16th IFAC Symposium on System Identification
The International Federation of Automatic Control
Brussels, Belgium. July 11-13, 2012

Parameter Estimation for Time Varying Dynamical Systems using Least Squares Support Vector Machines ⋆

Siamak Mehrkanoon * **Tillmann Falck** * **Johan A. K. Suykens** *

* *Katholieke Universiteit Leuven - ESAT - SCD/SISTA, Kasteelpark Arenberg 10, B-3001 Leuven (Heverlee), Belgium; e-mail: {siamak.mehrkanoon, tillmann.falck, johan.suykens}@esat.kuleuven.be*

Abstract: This paper develops a new approach based on Least Squares Support Vector Machines (LS-SVMs) for parameter estimation of time invariant as well as time varying dynamical SISO systems. Closed-form approximate models for the state and its derivative are first derived from the observed data by means of LS-SVMs. The time-derivative information is then substituted into the system of ODEs, converting the parameter estimation problem into an algebraic optimization problem. In the case of time invariant systems one can use least-squares to solve the obtained system of algebraic equations. The estimation of time-varying coefficients in SISO models, is obtained by assuming an LS-SVM model for it.

Keywords: Parameter estimation; deterministic dynamic models; Least squares support vector machines; time-varying parameters.

1. INTRODUCTION

Parameter estimation is widely used in modelling of dynamic processes in physics, engineering and biology. Various methods have been previously investigated in the literature for handling this problem. Mainly they fall into two categories. In the first category the approaches are based on a classical parameter estimator, usually the least square estimator [Biegler et al., 1986], or the maximum likelihood (MLE). First the dynamical system are simulated using initial guesses for parameters (if the initial condition is unavailable they will be appended to the parameters of the model). Then model predictions are compared with measured data and an optimization algorithm updates the parameters. The process of updating the parameters continues until no significant improvement in the objective function is observed. These approaches require numerical integration of differential equations for each update of the parameters. Therefore there is a large amount of computational work involved. Studies show that more than 90% of the computation time is consumed in the ODE solver during the identification process [Moles et al., 2003].

The second category includes methods, originally proposed by Varah [1982], that do not require repeated numerical integration and are referred to as two-step approaches. In [Varah, 1982] first a cubic spline is used to estimate the system dynamics from observational data. The predicted model then can be differentiated with respect to time to obtain the estimate of the derivative of the solution. 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.

Recently Dua [2011] proposed a method where an artificial neural network model is used to find numerical values for parameters of a time invariant dynamical systems. It is the purpose of this paper to introduce an approach based on least squares support vector machines for estimation of time invariant as well as time varying systems in state-space and transfer function form.

Throughout this paper, we assume that the dynamical system is uniquely solvable and that the parameters of the model are identifiable.

The paper is organized as follows: In section 2, the problem formulation for parameter estimation of time invariant systems is given. Section 3 describes an approach for time varying parameter estimation in dynamical systems described by ordinary differential equations. In section 4, examples are given in order to illustrate the effectiveness of the proposed method. Finally conclusions are discussed in section 5.

⋆ 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) 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. Johan Suykens is a professor at the KU Leuven, Belgium.

978-3-902823-06-9/12/$20.00 © 2012 IFAC

10.3182/20120711-3-BE-2027.00044

2. PARAMETER ESTIMATION FOR TIME INVARIANT NONLINEAR DYNAMICAL SYSTEMS

2.1 Problem Statement

Suppose that we are given a dynamical system in state-space form

$$\frac{dX}{dt} = F(t, X, \theta),\ X(0) = X_0, \tag{1}$$

subject to certain boundary or initial conditions which may be imposed on the basis of observation data. t denotes the independent variable (usually time). X is the state vector of the system where $\frac{d}{dt}\hat{X} = [\frac{d}{dt}\hat{x}_1, ..., \frac{d}{dt}\hat{x}_m]^T$, $X = [x_1, ..., x_m]^T$ and $F = [f_1, ..., f_m]^T$. $\theta = [\theta_1, ..., \theta_p]$ are unknown parameters of the system and X_0 are initial values.

In order to estimate the unknown parameters θ, the state variable $X(t)$ is observed at N time instants $\{t_1, ..., t_N\}$, so that we have

$$Y(t_i) = X(t_i) + E_i,\ i = 1, ..., N,$$

where $\{E_i\}_{i=1}^N$ are independent measurement errors with zero mean. The objective is to determine appropriate parameter values so that errors between the outputs of the estimated model and the measured data are minimized.

2.2 General Methodology

First we approximate the trajectory $\hat{X}(t) = [\hat{x}_1, ..., \hat{x}_m]^T$ on the basis of observations at N points $\{t_i, Y(t_i)\}_{i=1}^N$. Note that $Y(t_i)$ are the experimentally observed values of the state variables at time instant t_i, i.e. $Y(t_i) = [y_1(t_i), ..., y_m(t_i)]^T$. Then the estimation of the state derivative is obtained by differentiating the model with respect to time. In this paper we model the state x_k for $k = 1, ..., m$ as a Least-Squares Support Vector Machine [Suykens et al., 2002, 2001]. Therefore the goal is to find a model of the form $\hat{x}_k(t) = w_k^T\varphi(t) + b_k$. For the k-th state variable we formulate the following convex primal LS-SVM problem [Suykens et al., 2010],

$$\begin{aligned} \underset{w_k, b_k, e_k}{\text{minimize}} \quad & \frac{1}{2}w_k^T w_k + \frac{\gamma_k}{2}\|e_k\|_2^2 \\ \text{subject to} \quad & y_k(t_i) = w_k^T\varphi(t_i) + b_k + e_k^i, \quad i = 1, ..., N, \end{aligned}$$

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

$$\left[\begin{array}{c|c} \Omega + \gamma^{-1}I_N & 1_N \\ \hline 1_N^T & 0 \end{array}\right] \begin{bmatrix} \alpha^k \\ b_k \end{bmatrix} = \begin{bmatrix} y^k \\ 0 \end{bmatrix} \tag{2}$$

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. $1_N = [1; ...; 1] \in \mathbb{R}^N$, $\alpha^k = [\alpha_1^k; ...; \alpha_N^k]$, $y^k = [y_k(t_1); ...; y_k(t_N)]$ and I_N is the identity matrix. The model in dual form becomes:

$$\hat{x}_k(t) = w_k^T\varphi(t) + b_k = \sum_{i=1}^N \alpha_i^k K(t_i, t) + b_k \tag{3}$$

where K is the kernel function. Differentiating (3) with respect to t, one can obtain an analytical approximate expression for the derivative of the model

$$\frac{d}{dt}\hat{x}_k(t) = w_k^T\dot{\varphi}(t) = \sum_{i=1}^N \alpha_i^k \varphi(t_i)^T\dot{\varphi}(t). \tag{4}$$

Making use of Mercer's Theorem [Vapnik, 1998], derivatives of the feature map can be written in terms of derivatives of the kernel function [Lázaro et al., 2005]. Therefore $\varphi(t)^T\dot{\varphi}(s)$ is given by the derivative of $K(t, s)$ with respect to s. If we denote $K_s(t, s) = \frac{\partial K(t,s)}{\partial s}$, then equation (4) can be written as

$$\frac{d}{dt}\hat{x}_k(t) = w_k^T\dot{\varphi}(t) = \sum_{i=1}^N \alpha_i^k K_s(t_i, t). \tag{5}$$

Eqs. (3) and (5) are closed-form approximations for the k-th state in equation (1) and its derivative respectively. By applying the above procedure for all the state variables one can obtain the LS-SVM expression for $\hat{X} = [\hat{x}_1, ..., \hat{x}_m]^T$ and $\frac{d}{dt}\hat{X} = [\frac{d}{dt}\hat{x}_1, ..., \frac{d}{dt}\hat{x}_m]^T$. Therefore the values of the solution and time-derivative curves at some set of sample points $\{t_k\}_{k=1}^M$, which are not necessarily the same as the original points where the states are observed, can be obtained by evaluating the LS-SVM expressions for $\hat{X}$ and $\frac{d}{dt}\hat{X}$. These numerical values then are substituted into the system description (1), so that the unknown parameters appear in an algebraic expression, resulting in linear (if the system is linear in the parameters) or nonlinear (otherwise) least-squares estimation. Therefore the estimation of time invariant parameters is obtained by solving the following optimization problem

$$\begin{aligned} \underset{\boldsymbol{\theta}}{\text{minimize}} \quad & \frac{1}{2}\sum_i \|\Xi_i\|_2^2 \\ \text{subject to} \quad & \Xi_i = \frac{d}{dt}\hat{X}(t_i) - F(t_i, \hat{X}(t_i), \theta), \quad i = 1, ..., M. \end{aligned} \tag{6}$$

When the ODE model is linear in the parameters this problem is a convex optimization problem . As it has been remarked in [Varah, 1982], what we really are interested in to minimize is the error between the observed and model predicted values of the state variables i.e. the integrated residual errors

$$R_I(\theta_{est}) = \sum_{k=1}^M \left\| X(t_k) - \tilde{X}(t_k) \right\|_2^2 \tag{7}$$

where $\tilde{X}(t_k)$ is obtained by simulating the system with the estimated parameter θ_{est}.

3. TIME VARYING PARAMETER ESTIMATION

Consider a first order dynamical system of the form:

$$\frac{dx}{dt} + \theta(t)f(x(t)) = g(t),\ x(0) = x_0 \tag{8}$$

subject to certain initial conditions which may be imposed on the basis of observation data. The technique can be extended to higher order systems. f is an arbitrary known function and $\theta(t)$ is the time varying parameter of the system and is considered to be unknown. The state $x(t)$ has been measured at certain time instants $\{t_i\}_{i=1}^N$, which can be non-equidistant, i.e.

$$y_i = x(t_i) + \xi_i,\ i = 1, ..., N$$

where ξ_i's are i.i.d. random errors with zero mean and constant variance. $g(t)$ is the input signal whose values are known at data points $\{t_i\}_{i=1}^N$ i.e. $g_i = g(t_i)$ for $i = 1, ..., N$. The problem is to estimate the function $\theta(t)$ so that the solution of (8) with the estimated parameter $\theta(t)$ is as close as possible to the given data.

First we approximate functions $\hat{x}(t)$ and $\hat{g}(t)$ on the basis of observations at N points $\{t_i, y_i\}_{i=1}^N$, $\{t_i, g_i\}_{i=1}^N$ by means of least squares support vector regression. Therefore the convex primal LS-SVM model for $\hat{x}(t)$ can be written as follows,

$$\begin{aligned} &\underset{w,b,e}{\text{minimize}} \quad \frac{1}{2}w^T w + \frac{\gamma}{2}e^T e \\ &\text{subject to} \quad y_i = w^T\varphi(t_i) + b + e_i, \quad i = 1, ..., N \end{aligned} \tag{9}$$

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.

Lemma 1. Given a positive definite kernel function $K : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ with $K(t, s) = \varphi(t)^T\varphi(s)$ and a regularization constant $\gamma_1 \in \mathbb{R}^+$, the solution to (9) is given by the following dual problem

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

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. $1_N = [1; ...; 1] \in \mathbb{R}^N$, $\alpha = [\alpha_1; ...; \alpha_N]$, $y = [y_1; ...; y_N]$, and I_N is the identity matrix. The model in dual form becomes

$$\hat{x}(t) = w^T\varphi(t) + b = \sum_{i=1}^N \alpha_i K(t_i, t) + b \tag{10}$$

where K is the kernel function. The same procedure can be applied to obtain the LS-SVM approximation of the excitation $\hat{g}(t)$.

Note that the analytic LS-SVM expression for the state trajectory allows us to obtain a closed-form approximation for its derivative by differentiating (10) with respect to t,

$$\frac{d}{dt}\hat{x}(t) = w^T\dot{\varphi}(t) = \sum_{i=1}^N \alpha_i\varphi(t_i)^T\dot{\varphi}(t) = \sum_{i=1}^N \alpha_i K_s(t_i, t). \tag{11}$$

Here $K_s(t, s)$ is defined as previously. Eqs. (10) and (11) are approximations for the solution of the differential equation (8) and its derivative respectively. Therefore the derivative of the solution at some set of sample points $\{t_k\}_{k=1}^M$ can be obtained from (11). These time-derivative information together with values of the state variable at points $\{t_k\}_{k=1}^M$ are then substituted into the model description (8). But since the parameter present in (8) is time-varying, it can not be estimated by Eq. (6). Therefore let us assume an explicit LS-SVM model

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

as an approximation for the parameter $\theta(t)$. Having available the state and its derivative at $\{t_k\}_{k=1}^M$ points, we can estimate the time-varying coefficient $\theta(t)$ by solving the following optimization problem.

$$\begin{aligned} &\underset{v,b_\theta,e}{\text{minimize}} \quad \frac{1}{2}v^T v + \frac{\gamma}{2}\sum_{i=1}^M e_i^2 \\ &\text{subject to} \quad \frac{d}{dt}\hat{x}(t_i) + \left[v^T\psi(t_i) + b_\theta\right] f(\hat{x}(t_i)) = \\ &\qquad\qquad \hat{g}(t_i) + e_i, \text{ for } i = 1, ..., M. \end{aligned} \tag{12}$$

Lemma 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 (12) is given by the following dual problem

$$\left[\begin{array}{c|c} D\Omega D + \gamma^{-1}I_M & f(\hat{x}) \\ \hline f(\hat{x})^T & 0 \end{array}\right] \begin{bmatrix} \alpha \\ b_\theta \end{bmatrix} = \begin{bmatrix} \hat{g} - \frac{d\hat{x}}{dt} \\ 0 \end{bmatrix} \tag{13}$$

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

Proof. The Lagrangian of the constrained optimization problem (12) becomes

$$\mathscr{L}(v, b_\theta, e_i, \alpha_i) = \frac{1}{2}v^T v + \frac{\gamma}{2}\sum_{i=1}^M e_i^2 - \sum_{i=1}^M \alpha_i\left[\frac{d}{dt}\hat{x}_i + \left(v^T\psi(t_i) + b_\theta\right)f(\hat{x}_i) - \hat{g}_i - e_i\right]$$

where $\{\alpha_i\}_{i=1}^M$ are Lagrange multipliers. $\hat{g}_i = \hat{g}(t_i)$, $f(\hat{x}_i) = f(\hat{x}(t_i))$ and $\frac{d}{dt}\hat{x}_i = \frac{d}{dt}\hat{x}(t_i)$ for $i = 1, ..., M$. Then the Karush-Kuhn-Tucker (KKT) optimality conditions are as follows,

$$\begin{aligned} &\frac{\partial\mathscr{L}}{\partial v} = 0 \to v = \sum_{i=1}^M \alpha_i f(\hat{x}_i)\psi(t_i), \\ &\frac{\partial\mathscr{L}}{\partial b_\theta} = 0 \to \sum_{i=1}^M \alpha_i f(\hat{x}_i) = 0, \\ &\frac{\partial\mathscr{L}}{\partial e_i} = 0 \to e_i = -\frac{\alpha_i}{\gamma}, \quad i = 1, ..., M, \\ &\frac{\partial\mathscr{L}}{\partial \alpha_i} = 0 \to \left(v^T\psi(t_i) + b_\theta\right)f(\hat{x}_i) - e_i = \hat{g}_i - \frac{d}{dt}\hat{x}_i, \\ &\qquad\qquad \text{for } i = 1, ..., M. \end{aligned}$$

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

$$\begin{cases} \hat{g}_i - \frac{d}{dt}\hat{x}_i = \sum_{j=1}^M \alpha_j f(\hat{x}_j)\Omega_{ji}f(\hat{x}_i) + \frac{\alpha_i}{\gamma} + \\ \qquad b_\theta f(\hat{x}_i), \ i = 1, ..., M \\ 0 = \sum_{i=1}^M \alpha_i f(\hat{x}_i) \end{cases}$$

and writing these equations in matrix form gives the linear system in (13).

The model in the dual form becomes

$$\hat{\theta}(t) = v^T\psi(t) + b_\theta = \sum_{i=1}^M \alpha_i f(\hat{x}_i)\tilde{K}(t_i, t) + b_\theta \tag{14}$$

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

4. EXPERIMENTS

To illustrate the applicability of the proposed method, we list the computed results of the parameter estimation for three systems with time invariant coefficients and two first order systems with time varying parameter. For all the experiments, the RBF kernel is used, i.e. $K(x, y) = \exp(-\frac{(x-y)^2}{\sigma^2})$. The procedure is outlined in Algorithm 1.

Algorithm 1 Approximating the model's time varying parameter

1: Estimate the trajectories $\hat{X}$ from the observational data by using LS-SVM model, Eq. (2).
2: Differentiate the predicted model with respect to time to get an approximate model for the derivative of the state, Eq. (5).
3: Evaluate the state and its derivative model at time instants $\{t_i\}_{i=1}^{M}$.
4: **if** parameters are time invariant **then**
5: solve optimization problem (6)
6: **else**
7: solve Eq. (13) to get the estimate of the time varying parameter of the dynamical system.
8: **end if**

Example 1. Consider the nonlinear Bellman's problem originated from a chemical reaction [Bellman et al., 1967]

$$\frac{dx}{dt} = \theta_1(126.2 - x)(91.9 - x)^2 - \theta_2 x^2, x(1) = 0. \quad (15)$$

The observations of the state x with one decimal place accuracy are given in Table 1.

Table 1. Observations of state x for Bellman's problem (14) [Varah, 1982].

t	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0
x	0.0	1.4	6.3	10.4	14.2	17.6	21.4	23.0
t	10	12	15	20	25	30	40	
x	27.0	30.4	34.4	38.8	41.6	43.5	45.3	

Cross-validation is used to tune the regularization constant γ and kernel bandwidth σ, on a meaningful grid of possible (γ, σ) combinations. The estimated parameter values obtained by averaging over 50 simulation runs and the corresponding integrated residual R_I are as follows

$$[\hat{\theta}_1, \hat{\theta}_2, R_I] = [0.45 \times 10^{-6}, 0.28 \times 10^{-3}, 1.45]$$

which agree well with the true solution $[\theta_1, \theta_2] = [0.45 \times 10^{-6}, 0.27 \times 10^{-3}]$. The standard deviation of our approach for the parameters θ_1 and θ_2 are 8.92×10^{-8} and 1.30×10^{-5} respectively. It should be noted that in the described approach in [Varah, 1982] the spline knots have been chosen interactively. Whereas in our proposed method one does not need to work with the knots and instead the regularization constant γ is chosen automatically to avoid overfitting. Therefore in contrast with the approach of [Varah, 1982] in our proposed method less human effort is needed.

Example 2. Consider Barne's problem which is based on the Lotka-Voltra differential equations consisting of two ordinary differential equations with three parameters θ_1, θ_2 and θ_3 [Varah, 1982]

$$\frac{dx_1}{dt} = \theta_1 x_1 - \theta_2 x_1 x_2,\ x_1(0) = x_{10}$$
$$\frac{dx_2}{dt} = \theta_2 x_1 x_2 - \theta_3 x_2,\ x_2(0) = x_{20}.$$

The observed data values as given by [Varah, 1982] are reported in Table 2.

Table 2. Observations of the states x_1 and x_2 for Barne's problem [Varah, 1982].

t	0.0	0.5	1.0	1.5	2.0	2.5	3.0	3.5
$\mathbf{x_1}$	1.0	1.1	1.3	1.1	0.9	0.7	0.5	0.6
$\mathbf{x_2}$	0.3	0.35	0.4	0.5	0.5	0.4	0.3	0.25
t	4.0	4.5	5.0					
$\mathbf{x_1}$	0.7	0.8	1.0					
$\mathbf{x_2}$	0.25	0.3	0.35					

The estimated parameter values are obtained by taking the average over 50 simulation runs. Table 3, shows the values of the parameters reported in [Varah, 1982], [Shiang, 2009], Matlab *diffpar* [Edsberg et al., 1995] toolbox and the computed results obtained by the proposed method in this paper. It can be seen that they all are in good agreement.

Table 3. Estimated parameters of Barne's problem.

Method	$\widehat{\theta_1}$	$\widehat{\theta_2}$	$\widehat{\theta_3}$	$\widehat{x_{10}}$	$\widehat{x_{20}}$
[Varah, 1982]	0.85	2.13	1.91	1.02	0.25
[Edsberg et al., 1995]	0.81	2.29	2.00	0.99	0.21
[Shiang, 2009]	0.98	1.95	1.69	0.96	0.29
LS-SVM approach	0.84	2.14	1.96	0.99	0.29

The standard deviation of our approach for the parameters θ_1, θ_2 and θ_3 are 6.78×10^{-2}, 1.83×10^{-1} and 1.69×10^{-1} respectively. In our approach the $R_I(\theta_{est}) = 0.11$ which is also less than that (i.e. $R_I^2 = 0.35$) reported in [Varah, 1982].

Example 3. Consider the Lorenz equation [Lorenz, 1962] which form a system of three differential equations that are important in climate and weather predictions. It is well known that the Lorenz equation is an example of a nonlinear and chaotic system

$$\frac{dx_1}{dt} = a(x_2 - x_1)$$
$$\frac{dx_2}{dt} = x_1(b - x_3) - x_2$$
$$\frac{dx_3}{dt} = x_1 x_2 - c x_3$$

where a, b and c are the unknown parameters within the system. The initial condition at $t = 0$ is taken to be $(x_1(0), x_2(0), x_3(0)) = (-9.42, -9.34, 28.3)$. The correct parameters we are trying to reconstruct are $a = 10$, $b = 28$ and $c = 8/3$. The solution of the Lorenz system is prepared by numerically integrating the Lorenz equations using MATLAB built-in solver ode45, on domain [0,3] 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 level of noise (standard deviation of the noise) is denoted by η. In this problem η is considered to be $\eta = 0.0, 0.2$ and 0.5. The observation points are prepared within the domain of [0, 3] at every $\Delta t = 0.05$. After obtaining the closed-form approximation for the states x_1, x_2 and x_3 by means of LS-SVM, we used 301 equally spaced sample points in the interval [0, 3] to solve optimization problem (6). Table 4 reports the estimated parameters of the Lorenz system by averaging over 50 simulation runs.

Table 4. The values of parameters estimated of Lorenz model. Parameter η is the standard deviation of the noise.

η	Estimated parameters					
	a		b		c	
	a_{est}	$\|e_a\|$	b_{est}	$\|e_b\|$	c_{est}	$\|e_c\|$
0.0	9.99	0.0014	28.00	0.0062	2.67	0.0041
0.2	9.60	0.3919	28.03	0.0352	2.67	0.0042
0.5	9.34	0.6532	27.86	0.112	2.68	0.0147

The true values of model parameters a, b and c are 10.0, 28.0 and 8/3 respectively. Absolute errors are denoted by $|e_{\cdot}|$.

The average and standard deviation of our results after 50 simulations are depicted in Fig 1.

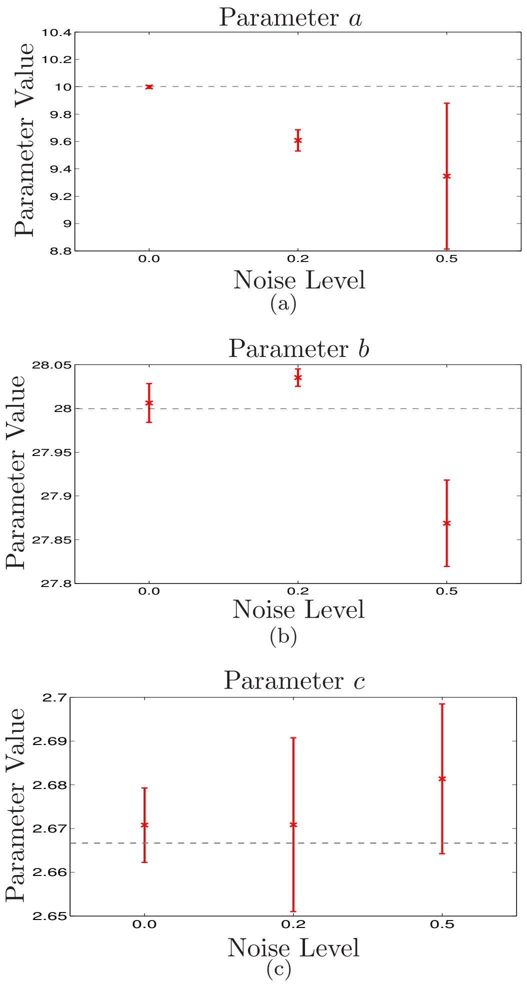

Fig. 1. Estimation of Lorenz parameters a, b and c from data with observational noise. The true value is indicated by the dashed line.

Example 4. Consider the following linear time varying system,

$$\frac{dx}{dt} - \frac{\sin(t)}{t+1} x(t) = \cos(t), \quad x(0) = 1. \tag{16}$$

The aim is to estimate the time varying coefficient $\frac{\sin(t)}{t+1}$ from measured data. For collecting the data, the solution of this equation has been obtained using Matlab built-in solver ode45, with the relative error tolerance RelTol= 10^{-6}, over the domain of [0, 20]. Thereafter the process is observed at N discrete time instants. Then we have artificially introduced random noise (Gaussian white noise with noise level η) to the true solution in order to create observational data. After obtaining the LS-SVM closed form approximation for the state x, we used $M = 200$ equally spaced sample points in the interval [0, 20] to solve optimization problem (12). The mean square error (MSE) for the test set (500 sample points in interval [0,20]) are tabulated in Table 5. Fig 2, shows the influence of noise level on the parameter estimation. 10-fold cross-validation is used for the model selection by choosing one of several models that has the smallest estimated generalization error.

Table 5. The influence of noise level and number of observed data on the parameter estimation. Parameter η is the standard deviation of the noise and N is the number of observed data.

N	η	MSE
100	0.0	2.02×10^{-8}
	0.1	3.9×10^{-4}
200	0.0	1.09×10^{-8}
	0.1	3.4×10^{-4}

Example 5. Consider the following nonlinear and time varying dynamical system,

$$\frac{dx}{dt} - \frac{\cos(t)}{\sin(t)+2} \cos(x(t)^2) = \cos(t), \quad x(0) = 1. \tag{17}$$

In order to estimate the time varying coefficient $\frac{\cos(t)}{\sin(t)+2}$, we generate the solution to (17) with the true parameter. Then random noise (Gaussian white noise with noise level η) is added to the true solution in order to create observational data. The mean square error (MSE) for the test set (500 sample points in interval [0,20]) is tabulated in Table 6. Fig 3, shows the influence of noise level on the parameter estimation. The kernel bandwidth σ and regularization constant γ are tuned by 10-fold cross validation, while the preference is given to less complex models.

Remark 3. In the case of nonlinear time varying dynamical systems, the result is sensitive to the choice of kernel parameter and regularization constant. Therefore the user should supply a suitable range for σ before applying cross validation.

Table 6. The influence of noise level and number of observed data on the parameter estimates. Parameter η is the standard deviation of the noise and N is the number of observed data.

N	η	MSE
100	0.0	8.34×10^{-5}
	0.05	3.5×10^{-3}
200	0.0	3.06×10^{-6}
	0.05	2.0×10^{-3}

5. CONCLUSIONS

In this paper a new approach based on LS-SVMs is proposed for parameter estimation of dynamical systems.

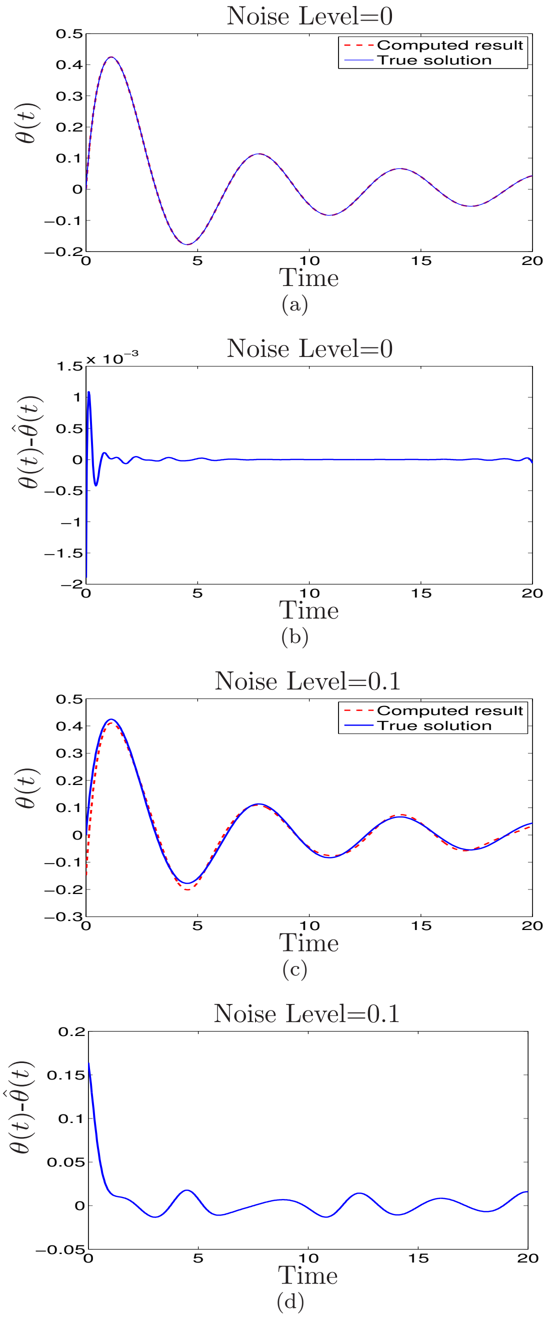

Fig. 2. Estimation of time varying parameter of dynamical system formulated in Example 4 using $M = 200$ sample points. 500 sample points in the interval $[0, 20]$ are used for the test set.

The method is applicable for both time invariant and time varying dynamics.

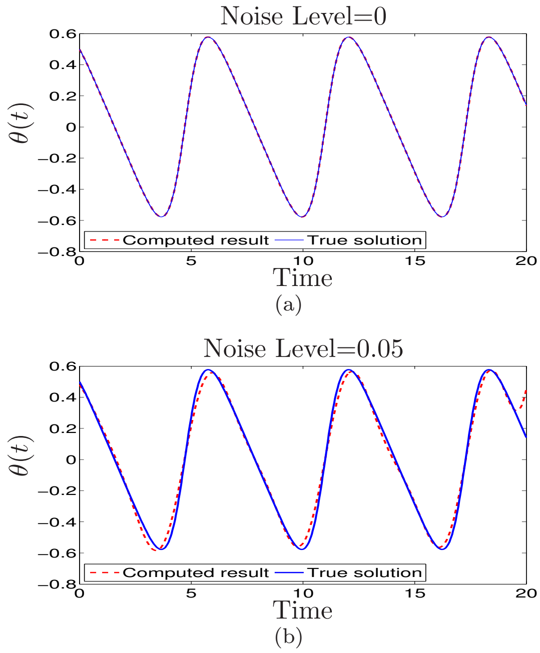

Fig. 3. Estimation of time varying parameter of dynamical system formulated in Example 5 using $M = 200$ sample points. 500 sample points in the interval $[0, 20]$ are used for the test set.

REFERENCES

R. Bellman, J. Jacquez, R. Kalaba and S. Schwimmer. Quasilinearization and the estimation of chemical rate constants from raw kinetic data. *Math. Biosci*, 1:71–76, 1967.

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

V. Dua. An Artificial Neural Network approximation based decomposition approach for parameter estimation of system of ordinary differential equations. *Computers and Chemical Engineering*, 35:545–555, 2011.

L. Edsberg, G. Wikström. Toolbox for parameter estimation and simulation in: dynamic systems with applications to chemical kinetics, *in: Proceedings of the Nordic MATLAB Conference*, COMSOL, Stockholm, 1995.

M. Lázaro, I. Santamaría, F. Pérez-Cruz, A. Artés-Rodriguez. Support vector regression for the simultaneous learning of a multivariate function and its derivative. *Neurocomputing*, 69(1-3):42–61, 2005.

E. Lorenz. Deterministic nonperiodic flow. *Journal of Atmospheric Sciences*, 20:130–141, 1963.

C. G. Moles, P. Mendes and J. R. Banga Parameter Estimation in Biochemical Pathways: A Comparison of Global Optimization Methods. *Genome Research*, 13: 2467–2474, 2003.

K. D. Shiang. A perturbation-based estimate algorithm for parameters of coupled ordinary differential equations, applications from chemical reactions to metabolic dynamics. *Computer Methods and Programs in Biomedicine*, 94:118–142, 2009.

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

J.A.K. Suykens, J. Vandewalle, B. De Moor. Optimal control by least squares support vector machines, *Neural Networks*, 14(1):23–35, 2001.

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

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

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