & Boutat, D. (2013). Identification of the delay parameter for nonlinear time-delay systems with unknown inputs. *Automatica*, 49(6), 1755-1760.