



Identification of nonlinear dynamical systems with time delay

Ghazaale Leylaz¹ · Shuo Wang² · Jian-Qiao Sun¹

Received: 6 January 2021 / Revised: 20 February 2021 / Accepted: 2 March 2021 / Published online: 1 April 2021
© The Author(s) 2021

Abstract

This paper proposes a technique to identify nonlinear dynamical systems with time delay. The sparse optimization algorithm is extended to nonlinear systems with time delay. The proposed algorithm combines cross-validation techniques from machine learning for automatic model selection and an algebraic operation for preprocessing signals to filter the noise and for removing the dependence on initial conditions. We further integrate the bootstrapping resampling technique with the sparse regression to obtain the statistical properties of estimation. We use Taylor expansion to parameterize time delay. The proposed algorithm in this paper is computationally efficient and robust to noise. A nonlinear Duffing oscillator is simulated to demonstrate the efficiency and accuracy of the proposed technique. An experimental example of a nonlinear rotary flexible joint is presented to further validate the proposed method.

Keywords Time delay estimation · Sparse regression · Algebraic operation · System identification · Bootstrapping resampling · Nonlinear dynamics

1 Introduction

Time delay exists in many engineering, physics, chemistry, biology and economics systems. In control systems, time delay due to sensor and actuator dynamics, signal transmission, and digital computations is an important factor that influences the stability and control performance. To make matter worse, time delay is often unknown. Time delay estimation in a control system is a challenging problem. It is even more challenging when the system dynamics is nonlinear and unknown. This paper presents a nonparametric identification technique to identify nonlinear dynamic systems and estimate time delay introduced by the feedback control.

There have been many studies of time delay identification of control systems. Richard presented an overview of time

delay estimation methods in [1]. The time delay estimation techniques based on pulsed inputs have been developed in [2,3]. The Padé approximation [4], the modified least square and recursive methods [5–9], instrumental variable identification [10], neural networks [11], algebraic estimation [12,13], adaptive techniques [14,15], and non-commutative rings [16] are just a few popular methods for time delay estimation.

The methods for time delay estimation can be in frequency or time domain [17,18]. In this paper, our focus is on the approaches in time domain. Since time delay usually appears in the system implicitly, the methods for conventional parameter estimation of dynamic systems cannot be directly applied to estimate time delay. Time delay τ usually appears in the exponential term $e^{-\tau s}$ in the transfer function of the system. The expansion techniques can parameterize it, including the classical Padé approximation, the Laguerre Fourier series, the Kautz series, the second-order Padé shift and the diagonal Padé shift. The main concern with the rational approximation is the truncation error and stability complications. Even though higher order expansions can reduce the truncation error, the system can become unstable even when the system is linear with a constant time delay [1]. In this paper, we employ the Taylor expansion. Xu demonstrates that the low order Taylor expansion gives promising estimation of small time delay [19].

✉ Jian-Qiao Sun
jqsun@ucmerced.edu

Ghazaale Leylaz
gleylaz@ucmerced.edu

Shuo Wang
shuowangcn@outlook.com

¹ Department of Mechanical Engineering, School of Engineering, University of California, Merced, CA 95343, USA

² Department of Mathematics, Information Engineering University, Zhengzhou, China

The work in [20] presents a nonlinear least square-based algorithm in which the instrumental variable method estimates the parameters of the transfer function of the system while an adaptive gradient-based iteration finds the optimal time delay from the filtered irregularly sampled data. The main problem with this algorithm is that the proposed cost function may have several local minima. Therefore, it highly depends on the initial guess for the parameters and especially the time delay. To deal with this issue, the authors use a low-pass filter to widen the convergence region around the global minimum. Another nonlinear recursive optimization algorithm is proposed in [21], which combines the linear method of Levenberg–Marquardt to compute the plant parameters with a modified Gauss–Newton algorithm to estimate time delays. A low-pass filter and a binary transformation are applied to the data corrupted by the white noise to create the regressor matrix for minimizing a quadratic cost function of the estimation error. The identification is online and is demonstrated on a MISO plant with multiple time delays. Similarly, the algorithm in [9] employs the Gauss–Newton method to estimate time delay when the simplified refined instrumental variable (*SRIVC*) method is used to find the plant parameters.

The Taylor expansion is used to parameterize the system with explicit time delay in [15]. An adaptive law for the parameter estimation is proposed such that the estimation error is converged. A recursive formula is introduced in [6] to improve the accuracy and convergence rate of another recursive algorithm for online estimation in [22]. The strategy in [8] aims at fractional time delay identification for discrete-time systems. It separates the influence of the system structure and the time delay by discretizing the system. With the help of the Kalman filter, the parameters are estimated recursively. The algorithm in [11] first parameterizes the system by a polynomial function and trains a neural network to estimate the parameters and time delay. A Schweizer and Wolff's σ measure denoted as σ_{SW} from the copula theory is introduced to study the relationship of input-output signals for SISO systems [23]. It is found that the measure reaches its maximum when the time delay is removed from the data. This property offers an approach to estimate time delay without the need of estimation of other parameters.

Most existing methods for time delay estimation rely on the knowledge of the system model. In this work, we propose the application of a recently developed nonparametric system identification technique to the time delay estimation. Brunton and colleagues proposed a method called the sparse identification of nonlinear dynamics (SINDy) in 2016 for creating a sparse representation of the unknown nonlinear function of the system [24]. The method has attracted a great attention from the community. It assumes that only a few important terms out of a library of functions are needed to describe the dynamics of the system. It combines machine learning and

sparsity-promoting techniques to find the sparse representation in the space of possible functions to model nonlinear dynamical systems. The connection of the SINDy method to the Akaike information criteria (AIC) for model selection has been studied in [25]. The promising results for system identification problems such as hybrid dynamical systems, chaotic Lorenz system and Burger's partial differential equation have been obtained [25,26]. The SINDy method applied to the model predictive control delivers better performance, requires significantly less data, and is more computationally efficient and robust to noise than the neural networks model [27]. Robustness to noise and requirement for measurements of derivatives are concerns with the SINDy approach [28]. The total variation regularized derivatives are commonly used to estimate the derivatives [29,30]. An integral form of equations of motion in combination with sparse regression is proposed in [31]. An application to model identification of nonlinear mechanical systems is reported in [32].

In practical cases, the data is often contaminated with noise. Fliess and Ramirez proposed a robust and fast algebraic identification technique to estimate time invariant linear systems without time delay [33]. Inspired by the algebraic operation, Belkoura in [12] and [13] first investigated the identifiability conditions for a general class of systems described by convolution equations. Then, an algebraic formulation is introduced for online estimation of time delay and parameters of structured and arbitrary input-outputs.

In this paper, we extend the SINDy approach by combining it with the algebraic signal processing method to deal with the issue of measurement noise, initial conditions and derivatives. The algebraic operation generates useful signals for system identification while filtering out the noise. We apply the Taylor expansion to make the time delay appear as a parameter of the model to identify. A nonlinear extended state estimator is adopted for derivative estimation. As a result, we arrive at a robust sparse regression combined with cross validation and bootstrapping techniques for nonparametric system identification. The simulation and experimental results illustrate that the proposed algorithm can overcome the following limitations of the existing techniques:

- *Multiple local minima* Achieving the global minimum is the main challenging for modified least square methods with recursive approach. The sparse regression is a convex optimization problem with a global minimum [34].
- *Nonparametric nonlinear identification* The only assumption of the proposed algorithm about the structure of the system is that the system is sparse in the space of base functions. The proposed algorithm relies on the data to make selection and is not limited to linear and SISO systems.
- *Noise resistant* The sparse regression is already robust. The algebraic operation offers additional filtering of the

noise. Furthermore, the proposed algorithm is equipped with bootstrapping to study the statistics of estimation such as mean and standard deviations.

- *Unstructured entries* Many classical strategies are designed for pulsed entries. The proposed approach analyzes the input and output data without any frame and structure assumption for entries.
- *Initial conditions* Another general assumption in time delay estimation is that the initial conditions are zero while in most of practical applications, it is not true. The algebraic operation and operational calculus make the proposed algorithm independent of the initial conditions.

The rest of the paper is organized as the follows. Section 2 introduces the assumptions and formulates the mathematical problem of system identification. The techniques of algebraic data preprocessing and derivative estimations, sparse regression in combination with bootstrapping resampling, and cross validation are explained in Sect. 3. Section 4 presents an example of a simulated nonlinear mass-spring-damper system under a proportional control with time delay. The experimental validation of the proposed algorithm on the rotary flexible joint made by Quanser is presented in Sect. 5. Finally, Sect. 6 concludes the paper.

2 Problem definition and assumptions

Consider a closed-loop second order system given by,

$$m\ddot{x} + g(x, \dot{x}) = f(t) + b_1\dot{x}(t - \tau) + b_0x(t - \tau) \tag{1}$$

where $m > 0$ is the mass of the system, the function $g(x, \dot{x})$ represents the nonlinear restoring or internal force of the system. The term $f(t)$ contains the reference information as well as external disturbances. The control consists of the output and its first-order derivative with feedback gains b_0 and b_1 , and a time delay τ . We make the following assumptions in this study.

Assumptions

1. Only the system response $x(t)$ is measured.
2. The system is exposed to random excitations included in $f(t)$.
3. The measurement contains Gaussian white noise with zero mean denoted as ϵ_x .
4. The system structure in terms of the function $g(x, \dot{x})$ is unknown.
5. The system is second order.

This paper is focused on nonparametric identification of closed-loop nonlinear dynamical systems with a control time delay.

3 The proposed method

For the restoring force $g(x, \dot{x})$ with an unknown structure, there are many candidate functions available to approximate it, such as polynomial, trigonometric, exponential functions or a combination of these functions. Because the polynomial is a popular base function to describe a wide range of dynamical systems, we use regular polynomials of x and \dot{x} with time-invariant coefficients to explain the proposed method,

$$g(x, \dot{x}) = \sum_{i=0}^N \sum_{j=0}^M c_{ij} x^i \dot{x}^j \tag{2}$$

where c_{ij} are the unknown coefficients of the polynomial. Other functions can also be considered with the proposed method.

For the system without any prior knowledge, the orders N and M of the polynomial are unknown. This study develops an algorithm to identify the polynomial defined in Eq. (2) from the response data such that it has a minimum number of terms and with the minimum orders N and M .

3.1 Algebraic operation

Noise robustness has been always an inevitable part of system identification techniques as the noise is unavoidable in real data. Time delay caused by sensors and actuators is usually so small that it is difficult or even impossible to estimate correct values of time delay from noisy data. In addition to noise robustness, dynamical system identification methods require derivatives of time series of the measured response. Traditional finite difference methods to estimate derivatives can amplify the noise. To deal with this issue, we propose to apply the algebraic operation to pre-process the data.

We use a nonlinear oscillator under a proportional control with gain k_p to illustrate the method.

$$\begin{aligned}
 m\ddot{x} + c_1\dot{x} + c_2\dot{x}^2 + c_3\dot{x}^3 \\
 + k_1x + k_3x^3 - k_px(t - \tau) = f(t) \\
 x(0) = x_0, \dot{x}(0) = v_0
 \end{aligned} \tag{3}$$

where m denotes mass, c_1 , c_2 and c_3 stand for damping coefficients, k_1 and k_3 are the stiffness coefficients. The system starts from initial conditions (x_0, v_0) which are generally unknown. We introduce several short-hands for the nonlin-

ear terms.

$$z_1 = \dot{x}x^2, \quad z_2 = \dot{x}^3, \quad z_3 = x^3 \quad (4)$$

Applying the Laplace transform to Eq. (3), we have

$$m \left\{ s^2 X(s) - sx_0 - v_0 \right\} + c_1 \{sX(s) - x_0\} + k_1 X(s) - k_p e^{-s\tau} X(s) = G(s) \quad (5)$$

where

$$G(s) = F(s) - c_2 Z_1(s) - c_3 Z_2(s) - k_3 Z_3(s), \quad (6)$$

$X(s)$ is the Laplace transform of $x(t)$, $Z_i(s)$ is the Laplace transform of $z_i(t)$ and $F(s)$ is the Laplace transform of $f(t)$.

Consider the Taylor expansion of the exponential term $e^{-s\tau}$

$$e^{-s\tau} = 1 - s\tau + \frac{s^2\tau^2}{2!} - \frac{s^3\tau^3}{3!} + \dots \quad (7)$$

We should point out that keeping too many higher order terms may not be beneficial to the identification process. We only need to keep a sufficient number of terms to generate enough equations to determine the unknown parameters including time delay.

As an example, we keep the terms up to the third order and substitute the Taylor expansion in Eq. (5).

$$m \left\{ s^2 X(s) - sx_0 - v_0 \right\} + c_1 \{sX(s) - x_0\} + k_1 X(s) - k_p \left\{ 1 - s\tau + \frac{s^2\tau^2}{2!} - \frac{s^3\tau^3}{3!} \right\} X(s) = G(s) \quad (8)$$

To eliminate the initial conditions, we differentiate Equation (8) with respect to s twice. To eliminate the derivative terms in time domain, we divide the resulting equation by s^3 . Back to time domain, we obtain an equation of signals

$$P_f(t) = k_p \frac{\tau^3}{3!} P_1(t) + \left\{ m - k_p \frac{\tau^2}{2} \right\} P_2(t) + \{c_1 + k_p \tau\} P_3(t) + c_2 P_4(t) + c_3 P_5(t) + \{k_1 - k_p\} P_6(t) + k_3 P_7(t) \quad (9)$$

where

$$\begin{aligned} P_1(t) &= 6 \int^{(2)} x(t) - 6 \int tx(t) + t^2 x(t) \\ P_2(t) &= 2 \int^{(3)} x(t) - 4 \int^{(2)} tx(t) + \int t^2 x(t) \\ P_3(t) &= -2 \int^{(3)} tx(t) + \int^{(2)} t^2 x(t) \end{aligned}$$

$$\begin{aligned} P_4(t) &= \int^{(3)} t^2 z_1(t), \quad P_5(t) = \int^{(3)} t^2 z_2(t) \\ P_6(t) &= \int^{(3)} t^2 x(t), \quad P_7(t) = \int^{(3)} t^2 z_3(t) \\ P_f(t) &= \int^{(3)} t^2 f(t) \end{aligned} \quad (10)$$

Note that $\int^{(n)} \phi(t)$ denotes the multiple integral $\int_0^t \int_0^{\sigma_1} \dots \int_0^{\sigma_{n-1}} \phi(\sigma_n) d\sigma_n \dots d\sigma_1$. Let t_k ($k = 1, 2, \dots, n_t$) be a set of sampled times. Define an error at time t_k as

$$\begin{aligned} e(k) &= k_p \frac{\tau^3}{3!} P_1(t_k) + \left\{ m - k_p \frac{\tau^2}{2} \right\} P_2(t_k) + \{c_1 + k_p \tau\} P_3(t_k) \\ &\quad + c_2 P_4(t_k) + c_3 P_5(t_k) + \{k_1 - k_p\} P_6(t_k) \\ &\quad + k_3 P_7(t_k) - P_f(t_k). \end{aligned} \quad (11)$$

Let us introduce an error vector \mathbf{e} , a parameter vector \mathbf{c} and a force vector \mathbf{p} as follows.

$$\begin{aligned} \mathbf{e} &= [e(1), e(2), \dots, e(n_t)]^T, \\ \mathbf{c} &= \left[k_p \frac{\tau^3}{3!}, m + k_p \frac{\tau^2}{2}, c_1 - k_p \tau, c_2, c_3, k_1 + k_p, k_3 \right]^T, \\ \mathbf{p} &= [P_f(t_1), P_f(t_2), \dots, P_f(t_{n_t})]^T. \end{aligned} \quad (12)$$

Let $\mathbf{P}(x)$ be the $n_t \times 7$ data matrix determined by the measurement of $x(t)$ such that its $(k, j)^{th}$ component is given by $P_{k,j} = P_j(t_k)$. Then, Eq. (11) can be written in the matrix form as

$$\mathbf{e}(\mathbf{c}) = \mathbf{P}(x)\mathbf{c} - \mathbf{p}. \quad (13)$$

To find the unknown parameters \mathbf{c} , we formulate a least mean square error problem as follows. For the following cost function,

$$J(\mathbf{c}) = \mathbf{e}^T \mathbf{e} = \|\mathbf{e}\|_2^2 \quad (14)$$

we find the parameter vector \mathbf{c} such that J is minimized. The solution of the parameters \mathbf{c} can be found by either the matrix inversion or by an iterative search algorithm. In general, when the numbers of unknown parameters and data points n_t are large, as is the case for multi-degree of freedom systems, it is better to use a global search algorithm to compute the parameters.

3.2 Sparse representation

Assume that the displacement measurement $x(t)$ of the system in Eq. (3) and external force $f(t)$ are available and contain random noises. We consider the polynomial model

for the restoring force in Eq. (2). Two questions immediately arise:

- What are the minimum orders N and M ?
- Are all the terms in the polynomial needed?

To investigate answers for these questions, we apply the methods in statistical learning [35]. One popular way is to start with big enough orders N and M or a large library of functions which may contain non-polynomial functions, and penalize the terms with small coefficients through a sparse regulator by using the least absolute shrinkage and selection operator (LASSO) [35]. The recent studies of sparse identification of nonlinear dynamics (SINDy) have developed a computationally efficient algorithm to compute the solution of the sparse regression. LASSO regression proposes to add an L^1 regularization term to the cost function in Eq. (14) to penalize the nonzero parameters that don't contribute to the system's dynamics

$$J_\lambda(\mathbf{c}) = \|\mathbf{e}\|_2^2 + \lambda \|\mathbf{c}\|_1$$

$$= \|\mathbf{P}(x)\mathbf{c} - \mathbf{p}\|_2^2 + \lambda \|\mathbf{c}\|_1 \tag{15}$$

where $\lambda > 0$ is a preselected positive number known as the sparse regulator and $\|\mathbf{c}\|_1 = \sum_i |c_i|$. The sparse regression algorithm in [24] is applied to find the parameter vector \mathbf{c} in order to minimize J_λ in Eq. (15). The algorithm is computationally efficient and robust to noise. It should be noted that the optimization problem in the framework of the LASSO is convex [34], which implies the existence of a unique optimal solution. The selection of the sparse regulator λ is critical. In Sect. 3.3, we explain how to employ cross-validation techniques to select a proper regulator value.

The LASSO penalizes the terms with small coefficients by regularization and keeps the important terms in the system model. In real world, time delay is usually small. Moreover, time delay appears in high order terms of the Taylor expansion. Therefore, the LASSO will penalize these terms, such as the coefficient of term $P_1(t)$ in Eq. (11).

We can recover the terms involving time delay in the following way. After each sparse regression computed with the SYNDy algorithm, we keep all the terms involving time delay as well as the terms selected by the LASSO regulation. The resulting data matrix denoted by $\mathbf{P}_s(x)$ is substituted in Eq. (15) to compute the updated coefficients \mathbf{c} .

3.3 Cross validation and bootstrapping

It is common to use cross-validation techniques from machine learning to determine the sparse regulation parameter λ . The value of λ in an finite interval is sampled. The cross-validation mean square error MSE_{CV} of the model over the test dataset is computed. The λ value which

minimizes the cross-validation error is selected. The corresponding model is chosen as an optimal model with a proper balance of complexity and accuracy.

Let T_{cv} denote the set of time instances of the test signal to be used for cross validation. For a given regulation parameter λ , the mean square cross validation error MSE_{CV} of the model over all the validation datasets is defined as

$$MSE_{CV}(\lambda) = \frac{1}{n_t} \|\mathbf{P}(x(t \in T_{cv}))\mathbf{c}_\lambda - \mathbf{p}(t \in T_{cv})\|_2^2 \tag{16}$$

where \mathbf{c}_λ denotes the model parameters for the regulation parameter λ . MSE_{cv} is an implicit function of λ . The SINDy algorithm attempts to select λ on the Pareto front of the multi-objective optimization problem with the objectives being accuracy and complexity of the model. The elbow of the Pareto front parameterized by λ is often the choice [24].

Unfortunately, in most of cases, the elbow of the Pareto front is ambiguous due to existence of a cluster of candidate models near the elbow. The information criteria (IC) for the candidate models can help to rank and select a model with a proper trade-off between the accuracy and complexity [25]. Popular statistical examples of the information criteria include the Akaike information criterion (AIC), Bayesian information criterion (BIC), deviance information criterion (DIC) and minimum description length. The work reported in [25] makes use of a big data matrix $\mathbf{P}(x)$. The sparse regression procedure is repeated over all possible combinatorial subsets of the data matrix. The resulting models are ranked through IC scores. The model with the smallest score is selected.

However, a big data matrix $\mathbf{P}(x)$ may not always be available for real-world applications. In this work, we propose a search algorithm to determine the sparse representation of the polynomial in Eq. (2). We start with linear model when $N = M = 1$, and increase the polynomial order until the prediction error of the model over the test dataset reaches an acceptable low level and begins to increase. The error of the test dataset is defined as,

$$MSE_{test} = \frac{1}{n_t} \|\mathbf{P}(x(t \in T_{cv}))\hat{\mathbf{c}} - \mathbf{p}(t \in T_{cv})\|_2^2 \tag{17}$$

where $\hat{\mathbf{c}}$ stands for the estimated vector of coefficients. There are different ways to generate training and test datasets. When the SINDy algorithm is applied to simulation examples, we can generate rich datasets to train and validate the model by considering the system responses for different initial conditions and excitations. In this work, we assume that the dataset consists of two long time series of the system response. One is used for training and another for cross-validation. The data matrices $\mathbf{P}(x)$ are generated for different orders of the polynomial. The SINDy algorithm selects the sparse model built

on the training data while the regularization parameter λ is selected with the help of cross validation on the test dataset.

Real measurements often contain noises, and can have outliers and missing data. Here, we propose to combine the sparse regression with bootstrapping in order to develop robust sparse regression. In particular, we consider K bootstrap sample vectors containing L elements of the original data points. Each vector is generated by uniform sampling of the data with replacement. For each bootstrap sample vector, the sparse regression is applied to identify the model. Finally, the parameters of the model is the average of the K estimated coefficient vectors $\hat{\mathbf{c}}_l$.

$$\tilde{\mathbf{c}} = \frac{1}{K} \sum_{l=1}^K \hat{\mathbf{c}}_l \quad (18)$$

The standard deviation of estimated coefficient vectors $\hat{\mathbf{c}}_l$ is computed to study the variation of parameter estimation.

Algorithm 1 summarizes the procedure.

Algorithm 1 Bootstrapping sparse regression algorithm for time-delay estimation.

```

1: procedure TIME-DELAY ESTIMATION( $X_{train}, f_{train}, X_{test}, f_{test},$ 
    $N, K, L, \lambda_{min}, \lambda_{max}$ )
2:   for  $PolyOrder \in [1, N; 1, M]$  do
3:      $g(x, \dot{x}) \leftarrow x, \dot{x}, PolyOrder$   $\triangleright$  Generate candidate
       polynomial restoring force
4:      $\mathbf{P}(x)_{train}, \mathbf{P}(x)_{test} \leftarrow$  library(Algebraic Operation( $g(x, \dot{x})$ ))
5:      $\mathbf{p}_{train}, \mathbf{p}_{test} \leftarrow$  Algebraic Operation( $f_{train}(t), f_{test}(t)$ )
6:     for  $k \in [1, K]$  do  $\triangleright$  Generate uniformly random sampling
       with replacement
7:        $\mathbf{P}(x)_{k,train}, \mathbf{p}_{k,train} \leftarrow$  Bootstrapping Sampling
       ( $\mathbf{P}(x)_{train}, \mathbf{p}_{train}, L$ )
8:       for  $\lambda \in [\lambda_{min}, \lambda_{max}]$  do  $\triangleright$  Search to find the sparse
       regulator
9:          $MSE(\lambda) \leftarrow$  SINDy( $\mathbf{P}(x)_{k,train}, \mathbf{p}_{k,train}, \lambda$ )
10:       end for
11:        $\lambda_{min} \leftarrow \min(MSE(\lambda))$ 
12:        $Model(PolyOrder, k) \leftarrow$  SINDy( $\mathbf{P}(x)_{k,train}, \mathbf{p}_{k,train},$ 
        $\lambda_{min}$ )
13:       if  $\mathbf{c}_\tau = 0$  then
14:          $Model(PolyOrder, k) \leftarrow$  OLS( $\mathbf{P}_s(x)_{k,train}, \mathbf{p}_{k,train}$ )
15:       end if
16:     end for
17:      $Model(PolyOrder) \leftarrow$  Average( $Model(PolyOrder, k)$ )
18:      $MSE_{test}(PolyOrder) \leftarrow$   $MSE(Model(PolyOrder),$ 
        $\mathbf{P}_{test}(x), \mathbf{p}_{test})$ 
19:   end for
20:    $Model(\mathbf{c}) \leftarrow$  Overall  $\min(MSE_{test})$   $\triangleright$  Select the model with
       minimum test error
21:   return  $Model(\mathbf{c})$ 
22: end procedure

```

Remark 3.1 Some remarks on the estimation error and the stability of the system are in order. The estimation error of the coefficients \mathbf{c} , in particular, time delay τ , can be attributed

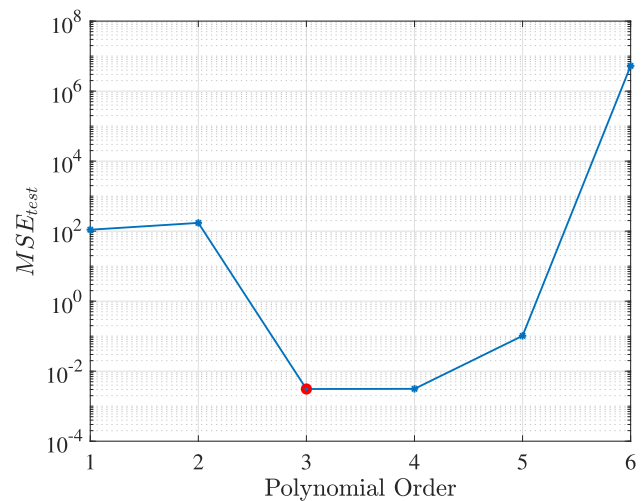


Fig. 1 The cross validation test error for the mass-spring-damper system

to two sources: the truncation error of the Taylor expansion and the regression error.

While keeping more higher order terms of the Taylor expansion helps to reduce truncation error, we may run the risk of instability of the truncated model. We have found that the third order term is a good compromise for accuracy, complexity and stability.

4 Simulated example

To demonstrate the proposed algorithm, we consider the second-order oscillatory system with nonlinear stiffness and damping in Equation (3). The following external forces are used to generate the training and test datasets.

$$f(t)_{Train} = 10 \sin(4t) + 40 \sin(4t^2) + \epsilon_f \quad (19)$$

$$f(t)_{Test} = 10 \sin(2t) + 40 \sin(8t) + \epsilon_f \quad (20)$$

For the training dataset, the system starts from the initial conditions $x_0 = 1$ and $v_0 = 0$. For the test dataset, the initial conditions are $x_0 = -1$ and $v_0 = 0.5$. The excitation contains a normally distributed random noise ϵ_f with zero mean and standard deviation $\sigma_{\epsilon_f} = 0.01$. We assume that the sensors have measurement noises such that the system output is given by $x(t) + \epsilon_x$ where ϵ_x is the normally distributed random noise with zero mean and standard deviation $\sigma_{\epsilon_x} = 0.01$.

Table 2 lists the system parameters used in the simulation. A proportional control with gain $k_p = 2$ is considered. We employ the third order of Taylor expansion of the time delay $x(t - \tau)$. The order of expansion is selected such that there are enough equations to solve for the unknown parameters of the system. Following Algorithm 1, we create the data matrix

Table 1 The true and estimated parameters of the simulation example

Terms	Parameters	Exact Value	Estimated	Standard Deviation *10 ⁻³
\ddot{x}	$k_p \frac{\tau^3}{6}$	0.0026667	0.001515	0.0125
\ddot{x}	$m - k_p \frac{\tau^2}{2}$	0.96	0.9611	0.0137
x	$k_1 - k_p$	4	3.9830	0.1612
x^2	0	0	0	0
x^3	k_3	2	2.0041	0.0943
\dot{x}	$c_1 + k_p \tau$	1.4	1.3922	0.2896
\dot{x}^2	0	0	0	0
\dot{x}^3	c_3	1	0.9973	0.0432
$x\dot{x}$	0	0	0	0
$x\dot{x}^2$	0	0	0	0
$x\dot{x}^3$	0	0	0	0
$x^2\dot{x}$	c_2	1	1.0074	0.0961
$x^2\dot{x}^2$	0	0	0	0
:	:	:	:	:
$x^3\dot{x}^3$	0	0	0	0

$\mathbf{P}(x)$ with polynomials of orders from 1 to 6 and $N = M$. For each polynomial model, 30 bootstrapping samples with the ratio of 50 percent of the total data are selected. 50 values of λ are sampled logarithmically in the range from 10^{-6} to 10^0 .

Because the first term includes the time delay, we constrain the LASSO algorithm to keep it from being penalized and removed in the process of searching for sparse representation. The average of sparse polynomials out of all the bootstrapping samples is taken as the final result.

Figure 1 shows the variation of the cross validation error MSE_{cv} as a function of the order N of the polynomial. It is seen from the figure that the cross validation error has a large drop when approaching $N = 3$ and stays at the same range for $N = 4$ and starts increasing from $N = 5$. We choose $N = 3$ as the optimal order of the polynomial when the minimum validation error occurs and the order is minimum. Table 1 lists the estimated coefficients of the signals in Eq. (9). The proposed algorithm has successfully detected the sparse terms and precisely estimated the coefficients. The time delay and parameters of the original system are calculated and listed in Table 2 together with the known values. The accuracy of the estimated parameters is quite acceptable.

Remark 4.1 A remark on small values of the standard deviation in Table 1 is in order. This fact is an indication that the randomness of the data is not significant so that the bootstrapping samples are not sufficiently different. We have checked the results several times to confirm this observation.

Table 2 The parameters of the mass-spring-damper system for the simulation example

Parameters	Selected value	Estimated value
m	1	0.9885
k_1	6	5.9830
k_3	2	2.0041
c_1	1	1.0609
c_2	1	1.0074
c_3	1	0.9973
τ	0.2	0.1656

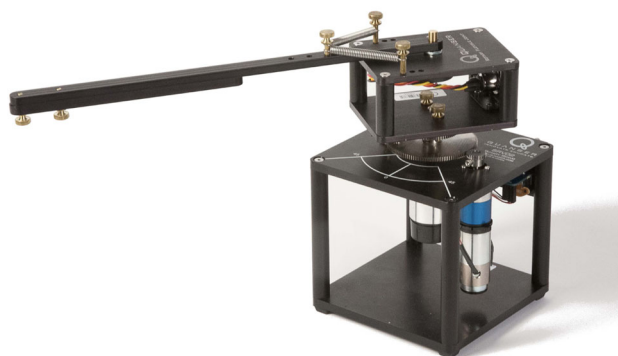


Fig. 2 The flexible joint set up by Quanser [36]

5 Experimental example

In this section, we employ a flexible joint experimental setup, made by Quanser, to prove the efficiency of the method

Table 3 The parameters of the rotary flexible link provided by Quanser [37]

Symbol	Description	Value
J_l	Link moment of inertia	0.0038 kg m ²
J_{eq}	Equivalent moment of inertia	9.76×10^{-5} kg m ²
B_l	Flexible link torsional damping	Not available
B_{eq}	Equivalent torsional viscous damping	0.015 N m/rad/s
k_s	Link torsional stiffness	$1.3 \frac{\text{N m}}{\text{rad}}$
k_g	High gear total gear ratio	70
k_t	Motor current torque constant	$7.68 \times 10^{-3} \frac{\text{N m}}{\text{A}}$
k_m	Motor back-emf constant	$7.68 \times 10^{-3} \frac{\text{V}}{\text{rad/s}}$
R_m	Motor armature resistance	2.6Ω
η_g	Gearbox efficiency	0.9
η_m	Motor efficiency	0.69

for experimental data from a single-input-multiple-outputs (SIMO) dynamical system. The joint is connected to the base through two springs with equal stiffness k_s and the base is fixed to the servo motor as Fig. 2 shows. The servo motor with input voltage V_m generates a torque M to turn the base with a rotation angle θ . The deflection angle of the joint relative to the base is α . The moment of inertia of the base is J_{eq} . The viscous damping coefficient of the base is B_{eq} . J_l stands for the moment of inertia of the joint. The equations of motion for the system can be derived as,

$$\begin{aligned}
 (J_{eq} + J_l)\ddot{\theta} + J_l\ddot{\alpha} + B_{eq}\dot{\theta} &= \tau_M, \\
 J_l\ddot{\alpha} + J_l\dot{\theta} + B_l\dot{\alpha} + k_s\alpha &= 0, \\
 \tau_M &= \frac{\eta_g k_g \eta_m k_t (V_m - k_g k_m \dot{\theta})}{R_m} \equiv a \cdot V_m + b \cdot \dot{\theta}, \\
 a &= \frac{\eta_g k_g \eta_m k_t}{R_m}, \quad b = -\frac{\eta_g k_g \eta_m k_t k_g k_m}{R_m},
 \end{aligned}
 \tag{21}$$

where torque τ_M is linearly dependent on the servo motor voltage V_m and $\dot{\theta}$. We assume that the motor parameters are known. Table 3 lists the parameters of the motor and joint provided by Quanser. We should point out that these numbers may not be exactly the same as the physical system parameters. Therefore, we should treat these numbers as a reference.

A proportional control with gain k_p adjusts the motor voltage to make the joint follow the desired trajectory of the angle θ while the deflection angle α stays minimum. The control is given by

$$V_m = k_p * (\theta_d - \theta) \tag{22}$$

The control is implemented in Simulink/MATLAB. The outputs of the system are θ and α . A time delay is introduced in the feedback signal θ .

In this example, we employ the second order Taylor expansion of the time delay term.

$$\theta(t - \tau) = \theta(t) - \tau\dot{\theta}(t) + \frac{\tau^2}{2}\ddot{\theta}(t) \tag{23}$$

Fig. 3 The system response to track a sinusoidal trajectory θ_d with amplitude 2 radian and frequency 0.66 Hz

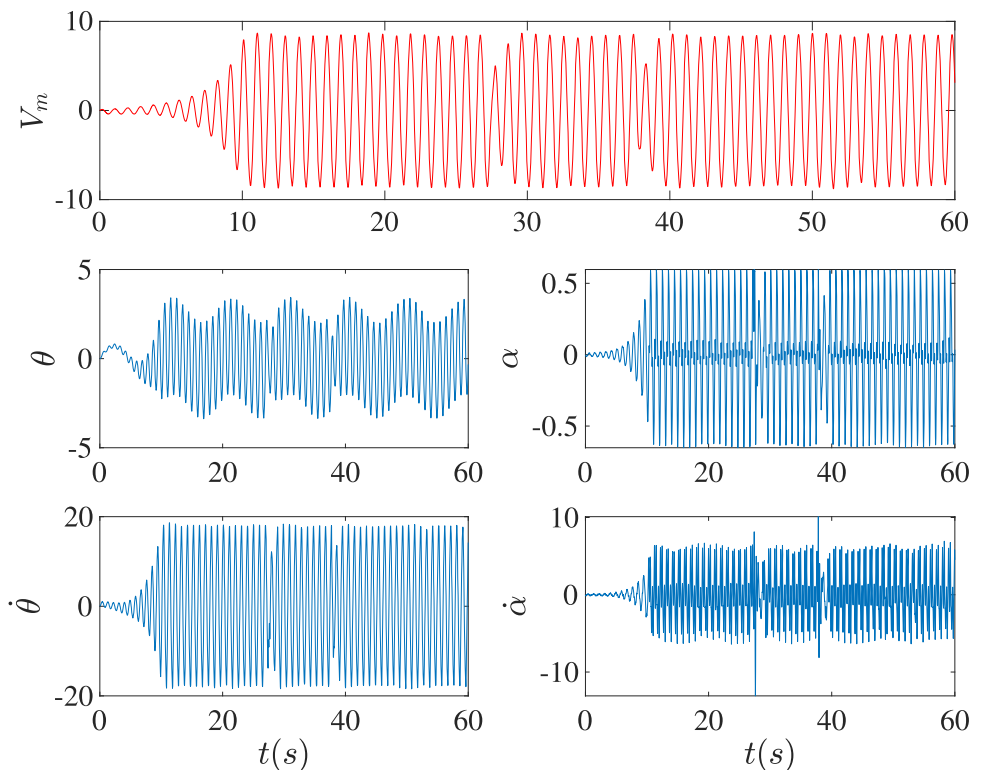


Fig. 4 The system response to a square-wave signal θ_d with amplitude 1 radian and frequency 0.66 Hz

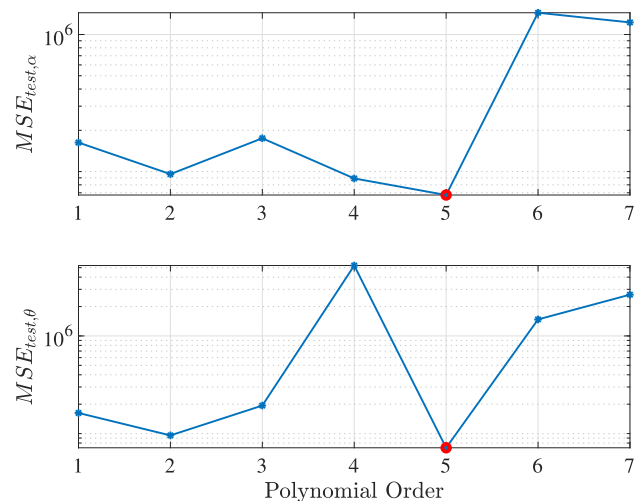
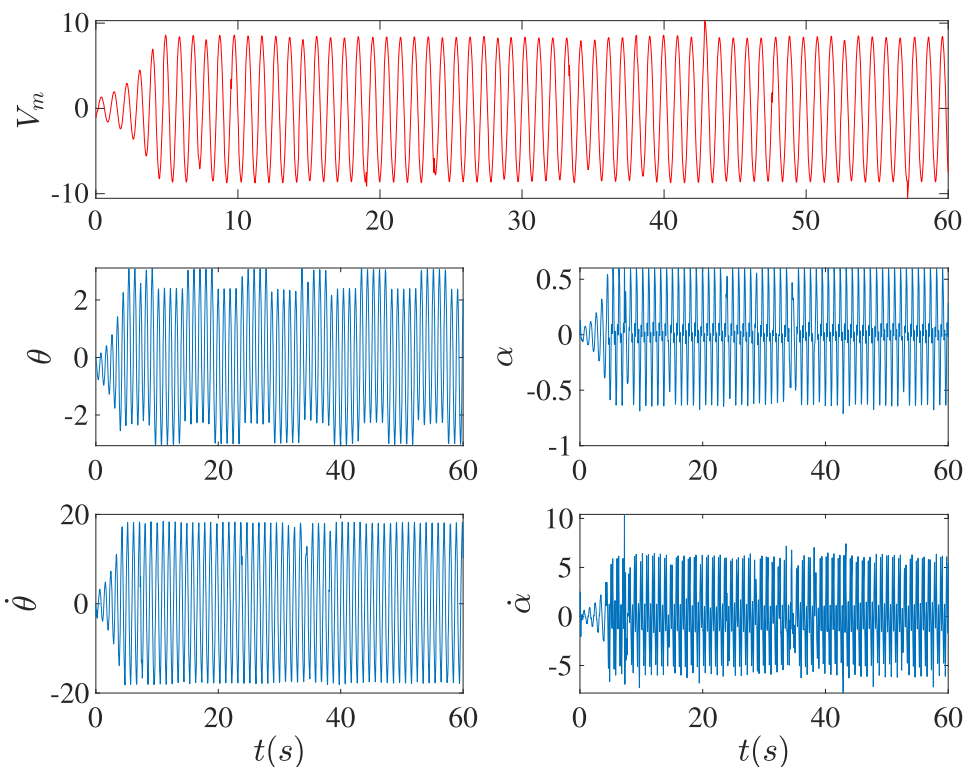


Fig. 5 The cross-validation error of the rotary flexible joint

The closed-loop system can be written as,

$$\left(J_{eq} + ak_p \frac{\tau^2}{2} \right) \ddot{\theta} + (B_{eq} - ak_p \tau) \dot{\theta} - B_l \dot{\alpha} - k_s \alpha = ak_p (\theta_d - \theta) + b \dot{\theta}, \tag{24}$$

$$J_{eq} \ddot{\alpha} - ak_p \frac{\tau^2}{2} \ddot{\theta} + (ak_p \tau - B_{eq}) \dot{\theta} + \frac{B_l (J_l + J_{eq})}{J_l} \dot{\alpha} + \frac{k_s (J_l + J_{eq})}{J_l} \alpha = -ak_p (\theta_d - \theta) - b \dot{\theta}. \tag{25}$$

We should point out that the tracking performance of the closed-loop system will not be as good as we would like to have. This is because the proportional control alone is not adequate and the time delay further deteriorates the performance. The purpose of choosing this control with time delay is simply for generating the data in a stable manner.

To generate the training dataset, we select a desired trajectory θ_d as

$$\theta_d = 2 \sin 2\pi f_1 t \tag{26}$$

where the frequency $f_1 = 0.66Hz$ is for the desired trajectory θ_d . To generate the test data, we choose a square-wave signal for the desired trajectory θ_d with the same amplitude and frequency as in Equation (26). We introduce a 0.2s delay in the control such that the closed-loop system is stable.

Figures 3 and 4 show the motor input voltage V_m and the sample responses of θ and α . The time series is one minute long with a sample time 0.001s. For simplicity, we assume that for each coordinate θ and α , the restoring force can be represented by a sum of four polynomials of a single variable, i.e. $\theta^i, \dot{\theta}^i, \alpha^j$, and $\dot{\alpha}^j$, together with a friction term related to $\dot{\theta}$, without considering the cross-product terms. The second coefficient in Equation (25), $ak_p \frac{\tau^2}{2}$, includes the time delay. We apply the constraint to the LASSO algorithm to keep it from being penalized and removed.

Some signals in the matrix $\mathbf{P}(x)$ contain the first order derivative of the angles $\dot{\theta}$ and $\dot{\alpha}$. Recall that the real mea-

Table 4 The θ and α terms, their coefficients and the standard deviations of the estimation for the rotary flexible joint

Terms in θ -equation	Estimated	Standard Deviation	Terms in α -equation	Estimated	Standard Deviation
$\ddot{\theta}$	0.0061	0.	$\ddot{\theta}$	0.0012	0.0001
$\ddot{\alpha}$	—	—	$\ddot{\alpha}$	−0.0061	0
θ	0.0625	0.0026	θ	−0.0603	0.0029
θ^2	0.0003	0.0026	θ^2	0.0003	0.0018
θ^3	−0.0254	0.0009	θ^3	0.024	0.0010
θ^4	−0.0004	0.0006	θ^4	0.0003	0
θ^5	0.0019	0.0034	θ^5	−0.0019	0
$\dot{\theta}$	−0.0922	0.0002	$\dot{\theta}$	0.0920	0.0036
$\dot{\theta}^2$	0.0002	0	$\dot{\theta}^2$	−0.0002	0.0002
$\dot{\theta}^3$	0.0004	0	$\dot{\theta}^3$	−0.0004	0
$\dot{\theta}^4$	0	0	$\dot{\theta}^4$	0	0
$\dot{\theta}^5$	0	0	$\dot{\theta}^5$	0	0
α	−1.1976	0.0139	α	1.2483	0.0128
α^2	0.9040	0.0896	α^2	−0.9458	0.0991
α^3	−2.3784	0.2446	α^3	2.4237	0.2634
α^4	−0.7031	0.2916	α^4	0.6930	0.3149
α^5	8.9105	0.5177	α^5	−9.0823	0.5629
$\dot{\alpha}$	−0.0650	0.0005	$\dot{\alpha}$	0.0644	0.0008
$\dot{\alpha}^2$	−0.0215	0	$\dot{\alpha}^2$	0.0216	0.0005
$\dot{\alpha}^3$	0.0001	0	$\dot{\alpha}^3$	−0.0001	0
$\dot{\alpha}^4$	0.0004	0	$\dot{\alpha}^4$	−0.0004	0
$\dot{\alpha}^5$	0	0	$\dot{\alpha}^5$	0	0
f_c	0.0516	0.0083	f_c	−0.0477	0.0084

measurements contain noises. We use the nonlinear state observer from the control studies to estimate the first-order derivatives without amplifying the noise in the computation [38,39]. The second order observer with gains β_1 and β_2 is defined by the following state equations.

$$\begin{aligned} \dot{z}_1 &= z_2 - \beta_1 e, \\ \dot{z}_2 &= -\beta_2 \text{fal}(e, \alpha, \delta) \end{aligned} \quad (27)$$

where

$$\text{fal}(e, \alpha, \delta) = \begin{cases} |e|^\alpha \cdot \text{sign}(e), & |e| > \delta \\ \frac{e}{\delta^\alpha}, & |e| \leq \delta \end{cases} \quad (28)$$

and the error $e = z_1 - \theta$. We have chosen $\alpha = 0.5$, $\delta = 0.05$, $\beta_1 = 100$ and $\beta_2 = 900$. According to [38,39], the estimation error e converges to zero quickly. Consequently, by definition, z_2 is an accurate estimate of the first order derivative $\dot{\theta}$. Figures 3 and 4 show the plot of estimated $\dot{\theta}$ and $\dot{\alpha}$ for both training and test dataset.

To train the model, we consider 10 bootstrapping samples with the length of 50 percent of the training dataset and search the polynomial order from 1 to 7. For the sparse regulator,

100 values of λ are sampled logarithmically in the range from 10^{-10} to 10^0 .

Figure 5 shows the variation of the cross validation error as a function of the order of polynomials. The results suggest that for both coordinates θ and α , the polynomial order $N = 5$ leads to the minimum cross validation error. This is the optimal order for the polynomials.

Table 4 lists the polynomial terms and the corresponding coefficients and associated standard deviation for the trained model of polynomial order $N = 5$. We should mention that the reported results have been round to 10^{-5} . The estimated value of Coulomb friction f_c is not negligible. It can play an important role in the system dynamics. We should point out that the linear model in Eq. (21) does not include the friction term, although the actual device always has friction. The friction is the primary reason for the discrepancy between the predicted response by Eq. (21) and the actual measurements.

Figures 3 and 4 indicate that the responses of the closed-loop system are not really tracking the references accurately. This is partly due to the effect of time delay and also the poor control design, as discussed earlier. The oscillatory responses of the closed-loop system are obviously responsible for the high order terms of the polynomials, particularly for the deflection angle α , as can be seen in Table 4. This highlights

Table 5 The nominal and estimated values of the parameters of the rotary flexible link by Quanser [37]

00 Symbol	Nominal values	Estimated values
$J_{eq} + ak_p \frac{\tau^2}{2}$	0.0102	0.0061
$B_{eq} - ak_p \tau$	-0.0662	-0.0922
k_s	1.3	1.1976
J_{eq}	0.00208	0.0012
B_l	Small unknown	0.0650
$ak_p \frac{\tau^2}{2}$	0.008121	0.0061
$ak_p \tau - B_{eq}$	0.0662	0.0920
$\frac{k_s(J_l + J_{eq})}{J_l}$	2.0501	1.2483
$\frac{B_l(J_l + J_{eq})}{J_l}$	Small unknown	0.064
τ	0.2	0.1734

the ability of the proposed system identification algorithm to estimate the time delay and to identify the nonlinearities in the system when the linear models are no longer adequate.

The motor parameter a is known and k_p is the given control gain. Hence, from the coefficient $-ak_p \frac{\tau^2}{2}$, the time delay can be calculated and is listed in Table 5 in comparison with the time delay we introduced to the control. The values of the parameters in Table 5 fall in a wide range from 0.002 to 2. Hence, the parameters can be several order of magnitudes apart, which makes it difficult to accurately estimate all the parameters in the presence of unwanted noises. This experimental study strongly demonstrates the robustness of the proposed algorithm to noises.

6 Conclusions

In summary, we have demonstrated that the proposed algorithm is effective to obtain governing equations of nonlinear dynamical systems with time delay from noisy experimental data. It can accurately estimate the time delay in the feedback control. For the first time, this study extends the sparse regression to nonlinear dynamical systems with time delay. We have equipped the sparse regression with an algebraic signal pre-processing and a nonlinear state observer. These operations are essential to compute the needed derivatives of measured time series and other signals without the need to know initial conditions, and to filter noises due to random excitations and measurements. Both simulation and experimental results have been used to validate the algorithm. The algorithm demonstrates excellent performances in identification.

The codes and all the data are available at: https://github.com/gleylaz/TimeDelay_SysID.

Acknowledgements The Shuo Wang thanks the Zhengzhou University for the fellowship to support his studies at the University of California, Merced, USA.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Richard JP (2003) Time-delay systems: an overview of some recent advances and open problems. *Automatica* 39(10):1667–1694
- Lima RBC, Barros PR (2015) Identification of time-delay systems: a state-space realization approach. *IFAC-PapersOnLine* 48(8):254–259
- Yan R, Chen F, Dong S, Liu T (2016) Gradient-based step response identification of low-order model for time delay systems. In: Proceedings of the 2016 Chinese control and decision conference (CCDC), Yinchuan, China, pp 5814–5819
- Jones RW (2004) Recursive estimation of time-delay using a high-order Pade approximant. *Proc SICE Annu Conf* 2:1328–1333
- Gawthrop PJ, Nihtilä MT (1985) Identification of time delays using a polynomial identification method. *Syst Control Lett* 5(4):267–271
- Ren XM, Rad AB, Chan PT, Lo WL (2005) Online identification of continuous-time systems with unknown time delay. *IEEE Trans Autom Control* 50(9):1418–1422
- Bedoui S, Ltaïef M, Abderrahim K (2013) Hierarchical gradient based identification of discrete-time delay systems. In: Proceedings of the 52nd IEEE conference on decision and control, pp 4565–4570
- Higo T, Kawaguchi T, Itamoto M, Adachi S (2018) Recursive identification of fractional time-delay systems using discrete-time model. *IFAC-PapersOnLine* 51(15):939–944
- Ha H, Welsh JS (2016) Parameter and delay estimation of continuous-time models utilizing multiple filtering. In: Proceedings of the IEEE 55th conference on decision and control (CDC), Las Vegas, NV, USA, pp 1205–1210
- Ha H, Welsh JS, Alamir M (2018) Useful redundancy in parameter and time delay estimation for continuous-time models. *Automatica* 95:455–462
- Yadaiah N, Deekshatulu BL, Sivakumar L, Sree Hari Rao V (2007) Neural network algorithm for parameter identification of dynamical systems involving time delays. *Appl Soft Comput* 7(3):1084–1091
- Belkoura L, Richard JP, Fliess M (2009) Parameters estimation of systems with delayed and structured entries. *Automatica* 45(5):1117–1125
- Belkoura L (2010) Identifiability and algebraic identification of time delay systems. *IFAC Proc* 43(2):1–8
- Jin F, Qiu T (2019) Adaptive time delay estimation based on the maximum correntropy criterion. *Digit Signal Proc* 88:23–32

15. Na J, Ren X, Xia Y (2014) Adaptive parameter identification of linear SISO systems with unknown time-delay. *Syst Control Lett* 66:43–50
16. Zheng G, Barbot JP, Boutat D (2013) Identification of the delay parameter for nonlinear time-delay systems with unknown inputs. *Automatica* 49(6):1755–1760
17. Beauduin T, Fujimoto H (2017) Identification of system dynamics with time delay: a two-stage frequency domain approach. *IFAC-PapersOnLine* 50(1):10870–10875
18. Yu L, Qiu TS, Song AM (2017) A time delay estimation algorithm based on the weighted correntropy spectral density. *Circuits Syst Signal Process* 36(3):1115–1128
19. Xu L (2014) A proportional differential control method for a time-delay system using the Taylor expansion approximation. *Appl Math Comput* 236:391–399
20. Chen F, Garnier H, Gilson M (2015) Robust identification of continuous-time models with arbitrary time-delay from irregularly sampled data. *J Process Control* 25:19–27
21. Ghoul Y, Ibn Taarit K, Ksouri M (2016) Online identification of continuous-time systems with multiple unknown time delays based on optimization approaches from sampled data. In: *Proceedings of the 7th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT), Hammamet, Tunisia*, pp 204–208
22. Gawthrop PJ, Nihtila MT, Rad AB (1989) Recursive parameter estimation of continuous systems with unknown time delay. *Control Theory Adv Technol* 5:227–248
23. Wang X, Su J, Zhang L (2018) Time-delay estimation for SISO systems using $SW\sigma$. *ISA Trans* 80:43–53
24. Brunton SL, Proctor JL, Kutz JN (2016) Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc Nat Acad Sci* 113(15):3932
25. Mangan N, Kutz J, Brunton S, Proctor J (2017) Model selection for dynamical systems via sparse regression and information criteria. *Proc R Soc A Math Phys Eng Sci* 473(20170):009
26. Mangan N, Askham T, Brunton S, Kutz J, Proctor J (2019) Model selection for hybrid dynamical systems via sparse regression. *Proc R Soc A Math Phys Eng Sci* 475(20180):534
27. Kaiser E, Kutz J, Brunton S (2017) Sparse identification of nonlinear dynamics for model predictive control in the low-data limit. *Proc R Soc A Math Phys Eng Sci* 474(20180):335
28. Champion K, Lusch B, Kutz JN, Brunton SL (2019) Data-driven discovery of coordinates and governing equations. *Proc Natl Acad Sci* 116(45):22445
29. Chartrand R (2011) Numerical differentiation of noisy, nonsmooth data. *ISRN Appl Math J* 2011(164):564
30. Rudin LI, Osher S, Fatemi E (1992) Nonlinear total variation based noise removal algorithms. *Physica D* 60:259–268
31. Schaeffer H, McCalla SG (2017) Sparse model selection via integral terms. *Phys Rev E* 96(2):23302
32. Leylaz G, Ma SF, Sun JQ (2021) An optimal model identification algorithm of nonlinear dynamical systems with the algebraic method. *J Vibr Acoust* 143, 021002–1–021002–8
33. Fliess M, SirafRamírez H (2003) An algebraic framework for linear identification. *ESAIM: Control. Optim Calculus Variat* 9:151–168
34. Lai Z, Nagarajaiah S (2019) Sparse structural system identification method for nonlinear dynamic systems with hysteresis/inelastic behavior. *Mech Syst Signal Process* 117:813–842
35. Hastie T, Tibshirani R, Friedman J (2009) *The elements of statistical learning: data mining, inference, and prediction*. Springer, Berlin
36. Quanser Inc (2011) *Rotary flexible joint workbook*. Markham, Ontario
37. Apkarian J, Lévis M, Gurocak H (2011) *User manual SRV02 Rotary servo base unit set up and configuration*. Quanser Inc., Markham
38. Han J (1995) A class of extended state observers for uncertain systems. *Control Decis* 10:85–88
39. Guo BZ, Zhao ZL (2016) *Active disturbance rejection control for nonlinear systems: an introduction*. Wiley, New York