International Conference on Control, Engineering & Information Technology (CEIT'13)
Proceedings Engineering & Technology - Vol.1, pp. 127-134, 2013

System identification based on Support Kernels Regression

Mounira TARHOUNI, Salah ZIDI*, Kaouther LAABIDI, Moufida KSOURI-LAHMARI
Laboratory of Research Analysis, Conception and Control of Systems (ACCS, ENIT)
BP 37, LE BELVEDERE 1002, Tunis, Tunisia
Email: mouniratarhouni@yahoo.fr, {kaouther.laabidi,moufida.ksouri }@enit.rnu.tn,
* LAGIS (USTL, Lille)
Villeneuve d'Ascq 59650 Lille-France
Email: salah_zidi@yahoo.fr

***Abstract*—This paper deals with the identification of nonlinear systems using multi-kernel approach. In this context, we have improved the Support Vector Regression (SVR) method in order to identify nonlinear complex system. Our idea consists in dividing the regressor vector in several blocks, and, for each one a kernel function is used. This blockwise SVR approach is called Support Kernel Regression (SKR). Furthermore, we have proposed two methods SKR(lin-rbf) and SKR(rbf-rbf). In these two methods we have divided the regressor into two Blocks. In the SKR(lin-rbf) based approach, the linear kernel and the Gaussian kernel are used, respectively, to identify the influence of the first block and of the second block on the model. However, in the SKR(rbf-rbf) approach two gaussian kernels are used. An example is presented for qualitative comparison with the classical SVR approach based on a single kernel function. The results reveal the accuracy and the robustness of the obtained model based on our proposed approaches.**

***Keywords*—Support Vector Regression; Support Kernel Regression; Nonlinear System Identification; Kernel Function**

I. INTRODUCTION

Recently, estimation techniques based on regularization and kernel methods play an important role. In this context, we mention the regression technique based on Support Vector Machines (SVM). SVM have received a great attention to be used to deal with this problem [1], [2]. These approaches obtain system models based on intelligent behavior and learn automatically from previous experiences.
The SVM was first proposed by Vapnik [3] in order to obtain maximum margin separating hyperplanes in classification problems but this technique has become a general learning theory and is applied in large a field of applications [1], [4], [5], [6], [7], [8], $\cdots$. A comprehensive description of this method for classification and regression problems can be found in [9] and [10], respectively. The basic idea is to map linear inseparable input data into a high dimensional linear separable feature space via a nonlinear mapping technique (kernel function) and to carry out linear classification or regression in feature space.
In this paper, we have been interested in the problem of non linear system identification based on SVM. This approach becomes inefficient when the system complexity becomes more and more important. In this context, the multi-kernel approach seems to be potential to deal with such problems.
In literature, several multi-kernel training based algorithms have been presented. Diosan and al [11] developed an Combined Kernels (ECKs). They considered a combination of multiple kernels and they used a Genetic algorithm (GA) for evolving these weights. This approach is applied to solving classification problems. They have compared their results to those obtained by [12] with a combined kernel learnt with convex methods (CCKs). In [13], the authors have proposed a componentwise Least Squares Support Vector Machines (LS-SVMs) for the estimation of additive models consisting of a sum of nonlinear components.
On our part, we have proposed in [14] a new approach called Least Squares Support Kernel Machines (LS-SKM). LS-SKM method which is a variant of Least Squares Support Vector Machines (LS-SVM) approach consists in dividing the regression vector into several sub-vectors, and, for each one a kernel function is used.
In this work, the proposed multi-kernel configuration is adopted to Support Vector Machines (SVM). Therefore, a new identification method called Support Kernel Regression (SKR) has been investigated. The basic idea consists in dividing the regressor on several blocks and for every one a kernel function is used. Indeed, the suggested approach reflects the fact that practical learning problems often involve multiple, heterogeneous data sources. Therefore, this blockwise training algorithm considers multiple kernels. Each kernel describes a similarity measure between data having the same origin or which sharing the common characteristics. In fact, this approach succeeds to improve the identification performance such as the accuracy of the elaborated model.
The outline of this paper is as follows: In Section 2, the problem statement is given. In section 3, we review the mathematical foundation of the SVR method. Section 4 introduces the new method Support Kernel Regression (SKR) for identification. In this part, first we present the context and the motivations, second, we present the SKR approach. In Section 5, the effectiveness of the proposed SKR based algorithm is presented through an illustrative example.

II. BLACK BOX SYSTEMS IDENTIFICATION BASED ON SUPPORT VECTOR REGRESSION (SVR)

Black box models are based on measurement data. Both model structure and model parameters are determined by using

experimental data. In this context, very little prior knowledge of the system behavior is exploited.
The steps of identification are illustrated in Figure 1. First, input-output data of the system are collected to constitute the training data samples $\{x_i, y_i\}_{i=1}^{N}$. $x_i = (u(i-1), \cdots, u(i-m_1), y(i-1), \cdots, y(i-n_1))$ is the regression vector at different sampling instants and y_i is the system output corresponding to x_i. m_1 and n_1 parameters represent, respectively, the input regression order and the output regression order. Second, the model structure is chosen. Then, the model parameters are selected. It is, usually, solved by the application of linear and nonlinear optimization technique. Finally, we make the necessary tests to validate the obtained parameters. In this

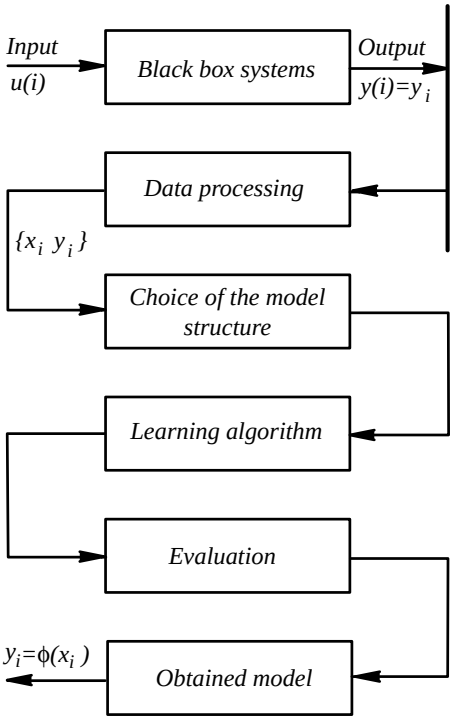

Fig. 1. Principle of black box systems identification.

work, we consider that the system is given by the following input/output representation:

$$\begin{aligned} y(i) &= \phi(u(i-1), \cdots, u(i-m_1), y(i-1), \cdots, y(i-n_1)) \\ &= y_i \end{aligned} \tag{1}$$

We will consider, according to the representer theorem [15], the assumption that every function can be described with a kernel function. Indeed, it is the basic idea of the regression based on SVR approach. This method uses a single kernel function to describe the system behavior. To obtain a more accurate model, a new method called " Support Kernel Regression (SKR) is proposed and developed in our work. The basic idea is to divide the regressor vector x_i in several blocks, and, for each block a kernel function is adopted. More details will be given in section IV.

III. Support Vector Regression (SVR)

A. Linear Regression with ε-insensitive loss function

Suppose we are given the training data $\{(x_1, y_1), \cdots, (x_N, y_N)\} \subset \chi \times \mathbf{R}$, where χ denote the space of the input patterns. In $\varepsilon - SV$ regression, the objective is to find a function $f(x)$ that has at most ε deviation from the actually obtained target y_i for all the training data, and at the same time is flat as possible [10]. In this case $f(x)$ has the following form:

$$f(x) = < w, x > + b \quad with \ w \in \chi, \ b \in \mathbf{R} \tag{2}$$

where $< ., . >$ denotes the dot product in χ. We can write this problem as a convex optimization problem:

$$\begin{aligned} &minimize \quad \tfrac{1}{2}\|w\|^2 \\ &subject \ \ to \quad \begin{cases} y_i - < w, x > - b \leq \varepsilon \\ < w, x > + b - y_i \leq \varepsilon \end{cases} \end{aligned} \tag{3}$$

The above convex optimization problem is feasible in cases where f actually exists and approximates all pairs (x_i, y_i) with ε precision. Sometimes, some errors are allowed. Introducing slack variables ξ_i, xi_i^* to cope with otherwise infeasible constraints of the optimization problem (3), the formulation becomes:

$$\begin{aligned} &minimize \quad \tfrac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \\ &subject \ \ to \quad \begin{cases} y_i - < w, x > - b \leq \varepsilon + \xi_i \\ < w, x > + b - y_i \leq \varepsilon + \xi_i^* \end{cases} \end{aligned} \tag{4}$$

The constant $C > 0$ determines the trade-off between the flatness of f and the amount up to which deviations larger than ε are tolerated. This corresponds to dealing with a so called ε-insensitive loss function described by

$$L_\varepsilon(y) = \begin{cases} 0 & if \ |f(x) - y| \leq \varepsilon \\ |f(x) - y| - \varepsilon & otherwise \end{cases} \tag{5}$$

Figure 2 show the idea graphically. The problem is solved by

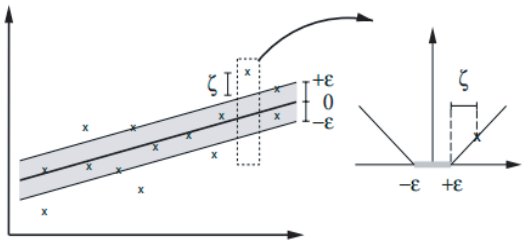

Fig. 2. The soft margin loss setting for a linear SVM.

minimizing the lagrangian L given by the following expression:

$$\begin{aligned} L := \tfrac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) - \sum_{i=1}^{N}(\eta_i\xi_i + \eta_i^*\xi_i^*) \\ - \sum_{i=1}^{N}\alpha_i(\varepsilon + \xi_i - y_i + < w, x_i > + b) \\ - \sum_{i=1}^{N}\alpha_i^*(\varepsilon + \xi_i^* + y_i - < w, x_i > - b) \end{aligned} \tag{6}$$

η_i^*, η_i, α_i, α_i^* are lagrange multipliers The optimality conditions are:

$$\begin{cases} \frac{\partial l}{\partial b} = 0 \mapsto \sum_{i=1}^{N}(\alpha_i - \alpha_i^*) = 0 \\ \frac{\partial L}{\partial w} = 0 \mapsto w - \sum_{i=1}^{N}(\alpha_i - \alpha_i^*) = 0 \\ \frac{\partial l}{\partial \xi_i} = 0 \mapsto C - \alpha_i - \eta_i = 0 \\ \frac{\partial l}{\partial \xi_i^*} = 0 \mapsto C - \alpha_i^* - \eta_i^* = 0 \\ i = 1, \cdots, N \end{cases} \tag{7}$$

Substituting these conditions into the equation (6) yields the dual optimization:

$$maximize \begin{cases} \frac{1}{2}\sum_{i,j=1}^{N}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) < x_i, x_j > \\ -\varepsilon\sum_{i,j=1}^{N}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{N} y_i(\alpha_i - \alpha_i^*) \end{cases}$$
$$subject\ to\ \sum_{i=1}^{N}(\alpha_i - \alpha_i^*) = 0\ and\ \alpha_i,\ \alpha_i^* \in [0, C] \tag{8}$$

Thus

$$f(x) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*) < x_i, x > +b \tag{9}$$

B. Nonlinear SVR

In this case, we suppose that the equation of the function presented in (2) that we will predict is described by the following equation:

$$f(x) =< \phi(x), x > +b \tag{10}$$

Using the kernel trick, we obtain the following optimization problem:

$$maximize \begin{cases} \frac{1}{2}\sum_{i,j=1}^{N}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)k(x_i, x_j) \\ -\varepsilon\sum_{i,j=1}^{N}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{N} y_i(\alpha_i - \alpha_i^*) \end{cases}$$
$$subject\ to\ \sum_{i=1}^{N}(\alpha_i - \alpha_i^*) = 0\ and\ \alpha_i,\ \alpha_i^* \in [0, C] \tag{11}$$

Solving α_i, α_i^* and b using KKT (Kurash-Kuhn-Tucker) conditions, the regression function of (9) becomes

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

We note that in the nonlinear setting, the optimization problem corresponds to finding the flattest function in feature space, not in the input space.

C. Loss Function

In literature several Loss Function are used [9], [10] such as quadratic loss function, ε-insensitive loss function, Huber loss function and Laplace loss function (figure 3).
The quadratic cost is optimal in a Maximum Likelihood (ML) sense, when the noise is Gaussian, whereas the linear cost is optimal for exponential noise. The use of the ε-Huber Loss Function has the ability to deal simultaneously with different kinds of noise. The use of ε-insensitive is not appropriate when Gaussian noise can be present in the data, whereas a Quadratic Loss Function does not produce sparse solutions [9].

Table I summarizes these loss functions [16]:

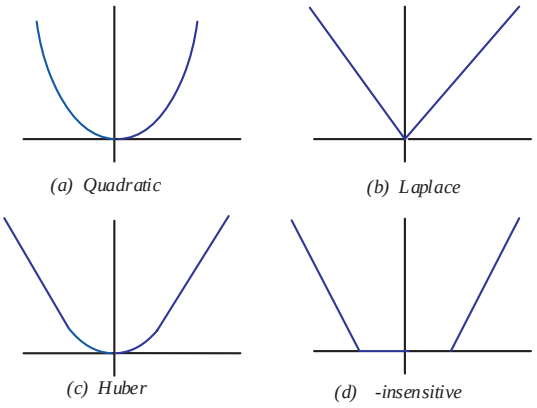

Fig. 3. Loss Function.

TABLE I. MOST POPULAR LOSS FUNCTIONS

	loss function
Laplacian	$c(\xi) = \lvert\xi\rvert$.
Gaussian	$c(\xi) = \frac{1}{2}\xi^2$.
Polynomial	$c(\xi) = \frac{1}{p}\lvert\xi\rvert^p$.
Huber's robust loss	$c(\xi) = \begin{cases} \frac{1}{2\sigma}\xi^2 & si\ \lvert\xi\rvert \le \sigma, \\ \lvert\xi\rvert - \frac{\sigma}{2} & sinon. \end{cases}$
Piecewise polynomial	$c(\xi) = \begin{cases} \frac{1}{p\sigma^{p-1}}(\xi)^p & si\ \lvert\xi\rvert \le \sigma, \\ \lvert\xi\rvert - \sigma\frac{p-1}{p} & sinon. \end{cases}$
ϵ-insensitive	$c(\xi) = \lvert\xi\rvert_\varepsilon$.

D. Problem of data distribution

The formulation of the model output based on SVR approach is given as a sum of term $k(x_i, x_j)$ weighted by the lagrange multipliers $(\alpha_i^* - \alpha_i)$, as given in (12). Every term describe a similarity measure between those data (x_i, x_j). This assumption has the disadvantage of considering that all the terms constituting the regressor have the common origin. Therefore, it does not reflect the reality of things. For that, In this work, we will focus on improving the above formulation. We recall, first, that the kernel function considers a scalar product in the new feature space. In this space, the kernel function characterizes a certain similarity $k(x_i, x_j)$ between two data x_i and x_j in initial space, as it is shown in the following kernel matrix (13).

$$\begin{pmatrix} k(x_1,x_1) & k(x_1,x_2) & \cdots & k(x_1,x_n) \\ k(x_2,x_1) & k(x_2,x_2) & \cdots & k(x_2,x_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_n,x_1) & k(x_n,x_2) & \cdots & k(x_n,x_n) \end{pmatrix}. \tag{13}$$

The data x_i is a vector, constituted by p terms. These terms, may be heterogeneous and having different origins. At this stage, we thought to divide the regression vector in several sub-vectors, and in this case, the search for similarity between two vectors is replaced by a sum of similarities between each two sub-vectors forming the database. This idea is applied to SVR approach, and consequently, a new approach called Support Kernel Regression (SKR), is proposed and applied for the identification of complex nonlinear systems.

IV. SUPPORT KERNEL REGRESSION (SKR)

A. Blockwise identification

Adopting our previously announced idea, we assume that each training data x_i will be decomposed on m sub-vectors

$x_{1,i},\ x_{2,i},\ \cdots, x_{m,i}$ as follows:

$$x_i = \begin{pmatrix} x_{1,i} \\ x_{2,i} \\ \vdots \\ x_{m,i} \end{pmatrix}. \tag{14}$$

Figure 4 depicts this graphically: with qj represent the dimension of the sub-vector x_{qj} and satisfying:

$$\sum_{j=1}^{m} qj = q,$$

where q is the dimension of the entire initial regressor x_i.

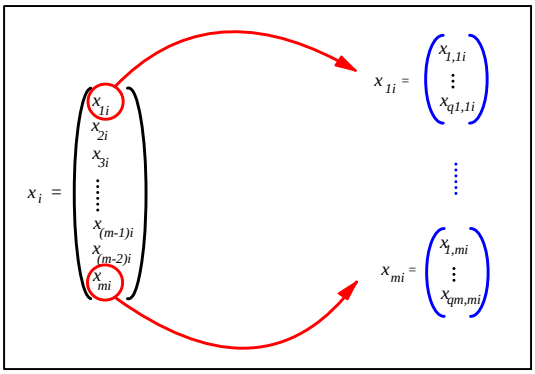

Fig. 4. Technique de rgression.

Thereafter, instead of using a single function φ adopted in SVR approach, we assume that the system output is expressed by several weighted functions $\varphi_1,\ \cdots,\ \varphi_m$. These functions are used to approximate the system output as follows:

$$y = < \begin{pmatrix} w_1 \\ \vdots \\ w_m \end{pmatrix}, \begin{pmatrix} a_1\varphi_1(x_1) \\ \vdots \\ a_m\varphi_m(x_m) \end{pmatrix} > +b. \tag{15}$$

We consider the input space χ will be associated with m different feature spaces $F_1, \cdots, F_m$ using functions: $\varphi_1, \cdots, \varphi_m$. We consider, also, that $w = (w_1, \cdots, w_m)$ and $x = (x_1, \cdots, x_m) \in \mathbb{R}^{q_1+\cdots+q_m}$.
Based on these assumptions, the approach SKR will be defined by the following optimization problem:

$$min \left\{ \begin{array}{l} \frac{1}{2}\|w_{1i}\|^2 + \cdots + \frac{1}{2}\|w_{mi}\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \\ subject\ to \left\{ \begin{array}{l} y_i - < w_{1i}, \phi_{1i}(x_{1i,i}) > - \cdots \\ - < w_{mi}, \phi_{mi}(x_{mi,i}) > -b \leq \varepsilon + \xi_i \\ < w_{1i}, \phi_{1i}(x_{1i,i}) > + \cdots + \\ < w_{mi}, \phi_{mi}(x_{mi,i}) > +b - y_i \leq \varepsilon + \xi_i^* \end{array} \right. \end{array} \right. \tag{16}$$

The corresponding lagrangian involving the dual variables α_i and η_i is, thus, described as follows:

$$\begin{array}{l} L := \frac{1}{2}\|w_{1i}\|^2 + \cdots + \frac{1}{2}\|w_{mi}\|^2 C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \\ -\sum_{i=1}^{N}(\eta_i\xi_i + \eta_i^*\xi_i^*) - \sum_{i=1}^{N}\alpha_i(\varepsilon + \xi_i - y_i \\ + < w_{1i}, \phi_{1i}(x_{1i,i}) > + \cdots + < w_{mi}, \phi_{mi}(x_{mi,i}) > +b) \\ -\sum_{i=1}^{N}\alpha_i^*(\varepsilon + \xi_i^* + y_i - < w_{1i}, \phi_{1i}(x_{1i,i}) > \\ + \cdots + < w_{mi}, \phi_{mi}(x_{mi,i}) > -b), \end{array} \tag{17}$$

with $\eta_i^*,\ \eta_i,\ \alpha_i,\ \alpha_i^*$ are Lagrange multipliers.
Optimality conditions are:

$$\left\{ \begin{array}{l} \frac{\partial l}{\partial b} = 0 \mapsto \sum_{i=1}^{N}(\alpha_i - \alpha_i^*) = 0, \\ \frac{\partial L}{\partial w_{1i}} = 0 \mapsto w_{1i} - \sum_{i=1}^{N}(\alpha_i - \alpha_i^*) = 0, \\ \vdots \\ \frac{\partial L}{\partial w_{mi}} = 0 \mapsto w_{mi} - \sum_{i=1}^{N}(\alpha_i - \alpha_i^*) = 0, \\ \frac{\partial l}{\partial \xi_i^*} = 0 \mapsto C - \alpha_i^* - \eta_i^* = 0, \\ \frac{\partial l}{\partial \xi_i} = 0 \mapsto C - \alpha_i - \eta_i = 0, \\ i = 1, \cdots, N. \end{array} \right. \tag{18}$$

$$\begin{array}{l} max \left\{ \begin{array}{l} \frac{1}{2}\sum_{i,j=1}^{N}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)(k_u(x_{1i,i}, x_{1j,j}) \\ + \cdots + k_{mi}(x_{mi,i}, x_{mj,j})) \\ -\varepsilon\sum_{i,j=1}^{N}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{N} y_i(\alpha_i - \alpha_i^*), \end{array} \right. \\ sc\ \sum_{i=1}^{N}(\alpha_i - \alpha_i^*) = 0\ and\ \alpha_i,\ \alpha_i^* \in [0, C]. \end{array} \tag{19}$$

Solve α_i, α_i^* and b using KKT (Kurash-Kuhn-Tucker) conditions. In this new approach, the proposed regression function is given by the formulation (20):

$$y(x) = \sum_{i=1}^{N}(\alpha_i^* - \alpha_i)[k_{1i}(x_{1i,i}, x_{1j,j}) + \cdots + k_{mi}(x_{mi,i}, x_{mj,j})] + b. \tag{20}$$

B. Blockwise separation: Examples

We have found that the model output SKR is given as a sum of multiple weighted kernels. Each kernel describes a similarity measure between the data which have the same origin or which share the common characteristics. The suggested idea is based on dividing the regressor in several sub-vectors (or blocks). In this level, the question is according to which criterion has this separation to be done?
We recall the assumption that the regressor is formed by: $x_i = (u(i-1), \cdots, u(i-m_1), y(i-1), \cdots, y(i-n_1))$. In this work and since we are confronted with the problem of nonlinear systems identification, we suggest that the separation of the input vector in blocs can be done based on two criteria: the sample time or the origin of every term which constitutes the regressor. The first separation is based on the assumption that the input and the output of the system at the iteration $(k-1)$, affect greatly, the model output at the current itration k compared to other oldest previous itrations. For this reason, we propose to use a kernel to identify the influence of the sub-vector formed by $u(k-1)$ and $y(k-1)$ and an other kernel for the rest of terms as shown in figure 5.
The second separation is based on the idea of dividing the regressor according to the signal origin. Since we are interested in the identification task, this division requires to take only control signals in the first group and the delayed outputs in the second group. Thereafter, for each group a kernel function will be adopted. The figure 6 shows it graphically:

V. Simulations and Results

The identification diagram structure based on SKR is shown in figure 7, y_i is the actual system output and $\hat{y}_i$ is the estimated output. First, input-output data of the system are collected to constitute the training data samples $\{x_i, y_i\}_{i=1}^{N}$. $x_i = (u(i-1), \cdots, u(i-m_1), y(i-1), \cdots, y(i-n_1))$ is the

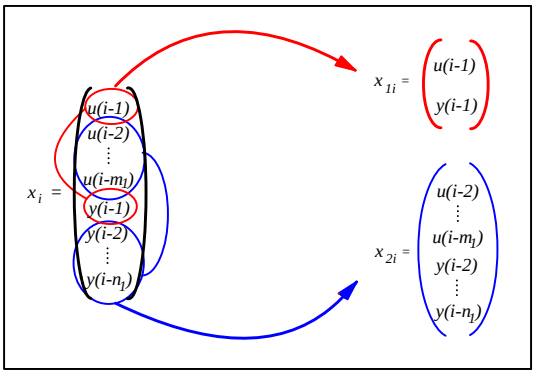

Fig. 5. Blockwise separation according to the time.

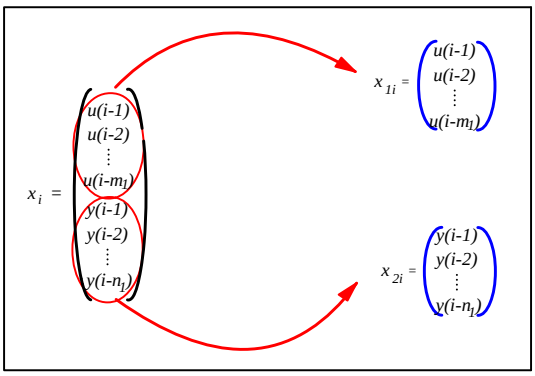

Fig. 6. Blockwise separation according to command signal, delayed output.

regression vector at different sampling instant and $y_i = y(i)$ is the system output corresponding to x_i. Second, the technique of regressor vector is used. Third, using the proposed SKR algorithms (SKR(lin-rbf), SKR(lin-rbf)), we can build the off-line model of the process. Finally, once the decision function is generated, the necessary tests are made to validate the chosen kernels and, also, the chosen kernel parameters. An

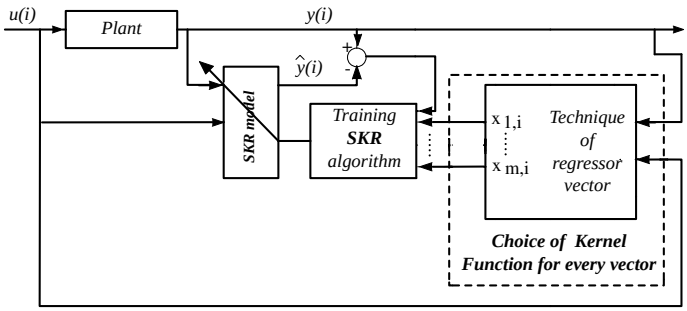

Fig. 7. Identification structure based on SKR techniques.

example is selected to demonstrate the good performances of our identification SKM based methods. For each example, we have generated 600 patterns. The first 200 samples are used as training set and the remaining 400 samples are used to validate the obtained model. The computation time mentioned in tables presents the resulted time during the training phase with 200 samples.

A. Example 1

In this session, we use the nonlinear system proposed in [17], [18] to test our SKR modeling method. The identified system is:

$$\begin{aligned} x_1(i+1) &= 0.1x_1(i) + 2\frac{u(i)+x_2(i)}{1+(u(i)+x_2(i))^2} \\ x_2(i+1) &= 0.1x_2(i) + u(i)(1+\frac{u(i)^2}{1+x_1(i)^2+x_2(i)^2} \\ y(i) &= x_1(i) + x_1(i). \end{aligned} \tag{21}$$

with an input excitation signal $u(i)$ chosen as a random value in $[-1,\ 1]$.

In this example, the model input vector is formed using $x_i = [u(i-1)\ u(i-2), \cdots,\ u(i-m_1),\ y(i-1)\ y(i-2), \cdots,\ y(i-n_1)]^T$ and as output $y_i = y(i)$. This vector will be divided into two groups $x_{1i} = [u(i-1)\ u(i-2), \cdots,\ u(i-m_1)]^T$ and $x_{2i} = [y(i-1)\ y(i-2), \cdots,\ y(i-n_1)]^T$. We have already announced that in the SKR (lin-rbf) method, the linear kernel and the Gaussian kernel were used, respectively, to identify the influence of x_{1i} and x_{2i} on the model. And, in the case of the SKR (rbf-rbf) based method two gaussian kernels were adopted.

The total correlation coefficient R^2_{tot}, the multiple correlation coefficient R^2_{mult} and the mean squares error (MSE) of the training data are calculated in order to decide if the obtained model is accepted or not. The model is acceptable if R^2_{tot}, R^2_{mult} is close to one and the MSE is close to zeros [19].

TABLE II. KERNEL PARAMETERS VALUES FOR DIFFERENT METHODS

Method	Kernels	p_1	p_2	C
Quadratic SVR	Linear	1	ϕ	1000
	RBF	ϕ	1.5	1000
Quadratic SKR	Linear-RBF	1	3	1000
	2RBF	2	4	1010
ε-insensitive SVR	Linear	1	ϕ	1000
	RBF	2	ϕ	1020
ε-insensitive SKR	Linear-RBF	1	4	1000
	2RBF	2	6	1000

To evaluate the proposed method performances, a comparison with two methods which are: linear SVR and Gaussian SVR will be done.

1) Input selection: It's trivial that in kernel machines, the size of model can increase when the model input dimension χ increases. Hence, input selection becomes a crucial task and it can, significantly, improve identification performance. For that, we will study the effect of the dimension of the regressor vector. Table III contains the values of R^2_{tot} according to the parameters m_1 and n_1. Based on these results, we found that the largest value of R^2_{tot} is obtained when $m_1 = 3$ and $n_1 = 5$. We notice also that if n_1 and m_1 becomes great the regressor vector contains some terms haven't any effect on the obtained model, contrariwise, using too many input terms may have undesirable effects on the identification performance and some input variables may be redundant or would become insignificant if some other input variables were present in the model. For example, the case when $m_1 = 10$ and $n_1 = 10$, the R^2_{tot} value, obtained based on quadratic SKR (lin-rbf) approach, decreases to 92.2% whereas it increases to 94.6908% when $m_1 = 3$ and $n_1 = 4$ and to 95.0638% when $m_1 = 3$ and $n_1 = 5$. Furthermore, when the model input

TABLE III. EVOLUTION OF R^2_{tot} ACCORDING TO THE PARAMETERS m_1, n_1

	Qaudratic SVR rbf	quadratic SKR (lin-rbf)	quadratic SKR (rbf-rbf)	ε-insensitive SVR rbf	ε-insensitive SKR (lin-rbf)	ε-insensitive SKR (rbf-rbf)
$m_1 = 2,\ n_1 = 2$	92.7216	92.8239	89.8832	92.7037	92.7114	89.4828
$m_1 = 2,\ n_1 = 4$	79.4039	92.9119	87.8897	81.8926	92.6129	88.3924
$m_1 = 2,\ n_1 = 5$	68.8681	92.1809	85.8544	74.7954	92.2881	86.7706
$m_1 = 3,\ n_1 = 3$	86.5163	92.8242	88.5696	86.8895	92.6100	88.8937
$m_1 = 3,\ n_1 = 4$	77.6569	94.6908	88.3008	78.4300	94.7375	88.9023
$m_1 = 3,\ n_1 = 5$	69.2320	95.0638	88.3919	76.7768	94.7368	88.3002
$m_1 = 4,\ n_1 = 5$	67.7843	94.8251	92.1101	76.2927	94.4972	92.5302
$m_1 = 4,\ n_1 = 6$	51.7861	92.8111	85.7681	69.2924	91.2839	85.8577
$m_1 = 10,\ n_1 = 10$	8.6875	92.2183	73.8575	23.2393	91.8490	81.1670

dimension χ (regressor vector dimension) becomes greater the computation time becomes more and more important. For example, it is equal to 7 *second* when we adopt the above based approach with $m_1 = 3$ and $n_1 = 5$ and it is equal to 8.1589 *second* when we adopt the same approach with $m_1 = 10$ and $n_1 = 10$.

The fitness criterion performance adopting our proposed methods: Quadratic SKR (lin-rbf) and ε-insensitive SKR (lin-rbf) are approximately the same, it reachs 95.0638% (with $m_1 = 3$, $n_1 = 5$), and 94.7375 (with $m_1 = 3$, $n_1 = 4$, respectively. However, the identification results based on the Gaussian quadratic SVR and on the Gaussian ε-insensitive SVR present a considerable identification errors. In fact, the best R^2_{tot} value is obtained when $m_1 = 2$, $n_1 = 2$ and it is equal to 92.9119 and 92.6129, respectively. The evolutions of the system output y with those of the identified model $\hat{y}$ based on Quadratic SKR (lin-rbf) proposed approach are compared. Figure 8 and Figure 9 presents respectively the output system/model and the identification errors. It is noticed that the modeling errors are very small compared to the results obtained using the classical quadratic SVR method (Figure 11). This may allow to conclude that the model obtained characterizes suitably the dynamic of the system and that our suggested method improves the model precision.

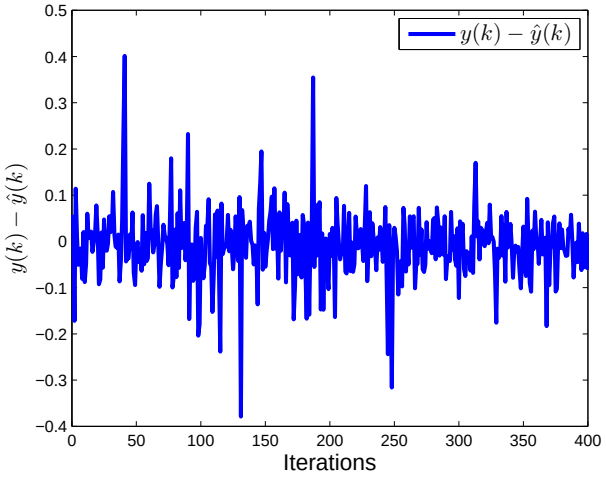

Fig. 9. Identification errors based on Quadratic SKR (lin-rbf), $m_1 = 3$, $n_1 = 5$.

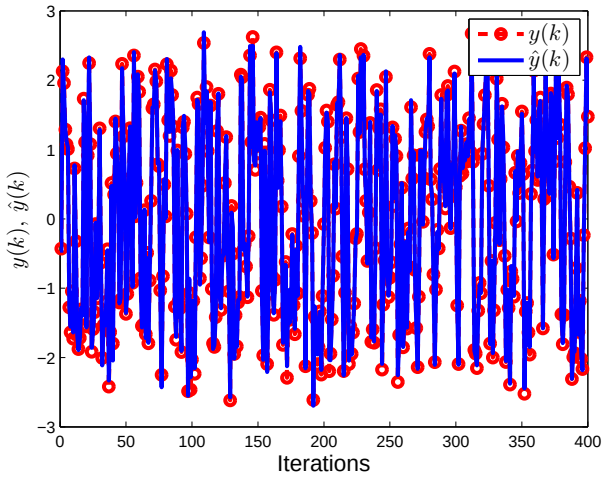

Fig. 8. Identification based on Quadratic SKR (lin-rbf), $m_1 = 3$, $n_1 = 5$.

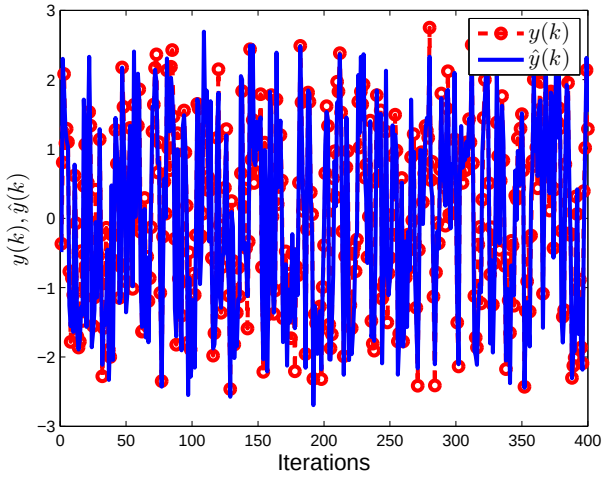

Fig. 10. Identification based on the gaussian Quadratic SVR, $m_1 = 2$, $n_1 = 2$.

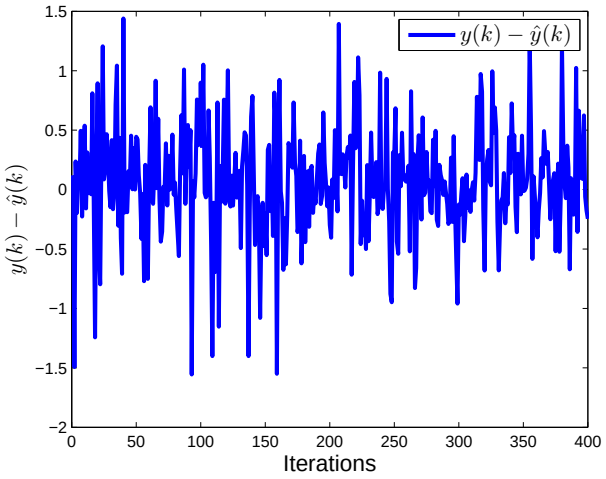

Fig. 11. Identification errors based on the gaussian Quadratic SVR, $m_1 = 2$, $n_1 = 2$.

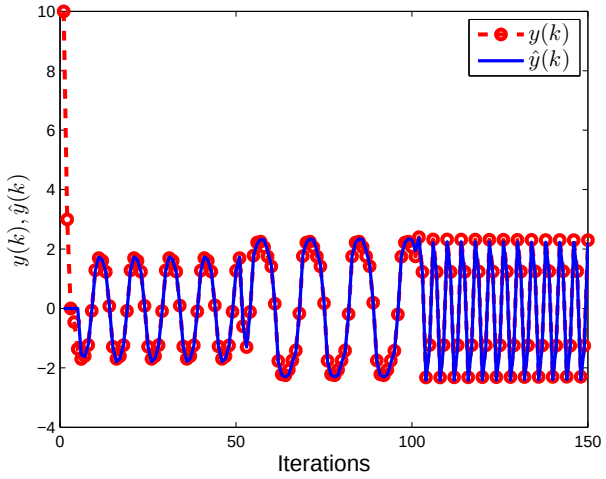

Fig. 12. Evolution of the system output y and the model output $\hat{y}$ based on Quadratic SKR (lin-rbf).

Based on these results, we notice that the suggested approach quadratic SKR (lin-rbf) presents the best performance in this complexe example. Therefore, in general two kernels were sufficient for all complexe non linear system. In fact, we have considered several systems and we have obtained that the identification performances using two kernel were acceptable in the almost taking examples.

2) Performance evaluation: In system identification, we know that the fundamental goal is to produce a model that captures the true dynamics and predicts accurately the output for unseen data. The model identified using a finite training datasets should not just have good accuracy over the training datasets, but also it must be tested on an independent datasets (ie. we search a method able to generalize). To guarantee this objective, we have taken the following command signal:

$$\begin{cases} u(i) = 0.5sin(\frac{\pi}{2} + \frac{i}{5\pi}) & for\ i = 1:50 \\ u(i) = sin(\frac{\pi}{2} + \frac{i}{7\pi}) & for\ i = 51:100 \\ u(i) = sin(i\frac{\pi}{2}) & for\ i = 101:150 \end{cases} \quad (22)$$

Figure 12 presents the evolution of the system output and the estimated output based on on Quadratic SKR (lin-rbf) method. We remark that the model keeps track, suitably, the system behavior.

To assess the effectiveness and robustness of the proposed method, a white noise has been added to the system output. We will discus the R^2_{tot} values with different signal to noise ratio (SNR) value given by the following definition:

$$SNR = 10log_{10}(\frac{\mu^2}{\sigma^2}) \quad (23)$$

where μ is the signal mean and σ is the standard deviation of the noise.

Table IV presents the total correlation coefficient values R^2_{tot} for different Signal-to-Noise Ratio (SNR) value. We remark that, the models are less accurate because R^2_{tot} values decreases gradually as the SNR decreases. Indeed, in the case of quadratic SKR (lin-rbf) method, R^2_{tot} is equal to 93.1945% when SNR is equal to $40\,dB$ and in the case of ε-insensitive SKR (lin-rbf) method, this criterion is equal to 92.9460%. Nevertheless, Based on quadratic SKR (lin-rbf), this parameter remainder equal to 84.2258% with a SNR reached $15\,dB$ and it is equal to 82.1785% adopting the ε-insensitive SKR (lin-rbf) technique. This demonstrates clearly that our proposed identification approach is robust.

We conclude, as well, that the quadratic SKR (lin-rbf) suggested method presents the best performance compared to others proposed methods (quadratic SKR (rbf-rbf), ε-insensitive SKR (lin-rbf) and ε-insensitive SKR (rbf-rbf)) in the presence or not of disturbance.

VI. Conclusion

The main contribution is to advocate the transition from Support Vector Regression (SVR) that uses to train one kernel to Support Kernel Regression (SKR) that uses to train multiple kernel. Thus, based on multi-kernel approach, a new SKR method was developed for regression and, on particularly, applied to nonlinear system identification. The basic idea consists in decomposing the regressor on several blocks and for each bloc a kernel function is adopted. Several kernel configuration were considered and consequently several method were suggested: quadratic SKR (lin-rbf), quadratic SKR (rbf-rbf), ε-insensitive SKR (lin-rbf) and ε-insensitive SKR (rbf-rbf). These blockwise training algorithms, which combine several type of kernel, contribute to good results and guarantee good identification performance with its remarkable robustness on disturbance rejection compared to performance obtained based on quadratic SVR and ε-insensitive SVR methods based on unique kernel function.

References

[1] Lauer, F. (2008) 'From Support Vector Machines to Hybrid System Identification', thèse doctorat de l'Université Henri Poincaré-Nancy1. spécialité automatique.

[2] Taouali O., Elaissi I., Garna T. and Hassani M., (2010) "A new approach for identification of MIMO non linear system with RKHS model", *WSEAS TRANSACTIONS on Information Science and Applications*.

[3] Vapnik V. (1995), "The Nature of Statistical Learning Theory", *Springer, N.Y.*, ISBN 0-387-94559-8.

TABLE IV. R^2_{tot} VALUES FOR DIFFERENT SNR VALUES

SNR	$m_1 = 2, n_1 = 2$		$m_1 = 3, n_1 = 5$		$m_1 = 3, n_1 = 4$	
	Quadratic SVR rbf	ε-insensitive SVR rbf	Quadratic SKR lin-rbf	Quadratic SKR rbf-rbf	ε-insensitive SKR lin-rbf	ε-insensitive SKR rbf-rbf
40	89.8398	90.0536	93.1945	85.9495	92.9460	89.2227
30	88.3805	88.5461	92.6527	85.7990	93.2189	88.7011
25	87.5586	87.0366	90.4649	85.5229	90.6071	87.0167
20	85.8702	85.7172	87.7633	81.8713	88.2621	87.2255
17.5	81.2848	82.6806	86.9762	82.7351	82.0828	82.6303
15	77.6714	78.4680	84.2258	75.7419	82.1785	81.1684
14	73.7845	72.4141	81.1170	74.1374	79.8626	78.3065
13	75.4385	75.2928	79.5095	74.8702	77.6537	75.8322
12	67.8851	71.3331	77.4600	71.2052	77.0699	75.2551
10	62.5669	67.9154	73.4359	63.4610	72.9823	69.0187

[4] Lee, K. L. and Billings, S. A. (2002)'Time series prediction using support vector machines, the orthogonal and the regularized orthogonal least-squares algorithms', *International Journal of Systems Science, Volume 33, Issue 10, pages 811–821*

[5] Zidi, S., Maouche, S. and Hammadi, S. (2006) 'Nouvelle approche pour la regulation des réseaux de transport multimodal', *6eme Conférence Francophone de MOdélisation et SIMulation*, MOSIM'06, Rabat, Maroc.

[6] ElFerchichi, S., Laabidi, K., Zidi, S. and Maouche S. (2009) 'Feature Selection using an SVM learning machine', *Signals, Circuits and Systems (SCS)*.

[7] Tarhouni, M., Laabidi, K., Zidi, S. and Lahmari-Ksouri, M. (2010a) 'Surveillance des systems complexes par les séparateurs à vaste marges, *8me Conférence Internationale de MOdélisation et SIMulation (MOSIM'10)*, Hammamet, Tunisie.

[8] Hacib, T., Acikgoz, H., Le Bihan, Y., Mekideche M.R., Meyer, O. and Pichon, L., (2010) "Support vector machines for measuring dielectric properties of materials", *COMPEL: The International Journal for Computation and Mathematics in Electrical and Electronic Engineering*, Vol. 29 Iss: 4, pp.1081 - 1089.

[9] Burge C., (1998) "A Tutorial on Support Vector Machines for Pattern Recognition, Data Mining and Knowledge Discovery ", *vol 2, pp 121–167*.

[10] Smola A. J., Scholkopf B., (1998) "A Tutorial on Support Vector Regression", NeuroCOLT2 Technical Report Series, NC2- TR-1998-030. Produced as part of the ESPRIT Working Group in Neural and Computational Learning II, NeuroCOLT2- 27150.

[11] Diosan, L., Oltean, M., Rogozan, A., and Pecuchet, J.P. (2007) 'Improving SVM Performance Using a Linear Combination of Kernels', B. Beliczynski et al. (Eds.): ICANNGA 2007, Part II, LNCS 4432, pp. 218-227,. Springer–Verlag Berlin Heidelberg.

[12] Lanckriet, G. R.G., Cristianini, N., Bartlett, P., El Ghaoui, L. and . Jordan, M.I (2004) 'Learning the Kernel Matrix with Semidefinite Programming', *Journal of Machine Learning Research*, Vol. 5, pp. 27-72.

[13] Pelckmans K., Goethals I. , De Brabanter J., Suykens J.A.K. and De Moor B., (2005) "Componentwise Least Squares Support Vector Machines", *arXiv:cs/0504086v1 [cs.LG]*.

[14] Tarhouni M., Zidi S., Laabidi K. and Ksouri-Lahmari M. (2012), "Least squares support kernel machines (LS-SKM) for identification", *Int. J. of Modelling, Identification and Control, Vol.17, No.1, pp.68 - 77*.

[15] Wahba G. (1990), "Splines Models for Observational Data", *Series in Applied Mathematics , Vol. 59, SIAM*.

[16] Smola A. J., Scholkopf B. (2004), A tutorial on support vector regression, Statistics and Computing 14:199222, 2004, Kluwer Academic Publishers, Manufactured in The Netherlands.

[17] Chen L. and Narendra K. S., (1999)'On a new Approach to the design of tracking controllers for nonlinear dynamical systems', *Proceedings of the American Control Conference. Sans Diego, California, p. 3534-3538.*

[18] Tlili, B., Bouani, F., Ksouri, M. (2008) 'Identification des systmes non linaires par des modles de type NARMA', *Journal Europendes Systmes Automatiss JESA, vol. 42, n10, pp. 1279–1309.*

[19] Haber, R. and Unbehauen, H. (1990) 'Structure identification of Non linear systemsA survey on Input/output Approaches," in *Automatica*, (Vol 26, issue 4), pp. 651–677.