On the embedding parameters in kernel identification problem of nonlinear dynamical systems

To cite this article: N R Antropov and E D Agafonov 2020 IOP Conf. Ser.: Mater. Sci. Eng. 734 012143

View the article online for updates and enhancements.
On the embedding parameters in kernel identification problem of nonlinear dynamical systems

N R Antropov and E D Agafonov
Reshetnev Siberian State University of Science and Technology, 31 Krasnoyarsky Rabochy Av., Krasnoyarsk, 660037, Russia
E-mail: nikita.antropov.92@mail.ru

Abstract. Dynamical models are of fundamental importance in many problems such as simulation, optimization and prediction. In identification problem of dynamical systems an input vector is typically considered as delayed vector of previous outputs. Embedding lag and embedding dimension should be chosen correctly. Identification of nonlinear systems is generally performed using kernel-based methods since they do not require any additional information about system structure. A common question in identification of dynamical system is sensitivity of kernel-based models to selected embedding lag and embedding dimension. The paper presents simulations of the kernel least mean squares algorithm on one-step prediction problem for various values of embedding lag and embedding dimension. It is shown that optimal embedding lag and embedding dimension is dependent on model parameters, while usage of optimal model parameters decreases the value of optimal embedding dimension.

1. Introduction
Solution of wide range of optimization, forecasting and control problems is associated with mathematical models. There is a well-established methodology for building models of linear systems, whereas for nonlinear systems such methodology is currently only in its infancy. One of the key issues in identification of nonlinear dynamical systems is determination of an appropriate embedding lag and embedding dimension.

Despite the fact, that there are significant amount of papers in the field of identification of nonlinear dynamical systems, there are actually no papers on the analysis of sensitivity of kernel-based models to optimal and non-optimal values of embedding parameters. In this paper, we propose an analysis of methods for identification of nonlinear dynamical systems and a series of computational experiments illustrating sensitivity of kernel-based models to various configurations of model parameters, including embedding lags and embedding dimensions.

2. Identification of nonlinear dynamical systems
2.1. Identification problem statement
In general, identification of nonlinear dynamical systems can be formalized in the framework of the following problem statement [1]. There is a nonlinear dynamical system influenced by random influences with zero mean and limited variance. There are observable and unobservable random inputs. Outputs of the system are synchronically measured together with observable inputs. Input and
variables and output variables are measured at discrete time instants with some periodicity. There is a random noise in measurement channels. Structure and order of the operator describing the system are unknown. The problem is to build a model of the nonlinear dynamical system that allows predicting the vector of output variables providing minimum identification error. The scheme of the proposed identification problem is presented in figure 1.

**Figure 1.** Identification setup of nonlinear dynamical system. Notation: $\xi(t)$ is vector of uncontrolled random influences; $u(t)$ is controlled input vector; $\omega(t)$ is uncontrolled input vector; $x(t)$ is output vector; $\eta(t)$, $\mu(t)$ are vectors of random noise in measurements channels; $u[t], x[t]$ are vectors of measured controlled input and output; $\hat{x}[t]$ is estimation of model output; $\theta[t]$ is vector of model parameters; $e[t]$ is vector of identification errors.

The identification problem involves seeking a function, connecting input and output variables in the form:

$$\hat{x}[t] = F(u[t], \theta[t]),$$

where $\hat{x}[t]$, $u[t]$, $\theta[t]$ are described above, $F(\cdot)$ is unknown function.

Nowadays there are many methods and algorithms for identification of nonlinear dynamical systems. Next subsection presents a review of basic approaches.

2.2. Methods of identification of nonlinear systems

Currently there are significant numbers of papers devoted to the problems of building models of nonlinear dynamical systems. One of the first approaches was probably concluded by Volterra [2]. This approach involves building of a model using the sum of convolution integrals. There are several disadvantages of this approach, such as uncertainty in the choice of the number of used time series
members and the need for additional experiments required to determine the response of the system to various test input influences, which requires a number of computational resources.

Subsequent research in this direction is devoted to the construction of the so-called Wiener [3] and Hammerstein [4] models, which consist of linear and nonlinear dynamical elements of the system. This approach allows to increase model completeness. Among the shortcomings of these models, it is necessary to single out their correspondence to the class of parametric ones, which does not allow solving the problem of uncertainty in choosing the structure of the model with subsequent estimation of the model parameters.

Solution of the uncertainty problem in determining the structure of the model of nonlinear dynamical systems became possible with the implementation of adaptive filtering methods [5] and kernel methods [6, 7]. At the same time, the key issue of kernel methods is not the choice of the structure of the model, but the estimation the model parameters of the methods providing acceptable modeling accuracy. Considered problem of the model parameters estimation, as well as the problem of the high computational complexity of kernel methods became obstacle to their practical application.

One approach to solving the problem of high computational complexity involves developing adaptive kernel filtering methods [8, 9], in which the model is constructed iteratively, using information at each step from the previous one. At the same time, the solution of this problem is facilitated by the intensive evolution of the idea of constructing compact and representative training samples, which allow a compromise between accuracy and computational complexity. These ideas were applied in dictionary learning methods that are currently widely for solving the problem of increasing computational efficiency of kernel identification methods [10].

Another approach to increasing the computational efficiency of kernel methods involves applying iterative optimization methods to estimation of parameters of kernel methods. In particular, using the stochastic gradient method in [11], a kernel least-squares algorithm was proposed. It provides higher computational efficiency but has lower accuracy and convergence rate. Later, in [12], a modified algorithm was proposed, which provides higher accuracy and convergence rate in comparison with the original algorithm. Increasing of computational efficiency of kernel methods is one of the main areas of research nowadays [13].

It is worth noting that most existing nonlinear dynamical systems operate under of non-stationary caused by time-varying effects associated with switching system modes, influence of uncontrolled influences, changing operating conditions and functioning parameters of the system, and so on [14]. In this regard, at present, one of the most important tasks of improving kernel methods for identification of nonlinear dynamical systems is development of approaches for tracking the change and drift of the parameters of nonlinear dynamical systems operating under conditions of non-stationarity.

Probably, the first algorithm for solving of the problem mentioned above was proposed in [15]. The main idea of this algorithm is to add the so-called forgetting parameter to the model, which determines the weight of each measurement in the training sample. The parameter is defined in such a way that the most relevant measurements have the greatest weight. As new measurements become available, the weight of previous measurements in the training set decreases. Despite the theoretical validity of this algorithm, it has a number of disadvantages limiting its practical application. Thus, the solutions obtained by using this algorithm within accumulation of measurements become more and more unstable leading to a decrease in the accuracy of identification.

This disadvantage of the algorithm [15] was eliminated in a new algorithm proposed in [16]. The paper gives a theoretical and experimental study of the algorithm based on the application of the probabilistic approach in the framework of the theory of Gaussian processes [17]. This algorithm contributed to the development of probabilistic approaches in the identification methods of nonlinear dynamical systems under conditions of non-stationarity. In particular, the probabilistic approach made possible to develop a modification of the kernel least-squares method for non-stationary systems [18] allows increasing the adaptability of the original algorithm. At present, the integration of probabilistic and kernel methods is one of the main areas of research in the field of identification methods for nonlinear dynamical systems operating under non-stationary conditions.
2.3. Embedding parameters estimation

One key feature of any dynamical system is dependence of the current state of the system on a certain set of its previous states. In this regard, one of the main parameters of discrete models of dynamical systems is the embedding lag and embedding dimension. Embedding dimension determines the number of previous states of a dynamical system necessary to predict its future state. The embedding lag and embedding dimension can be specified using the following representation of the vector of input variables:

\[ \mathbf{u}[t] = (\mathbf{u}[t-\tau], \mathbf{u}[t-2\tau], \ldots, \mathbf{u}[t-(d-1)\tau], \mathbf{x}[t-\tau], \mathbf{x}[t-2\tau], \ldots, \mathbf{x}[t-(d-1)\tau]) \]

where \( \tau \) is embedding lag, \( d \) is embedding dimension. In practice, these parameters should be different for input and output.

There are a number of algorithms and criterions for estimation of embedding lag like autocorrelation, mutual information and approximate period approach [19]. Embedding dimension can be estimated by computing invariants of the system, singular value decomposition method and false nearest neighbours algorithm [20].

Optimal values of the embedding lag and embedding dimension can change over time due to nonstationarity of the system caused by switching of operating modes of the system, influence of uncontrolled influences, as well as changes in operating conditions and functioning parameters of the system. Non-optimal values of the embedding lag and embedding dimension in model decreasing accuracy of identification, and, as a result, decreasing efficiency of prediction and control of the system. Thus, estimation of the optimal values of the embedding lag and embedding dimension is a relevant problem.

In the framework of this problem, one issue is the effect of the embedding lag and embedding dimension on the quality of the resulting models. It is worth noting that today in scientific literature there are actually no papers analyzing of influence of these parameters on the quality of the kernel-based models. In this regard, in the present work, it is proposed to carry out a series of computational experiments to evaluate the quality of kernel-based models of nonlinear dynamical systems for various configurations of model parameters, the embedding lags and embedding dimensions.

3. Simulations

Consider a Mackey-Glass time series, generated by the following equation [21]:

\[ \frac{dx(t)}{dt} = -bx(t) + \frac{ax(t-\Delta)}{1 + x_{10}^{10}(t-\Delta)}, \]

where \( a = 0.1, b = 0.2, \Delta = 30. \)

Simulations are performed on one-step prediction problem for time series, generated from nonlinear dynamical system mentioned above. Simulation is similar to one presented in [1]. Sample size \( N = 1250. \) First 1000 observations are used for training, whereas last 250 observations are used for testing. KLMS algorithm [11] with radial basis kernel function [8] is used for identification. Prediction error is calculated in terms of averaged normalized mean-squared error:

\[ \text{nMSE} = \log_{10} \left[ \frac{1}{N} \sum_{t=1}^{N} (x[t] - \hat{x}[t])^2 \right]. \]

Effect of the embedding lag and embedding dimension on the nMSE is shown on figures, illustrating contour plots of the nMSE for various combinations of the embedding lags and embedding dimensions. Figure 1 depicts nMSE for non-optimal configuration of model parameters, namely \( \sigma = 1 \) and KLMS step parameter \( \eta = 0.5 \), while nMSE on figure 2 is calculated for optimal kernel parameter \( \sigma \) and step parameter \( \eta \). The darker the area, the lower nMSE.
Figure 2. Contour plot of calculated steady-state nMSE for one-step prediction problem of Mackey-Glass time series generated from the nonlinear dynamical system for various combinations of the embedding lag $\tau$ and embedding dimension $d$. Identification is performed using KLMS algorithm with kernel parameter $\sigma = 1$ and KLMS step parameter $\eta = 0.5$.

Figure 3. Contour plot of calculated steady-state nMSE for one-step prediction problem of Mackey-Glass time series generated from the nonlinear dynamical system for various combinations of the embedding lag $\tau$ and embedding dimension $d$. Identification is performed using KLMS algorithm with optimal kernel parameter $\sigma$ and step parameter $\eta$.

On both figure 1 and figure 2 one can observe the presence of two regions with small nMSE. Figure 2 shows that for non-optimal model parameters configuration the minimum nMSE is at embedding lag $\tau = 1$ and embedding dimension $d = 6$, while for optimal model parameters configuration the minimum nMSE is at embedding lag $\tau = 5$ and embedding dimension $d = 4$. One can see that acceptable prediction accuracy can be obtained even for non-optimal model parameters at $\tau = 1$ and $d = 6$. It can be concluded that the optimal values of the embedding lag and embedding dimension are not independent of selected values of model parameters.

It is worth noting that for optimal model parameters the minimum of the nMSE can be obtained at smaller value of embedding dimension than for non-optimal model parameters allowing prediction with less computation resources. This fact is very important in the field of building of real-time and adaptive systems operating under restrictions on the computation time and memory.
Acknowledgments
The reported study was funded by RFBR, project number 19-37-90040.

References
[1] Antropov N and Agafonov E 2019 Adaptive kernel identification of nonlinear stochastic dynamical systems Applied Methods of Statistical Analysis Statistical Computation and Simulation: Proceedings of the International Workshop pp 445–52
[2] Alper P 1965 A consideration of the discrete volterra series IEEE Transactions on Automatic Control 10 322–7
[3] Billings S A and Fakhouri S Y 1977 Identification of nonlinear systems using the Wiener model Electronics Letters 13 502–4
[4] Billings S A and Fakhouri S 1979 Nonlinear system identification using the Hammerstein model International Journal of System Sciences 10 567–78
[5] Haykin S 2001 Adaptive Filter Theory (Englewood Cliffs, NJ: Prentice Hall) p 912
[6] Saunders C, Gammerman A and Vovk V 1998 Ridge regression learning algorithm in dual variables Proceedings of the 15th International Conference on Machine Learning (ICML) pp 515–21
[7] Schölkopf B, Herbrich R and Smola A J 2001 A generalized representer theorem Computational learning theory 2111 416–26
[8] Liu W, Príncipe J C and Haykin S 2010 Kernel Adaptive Filtering: A Comprehensive Introduction (John Wiley & Sons) p 240
[9] Engel Y, Mannor S and Meir R 2004 The kernel recursive least squares algorithm IEEE Transactions on Signal Processing 52 2275–85
[10] Saïde C, Lengelle R, Honeine P, Richard C and Achkar R 2015 Nonlinear adaptive filtering using kernel-based algorithms with dictionary adaptation International Journal of Adaptive Control and Signal Processing 29 1391–110
[11] Liu W, Pokharel P P and Príncipe J C 2008 The kernel least-mean-square algorithm IEEE Transactions on Signal Processing 56 543–54
[12] Chen B, Zhao S, Zhu P and Príncipe J C 2012 Quantized kernel least mean square algorithm IEEE Transactions on Neural Networks and Learning Systems 23 22–32
[13] Suykens J A K, Signoretto M and Argyriou A 2014 Regularization, Optimization, Kernels, and Support Vector Machines (London: Chapman & Hall/CRC) p 525
[14] Sugiyama M, Kawanabe M 2012 Machine Learning in Non-Stationary Environments: Introduction to Covariate Shift Adaptation (Cambridge MA: MIT Press) p 280
[15] Liu W, Park I, Wang Y and Príncipe J C 2009 Extended kernel recursive least squares algorithm IEEE Transactions on Signal Processing 57 3801–14
[16] Van Vaerenbergh S, Lazaro-Gredilla M and Santamaria I 2012 Kernel Recursive Least-Squares Tracker for Time-Varying Regression IEEE Transactions on Neural Networks and Learning Systems 23 1313–26
[17] Rasmussen C E and Williams C K I 2006 Gaussian Processes for Machine Learning (Cambridge MA: MIT Press) p 272
[18] Wang W, Wang S, Qian G and Yang B 2017 Kernel Least Mean Tracker Proceedings of the 36th Chinese Control Conference pp 5100–04
[19] Kantz H and Schreiber T Nonlinear 2004 Time Series Analysis (Cambridge U.K.: Cambridge University Press 2nd edition) p 388
[20] Soofi A S and Liangyue C 2002 Modelling and Forecasting Financial Data (Kluwer Academic Publishers) p 488
[21] Mukherjee S, Osuna E and Girosi F 1997 Nonlinear prediction of chaotic time series using a support vector machine Neural Networks for Signal Processing VII. Proceedings of the 1997 IEEE Signal Processing Society Workshop pp 511–20