The 8th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications
24-26 September 2015, Warsaw, Poland

Nonlinear Signal Reconstruction Based on Recursive Moving Window Kernel Method

Leonid Lyubchyk, Vladislav Kolbasin, Roman Shafeyev
National Technical University "Kharkiv Polytechnic Institute",
Frunze str. 21, Kharkov, 61002, Ukraine
lyubchyk@kpi.kharkiv.edu

***Abstract*—Reconstruction problem for signals generated by discrete nonlinear dynamic system is considered via unified approach to recurrent kernel-based dynamic systems. In order to prevent the model complexity increasing under on-line identification, the reduced order model kernel method is proposed and proper recurrent Least-Square identification algorithms are designed along with conventional regularization technique. The recurrent version of Moving Window Kernel Method is also considered and suitable identification algorithm is developed, which has tracking properties and may be successfully used for on-line identification of nonlinear and nonstationary signal reconstruction.**

Keywords—kernel methods, machine learning, nonlinear signal reconstruction, system identification, support vector machine, recurrent least-squares

I. INTRODUCTION

The problems of signal reconstruction from discrete samples are of a great importance in different applications, such as signal processing, automatic control, econometrics etc. [1]. At that recovered signal is usually treated as an output of the discrete dynamical system with unknown mapping function that allows reducing the signal reconstruction to an equivalent system identification problem using the signal measurements. The most difficult task is the nonstationary signal reconstruction generated by nonlinear models, for example, under the signal processing in data acquisition and information measurement systems with nonlinear sensors [2-4].

The efficient methods based on classical identification approaches were developed for linear signal generating model [5]. With respect to the nonlinear signal reconstruction the most difficult problem is a model structure choice. At that the dynamic nonlinear mapping recovery using the sample data in classical framework leads to the multiparameter estimation problem which becomes very complicated due to "curse of dimension" difficulties. It's stipulated the expedience of nonparametric methods and intelligent data analysis approaches such as artificial neural networks and machine learning, which usually needs a long sample for training. Most efficient under the short training sample are Support Vector Machine (SVM) [6] and Kernel Methods (KM) [7] approaches, which produce nonlinear and nonparametric versions of conventional identification algorithms.

Kernel identification methods, based on the idea of input data implicit nonlinear transformation into high-dimensional (theoretically infinite) feature space, ensure the possibility of nonlinear model approximation. Using the Mercer's theorem, the feature vectors are chosen so that its scalar products in feature space are the positive definite kernel functions. At that the identified model may be represented in a nonparametric form as linear combination of kernel function, though the weighting coefficients (auxiliary variables) may be computed without making direct reference to feature vectors [7]. Such an idea is proved to be very effective for nonlinear system identification [8].

In the initial version of KM the dimension of auxiliary variables vector as well as estimated model complexity increases proportionally to the measurements sample length, which make such approach unsuitable for on-line signal processing. In order the model complexity restriction, it is desirable to use the auxiliary vector of fixed dimension along with recurrent version of identification algorithm [9].

In [10] the SVM approach to Recursive Least-Square Kernel Method (RLSKM) has been considered, which seems to be very effective for nonstationary generating system identification. Sequential sparsification procedure was proposed in [11], and may be viewed as a form of regularization and ensure the restriction of the rate of model complexity increasing. In this way the resulting RLSKM algorithm reduces the order of the feature space. Another approach, known as a Moving Window Kernel Method (MWKM) [12], used at any time instant only fixed size subset of measured sample.

In this paper a unified approach to recurrent kernel identification algorithms design using moving window of measurements is proposed, which may be successfully used for nonlinear and non-stationary signal reconstruction. In order to fix the auxiliary vector dimension, the reduced order model KM is proposed and proper recurrent identification algorithms are designed. Finally, the recurrent version of MWKM regarding to auxiliary variables, is also developed along with suitable moving kernel matrix updating algorithms.

II. Problem Statement

Consider the signal model $\{x_k\}_{k=1}^n$, generated by nonlinear discrete dynamic system

$$x_{k+1} = f(x_k) + \varepsilon_k \,, k = 0, \ldots \tag{1}$$

where $f(\cdot)$ is unknown nonlinear mapping function and ε_k is a white noise discrete random process, $\mathbf{E}\{\varepsilon_k\} = 0,\ \mathbf{E}\{\varepsilon_k^2\} = \sigma^2$.

The problem of signal reconstruction is equivalent to the generating system (1) identification using signal measurements $\{y_k, x_k\}_{k=0}^n$, where $y_k = x_{k+1}$.

In the KM framework introduce the parameterized system identification model is:

$$y_k = \hat{f}(x_k) = \varphi^{\mathrm{T}}(x_k) w + \varepsilon_{k,} \quad k = \overline{0, n,} \tag{2}$$

where $\hat{f}(x)$ - model approximation, $\varphi : \mathbf{R}^1 \to \mathbf{R}^{\mathrm{M}}$ - nonlinear feature map, which transform the signal value into high-dimensional feature vector $\varphi(x) \in \mathbf{R}^{\mathrm{M}}$, and $\mathbf{w} \in \mathbf{R}^{\mathrm{M}}$ is an unknown model coefficient vector.

Equation (2) may be represented in matrix form as $\mathbf{y}_n = \mathbf{\Phi}_{n-1}^{\mathrm{T}} \mathbf{w} + \boldsymbol{\varepsilon}_n$, where $\mathbf{y}_n = (y_0\ y_1 \ldots y_n)^{\mathrm{T}}$ - measurement vector, $\boldsymbol{\varepsilon}_n = (\varepsilon_0\ \varepsilon_1 \ldots \varepsilon_n)^{\mathrm{T}}$ – noise vector, $\mathbf{\Phi}_{n-1} = \left(\varphi(x_0) \quad \varphi(x_1) \quad \ldots \quad \varphi(x_{n-1})\right)$ – feature matrix.

According with Mercer's theorem, feature vectors are taken hereby that its scalar products in feature space will be positive definite kernel functions $\varphi^{\mathrm{T}}(x_i)\varphi(x_j) = \kappa(x_i, x_j),\ i, j = \overline{1, n},$ for example, Gaussian kernel $\kappa(x, x') = \exp\{-\mu(x - x')^2\}$, where p, μ – tuning parameters.

In accordance with KM technique introduce the kernel matrix $\mathbf{K}_n = \mathbf{\Phi}_n^{\mathrm{T}} \mathbf{\Phi}_n$, which may be fined directly without reference to the feature vectors, because $\mathbf{K}_n = \left\| k_{i,j} \right\|, k_{i,j} = \kappa(x_i, x_j), i, j = \overline{1, n}.$ Also taking into consideration the auxiliary (dual) variables vector $\boldsymbol{\lambda}_n \in \mathbf{R}^n$, such as $\mathbf{w}_n = \mathbf{\Phi}_{n-1} \boldsymbol{\lambda}_n$. At that the nonlinear system (1) model estimation and one step signal prediction may be represented as

$$\hat{x}_{k+1} = \hat{f}(x_n) = \phi^{\mathrm{T}}(x_n) \mathbf{\Phi}_{n-1} \mathbf{w}_n = \mathbf{k}_{n-1}^{\mathrm{T}}(x_n) \boldsymbol{\lambda}_n, \tag{3}$$

where $\hat{x}_{k+1}$ – one step ahead signal prediction, $\mathbf{k}_{n-1}^{\mathrm{T}}(x_n) = (\kappa(x_n, x_0)\ \kappa(x_n, x_1) \ldots \kappa(x_n, x_{n-1}))$ - kernel vector and $\boldsymbol{\lambda}_n$ is an auxiliary variables vector estimate at instant k, which should be obtained by the measurements sample $\{y_k, x_k\}_{k=0}^{n-1}$.

The purpose of this paper is the recurrent KM identification algorithms design, which ensures on-line $\boldsymbol{\lambda}_n$ estimates. We will consider the following alternatives of recurrent KM identification:

- *Recurrent KM with growing window* $\boldsymbol{\lambda}_{n+1} = F(\boldsymbol{\lambda}_{n+1}, y_{n+1}, \mathbf{K}_n)$, which use the complete training sample $\{y_k, x_k\}_{k=0}^n$ for dual variable $\boldsymbol{\lambda}_n$ estimation.
- *Recurrent KM reduced order model* $\widetilde{\boldsymbol{\lambda}}_{n+1} = F(\widetilde{\boldsymbol{\lambda}}_{n+1}, y_{n+1}, \widetilde{\mathbf{K}}_{n,r})$, which also use the complete measurements sample but with respect to fix dimension dual variables $\widetilde{\boldsymbol{\lambda}}_n \in \mathbf{R}^r$ and related kernel matrix $\overline{\mathbf{K}}_{n,r}$.
- *Recurrent KM with moving widow* $\overline{\boldsymbol{\lambda}}_{n+1} = F(\overline{\boldsymbol{\lambda}}_{n+1}, y_{n+1}, \overline{\mathbf{K}}_{n,s}), \overline{\boldsymbol{\lambda}}_n \in \mathbf{R}^{\mathrm{s}},$ which uses the moving window measurements sample $\{y_k, x_k\}_{k=n-s}^n$ and kernel matrix $\overline{\mathbf{K}}_{n,s}$, built up on the respective measurements.

III. Recurrent Kernel Identification With Growing Window

In accordance with general SVM approach [7], system (1) identification problem using the full measurement sample (growing window) may be reduced to the following constrained optimization problem with regularized estimation cost function with regularization parameter $\gamma > 0$:

$$\begin{aligned} & J_n(\mathbf{w}) = \frac{1}{2} \mathbf{w}^{\mathrm{T}} \mathbf{w} + \frac{1}{2} \gamma \cdot \boldsymbol{\varepsilon}^{\mathrm{T}} \boldsymbol{\varepsilon} \to \min_{\mathbf{w}, \boldsymbol{\varepsilon}}, \\ & \mathbf{y}_n = \mathbf{\Phi}_{n-1}^{\mathrm{T}} \mathbf{w} + \boldsymbol{\varepsilon}_n. \end{aligned} \tag{4}$$

The optimization problem (4) is solved using Lagrange function with dual multipliers $\boldsymbol{\lambda} \in \mathbf{R}^n$

$$\begin{aligned} L_n(\mathbf{w}, \boldsymbol{\lambda}) = & \frac{1}{2} \mathbf{w}^{\mathrm{T}} \mathbf{w} + \frac{1}{2} \gamma \cdot \boldsymbol{\varepsilon}^{\mathrm{T}} \boldsymbol{\varepsilon} + \\ & + \boldsymbol{\lambda}^{\mathrm{T}} (\mathbf{y}_n - \mathbf{\Phi}_{n-1}^{\mathrm{T}} \mathbf{w} - \boldsymbol{\varepsilon}_n). \end{aligned} \tag{5}$$

In such a way a solution may be obtained in the explicit form $\mathbf{w}_n = \mathbf{\Phi}_{n-1}\boldsymbol{\lambda}_n$, $\boldsymbol{\varepsilon}_n = \gamma^{-1}\boldsymbol{\lambda}_n$, therefore the dual variables estimate $\boldsymbol{\lambda}_n$ takes the form of ridge regression:

$$\hat{\boldsymbol{\lambda}}_n = \left(\gamma^{-1}\mathbf{I}_n + \mathbf{K}_{n-1}\right)^{-1}\mathbf{y}_n = \mathbf{K}_{n-1}^{-1}(\gamma)\mathbf{y}_n, \qquad (6)$$

where $\mathbf{I}_n$ is an identity $n \times n$ matrix, and $\mathbf{K}_{n-1}(\gamma)$ is a regularized kernel matrix.

The recurrent estimation for dual variables $\boldsymbol{\lambda}_{n+1} = \mathbf{K}_n^{-1}(\gamma)\mathbf{y}_{n+1}$ at instant $n+1$ may be easily obtained using Sherman-Morrison-Woodbury formula [12] for the regularized kernel matrix $\mathbf{K}_n(\gamma)$:

$$\begin{aligned}
\mathbf{K}_n(\gamma) &= \left(\begin{array}{c|c} \mathbf{K}_{n-1}(\gamma) & \mathbf{k}_{n-1}(x_n) \\ \hline \mathbf{k}_{n-1}^{\mathrm{T}}(x_n) & \gamma^{-1} + k_{n,n} \end{array}\right), \\
\mathbf{K}_n^{-1}(\gamma) &= \left(\begin{array}{c|c} \mathbf{K}_{n-1}^{-1}(\gamma) + \delta_n^{-1}\mathbf{Q}(x_n) & \mathbf{P}(x_n) \\ \hline \mathbf{P}(x_n) & \delta_n^{-1} \end{array}\right),
\end{aligned} \qquad (7)$$

where

$$\begin{aligned}
\mathbf{Q}(x_n) &= \mathbf{K}_{n-1}^{-1}(\gamma)\mathbf{k}_{n-1}(x_n)\mathbf{k}_{n-1}^{\mathrm{T}}(x_n)\mathbf{K}_{n-1}^{-1}(\gamma), \\
\mathbf{P}(x_n) &= -\delta_n^{-1}\mathbf{K}_{n-1}^{-1}(\gamma)\mathbf{k}_{n-1}(x_n), \\
\delta_n &= \gamma^{-1} + k_{n,n} - \mathbf{k}_{n-1}^{\mathrm{T}}(x_n)\mathbf{K}_{n-1}^{-1}(\gamma)\mathbf{k}_{n-1}(x_n).
\end{aligned} \qquad (8)$$

Consequently the recurrent algorithm for dual variables estimates may be represented as

$$\begin{aligned}
\hat{\boldsymbol{\lambda}}_{n+1} &= \left(\begin{array}{c} \hat{\boldsymbol{\lambda}}_n - \delta_n^{-1} z_{n+1}(\hat{\boldsymbol{\lambda}}_n)\mathbf{K}_{n-1}^{-1}(\gamma)\mathbf{k}_{n-1}(x_n) \\ \hline \delta_n^{-1} z_{n+1}(\hat{\boldsymbol{\lambda}}_n) \end{array}\right), \\
z_{n+1}(\hat{\boldsymbol{\lambda}}_n) &= y_{n+1} - \omega_n(\hat{\boldsymbol{\lambda}}_n),
\end{aligned} \qquad (9)$$

where $\omega_n(\boldsymbol{\lambda}_n) = \mathbf{k}_{n-1}^{\mathrm{T}}(x_n)\hat{\boldsymbol{\lambda}}_n$, along with regularized kernel matrix $\mathbf{K}_n(\gamma)$ updating procedure (8).

As a result the time-series model estimate and one-step ahead prediction may be obtained as (3) with corresponding dual auxiliary variables estimates $\hat{\boldsymbol{\lambda}}_n$.

It is obvious that the dimension of dual variables estimates $\hat{\boldsymbol{\lambda}}_n$ and thereafter the estimated model $\hat{f}(x)$ complexity grows proportionally to the measurements sample length, which make the considered scheme unsuitable for on-line application.

IV. Reduced Kernel Model Recurrent Identification

It is desirable to restrict kernel matrix dimension with the purpose of over-fitting preventing and to keep the complexity of model bounded. For such a purpose a sparsification process was proposed [11], in which a next sample is only admitted into the kernel matrix if its image in feature space cannot be sufficiently well approximated by combining the previously admitted samples. We consider another approach in which the reduced order model is formed from the pre-established linear independent feature vectors, corresponding to the fixed input vectors. In such a way the dimension of kernel matrix is fixed in advance so the model complexity doesn't growing under identification process.

Consider the reduced order feature matrix $\widetilde{\mathbf{\Phi}}_r$ consist of r constant linear independent basic (support) feature vectors $\widetilde{\mathbf{\Phi}}_r = \left(\widetilde{\varphi}(x_1)\ \widetilde{\varphi}(x_2)\ldots\widetilde{\varphi}(x_r)\right)^{\mathrm{T}}$, whish has been initially constructed from pre-established input samples $\widetilde{x}_i, i = \overline{1,r}$, selected in such a way, that $\operatorname{rang}(\widetilde{\mathbf{\Phi}}_r) = r$. In practice, such a "feature" condition may be easily verified using equivalent "kernel" condition $\operatorname{rang}(\widetilde{\mathbf{K}}_r) = r$ using appropriate kernel matrix $\widetilde{\mathbf{K}}_r = \widetilde{\mathbf{\Phi}}_r^{\mathrm{T}}\widetilde{\mathbf{\Phi}}_r$.

Any feature vector from measured sample may be represent as linear combination of basic (support) feature vectors $\varphi(x_i) = \sum_{j=1}^{r} a_{ij}\varphi(\widetilde{x}_j),\ i = \overline{1,n}$, or, in matrix form, as $\mathbf{\Phi}_n = \widetilde{\mathbf{\Phi}}_r\mathbf{A}_n^{\mathrm{T}}$, where $\mathbf{A}_n = \left\|a_{ij}\right\|$ is a matrix of corresponding expansion coefficients, which may be obtained by the minimum least-squares approximation :

$$\Delta_{\mathbf{A}} = \left\|\mathbf{\Phi}_n - \widetilde{\mathbf{\Phi}}_r\mathbf{A}_n^{\mathrm{T}}\right\|^2 \to \min_{\mathbf{A}_n}. \qquad (10)$$

The solution of (10) may be easily obtained in the explicit form as $\mathbf{A}_n = \widetilde{\mathbf{K}}_{n,r}\widetilde{\mathbf{K}}_r^{-1}$, where $\widetilde{\mathbf{K}}_{n,r} = \mathbf{\Phi}_n^{\mathrm{T}}\widetilde{\mathbf{\Phi}}_r$.

At that the attainable approximation accuracy are determined by $\min\Delta_{\mathrm{A}} = \left\|\mathbf{K}_n - \widetilde{\mathbf{K}}_{n,r}\mathbf{K}_r^{-1}\widetilde{\mathbf{K}}_{n,r}^{\mathrm{T}}\right\|^2$.

Furthermore, the model parameters vector may be express from reduced order dual variables $\widetilde{\boldsymbol{\lambda}}_n \in \mathbf{R}^r$

$$\mathbf{w}_n = \mathbf{\Phi}_{n\text{-}1}\boldsymbol{\lambda}_n = \widetilde{\mathbf{\Phi}}_r\mathbf{A}_{n-1}^{\mathrm{T}}\boldsymbol{\lambda}_n = \mathbf{\Phi}_r\widetilde{\boldsymbol{\lambda}}_n, \qquad (11)$$

where $\widetilde{\boldsymbol{\lambda}}_n = \mathbf{A}_{n-1}^{\mathrm{T}}\boldsymbol{\lambda}_n$.

Corresponding reduced order identified model and prediction function are

$$\hat{x}_{n+1} = \hat{f}_n(x_n) = \phi^{\mathrm{T}}(x_n)\boldsymbol{\Phi}_r\tilde{\boldsymbol{\lambda}}_n = \mathbf{k}_r^{\mathrm{T}}(x_n)\tilde{\boldsymbol{\lambda}}_n,$$
$$\mathbf{k}_r(x_n) = \left(\kappa(x_n,\tilde{x}_1)\ldots\kappa(x_n,\tilde{x}_r)\right)^{\mathrm{T}}. \quad (12)$$

In such a way reduced order vector of dual variables estimate $\widetilde{\boldsymbol{\lambda}}_n$ may be obtained via SVM approach as a solution of suitable optimization problem.

A. Nonregularized case $\gamma^{-1} = 0$.

Using representation (11), the appropriate cost function takes the form

$$J_n = \frac{1}{2}\left\|\mathbf{y}_n - \boldsymbol{\Phi}_{n-1}^{\mathrm{T}}\mathbf{w}\right\|^2 = \\ = \frac{1}{2}\left\|\mathbf{y}_n - \mathbf{A}_{n-1}\widetilde{\boldsymbol{\Phi}}_r^{\mathrm{T}}\widetilde{\boldsymbol{\Phi}}_{n-1}\widetilde{\boldsymbol{\lambda}}\right\|^2 \to \min_{\widetilde{\boldsymbol{\lambda}}}. \quad (13)$$

The solution of (13) is $\widetilde{\boldsymbol{\lambda}}_n = \left(\mathbf{A}_{n-1}\widetilde{\mathbf{K}}_r\right)^{+}\mathbf{y}_n$, where "+" denotes Moor-Penrose generalized inversion. Taking into account that $\mathbf{A}_{n-1} = \widetilde{\mathbf{K}}_{n-1,r}\widetilde{\mathbf{K}}_r^{-1}$, due to the generalized inversion properties, the reduced order dual variables estimate may be obtained as following: $\widetilde{\boldsymbol{\lambda}}_n = \widetilde{\mathbf{K}}_r^{-1}\left(\mathbf{A}_{n-1}\right)^{+}\mathbf{y}_n = \left(\widetilde{\mathbf{K}}_{n-1,r}\right)^{+}\mathbf{y}_n$.

In such a way a recurrent algorithm for $\tilde{\boldsymbol{\lambda}}_{n+1} = \left(\tilde{\mathbf{K}}_{n,r}\right)^{+}\mathbf{y}_{n+1}$ updating may be easily derived.

As far as $\widetilde{\mathbf{K}}_{n,r}^{\mathrm{T}} = \left(\widetilde{\mathbf{K}}_{n-1,r}^{\mathrm{T}} : \mathbf{k}_n(x_n)\right)$, one can use the known Greville formula [13] for reduced order Moor-Penrose inverse kernel matrix updating:

$$\widetilde{\mathbf{K}}_{n,r}^{+} = \left((\mathbf{I}_r - \mathbf{q}_n\mathbf{k}_n^{\mathrm{T}}(x_n))\widetilde{\mathbf{K}}_{n-1,r}^{+} : \mathbf{q}_n\right),$$
$$\mathbf{q}_n = \mathbf{Y}(x_n)\mathbf{Z}(\widetilde{\mathbf{K}}_{n-1,r})\mathbf{k}_r(x_n),$$
$$\mathbf{Y}(x_n) = \left(\alpha_n\mathbf{I} + \mathbf{K}_r^{\mathrm{T}}(x_n)\mathbf{Z}(\widetilde{\mathbf{K}}_{n-1,r})\mathbf{k}_r(x_n)\right)^{-1} \quad (14)$$
$$\mathbf{Z}(\widetilde{\mathbf{K}}_{n-1,r}) = \mathbf{I}_r - \widetilde{\mathbf{K}}_{n-1,r}^{+}\widetilde{\mathbf{K}}_{n-1,r},$$
$$\alpha_n = 1 - \mathrm{sgn}\left(\mathbf{k}_r^{\mathrm{T}}(x_n)\mathbf{Z}(\widetilde{\mathbf{K}}_{n-1,r})\mathbf{k}_r(x_n)\right)$$

Thereafter recurrent identification algorithm for dual variables estimates is

$$\widetilde{\boldsymbol{\lambda}}_{n+1} = \widetilde{\boldsymbol{\lambda}}_n + \mathbf{q}_n(\mathbf{y}_{n+1} - \mathbf{k}_r^{\mathrm{T}}(x_n)\widetilde{\boldsymbol{\lambda}}_n). \quad (15)$$

B. Regularized case $\gamma > 0$.

Using the introduced representation for unknown model parameters vector $\mathbf{w}$, the regularized estimation cost function for reduced order model will be taken in the following form:

$$J_n = \frac{1}{2}\left\|\mathbf{y}_n - \boldsymbol{\Phi}_{n-1}^{\mathrm{T}}\mathbf{w}\right\|^2 + \gamma^{-1}\mathbf{w}^{\mathrm{T}}\mathbf{w} = \\ = \frac{1}{2}\left\|\mathbf{y}_n - \widetilde{\mathbf{K}}_{n-1,r}\widetilde{\boldsymbol{\lambda}}\right\|^2 + \gamma^{-1}\widetilde{\boldsymbol{\lambda}}^{\mathrm{T}}\widetilde{\mathbf{K}}_r\widetilde{\boldsymbol{\lambda}} \to \min_{\widetilde{\boldsymbol{\lambda}}}. \quad (16)$$

The explicit solution is $\widetilde{\boldsymbol{\lambda}}_n = \mathbf{P}_{n-q,r}^{-1}\mathbf{K}_{n-1,r}^{\mathrm{T}}\mathbf{y}_n$, where $\mathbf{P}_{n-q,r} = \gamma^{-1}\widetilde{\mathbf{K}}_r + \widetilde{\mathbf{K}}_{n-1,r}^{\mathrm{T}}\widetilde{\mathbf{K}}_{n-1,r}$.

Taking into account that

$$\widetilde{\mathbf{K}}_{n,r}^{\mathrm{T}}\widetilde{\mathbf{K}}_{n,r} = \widetilde{\mathbf{K}}_{n-1,r}^{\mathrm{T}}\widetilde{\mathbf{K}}_{n-1,r} + \mathbf{k}_r(x_n)\mathbf{k}_r^{\mathrm{T}}(x_n), \quad (17)$$

the recurrent form for reduced order dual vector estimate $\widetilde{\boldsymbol{\lambda}}_{n+1} = \mathbf{P}_{n,r}^{-1}\widetilde{\mathbf{K}}_{n,r}\mathbf{y}_{n+1}$ may be represented as

$$\widetilde{\boldsymbol{\lambda}}_{n+1} = \widetilde{\boldsymbol{\lambda}}_n + \gamma_n\mathbf{P}_{n-1,r}^{-1}\mathbf{k}_r(x_n)\left(\mathbf{y}_{n+1} - \mathbf{k}_r^{\mathrm{T}}(x_n)\widetilde{\boldsymbol{\lambda}}_n\right),$$
$$\gamma_n = \left(1 + \mathbf{k}_r^{\mathrm{T}}(x_n)\mathbf{P}_{n-1,r}^{-1}\mathbf{k}_r(x_n)\right)^{-1}. \quad (18)$$

Using the matrix inversion lemma [14], the recurrent procedure for inverse matrix updating

$$\mathbf{P}_{n,r} = \mathbf{P}_{n-1,r} + \mathbf{k}_r(x_n)\mathbf{k}_r^{\mathrm{T}}(x_n),$$
$$\mathbf{P}_{n,r}^{-1} = \mathbf{P}_{n-1,r}^{-1} - \gamma_n^{-1}\mathbf{P}_{n-1,r}^{-1}\mathbf{k}_r(x_n)\mathbf{k}_r^{\mathrm{T}}(x_n)\mathbf{P}_{n-1,r}^{-1}. \quad (19)$$

V. Recurrent Kernel Identification Using Moving Window

Moving window KM approach consider for estimation at instant n only last $n-s$ measurement $\mathbf{y}_{n,s} = (y_{n-s+1}\ldots y_n)^{\mathrm{T}}$. consequently $\bar{\boldsymbol{\lambda}}_n \in \mathbf{R}^s$. and the kernel matrix $\overline{\mathbf{K}}_{n,s} = \boldsymbol{\Phi}_{n,s}^{\mathrm{T}}\boldsymbol{\Phi}_{n,s}$ has a fixed dimension $(s \times s)$ and measurement equation takes the form

$$\mathbf{y}_{n,s} = \boldsymbol{\Phi}_{n-1,s}^{\mathrm{T}}\mathbf{w} + \boldsymbol{\varepsilon}_{n,s} = \mathbf{K}_{n-1,s}\bar{\boldsymbol{\lambda}}_n + \boldsymbol{\varepsilon}_{n,s},$$
$$\mathbf{w} = \boldsymbol{\Phi}_{n,s}\bar{\boldsymbol{\lambda}}. \quad (20)$$

Consider the moving estimation cost function includes at any instant $n+1$ *a priori* information term determined by previously estimate at instant n.

The optimization problem is

$$J_{n,s} = \left\| \mathbf{y}_{n+1,s} - \mathbf{K}_{n,s}\bar{\boldsymbol{\lambda}} \right\|^2 + \\ + \gamma^{-1}(\bar{\boldsymbol{\lambda}} - \bar{\boldsymbol{\lambda}}_n)^{\mathrm{T}}\mathbf{K}_{n,s}(\bar{\boldsymbol{\lambda}} - \bar{\boldsymbol{\lambda}}_n) \to \min_{\bar{\boldsymbol{\lambda}}}. \quad (21)$$

Condition of optimality leads to the normal equations:

$$\left(\overline{\mathbf{K}}_{n,s}^{\mathrm{T}}\overline{\mathbf{K}}_{n,s} + \gamma^{-1}\overline{\mathbf{K}}_{n,s}\right)\bar{\boldsymbol{\lambda}}_n = \\ = \overline{\mathbf{K}}_{n,s}^{\mathrm{T}}\mathbf{y}_{n+1,s} + \gamma^{-1}\overline{\mathbf{K}}_{n,s}\bar{\boldsymbol{\lambda}}_n. \quad (22)$$

By identity $\mathbf{A}^{-1}\left(\gamma^{-1}\mathbf{I}_s + \mathbf{A}\right)^{-1}\mathbf{A} = \left(\gamma^{-1}\mathbf{I}_s + \mathbf{A}\right)^{-1}$ the recurrent dual vector estimate takes the form:

$$\bar{\boldsymbol{\lambda}}_{n+1} = \left(\gamma^{-1}\mathbf{I}_s + \overline{\mathbf{K}}_{n,s}\right)^{-1}\left(\gamma^{-1}\bar{\boldsymbol{\lambda}}_n + \mathbf{y}_{n+1}\right). \quad (23)$$

At last, it is necessary to put forward the updating algorithm for inverse regularized moving kernel matrix $\overline{\mathbf{K}}_{n,s}^{-1}(\gamma)$. Using the approach, proposed by [15], consider two step inverse regularized kernel matrix updating algorithm $\overline{\mathbf{K}}_{n-1,s}^{-1}(\gamma) \to \overline{\mathbf{K}}_{n-1,s-1}^{-1}(\gamma) \to \overline{\mathbf{K}}_{n,s}^{-1}(\gamma)$, which use auxiliary "down-sizing" matrix $\overline{\mathbf{K}}_{n-1,s-1}(\gamma)$:

$$\overline{\mathbf{K}}_{n-1,s}(\gamma) = \left(\begin{array}{c|c} \gamma^{-1} + k_{n-s,n-s} & \mathbf{k}_{n-1,s-1}^{\mathrm{T}}(x_{n-s}) \\ \hline \mathbf{k}_{n-1,s-1}(x_{n-s}) & \overline{\mathbf{K}}_{n-1,s-1}(\gamma) \end{array}\right), \quad (24)$$

$$\mathbf{k}_{n-1,s-1}(x_{n-s}) = \left(\kappa_{n-1}(x_{n-s}) \ \ldots \ \kappa_{n-s+1}(x_{n-s})\right)^{\mathrm{T}}.$$

Then at 1-st step of the algorithm including the "downsizing" matrix inverse is:

$$\overline{\mathbf{K}}_{n-1,s-1}^{-1} = \mathbf{R}_s\overline{\mathrm{K}}_{n-1,s}^{-1}\mathbf{R}_s^{\mathrm{T}} - \\ - (\mathrm{e}_1^{\mathrm{T}}\overline{\mathbf{K}}_{n-1,s}^{-1}\mathrm{e}_1)^{-1}\mathbf{R}_s\overline{\mathbf{K}}_{n-1,s}^{-1}\mathbf{e}_1\mathbf{e}_1^{\mathrm{T}}\overline{\mathbf{K}}_{n-1,s}^{-1}\mathbf{R}_s^{\mathrm{T}}, \quad (25)$$

where $\mathbf{R}_s = \left(0_s \vdots \mathbf{I}_{s-1}\right)$, $\mathbf{e}_1 = (1\ldots0)^{\mathrm{T}}$.

Using moving kernel matrix $\mathbf{K}_{n,s}^{-1}(\gamma)$ representation

$$\overline{\mathbf{K}}_{n,s}(\gamma) = \left(\begin{array}{c|c} \overline{\mathbf{K}}_{n-1,s-1}(\gamma) & \mathbf{k}_{n-1,s-1}(x_n) \\ \hline \mathbf{k}_{n-1,s-1}^{\mathrm{T}}(x_n) & \gamma^{-1} + k_{n,n} \end{array}\right), \quad (26)$$

where $\mathbf{k}_{n-1,s-1}(x_n) = \left(\kappa_{n-1}(x_n) \ \ldots \ \kappa_{n-s+1}(x_n)\right)^{\mathrm{T}}$, the 2-nd step of $\overline{\mathbf{K}}_{n,s}^{-1}(\gamma)$. updating is the following:

$$\overline{\mathbf{K}}_{n,s}(\gamma) = \left(\begin{array}{c|c} \overline{\mathbf{K}}_{n-1,s-1}(\gamma) + \mathbf{R}(x_n) & \mathbf{S}(x_n) \\ \hline -\mathbf{S}(x_n) & \delta_n^{-1} \end{array}\right), \quad (27)$$

$$\mathbf{R}(x_n) = \delta_n^{-1}\mathbf{K}_{n-1,s-1}^{-1}(\gamma)\mathbf{k}_{n-1}(x_n)\mathbf{k}_{n-1}^{\mathrm{T}}(x_n)\mathbf{K}_{n-1,s-1}^{-1}(\gamma),$$

$$\mathbf{S}(x_n) = -\delta_n^{-1}\mathbf{K}_{n-1,s-1}^{-1}(\gamma)\mathbf{k}_{n-1,s-1}^{\mathrm{T}}(x_n),$$

$$\delta_n = \gamma^{-1} + k_{n,n} - \mathbf{k}_{n-1,s-1}^{\mathrm{T}}(x_n)\overline{\mathbf{K}}_{n-1,s-1}^{-1}(\gamma)\mathbf{k}_{n-1,s-1}(x_n).$$

VI. CONCLUSION

Recurrent KM approach for nonlinear signal reconstruction combining with model reduction technique leads to identification algorithms efficiency improvement. The advantage of such an approach is that the identified model complexity does not increase as the number of measurements increases, so recurrent KM algorithms may be used for non-stationary signals reconstruction. The algorithm regularization parameter and kernel tuning parameter optimization may be performed via cross-validation technique.

REFERENCES

[1] W. J. Fitzerald, R. L. Smith, A. T. Walden and P. C. Young (Eds.), *Nonlinear and Nonstationary Signal Processing,* Cambridge,University Press, New York, 2000.

[2] T. Bakó, T. Dabóczi, "Reconstruction of nonlinearly distorted signals with regularized inverse characteristics," *IEEE Instrumentation and Measurement Technology Conference*, Budapest, Hungary, May 21-23, 2001.

[3] X. Liu, G. Wei, J. Sun, D. Liu, "Nonlinear multifunctional sensor signal reconstruction based on least squares support vector machines and total least squares algorithm," *Journal of Zhejiang University – Science A: Applied Physics & Engineering*, vol. 4. no. 10, 2009, pp. 497-503.

[4] J. Sun, D. Liu, X. Liu, G. Wei, "Input signal reconstruction based on improved moving least squares for nonlinear multiple-input multiple-output sensor," in *Proceedings of the International Joint Conference ICROS-SICE*, 2009.

[5] L. Ljung, *System Identification: Theory for the User*, Second ed., Prentice Hall, New Jersey, 1999.

[6] V. Vapnik, *Statistical Learning Theory*, New-York, Wiley, 1998.

[7] B. Scholkopf, A. Smola, *Learning with Kernels*, Cambridge, MA: MIT Press, 2002.

[8] M. Espinoza, J. Suykens, and B. Moor, "Kernel based partially linear models and nonlinear identification," *IEEE Transactions on Automatic Control*, vol. 50, issue 10, 2005, pp. 1602-1606.

[9] J. Kivinen, A. Smola, R. Williamson, "Online learning with kernels," *IEEE Transactions on Signal Processing*, vol. 52, issue 2, 2004, pp. 2165-2176.

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

[11] Y. Engel, S. Mannor, R. Meir, "The kernel recursive least-squares algorithm", *IEEE Transactions on Signal Processing*, vol. 52, issue 8, 2004, pp. 2275-2285.

[12] G. H. Golub, C. F. Van Loan, *Matrix Computations*, Baltimore MD, Johns Hopkins University Press, 1989.

[13] A. Ben-Israel, T. Greville, *Generalized Inverses. Theory and Applications*, Springer, 2003.

[14] S. Haykin, *Adaptive Filter Theory*, 3rd ed., Prentice Hall, 1996.

[15] S. Van Vaerenbergh, V. Javier, I. Santamar, "Nonlinear system identification using a new sliding window kernel algorithm," *Journal of Communications*, vol. 2, no. 3, 2007, pp. 1-8.