Parameter estimation of delay differential equations: an integration-free LS-SVM approach

Siamak Mehrkanoon[a,*], Saeid Mehrkanoon[b], Johan A.K. Suykens[a]

[a] *KU Leuven, ESAT-SCD, Kasteelpark Arenberg 10, B-3001 Leuven (Heverlee), Belgium, Email: {siamak.mehrkanoon,johan.suykens}@esat.kuleuven.be*

[b] *University of New South Wales, Black Dog Institute, Hospital Rd, Randwick, NSW 2031, Australia, Email: s.mehrkanoon@unsw.edu.au*

Abstract

This paper introduces an estimation method based on Least Squares Support Vector Machines (LS-SVMs) for approximating time-varying as well as constant parameters in deterministic parameter-affine delay differential equations (DDEs). The proposed method reduces the parameter estimation problem to an algebraic optimization problem. Thus, as opposed to conventional approaches, it avoids iterative simulation of the given dynamical system and therefore a significant speedup can be achieved in the parameter estimation procedure. The solution obtained by the proposed approach can be further utilized for initialization of the conventional nonconvex optimization methods for parameter estimation of DDEs. Approximate LS-SVM based models for the state and its derivative are first estimated from the observed data. These estimates are then used for estimation of the unknown parameters of the model. Numerical results are presented and discussed for demonstrating the applicability of the proposed method.

Keywords: Delay differential equations, Parameter identification, Least squares support vector machines, Closed-form approximation

1. Introduction

Delay differential equations (DDEs) have been successfully used in the mathematical formulation of real life phenomena in a wide variety of applications especially in science and engineering such as population dynamics, infectious disease, control problems, secure communication, traffic control and economics [1, 2, 3]. In contrast with ordinary differential equations (ODEs) where the unknown function and its derivatives are evaluated at the same time instant, in a DDE the evolution of the system at a certain time instant, depends on the state of the system at an earlier time. A typical first order single-delay scalar DDE model may be expressed as:

$$\begin{aligned} \dot{x}(t) &= f_1(t, x(t), x(t-\tau_1), \theta(t)), \quad t \geq t_{in}, \\ x(t) &= \mathcal{H}_1(t), \quad \rho \leq t \leq t_{in} \end{aligned} \tag{1}$$

where $\mathcal{H}_1(t)$ is the initial function (history function), τ_1 is the delay or lag which is non-negative and can in general be constant, time dependent or state dependent i.e. $\tau_1 = \tau_1(t, x(t))$ and $\rho = \min_{t \geq t_{in}}\{t - \tau_1\}$. The term $x(t-\tau_1)$ is called the delay term. In more general models, the derivative $\dot{x}(t)$ may depend on $x(t)$ and $\dot{x}(t)$ itself at some past value $t - \tau_1$. In this case equation (1) can be rewritten in a more general form as follows

$$\begin{aligned} \dot{x}(t) &= f_2(t, x(t), x(t-\tau_1), \dot{x}(t-\tau_2), \theta(t)), \quad t \geq t_{in}, \\ x(t) &= \mathcal{H}_2(t), \quad \rho \leq t \leq t_{in} \end{aligned} \tag{2}$$

*Corresponding author

Email address: Siamak.Mehrkanoon@esat.kuleuven.be (Siamak Mehrkanoon)

where $\rho = \min_{1\le i\le 2}\{\min_{t\ge t_{in}}(t-\tau_i)\}$. Equation (2) is called delay differential equation of neutral type (NDDE). Models (1) and (2) usually involve some unknown parameters that require to be estimated from the observational data. We consider sets $\{\theta(t), \mathcal{H}_1(t), \tau_1\}$ and $\{\theta(t), \mathcal{H}_2(t), \tau_1, \tau_2\}$ as parameters of the models (1) and (2) respectively.

Identification of unknown parameters in differential equations has been studied and addressed by many authors (see [4, 5, 6, 7, 8, 9]). Most of the available approaches utilize the classical parametric inference such as the least squares estimator or the maximum likelihood estimation [10]. In these approaches first the dynamical system is simulated using initial guesses for the parameters. Then model predictions are compared with measured data and an optimization algorithm updates the parameters. Therefore considering the dynamical system (1) one has to solve the following optimization problem:

$$\underset{\theta(t),\tau_1}{\operatorname{argmin}}\, J(\theta(t), \tau_1) = \sum_{k=1}^{N} (y^m(t_k) - y^p(t_k))^2, \tag{3}$$

where $y^m(t)$ and $y^p(t)$ are the measured data and model prediction respectively. It should be noted that the objective function of the optimization problem for DDE differs from that of ODE. The cost function $J(\theta(t), \tau_1)$ in (3) might be non-smooth because the state trajectory might be non-smooth in the parameter and this will make the optimization problem more complicated.

Solving (3) requires repeated simulation of the system of DDE under study. Since the analytic solution of DDE is usually not available, therefore one needs to apply a numerical algorithm to simulate the given dynamic system. Although quite efficient numerical routines for solving differential equations are available they usually slow down the parameterization process dramatically and this situation is even more sensible when the underlying dynamic is described by delay differential equations. That is due to the existence of delay terms that force the solver to use an interpolation technique in order to advance the solution. It should also be noted that, as opposed to ordinary differential equation, the numerical solution of DDE not only depends on the parameter values, but also on the history function, $H_1(t)$ for $t \in [\rho, t_{in}]$, which is usually unknown. Given that the initial function is an infinite-dimensional set, the problem becomes an infinite-dimensional optimization problem and very difficult to solve [11]. Consequently, it would be of great benefit to eliminate any need of numerical DDE solvers.

Varah [13], proposed an approach for time-invariant parameter estimation of ODEs that does not require repeated numerical integration and is referred to as a two-step approach. First a cubic spline is used to estimate the system dynamics and its derivative from observational data. In the second step these estimates are plugged into a given differential equation and the unknown parameters are found by minimizing the squared difference of both sides of the differential equation.

The authors in [12] first estimate the derivative $\dot{x}(t)$ from the noisy data using nonparametric smoothing methods and then inferred the constant delay τ, for a special DDE model, in the framework of the generalized additive model.

The author in [14] proposed a method where an artificial neural network model is used to estimate the time invariant parameters of a dynamical systems governed by ordinary differential equations. Despite the fact that the classical neural networks have nice properties such as universal approximation, they still suffer from having two persistent drawbacks. The first problem is the existence of many local minima solutions. The second problem is how to choose the number of hidden units.

The parameter estimation in ordinary differential equations using least squares support vector machines is studied in [15]. It is the aim of this paper to extend the method proposed in [15] for estimating the unknown time varying/invariant parameters in parameter-affine delay differential equations for both non-neutral and neutral cases. Throughout this paper, we assume that the dynamical system is uniquely solvable and that the parameters of the model are identifiable. For stability of the solutions of systems with delays one may refer to [16, 17].

The paper is organized as follows. In section 2, the problem statement is given. In section 3, estimation of the state trajectory and its derivative by means of least squares support vector machines is discussed. Section 4 describes least squares support vector machines formulation to approximate the time varying/invariant parameters in DDEs and NDDEs. In section 5, examples are given in order to confirm the validity and applicability of the proposed method.

2. Problem statement

2.1. Reconstruction of fixed delays

Consider the dynamics of a process during a given time interval modeled by a system of nonlinear DDEs with associated history functions $\mathcal{H}(t)$ of the form:

$$\begin{aligned} \dot{x}(t) &= f(t, x(t), x(t-\tau_1), x(t-\tau_2), \ldots, x(t-\tau_p)), \quad t \geq t_{in}, \\ x(t) &= \mathcal{H}(t), \quad \rho \leq t \leq t_{in} \end{aligned} \tag{4}$$

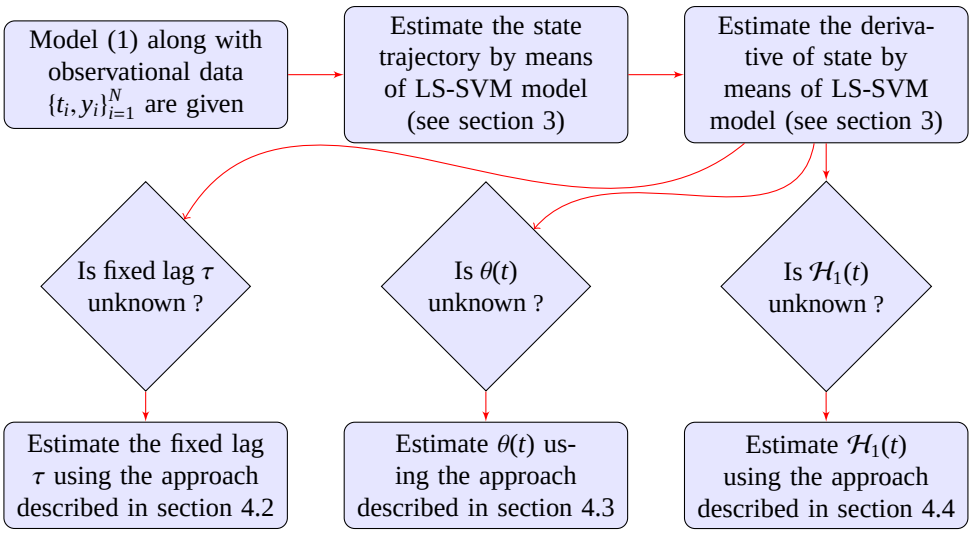
where $\rho = \min_{1\leq i\leq p} \{\min_{t\geq t_{in}}(t-\tau_i)\}$, $x(t) \in \mathbb{R}^n$ and the delays $\{\tau_i\}_{i=1}^p$ are constant and unknown. In order to estimate the model parameters, all the states of the system are measured i.e. $y(t_i) = x(t_i) + e(t_i)$ where $\{e(t_i)\}_{i=1}^N$ are independent measurement errors with zero mean. Throughout this paper a particular structure of (4) is considered. It is assumed that nonlinear model (4) exhibits the parameter-affine form i.e. it is affine in the $x(t-\tau_i)$ for $i = 1, \ldots, p$.

2.2. Reconstruction of time varying parameters

Consider the nonlinear state-dependent delay differential equation given in (1) with associated history function $\mathcal{H}_1(t)$. In order to estimate the unknown parameters, a set of measurements $y(t_i)$ are collected. In general the set of measurements $y(t_i)$ do not necessarily correspond to the model states $x(t_i)$. However here it is assumed that the system states are measured with measurement error $e(t_i)$, therefore the sate space model has the following form:

$$\begin{aligned} \dot{x}(t) &= f_1(t, x(t), x(t-\tau_1), \theta(t)), \quad t \geq t_{in}, \\ y(t_i) &= x(t_i) + e_i, \ i = 1, \ldots, N \end{aligned} \tag{5}$$

where $y(t)$ is the output of the system which has been observed at N time instants and $\{e_i\}_{i=1}^N$ are independent measurement errors with zero mean. The unknown $\{\mathcal{H}_1(t), \theta(t)\}$ are time dependent. In order to keep the model affine in the unknown time varying parameters we do not assume that both of them are unknown at the same time. Therefore as in [7, 8], we consider the case that one of them is unknown at the time of applying the estimation procedure. Hence the following cases can be studied: (i) $\mathcal{H}_1(t)$ is known and $\theta(t)$ is unknown, (ii) $\theta(t)$ is known and the history function $\mathcal{H}_1(t)$ is unknown, The same assumption is made for parameter estimation of the neutral delay differential equation (2). The general stages of the procedure when the dynamic system follows model (1) is described by the following flow-chart:



3. Estimation of the state trajectory and its derivative

Let us consider a given training set $\{t_i, y_i\}_{i=1}^N$ with input data $t_i \in \mathbb{R}$ and output data $y_i \in \mathbb{R}$ that are obtained from (5). The goal in regression is to estimate a model of the form $\hat{x}(t) = w^T\varphi(t) + b$. The primal LS-SVM model for regression can be written as follows [18, 19]

$$\begin{aligned} \underset{w,b,e}{\text{minimize}} \quad & \frac{1}{2}w^T w + \frac{\gamma}{2}e^T e \\ \text{subject to} \quad & y_i = w^T\varphi(t_i) + b + e_i \ , \ \ i = 1, ..., N \end{aligned} \tag{6}$$

where $\gamma \in \mathbb{R}^+, b \in \mathbb{R}, w \in \mathbb{R}^h$. $\varphi(\cdot) : \mathbb{R} \to \mathbb{R}^h$ is the feature map and h is the dimension of the feature space. The dual solution is then given by

$$\left[\begin{array}{c|c} \Omega + I_N/\gamma & 1_N \\ \hline 1_N^T & 0 \end{array}\right]\left[\begin{array}{c} \alpha \\ \hline b \end{array}\right] = \left[\begin{array}{c} y \\ \hline 0 \end{array}\right]$$

where $\Omega_{ij} = K(t_i, t_j) = \varphi(t_i)^T\varphi(t_j)$ is the (i, j)-th entry of the positive definite kernel matrix. $1_N = [1, \ldots, 1]^T \in \mathbb{R}^N$, $\alpha = [\alpha_1, \ldots, \alpha_N]^T, y = [y_1, \ldots, y_N]^T$ and I_N is the identity matrix. The model in dual form becomes:

$$\hat{x}(t) = w^T\varphi(t) + b = \sum_{i=1}^{N} \alpha_i K(t_i, t) + b, \tag{7}$$

where K is the kernel function. Making use of Mercer's theorem [20], derivatives of the feature map can be written in terms of derivatives of the kernel function. Therefore one can obtain a closed-form approximate expression for the derivative of the model (7) with respect to time as follows [21],

$$\frac{d}{dt}\hat{x}(t) = w^T\dot{\varphi}(t) = \sum_{i=1}^{N} \alpha_i K_s(t_i, t), \tag{8}$$

where $K_s(t, s) = \frac{\partial(\varphi(t)^T\varphi(s))}{\partial s}$.

4. Parameter estimation of DDE

4.1. General Methodology

The proposed scheme will make use of the LS-SVM ability to provide a closed-form approximation for the state trajectory and its derivative from measured data. We approximate the trajectory $\hat{x}(t)$ on the basis of observations at N points $\{t_i, y(t_i)\}_{i=1}^N$ using (7). Then (8) is utilized for approximating the state derivative. These closed-form expressions will be used later in the process of parameter estimation.

4.2. Fixed delay τ is unknown

For the sake of simplicity the methodology is described for a scalar DDE with single delay, but the approach is applicable for identifying multi-delays in a system of DDEs provided that they are identifiable. Consider the following single delay parameter-affine DDE:

$$\dot{x}(t) = f(t, x(t))x(t - \tau), \quad t \geq t_{in}, \tag{9}$$

where $f(\cdot) : \mathbb{R}^2 \longrightarrow \mathbb{R}$ is an arbitrary nonlinear function and τ is the constant parameter of the system which is unknown. In order to estimate the unknown τ value, the state of the system is measured i.e. $y(t_i) = x(t_i) + e(t_i)$ where $\{e(t_i)\}_{i=1}^N$ are independent measurement errors with zero mean. Let us assume an explicit LS-SVM model

$$\hat{x}_\tau(t) = v^T\psi(t) + d,$$

as an approximation for the term $x(t-\tau)$ where $\psi(\cdot) : \mathbb{R} \to \mathbb{R}^h$ is the feature map. Substituting the closed-form expressions for the state and its derivative, $\frac{d}{dt}\hat{x}(t)$ and $\hat{x}(t)$ obtained from (7) and (8) respectively, into the model description (9), the sought parameters v and d are identified as those minimizing the following optimization problem:

$$\begin{aligned} \underset{v,d,e}{\text{minimize}} \quad & \frac{1}{2}v^T v + \frac{\gamma}{2}\sum_{i=1}^{M} e_i^2 \\ \text{subject to} \quad & \frac{d}{dt}\hat{x}(t_i) = \Big(v^T\psi(t_i) + d\Big)f(t_i, \hat{x}(t_i)) + e_i, \text{ for } i = 1, ..., M. \end{aligned} \tag{10}$$

Remark 4.1. *Since closed-form expressions for the state and its derivative are available we are not limited to choose $M = N$, i.e. we can evaluate the constraint of the above optimization problem at the time instant t_i which is not necessarily the same as time instants that the system is measured.*

Lemma 4.1. *Given a positive definite kernel function $\tilde{K} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ with $\tilde{K}(t,s) = \psi(t)^T\psi(s)$ and a regularization constant $\gamma \in \mathbb{R}^+$, the solution to (10) is given by the following dual problem*

$$\left[\begin{array}{c|c} D\tilde{\Omega}D + \gamma^{-1}I & F \\ \hline F^T & 0 \end{array}\right]\left[\begin{array}{c} \alpha \\ \hline d \end{array}\right] = \left[\begin{array}{c} \frac{d\hat{x}}{dt} \\ \hline 0 \end{array}\right] \tag{11}$$

where $\tilde{\Omega}(i,j) = \tilde{K}(t_i, t_j) = \psi(t_i)^T\psi(t_j)$ is the (i,j)-th entry of the positive definite kernel matrix and I is the identity matrix. Also $\alpha = [\alpha_1, \ldots, \alpha_M]^T$, $F = [f(t_1, \hat{x}(t_1)), \ldots, f(t_M, \hat{x}(t_M))]^T$, $\frac{d\hat{x}}{dt} = [\frac{d}{dt}\hat{x}(t_1), \ldots, \frac{d}{dt}\hat{x}(t_M)]^T$. D is a diagonal matrix with the elements of F on the main diagonal.

Proof 4.1. *The Lagrangian of the constrained optimization problem (10) becomes*

$$\mathcal{L}(v, d, e_i, \alpha_i) = \frac{1}{2}v^T v + \frac{\gamma}{2}\sum_{i=1}^{M} e_i^2 - \sum_{i=1}^{M}\alpha_i\Big[\Big(v^T\psi(t_i) + d\Big)f(t_i, \hat{x}(t_i)) + e_i - \frac{d}{dt}\hat{x}(t_i)\Big],$$

where $\{\alpha_i\}_{i=1}^M$ are Lagrange multipliers. Then the Karush-Kuhn-Tucker (KKT) optimality conditions are as follows,

$$\begin{aligned} &\frac{\partial\mathcal{L}}{\partial v} = 0 \to v = \sum_{i=1}^{M}\alpha_i f(t_i, \hat{x}(t_i))\psi(t_i), \\ &\frac{\partial\mathcal{L}}{\partial d} = 0 \to \sum_{i=1}^{M}\alpha_i f(t_i, \hat{x}(t_i)) = 0, \\ &\frac{\partial\mathcal{L}}{\partial e_i} = 0 \to e_i = \frac{\alpha_i}{\gamma}, \quad i = 1, \ldots, M, \\ &\frac{\partial\mathcal{L}}{\partial \alpha_i} = 0 \to \Big(v^T\psi(t_i) + d\Big)f(t_i, \hat{x}(t_i)) + e_i = \frac{d}{dt}\hat{x}(t_i), \text{for } \; i = 1, \ldots, M. \end{aligned}$$

After elimination of the primal variables v and $\{e_i\}_{i=1}^M$ and making use of Mercer's Theorem, the solution is given in the dual by

$$\begin{cases} \frac{d}{dt}\hat{x}(t_i) = \sum_{j=1}^{M}\alpha_j f(t_j, \hat{x}(t_j))\Omega_{ji} f(t_i, \hat{x}(t_i)) + \frac{\alpha_i}{\gamma} + d\, f(t_i, \hat{x}(t_i)), \; i = 1, \ldots, M \\ 0 = \sum_{i=1}^{M}\alpha_i f(t_i, \hat{x}(t_i)) \end{cases}$$

Writing these equations in matrix form gives the linear system in (11). □

The model in the dual form becomes

$$\hat{x}_\tau(t) = v^T\psi(t) + d = \sum_{i=1}^{M}\alpha_i f(t_i, \hat{x}(t_i))\tilde{K}(t_i, t) + d, \tag{12}$$

where $\tilde{K}$ is the kernel function.

Remark 4.2. *If one is not interested in having a closed-form approximation to the term $x(t-\tau)$, an alternative way to obtain an approximation for $x(t-\tau)$ at the time instant t_i is by using (9) directly, i.e. $x(t_i-\tau) = \frac{d}{dt}\hat{x}(t_i)^{-1} f(t_i, \hat{x}(t_i))$. A similar strategy can be applied in the case that the dynamics of the process is described by a system of delay differential equations. After substituting the closed-form expressions for the states and their derivatives into the model, then one has to solve a system of linear equations (provided that the underlying system is affine in the unknown parameter) to obtain the approximation of the delay terms $x(t-\tau_j)$ for $j = 1, \ldots, p$ at time instants $t = t_i$, for $i = 1, \ldots, N$.*

After obtaining the estimation $\hat{x}_\tau(t)$, the task is to estimate the fixed delay τ. To this end, let us first define a shifting operator $\Delta_m(\cdot)$ which will be used in the process of estimation of the delay τ. Operator $\Delta_m(\cdot)$ shifts the given time series, which in our problem setting can for example be $\hat{x}(t)$ or $\hat{\dot{x}}(t)$, m steps forward in time in a certain manner, while keeping the length of the time series unchanged. This is done by adding a constant vector of size m (whose values will be clarified later) from the left to the time series and removing the m last elements of the time series simultaneously. Therefore, given the time series $\hat{x}(t) = [\hat{x}(t_1), \hat{x}(t_2), \ldots, \hat{x}(t_N)]^T$, operator $\Delta_m(\cdot)$ is defined as follows:

$$z(t) = \Delta_m(\hat{x}(t)) = \begin{cases} [\underbrace{z(t_1), z(t_2), \ldots, z(t_m)}_{\text{Constant vector}}, \hat{x}(t_1), \hat{x}(t_2), \ldots, \hat{x}(t_{N-m})]^T, & \text{for } 1 \le m \le N-1 \\ \hat{x}(t), & \text{for } m = 0 \end{cases} \tag{13}$$

with $z(t_1) = z(t_2) = \ldots, z(t_m) = c$ where c is a constant. Noting that in an ideal case (noise free) one can expect a delay differential equation to have the following property

$$\hat{x}_\tau(t)\Big|_{t=\tau} = \hat{x}(t)\Big|_{t=t_{in}}, \quad \text{for } \tau \ge 0, \tag{14}$$

it is natural to utilize the first element of $\hat{x}(t)$, i.e., $\hat{x}(t_1)$ as a constant c used in operator $\Delta_m(\cdot)$. In order to estimate the delay τ, we use the sample correlation coefficient function defined as:

$$r_{z\hat{x}_\tau} = \frac{\sum_{i=1}^{N}(z(t_i)-\mu_1)(\hat{x}_\tau(t_i)-\mu_2)}{\sqrt{\sum_{i=1}^{N}(z(t_i)-\mu_1)^2}\sqrt{\sum_{i=1}^{N}(\hat{x}_\tau(t_i)-\mu_2)^2}}, \tag{15}$$

where μ_1 and μ_2 denote the sample mean of time series $z(t)$ and $\hat{x}_\tau(t)$ respectively. Given $\hat{x}(t)$ and $\hat{x}_\tau(t)$ the process of estimating the unknown delay τ is described in Algorithm 1.

Algorithm 1: Approximating the constant delay of a given DDE

Input: Time series $\hat{x}(t)$ and $\hat{x}_\tau(t)$ of size N; sampling time T_s (in seconds).
Output: Time delay τ

for $m \leftarrow 0$ ***to*** $N-1$ **do**
  $z(t) \leftarrow \Delta_m(\hat{x}(t))$
  $R(m) \leftarrow$ `Corrcoef`$(z(t), \hat{x}_\tau(t))$
$\tau \leftarrow T_s \times \underset{m}{\text{argmax}}\, R(m)$
return τ

In Algorithm 1, `Corrcoef` is a Matlab built-in function that computes the correlation coefficient of two signals and $R(m)$ corresponds to $r_{z\hat{x}_\tau}$. One may notice that in this approach we are not using the history function for estimating the time delay τ. But if the history function is known a priori, one may use it for constructing the constant vector used in operator $\Delta_m(\cdot)$ by taking the value of history function at time t_{in}.

4.3. Parameter $\theta(t)$ is unknown

Consider model (1) and case (i) where the time varying parameter $\theta(t)$ is unknown and delay τ_1 is known. Therefore with a slight abuse of notation, let us assume an explicit LS-SVM model

$$\hat{\theta}(t) = v^T\psi(t) + d,$$

as an approximation for the parameter $\theta(t)$. The adjustable parameters v and d are to be found by solving the following optimization problem

$$\begin{aligned} \underset{v,d,e,\epsilon,\theta_i}{\text{minimize}} \quad & \frac{1}{2}v^T v + \frac{\gamma}{2}(\sum_{i=1}^{M} e_i^2 + \sum_{i=1}^{M} \epsilon_i^2) \\ \text{subject to} \quad & \frac{d}{dt}\hat{x}(t_i) = f_1(t_i, \hat{x}(t_i), \hat{x}(t_i - \tau_1), \theta_i) + e_i, \text{ for } i = 1, \ldots, M, \\ & \theta_i = v^T \psi(t_i) + d + \epsilon_i, \text{ for } i = 1, \ldots, M. \end{aligned} \tag{16}$$

Here the obtained closed-form expressions for the state and its derivative, $\frac{d}{dt}\hat{x}(t)$ and $\hat{x}(t)$ obtained from (7) and (8), are substituted into the model description (1). If f_1 is nonlinear in $\theta(t)$ then the above optimization problem is non-convex. The solution of (16) in the dual can be obtained by solving a system of nonlinear equations. However, throughout this paper, we present our results for the case that the nonlinear model (1) is affine in the parameter $\theta(t)$. More precisely we consider the following parameter-affine form of (1)

$$\begin{aligned} \dot{x}(t) &= \theta(t) f_1(t, x(t), x(t - \tau_1)), \quad t \geq t_{in}, \\ x(t) &= \mathcal{H}_1(t), \quad t \leq t_{in}. \end{aligned}$$

This will result in the following convex optimization problem:

$$\begin{aligned} \underset{v,d,e}{\text{minimize}} \quad & \frac{1}{2}v^T v + \frac{\gamma}{2}\sum_{i=1}^{M} e_i^2 \\ \text{subject to} \quad & \frac{d}{dt}\hat{x}(t_i) = \Big(v^T \psi(t_i) + d\Big) f_1(t_i, \hat{x}(t_i), \hat{x}(t_i - \tau_1)) + e_i, \text{ for } i = 1, \ldots, M. \end{aligned} \tag{17}$$

Lemma 4.2. *Given a positive definite kernel function $\tilde{K} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ with $\tilde{K}(t, s) = \psi(t)^T \psi(s)$ and a regularization constant $\gamma \in \mathbb{R}^+$, the solution to (17) is given by the following dual problem*

$$\left[\begin{array}{c|c} D\tilde{\Omega}D + \gamma^{-1}I & F_1 \\ \hline F_1^T & 0 \end{array}\right]\left[\begin{array}{c} \alpha \\ \hline d \end{array}\right] = \left[\begin{array}{c} \frac{d\hat{x}}{dt} \\ \hline 0 \end{array}\right] \tag{18}$$

where $\tilde{\Omega}(i, j) = \tilde{K}(t_i, t_j) = \psi(t_i)^T \psi(t_j)$ is the (i, j)-th entry of the positive definite kernel matrix and I is the identity matrix. Also $\alpha = [\alpha_1, \ldots, \alpha_M]^T$, $F_1 = [f_1(t_1, \hat{x}(t_1), \hat{x}(t_1 - \tau_1)), \ldots, f_1(t_M, \hat{x}(t_M), \hat{x}(t_M - \tau_1))]^T$, $\frac{d\hat{x}}{dt} = [\frac{d}{dt}\hat{x}(t_1), \ldots, \frac{d}{dt}\hat{x}(t_M)]^T$. D is a diagonal matrix with the elements of F_1 on the main diagonal.

The model in the dual form becomes:

$$\hat{\theta}(t) = \sum_{i=1}^{M} \alpha_i f_1(t_i, \hat{x}(t_i), \hat{x}(t_i - \tau_1))\tilde{K}(t_i, t) + d, \tag{19}$$

where $\tilde{K}$ is the kernel function.

Proof 4.2. *The approach is the same as in proof of Lemma 4.1.* □

It should be noted that in the process of estimating $\theta(t)$, the values of the history function $\mathcal{H}_1(t)$ are not used. Therefore $\mathcal{H}_1(t)$ can also be unknown while $\theta(t)$ is being estimated which is the advantage of the proposed method compared with conventional approaches that require the history function for simulating the underlying model.

Remark 4.3. *The same procedure can be applied for estimating the unknown parameter $\theta(t)$ in parameter-affine form of model (2).*

4.4. History function $\mathcal{H}_1(t)$ is unknown

Consider model (1) and case (ii) where the parameter $\mathcal{H}_1(t)$ is unknown and all the other parameters are known. It is assumed that the nonlinear function f_1 is affine in $x(t-\tau_1)$. More precisely we consider the following form of (1):

$$\begin{aligned} \dot{x}(t) &= x(t-\tau_1)\, f_1(t, x(t), \theta(t)), \quad t \geq t_{in}, \\ x(t) &= \mathcal{H}_1(t), \quad t \leq t_{in} \end{aligned} \tag{20}$$

where τ_1 can be time and state dependent. Since the history function is time varying let us, with a slight abuse of notation, assume an explicit LS-SVM model

$$\hat{\mathcal{H}}_1(t) = v^T \psi(t) + d,$$

as an approximation to the true $\mathcal{H}_1(t)$. Optimal value for v and d can be obtained by solving the following convex optimization problem:

$$\begin{aligned} \underset{v,d,e}{\text{minimize}} \quad & \frac{1}{2} v^T v + \frac{\gamma}{2} \sum_{i=1}^{|\mathcal{T}|} e_i^2 \\ \text{subject to} \quad & \frac{d}{dt}\hat{x}(t^i_{sel}) = \Big(v^T \psi(t^i_{sel}) + d\Big) f_1(t^i_{sel}, \hat{x}(t^i_{sel}), \theta(t^i_{sel})) + e_i, \text{ for } i = 1, \ldots, |\mathcal{T}|, \end{aligned} \tag{21}$$

where $\frac{d}{dt}\hat{x}(t^i_{sel})$ and $\hat{x}(t^i_{sel})$ are estimations of the state trajectory and its derivative obtained by using LS-SVM models (7) and (8) respectively. $|\mathcal{T}|$ is the cardinality of the ordered set $\mathcal{T} = \{t^1_{sel}, t^2_{sel}, \ldots, t^{|\mathcal{T}|}_{sel}\}$ whose elements are selected using Algorithm 2.

Algorithm 2: Approximating the model's time varying history function

Input: Vector T consists of time instants $\{t_i\}_{i=1}^N$ and the delay τ_1
Output: set $\mathcal{T}$

for $i \leftarrow 1$ ***to*** N **do**
  $tlag(i) \leftarrow t_i - \tau_1(t_i)$
Find a vector of indices of elements of *tlag* whose values are less than t_{in} (assuming that $t_{in} = t_1$)
$\mathcal{T} \leftarrow$ elements of T corresponding to the indices found in step 2.
return $\mathcal{T}$

The solution to (21) in the dual can be obtained by solving linear system (18) with $\alpha = [\alpha_1, \ldots, \alpha_{|\mathcal{T}|}]^T$, $F_1 = [f_1(t^1_{sel}, \hat{x}(t^1_{sel}), \theta(t^1_{sel})), \ldots, f_1(t^{|\mathcal{T}|}_{sel}, \hat{x}(t^{|\mathcal{T}|}_{sel}), \theta(t^{|\mathcal{T}|}_{sel}))]^T$ and $\frac{d\hat{x}}{dt} = [\frac{d}{dt}\hat{x}(t^1_{sel}), \ldots, \frac{d}{dt}\hat{x}(t^{|\mathcal{T}|}_{sel})]^T$. D is a diagonal matrix with the elements of F_1 on the main diagonal. The model in the dual form becomes:

$$\hat{\mathcal{H}}_1(t) = \sum_{i=1}^{|\mathcal{T}|} \alpha_i\, f_1(t^i_{sel}, \hat{x}(t^i_{sel}), \theta(t^i_{sel}))\tilde{K}(t_i, t) + d, \tag{22}$$

where $\tilde{K}$ is the kernel function. If delay τ_1 in the model (20) is constant, one can first utilize Algorithm 1 to estimate the delay τ_1 and then apply Algorithm 2 to obtain a closed-form approximation to the history function $\mathcal{H}_1(t)$.

Remark 4.4. *The same procedure can be applied for estimating the unknown history function $\mathcal{H}_2(t)$ in a parameter-affine form of model (2).*

5. Experiments

In this section, six experiments are performed to demonstrate the capability of the proposed method for time varying/invariant parameters of parameter-affine non-neutral DDEs and neutral DDEs. The last three test problems are taken from [7] and [8], but in contrast with the approach given in these references, we allow to have measurement

errors. The performance of the LS-SVM model depends on the choice of the tuning parameters. In this paper, for all experiments, the Gaussian RBF kernel i.e. $K(x,y) = \exp(-\frac{\|x-y\|_2^2}{\sigma^2})$ is used. Therefore, a model is determined by the regularization parameter γ and the kernel bandwidth σ. The 10-fold cross validation criterion is used to tune these parameters. The SNR stands for signal to noise ratio which is calculated using $20\log_{10}(\frac{A_{signal}}{A_{noise}})$ where A_{signal} and A_{noise} are the root mean square of the signal and noise respectively. The estimated parameter values are obtained by averaging over 10 simulation runs. As error bounds we used about twice the standard deviation of the error.

5.1. Constant parameters

Problem 5.1. Consider a Kermack-McKendrick model of an infectious disease with periodic outbreak [22, Example 1]

$$\begin{aligned}\dot{x}_1(t) &= -x_1(t)x_2(t-\tau_1) + x_2(t-\tau_2)\\ \dot{x}_2(t) &= x_1(t)x_2(t-\tau_1) - x_2(t)\\ \dot{x}_3(t) &= x_2(t) - x_2(t-\tau_2)\end{aligned} \tag{23}$$

on $[0, 20]$ with history $x_1(t) = 5, x_2(t) = 0.1$ and $x_3(t) = 1$ for $t \le 0$. The true value of the delays are $\tau_1 = 1$ and $\tau_2 = 10$. For collecting the data, the solution of the this system is prepared by numerically integrating the differential equation (23) using MATLAB built-in solver *dde23*, on domain $[0, 20]$ with the relative error tolerance RelTol= 10^{-6}. Then the model observation data are constructed by adding Gaussian white noise with zero mean to the true solution. The observation data points are prepared within the domain of $[0, 20]$ with sampling time $T_s = 100$ ms (i.e. 201 data points). The obtained results are shown in Fig. 1. As Fig. 1(e) and (g) suggest the peaks of the correlation coefficients occurred nearly at indices 10 and 100. Multiplying these indices with sampling time T_s (in seconds), yields an estimate of the unknown delays τ_1 and τ_2, respectively. Fig. 2(a) and (b) show the influence of noise level on the parameter estimation. It should be noted that as the value of signal to noise ratio increases, the standard deviation of the estimation error decreases.

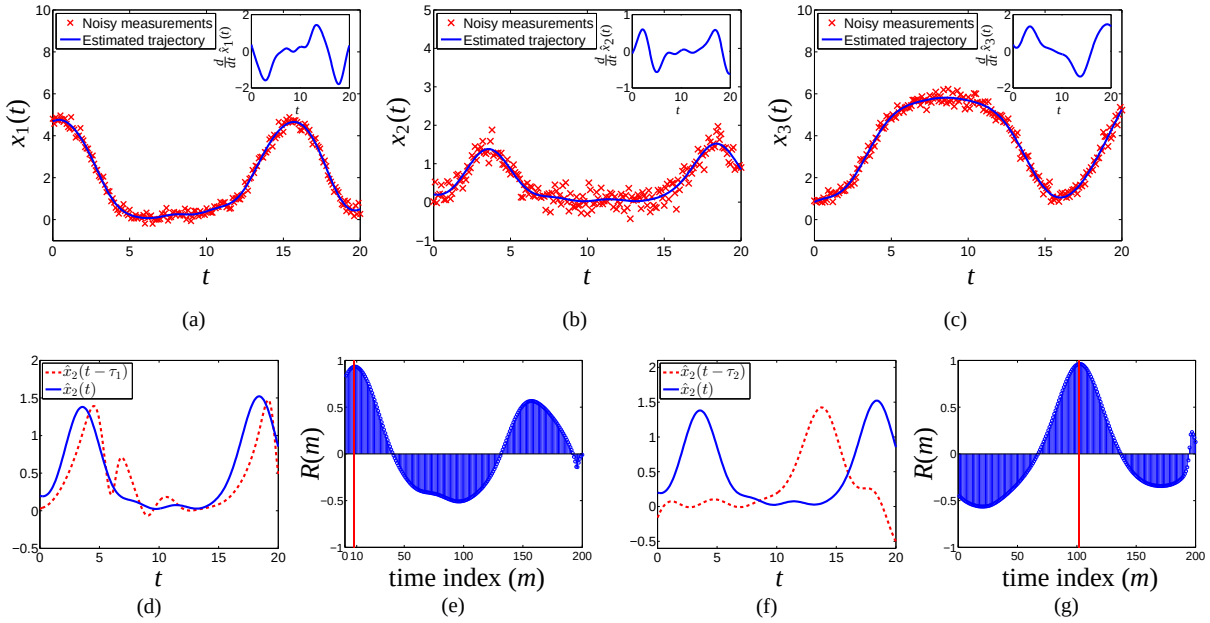
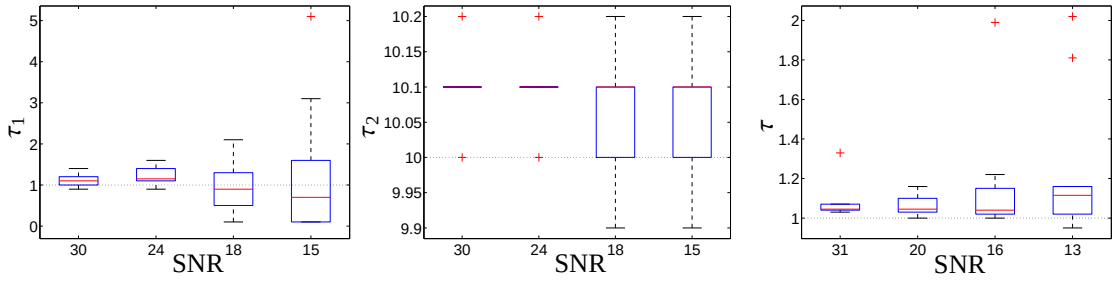


Figure 1: Estimation of constant delays τ_1 and τ_2 in Problem 5.1 from observational data. (a) Estimation of the first state $x_1(t)$ and its derivative from the observational data. (b) Estimation of the second state $x_2(t)$ and its derivative from observational data. (c) Estimation of the third state $x_3(t)$ and its derivative from observational data. (d) Estimation of $x_2(t-\tau_1)$ and $x_2(t)$. (e) Correlation-coefficient values as a function of time index m for two time series $x_2(t)$ and $x_2(t-\tau_1)$ as computed in Algorithm 1. (f) Estimation of $x_2(t-\tau_2)$ and $x_2(t)$. (g) Correlation-coefficient values as a function of time index m for two time series $x_2(t)$ and $x_2(t-\tau_2)$, as computed in Algorithm 1.



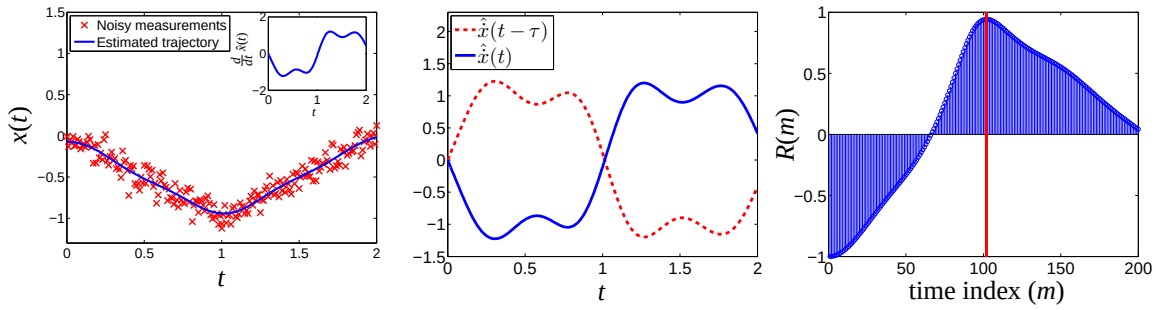
(a) (b) (c)

Figure 2: Estimation of constant delays τ_1 and τ_2 in Problem 5.1 and delay τ in Problem 5.2 from observational data for different values of signal to noise ratio. The exact value of the lags are denoted by the dashed lines. (a) Estimation of delay τ_1 for problem 5.1. (b) Estimation of the delay τ_2 for problem 5.1. (c) Estimation of the delay τ for problem 5.2.

Problem 5.2 Consider a triangle wave defined by the following scalar NDDE:

$$\begin{aligned} \dot{x}(t) &= -\dot{x}(t-\tau) \\ x(t) &= t, \ -\tau \le t \le 0. \end{aligned} \tag{24}$$

In order to prepare the observational data, the solution to (24) is generated, with the true delay $\tau = 1$, by using MATLAB built-in solver *ddesd*, on domain [0,2] with the relative error tolerance RelTol= 10^{-6}. Then the model observation data are constructed by adding Gaussian white noise with zero mean to the true solution. The observation points are prepared within the domain of $[0, 2]$ with sampling time $T_s = 10$ ms (i.e. 201 data points). Fig. 3 represents the results obtained by applying the proposed method for estimating the unknown delay τ. The result of parameter estimation for different values of signal to noise ratio is depicted in Fig. 2(c). From Fig. 2(c), one may notice that as the value of signal to noise ratio increases, the standard deviation of the estimation error decreases.



(a) (b) (c)

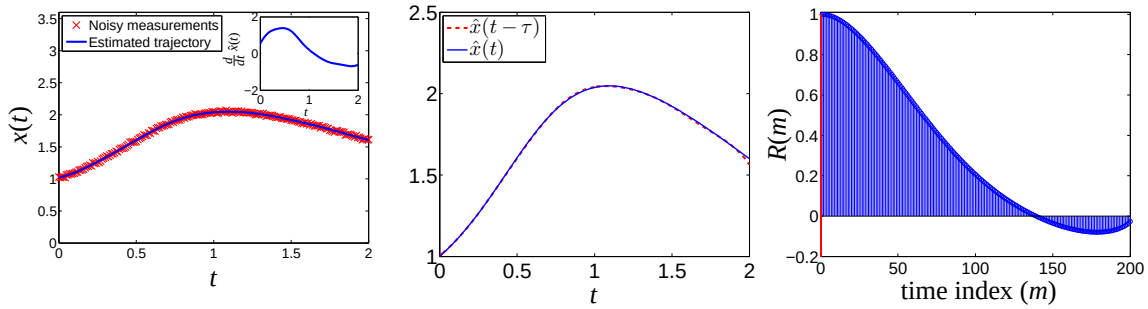
Figure 3: Estimation of constant lag τ in Problem 5.2 from observational data. (a) Estimation of the state $x(t)$ and its derivative from observational data. (b) Estimation of $\hat{\dot{x}}(t-\tau)$ and $\hat{\dot{x}}(t)$. (c) Correlation-coefficient values as a function of time index m for two time series $\hat{\dot{x}}(t)$ and $\hat{\dot{x}}(t-\tau)$, as computed in Algorithm 1.

Problem 5.3 Consider an artificial example:

$$\begin{aligned} \dot{x}(t) &= \sin(x(t)+t)x(t-\tau), \quad t \in [0,2] \\ x(0) &= 1, \end{aligned} \tag{25}$$

where the true delay $\tau = 0$. The solution to (25) is generated, with the true delay $\tau = 0$, by using MATLAB built-in solver *ode45*, on domain [0,2] with the relative error tolerance RelTol= 10^{-6}. Then the model observation data are

constructed by adding Gaussian white noise with zero mean to the true solution. The observation points are prepared within the domain of $[0, 2]$ with sampling time $T_s = 10$ ms (i.e. 201 data points). The obtained results for estimating the unknown delay τ are shown in Fig. 4. As Fig. 4(c) suggests the peak of the correlation coefficient occurred at index $m = 0$. Based on Algorithm 1, multiplying this index with sampling time T_s (in seconds), yields an estimate of the unknown delays τ. Thus the estimated lag τ is zero.



(a) (b) (c)

Figure 4: Estimation of constant lag τ in Problem 5.3 from observational data. (a) Estimation of the state $x(t)$ and its derivative from observational data. (b) Estimation of $\hat{x}(t-\tau)$ and $\hat{x}(t)$. (c) Correlation-coefficient values as a function of time index m for two time series $\hat{x}(t)$ and $\hat{x}(t-\tau)$, as computed in Algorithm 1.

5.2. Time varying parameters

Problem 5.4. Consider the linear delay equation [7, Problem 2]

$$\begin{aligned} \dot{x}(t) &= \theta(t)x(t-\xi(t)),\ t \in [0,2] \\ x(t) &= \mathcal{H}_1(t),\ t \in [-2,0] \end{aligned} \tag{26}$$

where

$$\xi(t) = \begin{cases} 2-t^2, & t \in [0,1] \\ 1, & t \in [1,2] \end{cases}, \quad \theta(t) = \begin{cases} \frac{-t}{t+1}, & t \in [0,1] \\ -\frac{1}{2}, & t \in [1,2] \end{cases}$$

and $\mathcal{H}_1(t) = t^2$. It is assumed that the initial function $\mathcal{H}_1(t)$ and $\xi(t)$ are known and we aim at estimating the unknown parameter $\theta(t)$ from measured data. For collecting the data, the solution of this system is prepared by numerically integrating the differential equation (26) using MATLAB built-in solver *ddesd*, on domain [0,2] with the relative error tolerance RelTol= 10^{-6}. Then the model observation data are constructed by adding Gaussian white noise with zero mean to the true solution. The observation points are prepared within the domain of $[0, 2]$ with sampling time $T_s = 10$ ms (i.e. 201 data points). Applying the presented scheme in section 4.3, an estimation $\hat{\theta}(t)$ is obtained and the results are depicted in Fig. 5(a) and (b). The root mean square errors (RMSE) for different values of signal to noise ratio are also tabulated in Table 1. From Table 1, it is apparent that as the value of signal to noise ratio (SNR) increases, the estimation error decreases.

Problem 5.5. Consider the linear delay equation (26). In this problem we assume that $\theta(t)$ and $\xi(t)$ are known and we aim at estimating the initial function from measured data [7, Problem 1]

$$\begin{aligned} \dot{x}(t) &= \theta(t)x(t-\xi(t)),\ t \in [0,2] \\ x(t) &= \mathcal{H}_1(t),\ t \in [-2,0]. \end{aligned} \tag{27}$$

As in Problem 5.4, the observational data are prepared within the domain of $[0, 2]$ with sampling time $T_s = 10$ ms (i.e. 201 data points). Fig. 5(c) and (d), shows the obtained approximation $\hat{\mathcal{H}}_1(t)$ for the history function when the scheme described in section 4.4 is utilized. The root mean square errors (RMSE) for different values of signal to noise ratio are recorded in Table 1. From Table 1, it is apparent that as the value of signal to noise ratio (SNR) increases, the estimated parameter converges to the true parameter.

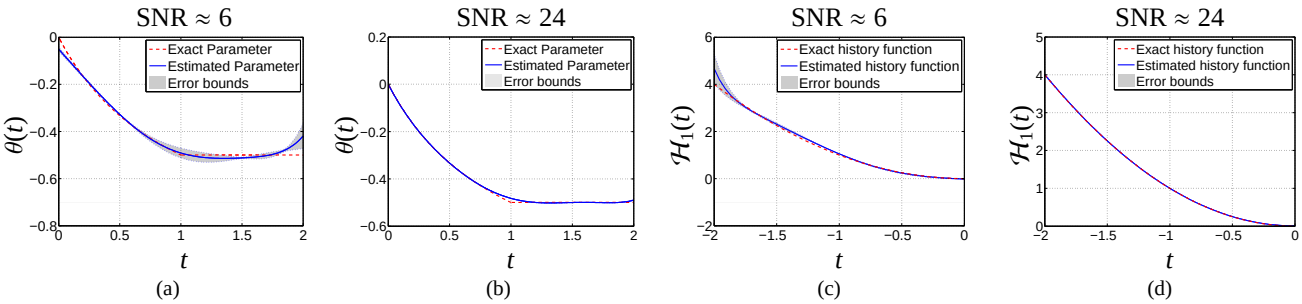


Figure 5: (a) and (b) Estimation of time varying parameter $\theta(t)$ in Problem 5.4 from observational data for different values of signal to noise ratio. (c) and (d) Estimation of History function $\mathcal{H}_1(t)$ in Problem 5.5 from observational data for different values of signal to noise ratio.

Table 1: The influence of signal to noise ratio on the parameter estimates for problems 5.4, 5.5, 5.6 when 201 data points is used.

	RMS Error		
SNR	**Problem 5.4**	**Problem 5.5**	**Problem 5.6**
6	$1.72e-2$	$2.87e-1$	$1.13e-1$
11	$1.32e-2$	$2.12e-2$	$1.17e-2$
18	$7.01e-3$	$4.02e-3$	$3.14e-3$
24	$2.10e-3$	$2.03e-3$	$1.01e-3$

SNR stands for signal to noise ratio.

Problem 5.6. Consider the following state dependent delay neutral delay differential equations [8, Problem 1]

$$\begin{aligned} \dot{x}(t) &= \theta(t) + \dot{x}(t - \frac{t^2}{t^2+4}|x(t)| - 1), \ t \in [0,1] \\ x(t) &= \frac{1}{4}t^2 + 1, \ t \leq 0. \end{aligned} \tag{28}$$

It is assumed that the time varying parameter $\theta(t)$ is unknown and has to be estimated from measured data. The true parameter is $\theta(t) = \frac{1}{8}t^2 + \frac{1}{2}$. It is easy to check that the true solution of for the given $\theta(t)$ is $x(t) = \frac{1}{4}t^2 + 1$.

The model observation data are constructed by adding Gaussian white noise with zero mean to the true solution. The observation points are prepared within the domain of $[0, 1]$ with sampling time $T_s = 5$ ms (i.e. 201 data points). The obtained results are shown in Fig. 6. The root mean square errors (RMSE) for different values of signal to noise ratio are given in Table 1. The results reveal that higher order accuracy can be achieved by increasing the value of signal to noise ratio.

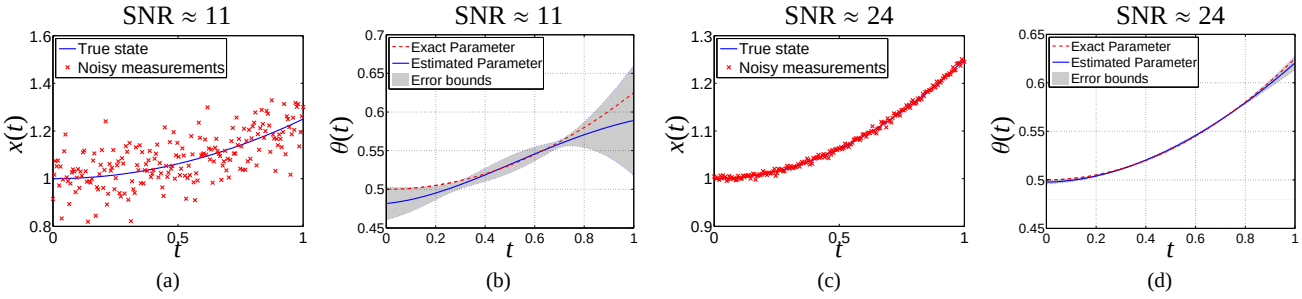


Figure 6: Estimation of time varying parameter $\theta(t)$ in Problem 5.6 from observational data for different values of signal to noise ratio.

6. Conclusions

In this paper a new approach based on LS-SVMs has been proposed for estimation of constant as well as time varying parameters of dynamical system governed by non-neutral and neutral delay differential equations from observational data in the presence of measurements noise. The method provides a fast approximation for the unknown parameters of the model without requiring numerical integration of the given dynamic system. Therefore it makes a suitable candidate for online parameter estimation. In addition the obtained results can be used in initialization of the conventional optimization approach where repeated integration of the dynamic system is required.

Acknowledgments

This work was supported by Research Council KUL: GOA/11/05 Ambiorics, GOA/10/09 MaNet, CoE EF/05/006 Optimization in Engineering(OPTEC), IOF-SCORES4CHEM, several PhD/postdoc & fellow grants;Flemish Government:FWO: PhD/postdoc grants, projects: G0226.06 (cooperative systems and optimization), G0321.06 (Tensors), G.0302.07 (SVM/Kernel), G.0320.08 (convex MPC), G.0558.08 (Robust MHE), G.0557.08 (Glycemia2), G.0588.09 (Brain-machine), G.0377.12 (structured models) research communities (WOG: ICCoS, ANMMM, MLDM); G.0377.09 (Mechatronics MPC) IWT: PhD Grants, Eureka-Flite+, SBO LeCoPro, SBO Climaqs, SBO POM, O&O-Dsquare; Belgian Federal Science Policy Office: IUAP P6/04 (DYSCO, Dynamical systems, control and optimization, 2007-2011); EU: ERNSI; FP7-HD-MPC (INFSO-ICT-223854), COST intelliCIS, FP7-EMBOCON (ICT-248940); Contract Research: AMINAL; Other:Helmholtz: viCERP, ACCM, Bauknecht, Hoerbiger, ERC AdG A-DATADRIVE-B. Johan Suykens is a professor at the KU Leuven, Belgium.

References

[1] A. Bellen and M. Zennaro, *Numerical methods for delay differential equations*, Numerical Mathematics and Scientific Computation, Clarendon Press, Oxford University Press, New York, 2003.

[2] J.J. Batzel , H.T. Tran, Stability of the human respiratory control system I. Analysis of a two-dimensional delay state-space model. *J Math Biol*, 2000, 41:45-79.

[3] T.K. Nagy, G. Stépán, F.C. Moon. Subcritical Hopf bifurcation in the delay equation model for machine tool vibrations. *Nonlinear Dynamics* 2001;26:121-42.

[4] H.T. Banks, J.A. Burns and E.M. Cliff, Parameter estimation and identification for systems with delays, *SIAM J. Control & Opt*, (19), 1981, pp. 791-828.

[5] H.T. Banks, P.L. Daniel, Estimation of delays with other parameters in nonlinear functional differential equations, *SIAM J. Control & Opt*, (21), 1983, pp. 895-915.

[6] S. Ahmed, B. Huang, and S.L. Shah, Parameter and delay estimation of continuous-time models using a linear filter, *Journal of Process Control*, (16), 2006, pp. 323-331.

[7] F. Hartung, Parameter estimation by quasilinearization in functional differential equations with state-dependent delays: a numerical study, *Nonlinear Analysis*, (47), 2001, pp. 4557-4566.

[8] F. Hartung and J. Turi, Identification of parameters in neutral functional differential equations with state-dependent delays, *Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference 2005 Seville*, Spain, December 12-15, 2005.

[9] S.N. Wood, Partially specified ecological models, *Ecological Monographs*, (71), 2001, pp. 1-25.

[10] L.T. Biegler, J.J. Damiano, G.E. Blau, Nonlinear parameter estimation: a case-study comparison, *AIChE Journal*, 32(1), 1986, pp. 29-45.

[11] L. Wang and J. Cao, Estimating parameters in delay differential equation models, *Journal of Agricultural, Biological, and Environmental Statistics*, (17), 2012, pp. 68-83.

[12] S.P. Ellner, B.E. Kendall, S.N. Wood, E. McCauley, and C.J. Briggs, Inferring mechanism from time-series data: delay-differential equations, *Physica D*, 110, (1997), pp. 182-194.

[13] J. M. Varah. A spline least squares method for numerical parameter estimation in differential equations, *SIAM Journal on Scientific and Statistical Computing*, 3(1), 1982, pp. 28-46.

[14] V. Dua. An Artificial Neural Network approximation based decomposition approach for parameter estimation of system of ordinary differential equations, *Computers and Chemical Engineering*, (35), 2011, pp. 545-555.

[15] S. Mehrkanoon, T. Falck, J.A.K. Suykens, Parameter Estimation for Time Varying Dynamical Systems using Least Squares Support Vector Machines, *in Proc. of the 16th IFAC Symposium on System Identification (SYSID 2012), Brussels, Belgium*, Jul. 2012, pp. 1300-1305.

[16] V.B. Kolmanovskii and A.D. Myshkis. *Applied Theory of Functional Differential Equations*, Kluwer, Dordrecht 1992.

[17] A.D. Myshkis. Razumikhin's method in the qualitative theory of processes with delay, *Journal of Applied Mathematics and Stochastic Analysis*, 8(3), 1995, pp. 233-247.

[18] J.A.K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, J. Vandewalle. *Least Squares Support Vector Machines*, World Scientific, Singapore, 2002.

[19] J.A.K. Suykens, C. Alzate, K. Pelckmans, Primal and dual model representations in kernel-based learning, *Statistics Surveys*, 4:148-183, 2010.

[20] V. Vapnik, *Statistical learning theory*, New York, Wiley 1998.

[21] S. Mehrkanoon, J.A.K. Suykens, LS-SVM approximate solution to linear time varying descriptor systems, *Automatica*, vol. 48, no. 10, 2012, pp. 2502-2511.

[22] L.F. Shampine and S. Thompson, Solving DDEs in MATLAB, *Applied Numerical Mathematics* 37, (2001), 441-458.