A new Online Kernel method identification on RKHS Space

Okba Taouali, Ines Zakraoui

*Laboratory of Automatic Signal and Image Processing,
National School of Engineers of Monastir, University of
Monastir, 5019, Tunisia*

taoualiokba@yahoo.fr, zakraoui_iness@msn.com

Ilyes Elaissi, Hassani Messaoud

*Laboratory of Automatic Signal and Image Processing,
National School of Engineers of Monastir, University of
Monastir, 5019, Tunisia*

elaissi_ilyes@yahoo.fr, hassani.messaoud@enim.rnu.tn

Abstract—This paper proposes a new kernel method for online identification of nonlinear system. The proposed Support Vector Regression- Regularized Network (SVR-RN) method uses the technique SVR in an offline phase to reduce the parameters number of the RKHS. Then the RN method is used to update theses reduced parameters.

Keywords—RKHS, Kernel method, online identification, SVR, RN, Tennessee process

I. INTRODUCTION

kernel methods such as support vector machines, SVM [19], regularization networks (RN) [4], kernel principal component analysis K-PCA [10] and Kernel partial least square KPLS [9] suffer from the increasing number of the RKHS model parameters making these methods not suitable for use in an on-line scenario. Recently online RKHS identification techniques were proposed in literature such as Sliding window kernel recursive least squares SWKRLS [18], [24] online RKPCA [17], online SVD-KPCA [5], online SVM [25] and Fast Iterative KPCA [12].... Theses online methods are more useful when the system to be identified is time-varying.

In this paper, we propose a new online method based on Support Vector Regression (SVR) and RN techniques. First in an offline phase, the proposed technique uses the SVR algorithm to obtain a RKHS model with a reduced number of parameters. Then we proceed the online phase using the RN method to update the parameter number of RKHS model obtained by the SVR technique.

The paper is organized as follows. In section 2, we present the Reproducing Kernel Hilbert Space (RKHS). Section 3 is devoted to the modelling in RKHS. The Learning methods SVR and RN are presented in section 4. In section 5, we propose the new online SVR-RN method. The proposed algorithm has been tested to identify the Tennessee Eastman process [1].

II. RKHS MODEL

Let $k : E \times E \rightarrow \mathbb{R}$ be a continuous positive definite kernel

$F_k \subset L^2(E)$ be a Hilbert space associated to the kernel k and defined by:

$$F_k = \left\{ f \in L^2(E) / f = \sum_{i=1}^l w_i \varphi_i \text{ and } \sum_{j=1}^l \frac{w_j^2}{\sigma_j} < +\infty \right\} \quad (1)$$

Let consider a set of observations $\{x_i, y_i\}_{i=1, \dots, n}$ with $x_i \in \mathbb{R}^d$, $y_i \in \mathbb{R}$ are respectively the system input and output. According to the statistical learning theory (SLT) [22] the identification problem in the RKHS F_k can be formulated as a minimization of the regularized empirical risk. Thus it consists in finding the function $f^* \in F_k$ such that:

$$f^* = \min_{f \in F_k} \frac{1}{n} \sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \|f\|_{F_k}^2 \quad (2)$$

Where n is the measurement number and λ is a regularization parameter chosen in order to ensure a generalization ability to the solution f^* .

Let's define the application Φ :

$$\Phi : E \rightarrow \mathbb{R}^l$$

$$x \mapsto \Phi(x) = \begin{pmatrix} \varphi_1(x) \\ \vdots \\ \varphi_l(x) \end{pmatrix} \quad (3)$$

The Gram matrix K associated to the kernel k is an n -dimensional square matrix, so that:

$$K_{i,j} = k(x_i, x_j) \text{ for } i, j = 1, \dots, n \quad (4)$$

The kernel trick [8] is so that:

$$\langle \Phi(x), \Phi(x') \rangle = k(x, x') \text{ } x, x' \in E \quad (5)$$

According to the representer theorem [11], the solution f^* of the optimization problem (2) is a linear combination of the kernel k applied to the n measurements x_i , $i = 1, \dots, n$, as:

$$f^*(x) = \sum_{i=1}^n a_i^* k(x_i, x) \quad (6)$$

To solve the optimization problem (2) we can use some kernel methods such that Support Vector Machine (SVM), Least Square Support Vector Machine (LSSVM) [7], Regularization Network (RN), Kernel Partial Least Square (KPLS) and Kernel Principal Component Analysis KPCA.

In the next section we present the Support vector machine on regression version (SVR) and the regularized Network methods

III. LEARNING MACHINES

The algorithm used to estimate the parameters number of the RKHS model in (6) is called learning machines such as support vector machines and regularization network.

A. Support vector machines

Support Vector Machines (SVM) have been recently developed in the framework of statistical learning theory [22] and have been successfully applied to a number of applications, ranging from time series prediction, to face recognition [26], to biological data processing for medical diagnosis [23]. Their theoretical foundations and their experimental success encourage further research on their characteristics, as well as their further use. Based on the theory of Support Vector Machines, Support vector Regression SVR is now a well established method for designing black-box models in engineering. The aim of SVR is to build a model $f: \mathbb{R}^d \rightarrow \mathbb{R}$ of the output of system that depends on a set of factors.

$$f(x) = \sum_{i=1}^n w_i \Phi(x_i) + b \quad (7)$$

where $\{\Phi(x_i)\}_{i=1, \dots, n}$ are the data in the features space, $\{w_i\}_{i=1, \dots, n}$ and b are coefficients. They can be estimated by minimizing the regularized risk function

$$J_r(w) = \frac{C}{n} \sum_{i=1}^n V_{\mathcal{E}}(y_i, f(x_i)) + \frac{1}{2} \|w\|^2 \quad (8)$$

Where: (7)

$V_{\mathcal{E}}(y_i, f(x_i))$ is the so-called loss function measuring the approximate errors between expected output y_i and the calculated output $f(x_i)$.

C is a regularization constant determining the trade-off between the training error and the generalization performance.

$\frac{1}{2} \|w\|^2$ is used as a measurement of function flatness.

By Introduction of slack variables ξ, ξ^* leads (8) to the following constrained function.

$$\begin{aligned} \text{Min } J(w, \xi^*) &= \text{Min} \left(\frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \right) \\ \text{s.t.} \end{aligned} \quad (9)$$

$$\begin{cases} y_i - \langle w, \Phi(x_i) \rangle - b \leq \varepsilon + \xi_i \\ \langle w, \Phi(x_i) \rangle + b - y_i \leq \varepsilon + \xi_i^* \end{cases} \quad (10)$$

$$\xi_i, \xi_i^* \geq 0, i = 1, \dots, n$$

This formulation of the problem comes back to use $\mathcal{E}$ -insensitive loss function of the following shape:

$$|y - f(x)|_{\mathcal{E}} = \begin{cases} 0 & \text{if } |y - f(x)| \leq \varepsilon \\ |y - f(x)| - \varepsilon & \text{if } |y - f(x)| > \varepsilon \end{cases} \quad (11)$$

According the representer theorem [8] the solution of the problem is:

$$f(x, \alpha_i, \alpha_i^*) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) k(x_i, x) + b \quad (12)$$

In (12), Lagrange multipliers α_i and α_i^* satisfy the equality $\alpha_i \times \alpha_i^* = 0$, $\alpha_i \geq 0$, $\alpha_i^* \geq 0$, $i = 1, \dots, n$

Those vectors with $\alpha_i \neq 0$ are called support vectors, which contribute to the final solution.

$$f(x) = \sum_{i=1}^{n_{SV}} \alpha_i k(x_i^{n_{SV}}, x) \quad (13)$$

B. Regularized Network

In this case the cost function is:

$$V(y^{(i)}, f(x^{(i)})) = (y^{(i)} - f(x^{(i)}))^2 \quad (14)$$

Therefore the criterion J_r is

$$J_r(A) = \frac{1}{2} \sum_{i=1}^n \left(y_i - \sum_{j=1}^n a_j k(x_i, x_j) \right)^2 + \frac{\rho}{2} \|A\|^2 \quad (15)$$

Where:

$$A = (a_1 \ a_2 \ \dots \ a_n)^T$$

So the solution provides the parameters $\{a_i\}$ which verify:

$$a_i = \sum_{j=1}^n (K + \lambda n I_n)^{-1}_{i,j} y_j \quad (16)$$

In a matrix form, the relation (16) is:

$$(K + \lambda n I_n) A = Y \quad (17)$$

Where: K is the Gram matrix and $Y = (y_1, y_2, \dots, y_n)^T$

IV. THE PROPOSED ONLINE SVR-RN METHOD

The main idea of the proposed method is to use the SVR technique on the offline phase to reduce the parameter number of the RKHS model and then we use the Regularized Network method to update theses parameters.

First step, to determine the optimal value of the parameters σ of the kernel associated to the RKHS model, an offline learning step on a reduced observation set $D = \{(x_1, y_1), \dots, (x_n, y_n)\}$ is carried out till the provided RKHS model approximates correctly the nonlinear system. In the offline phase we apply the SVR technique to reduce the number of the parameters of the RKHS model and the resulting model is written as:

$$\tilde{y}(x) = \sum_{j=1}^{nsv} a_j k(x_j^{svi}, x) \quad (18)$$

The on line identification phase begins from the instant $n+1$. At this instant a new couple of observations is available $\{x_{n+1}, y_{n+1}\}$.

From the equation (18), we calculate the output RKHS model given by:

$$\tilde{y}_{n+1} = \sum_{j=1}^{nsv} a_j k(x_j^{svi}, x_{n+1}) \quad (19)$$

The error between the estimated output and the measured on actual one is:

$$e_{n+1} = (\tilde{y}_{n+1} - y_{n+1}) \quad (20)$$

If

$$|e_{n+1}| < \varepsilon \quad (21)$$

where ε is a given threshold, we can say the model approaches sufficiently the system behavior. If not, we update the parameters $\{a_j\}$ by minimizing the criterion $J_{r,n+1}$

$$J_{r,n+1}(A_{n+1}) = \frac{1}{2} \sum_{i=1}^{nsv} \left(y_i^{svi} - \sum_{j=1}^{nsv} a_j k(x_j^{svi}, x_n) \right)^2 + \frac{1}{2} \left(y_{n+1} - \sum_{j=1}^{nsv} a_j k(x_j^{svi}, x_n) \right)^2 + \frac{\rho}{2} \|A_{n+1}\|^2 \quad (22)$$

Where

$$A_{n+1} = \begin{pmatrix} a_1 \\ \vdots \\ a_{nsv} \end{pmatrix} \in IR^{nsv}$$

We can write in a matrix form,

$$J_{r,n+1}(A_{n+1}) = \frac{1}{2} \left(\|K_{(n+1)nsv} A_{n+1} - Y_{n+1}\|^2 + \rho \|A_{n+1}\|^2 \right) \quad (23)$$

Where

$$K_{(n+1)nsv} = \begin{pmatrix} K(x_1^{svi}, x_1) & \dots & K(x_1^{svi}, x_{nsv}) \\ \vdots & \ddots & \vdots \\ K(x_{nsv}^{svi}, x_{nsv}) & \dots & K(x_{nsv}^{svi}, x_{n+1}) \end{pmatrix} \in IR^{(nsv+1) \times nsv} \text{ et}$$

$$Y_{n+1} = \begin{pmatrix} y_1^{svi} \\ \vdots \\ y_{nsv}^{svi} \\ y_{n+1} \end{pmatrix} \in IR^{P+1} \quad (24)$$

The criterion (23) can be written as the following :

$$\begin{aligned} \min_{A_{n+1} \in IR^{nsv}} J_{r,n+1}(A_{n+1}) &= \min_{A_{n+1} \in IR^{nsv}} \frac{1}{2} \left(\|K_{(n+1)nsv} A_{n+1} - Y_{n+1}\|^2 + \rho \|A_{n+1}\|^2 \right) \\ &= \min_{A_{n+1} \in IR^{nsv}} \left(\frac{1}{2} (K_{(n+1)nsv} A_{n+1} - Y_{n+1})^T (K_{(n+1)nsv} A_{n+1} - Y_{n+1}) + \frac{\rho}{2} A_{n+1}^T A_{n+1} \right) \end{aligned} \quad (28)$$

The minimum of $J_{r,n+1}$ is reached for:

$$\frac{\partial J_{r,n+1}}{\partial A_{n+1}} = 0 \quad (25)$$

Therefore

$$\frac{\partial J_{r,n+1}}{\partial A_{n+1}} = \frac{\partial}{\partial A} \left[\frac{1}{2} (A_{n+1}^T K_{(n+1)nsv}^T - Y_{n+1}^T) (K_{(n+1)nsv} A_{n+1} - Y_{n+1}) \right] + \frac{\rho}{2} A_{n+1}^T A_{n+1}$$

$$= \frac{1}{2} \frac{\partial}{\partial A_{n+1}} \left[A_{n+1}^T K_{(n+1)nsv}^T K_{(n+1)nsv} A_{n+1} - A_{n+1}^T K_{(n+1)nsv}^T Y_{n+1} - Y_{n+1}^T K_{(n+1)nsv} A_{n+1} - Y_{n+1}^T Y_{n+1} + \rho A_{n+1}^T A_{n+1} \right]$$

(26)

As $A_{n+1}^T K_{(n+1)nsv} Y_{n+1}$ et $Y_{n+1}^T K_{(n+1)nsv} A_{n+1}$ are so scalar and transposed, then:

$$\frac{\partial J_{r,n+1}}{\partial A_{n+1}} = \frac{\partial}{\partial A_{n+1}} \left[A_{n+1}^T K_{(n+1)nsv}^T K_{(n+1)nsv} A_{n+1} - 2 Y_{n+1}^T K_{(n+1)nsv} A_{n+1} - Y_{n+1}^T Y_{n+1} + \rho A_{n+1}^T A_{n+1} \right]$$

(27)

Therefore

$$\frac{\partial J_{r,n+1}}{\partial A_{n+1}} = K_{(n+1)nsv}^T K_{(n+1)nsv} A_{n+1} - K_{(n+1)nsv}^T Y_{n+1} + \rho A_{n+1}$$

$$= (K_{(n+1)nsv}^T K_{(n+1)nsv} + \rho I_{nsv}) A_{n+1} - K_{(n+1)nsv}^T Y_{n+1}$$

Then

$$A_{n+1} = (K_{(n+1)nsv}^T K_{(n+1)nsv} + \rho I_{nsv})^{-1} K_{(n+1)nsv}^T Y_{n+1}$$

To every new observation, we update the parameters of the model (19) using the relation (29).

The Online SVR-RN algorithm

Offline phase:

According to (16) we determine the nsv support vector resulting from the processing of a n measurement set. Then we write RKHS model obtained by SVR method

$$\tilde{y}_n = \sum_{j=1}^{nsv} a_j k(x_j^{svi}, x_n)$$

Online phase:

For a new couple of observation $\{x_{n+1}, y_{n+1}\}$

1 - Calculate the output of the model RKHS

$$\tilde{y}_{n+1} = \sum_{j=1}^{nsv} a_j k(x_j^{svi}, x_{n+1})$$

2- Calculate the value of e_{n+1} .

3- If the condition (21) is satisfied $\Rightarrow$ we comes back to 1 for a new observation $\{x_{n+2}, y_{n+2}\}$

Otherwise:

Estimate the parameters $\{a_j\} j=1, \dots, nsv$ using the relation (29).

V. SIMULATIONS

A. Tennessee Eastman Process

The Tennessee Eastman (TE) process [1] is a highly nonlinear, non-minimum phase, and open-loop unstable chemical process. It consists of a reactor/separator/recycle arrangement that provides two products G and H from four reactants A, C, D and E. Also a byproduct F is present in the process. The simultaneous, irreversible and exothermic gas-liquid reactions are:

The process has 12 valves available for manipulation and 41 measurements available for monitoring or control. The detailed description of these variables, process disturbances and operating conditions, is given in [1]. The process flowsheet is presented in Figure 1.

Fig. 1. Tennessee Process

The modelling and the identification of the Tennessee Eastman process represent a challenge for the control community. They have been the subject of several studies [15], [16], [19] and [20]. Srinivas and Arkun in 1997 [16] have used input/ output process data to identify an autoregressive (AR) model parameters.

The input/output data used to build the model were generated from the model of Tennessee implemented for the program Matlab © in the toolbox Simulink [1]. The process has 12 inputs and 41 outputs.

We are interested to model the Temperature of separator. To generate the data from Simulink, the simulation step size was 0.0005 s and the data were collected every 0.01s. We present in the Figures 2 the separator temperature of the process.

To build the RKHS model we use the RBF Kernel (Radial Basis Function) given by

$$k(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right)$$

Where the parameter $\sigma = 100$ and the regularization parameter $\lambda = 10^{-10}$ are determined using the cross validation in the offline phase that we have used a training set of 200 observations. The parameters number of RKHS model given by the SVR technique is equal to 11.

In the online phase, the total number of observations is 3300. In Figures 2 and 3 we represent the online SVR-RN output as well as the system output and the evolution of the Mean Square Error (NMSE). We remark that the model output is in concordance with the system output, indeed the Normalized Mean Square Error is equal to 6.38×10^{-6} .

Fig. 2. System and model outputs during the online identification

Fig. 3. Evolution of NMSE

VI. CONCLUSION

In this paper, we have proposed an online kernel method for nonlinear system parameter identification. Through several experiments, we showed the accuracy and good scaling properties of the proposed SVR-RN method. This algorithm has been tested for identifying a Tennessee Eastman process and the results were satisfactory.

REFERENCES

- [1] A. M.S. Cruz. Tennessee Eastman Plant –wide Industrial Process. Technical report, CAPEC, Technical University of Denmark, 2004
- [2] R. Cedric Senior, José Carlos M. Bermudez, Paul Honeine, Online prediction of time series data with kernels, IEEE TRANSACTIONS ON SIGNAL PROCESSING, 57, pp.1058-1067, 2009.
- [3] D.Kuzmin, M.K. Warmuth., Online Kernel PCA with Entropic Matrix Updates, in Proceedings of the 24 the International Conference on Machine Learning, Corvallis, OR, 2007.
- [4] I. Elaissi, Modélisation, Identification et Commande Prédictive des systèmes non linéaires par utilisations des espaces RKHS, thèse de doctorat, ENIT Tunisie, 2009.
- [5] I. Elaissi, Ines jaffal, O. Taouali and H. Messaoud. Online Prediction Model based SVD-KPCA methos ISA transactions, Vol 12, 2012.
- [6] Y. Engel, S. Mannor and R. Meir. The Kernel recursive Least squares Algorithm. Rapport technique, 2003.
- [7] J. Suykens, A. K., Van Gestel, T., De Brabanter, J., De Moor, B. et Vandewalle, J, Least squares support vector machines, (Singapore: WorldScientific, 2002).
- [8] G. Wahba, An introduction to model building with Reproducing Kernel Hilbert Spaces. Technical report No 1020, (Department of Statistics, University of Wisconsin-Madison, 2000).
- [9] R. Rosipal and L.J. Trecho, Kernel Partial Least Squares in Reproducing Kernel Hilbert Spaces, Journal of machine learning research 2, (2001), 97-123
- [10] B. Scholkopf, A. Smola et K-R Muller, Nonlinear component analysis as Kernel eigenvalue problem, Neural computation. 10, (1998), 1299-1319.
- [11] B. Scholkopf and A. Smola, Learning with kernels. (The MIT press, 2002).

- [12] G. Simon, Nicol N. Schraudolph, S.V. N. Vishwanathan, Fast Iterative Kernel Principal Component Analysis, *Journal of Machine Learning Research* .8, 1893-1918. 2007.
- [13] C.Tat-Jun, K.Schindler, D. Suter, Incremental Kernel SVD for Face Recognition with Image Sets, *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition FGR*, 2006.
- [14] F. Lauer, *Machines à Vecteurs de Support et Identification de Systèmes Hybrides*, thèse de doctorat de l'Université Henri Poincaré – Nancy 1, octobre 2008, Nancy, France.
- [15] N.L. Ricker and J.H. Lee. Nonlinear modeling and state estimation for the Tennessee Eastman challenge process. *Computers and Chemical Engineering*, 19, pp. 983-1005, 1995b.
- [16] R. Srinivas and Y. Arkun. Control of the Tennessee Eastman process using input–output models. *Journal of Process Control*, 7, pp. 387-400, 1997.
- [17] O. Taouali, I. Elaissi, and H. Messaoud. Online Prediction Model based on Reduced Kernel Principal Component Analysis. *Neural Computing and Application journal*, Vol 19, 2010.
- [18] O. Taouali, I. Elaissi, and H. Messaoud. Design and comparative study of online kernel methods identification of nonlinear system in RKHS space. *Artificial Intelligence Review*, May, 2011 .
- [19] O. Taouali. *Apprentissage statistique pour la Modélisation et l'Identification hors ligne et en Ligne des Systèmes Non Linéaires dans des Espaces de Hilbert à Noyau Reproductible (RKHS)*, thèse de doctorat, ENIM, Tunisie, 2010
- [20] O. Taouali, I. Aissi, N. Villa and H. Messaoud, Identification of nonlinear Multivariable Processes Modelled on Reproducing Kernel Hilbert Space: Application to Tennessee Process. *ICONS, 2nd IFAC International Conference on Intelligent Control Systems and Signal Processing*, pp.1-6, Istanbul 2009.
- [21] V. Vovk; Leading strategies in competitive on-line prediction, *Theoretical Computer Science* . 405, 285-296, 2008
- [22] V.N. Vapnik, *Statistical Learning Theory*, Wiley New York.1998
- [23] K. Veropoulos. N. Cristianini, and Campbell. C, *The Application of Support Vector Machines to Medical Decision Support: A Case Study*, *ACAI conference*, 1999.
- [24] S. Van Vaerenbergh, Javier Via and Ignacio Santamaria. Nonlinear System Identification using a New Sliding-Window Kernel RLS Algorithm, *journal of Communications*, vol 2, 2007.
- [25] W. Wanga, C. Mena, W. Lub, Online prediction model based on support vector machine, *Neurocomputing* 71,550-558 , (2008),
- [26] M.H. Yang, Kernel Eigenfaces vs kernel First faces :Face recognition using kernel method, In *IEEE FGR* pp. 215-220, 2002.