Knowledge Discovery and Data Mining

KERNEL-BASED METHODS FOR NON-STATIONARY TIME-SERIES IDENTIFICATION AND PREDICTION

Leonid Lyubchyk, Vladyslav Kolbasin

Abstract: Identification and prediction problem of nonlinear time-series generated by discrete dynamic system is considered via Kernel Method approach. A unified approach to recurrent kernel identification algorithms design is proposed. In such a way a recurrent modification of initial Kernel Method with growing windows is considered. In order to prevent the model complexity increasing under on-line identification, the reduced order model kernel method is proposed and proper recurrent identification algorithms are designed along with conventional regularization technique. Such an approach leads to a new type of Recursive Least-Square Kernel Method identification algorithms. Finally, the recurrent version of Sliding Window Kernel Method is also developed along with suitable identification algorithms. The proposed algorithm has tracking properties and may be successfully used for on-line identification of nonlinear non-stationary time-series.

Keywords: identification, kernel methods, machine learning, nonlinear model, prediction, recurrent least-squares, support vector machine, time-series

ACM Classification Keywords: G. Mathematics of Computing: G.1 Numerical Analysis: Least squares approximation, nonlinear approximation, G.3 Probability and Statistics: Time series analysis

Introduction

The problems of time-series identification and prediction are of a great importance in different applications, such as signal processing, automatic control, econometrics etc. Most efficient methods based on classical identification approaches [Ljung, 1999] were developed for linear time-series, described by autoregressive - moving average (ARMAX) or, in most general case, by discrete-time state-space dynamic model. With respect to the nonlinear time-series, for instance, discrete-time chaotic processes, the most difficult problem is the admissible model selection and a model structure choice [Dorffner, 1996]. At that the complex dynamic nonlinear mapping recovery using the sample data in the classical framework of parameterized model application leads to the multi-parameter 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) [Vapnik, 1998] and Kernel Methods (KM) [Scholkopf, Smola, 2002] approaches, which produce non-linear and non-parametric 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 complex nonlinear model high quality 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 ("kernel trick") [Scholkopf, Smola, 2002]. Such an idea is proved to be very effective for non-linear identification [Espinoza, Suykens, De Moor, 2005].

In the initial version of Kernel Method the dimension of auxiliary variables vector as well as estimated model complexity increases proportionally to the training sample length, which make it unsuitable for on-line application. In order the model complexity restriction and simplification of computations, it is desirable to use the auxiliary vector of fixed dimension along with recurrent version of identification algorithm [Kivinen, Smola, Williamson, 2004]. Moreover, the recurrent on-line learning successfully used for non-stationary time-series identification.

In [Suykens, Van Gestel, De Brabanter, De Moor, Vandewalle, 2002] the SVM approach to Recursive Least-Square Kernel Method (RLSKM) has been considered. Sequential sparsification procedure was proposed in [Engel, Mannor, Meir, 2004], which 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 Sliding Window Kernel Method (SWKM) [Vaerenbergh, Javier, Santamar, 2007], used at any time instant only fixed size subset of training sample.

In this paper a unified approach to recurrent kernel identification algorithms design is proposed. At first, we consider a recurrent modification of initial KM with “growing” windows. In order to fix the auxiliary vector dimension, the reduced order model KM is proposed and proper recurrent identification algorithms are designed. Finally, the full recurrent version of SWKM regarding to auxiliary variables, is also developed along with suitable sliding kernel matrix updating algorithms.

Problem Statement

Consider the time-series $\{x_k\}_{k=1}^{n}$, 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 function and ε_k is a noise discrete random process, $\mathbf{E}\{\varepsilon_k\} = 0,\ \mathbf{E}\{\varepsilon_k^2\} = \sigma^2$. The problem is the nonlinear time-series (1) identification and prediction using the observed sequence $\{y_k, x_k\}_{k=0}^{n}$, where $y_k = x_{k+1}$. In the KM framework the parameterized time-series 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, $\varphi : \mathbf{R}^1 \to \mathbf{R}^{\mathrm{M}}$ is a nonlinear feature map, which transform the original inputs into high-dimensional feature vector $\varphi(x) \in \mathbf{R}^{\mathrm{M}}$, and $\mathbf{w} \in \mathbf{R}^{\mathrm{M}}$ is an unknown 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 $\mathbf{y}_n = (y_0\ y_1 \ldots y_n)^{\mathrm{T}}$ - observation vector, $\boldsymbol{\varepsilon}_n = (\varepsilon_0\ \varepsilon_1 \ldots \varepsilon_n)^{\mathrm{T}}$ - noise vector, and feature matrix $\mathbf{\Phi}_{n-1} = \big(\varphi(x_0) \quad \varphi(x_1) \quad \ldots \quad \varphi(x_{n-1})\big)$.

In according with Mercer’s theorem, feature vectors are taken hereby that its scalar products in feature space will be positive definite kernel functions $\varphi^{\mathrm{T}}(x_i)\varphi(x_j) = \kappa(x_i, x_j),\ i, j = \overline{1,n}.$ Commonly used polynomial kernels $\kappa(x, x') = (\mu + x \cdot x')^p$ of degree p or Gaussian kernels $\kappa(x, x') = \exp\{-\mu(x - x')^2\}$, where p, μ – tuning parameters of kernel model.

In accordance with KM technique introduce further the kernel matrix $\mathbf{K}_n = \mathbf{\Phi}_n^{\mathrm{T}}\mathbf{\Phi}_n$, which may be computed directly without reference to the feature vectors, because $\mathbf{K}_n = \left\| k_{i,j} \right\|, k_{i,j} = \kappa(x_i, x_j), i, j = \overline{1,n},$ and 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 time-series (1) model estimation and one step prediction may be represented as

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

where $\hat{x}_{k+1}$ - one step ahead time-series 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 training sample $\{y_k, x_k\}_{k=0}^{n-1}$, and in accordance with "kernel trick" express in terms of only kernel matrix $\mathbf{K}_{n-1}$.

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 identification 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 identification* $\widetilde{\boldsymbol{\lambda}}_{n+1} = F(\widetilde{\boldsymbol{\lambda}}_{n+1}, y_{n+1}, \widetilde{\mathbf{K}}_{n,r})$, which also use the complete training sample but with respect to fix dimension dual variables $\widetilde{\boldsymbol{\lambda}}_n \in \mathbf{R}^r$ and kernel matrix $\overline{\mathbf{K}}_{n,r}$.

– *Recurrent KM identification with sliding widow* $\overline{\boldsymbol{\lambda}}_{n+1} = F(\overline{\boldsymbol{\lambda}}_{n+1}, y_{n+1}, \overline{\mathbf{K}}_{n,s}), \overline{\boldsymbol{\lambda}}_n \in \mathbf{R}^s$, which uses the sliding window training sample $\{y_k, x_k\}_{k=n-s}^{n}$ and kernel matrix $\overline{\mathbf{K}}_{n,s}$, built up on the respective observations.

Recurrent Kernel Identification with Growing Window

In accordance with general SVM approach [Vapnik, 1998] nonlinear time-series (1) identification problem using the complete training sample (growing window) may be reduced to the following constrained optimization problem with regularized estimation cost function with regularization parameter $\gamma > 0$:

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

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

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

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

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

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 [Golub, Van Loan, 1998] for the regularized kernel matrix $\mathbf{K}_n(\gamma)$:

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

$$\mathbf{K}_n^{-1}(\gamma) = \left(\begin{array}{c|c} \mathbf{K}_{n-1}^{-1}(\gamma) + \delta_n^{-1}\mathbf{K}_{n-1}^{-1}(\gamma)\mathbf{k}_{n-1}(x_n)\mathbf{k}_{n-1}^{\mathrm{T}}(x_n)\mathbf{K}_{n-1}^{-1}(\gamma) & -\delta_n^{-1}\mathbf{K}_{n-1}^{-1}(\gamma)\mathbf{k}_{n-1}(x_n) \\ \hline -\delta_n^{-1}\mathbf{k}_{n-1}^{\mathrm{T}}(x_n)\mathbf{K}_{n-1}^{-1}(\gamma) & \delta_n^{-1} \end{array}\right), \tag{8}$$

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

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

$$\boldsymbol{\lambda}_{n+1} = \begin{pmatrix} \boldsymbol{\lambda}_n - \delta_n^{-1}[y_{n+1} - \omega_n(\boldsymbol{\lambda}_n)]\mathbf{K}_{n-1}^{-1}(\gamma)\mathbf{k}_{n-1}(x_n) \\ \delta_n^{-1}[y_{n+1} - \omega_n(\boldsymbol{\lambda}_n)] \end{pmatrix}, \tag{9}$$

where $\omega_n(\boldsymbol{\lambda}_n) = \mathbf{k}_{n-1}^{\mathrm{T}}(x_n)\boldsymbol{\lambda}_n$, 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). It is obvious that the dimension of dual variables estimates $\boldsymbol{\lambda}_n$ and thereafter the estimated model $\hat{f}(x)$ complexity grows proportionally to the training sample length, which make the considered scheme unsuitable for on-line application.

Reduced Kernel Model Recurrent Identification

From the computational point of view for on-line application it is desirable to limit the number of data vectors from which the kernel matrix is calculated. It allows both to reduce the order of the feature space (which prevents overfitting) and to keep the complexity of model bounded. In order to limit the size of the kernel matrix a sparsification process was proposed [Engel, Mannor, Meir, 2004], in which an input 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 size of kernel matrix is fixed in advance so the model complexity doesn't growing under identification process.

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

Any feature vector from training 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 $\boldsymbol{\Phi}_n = \widetilde{\boldsymbol{\Phi}}_r\mathbf{A}_n^{\mathrm{T}}$, where $\mathbf{A}_n = \|a_{ij}\|$ is a matrix of corresponding expansion coefficients, which may be obtained by the minimum least-squares approximation :

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

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 $\widetilde{\mathbf{K}}_{n,r} = \boldsymbol{\Phi}_n^{\mathrm{T}}\widetilde{\boldsymbol{\Phi}}_r$. 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 $\widetilde{\boldsymbol{\lambda}}_n \in \mathbf{R}^r$

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

where $\widetilde{\boldsymbol{\lambda}}_n = \mathbf{A}_{n-1}^{\mathrm{T}}\boldsymbol{\lambda}_n$. Corresponding reduced order identified model and prediction function are

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

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.

1. *Non-regularized case* $\gamma^{-1}=0$. Using representation (11), the appropriate cost function takes the form

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

The solution of (12) 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 $\widetilde{\boldsymbol{\lambda}}_n=\widetilde{\mathbf{K}}_r^{-1}\left(\mathbf{A}_{n-1}\right)^{+}\mathbf{y}_n=\left(\widetilde{\mathbf{K}}_{n-1,r}\right)^{+}\mathbf{y}_n$.

In the following 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 $\widetilde{\mathbf{K}}_{n,r}^{\mathrm{T}}=\left(\widetilde{\mathbf{K}}_{n-1,r}^{\mathrm{T}}\vdots\mathbf{k}_n(x_n)\right)$, one can use the known Greville formula [Ben-Israel, Greville, 2003] for reduced order Moor-Penrose inverse kernel matrix updating:

$$\widetilde{\mathbf{K}}_{n,r}^{+}=\left((\mathbf{I}_r-\mathbf{q}_n\mathbf{k}_n^{\mathrm{T}}(x_n))\widetilde{\mathbf{K}}_{n-1,r}^{+}\vdots\mathbf{q}_n\right) \tag{14}$$

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

$\alpha_n=1-\mathrm{sgn}\left(\mathbf{k}_r^{\mathrm{T}}(x_n)\mathbf{Z}(\widetilde{\mathbf{K}}_{n-1,r})\mathbf{k}_r(x_n)\right)$

Thereafter recurrent identification algorithm for dual variables estimates is

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

The proper initial conditions at instant $n=r$ are $\widetilde{\mathbf{K}}_{r,r}^{+}=\widetilde{\mathbf{K}}_{r,r}^{-1}$, $\widetilde{\mathbf{K}}_{r,r}=\widetilde{\mathbf{\Phi}}_r^{\mathrm{T}}\widetilde{\mathbf{\Phi}}_r$.

2. *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 form:

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

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 $\mathbf{P}_{n-q,r}=\gamma^{-1}\widetilde{\mathbf{K}}_r+\widetilde{\mathbf{K}}_{n-1,r}^{\mathrm{T}}\widetilde{\mathbf{K}}_{n-1,r}$.

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

$$\widetilde{\boldsymbol{\lambda}}_{n+1}=\widetilde{\boldsymbol{\lambda}}_n+\gamma_n\mathbf{P}_{n-1,r}^{-1}\mathbf{k}_r(x_n)\left(\mathbf{y}_{\mathrm{n}+1}-\mathbf{k}_r^{\mathrm{T}}(x_n)\widetilde{\boldsymbol{\lambda}}_n\right),\quad\gamma_n=\left(1+\mathbf{k}_r^{\mathrm{T}}(x_n)\mathbf{P}_{n-1,r}^{-1}\mathbf{k}_r(x_n)\right)^{-1}. \tag{17}$$

Using the matrix inversion lemma [Haykin, 1996], the recurrent procedure for inverse matrix $\mathbf{P}_{n,r}=\mathbf{P}_{n-1,r}+\mathbf{k}_r(x_n)\mathbf{k}_r^{\mathrm{T}}(x_n)$ updating takes the following form:

$$\mathbf{P}_{n,r}^{-\backprime1}=\mathbf{P}_{n-1,r}^{-\backprime1}-\gamma_n^{-1}\mathbf{P}_{n-1,r}^{-\backprime1}\mathbf{k}_r(x_n)\mathbf{k}_r^{\mathrm{T}}(x_n)\mathbf{P}_{n-1,r}^{-\backprime1}. \tag{18}$$

Thereby equations (15), (16) may be treated as RLSKM version for nonlinear time-series identification based on reduced order model.

Recurrent Kernel Identification with Sliding Window

Sliding window KM approach consider for estimation at instant n only last $n-s$ observations, so observation vector is $\mathbf{y}_{n,s}=(y_{n-s+1}\ldots y_n)^{\mathrm{T}}$. Consequently, the kernel matrix $\overline{\mathbf{K}}_{n,s}=\mathbf{\Phi}_{n,s}^{\mathrm{T}}\mathbf{\Phi}_{n,s}$ has a fixed dimension $(s\times s)$ and observation equation takes the form $\mathbf{y}_{n,s}=\mathbf{\Phi}_{n-1,s}^{\mathrm{T}}\mathbf{w}+\boldsymbol{\varepsilon}_{n,s}=\mathbf{K}_{n-1,s}\overline{\boldsymbol{\lambda}}_n+\boldsymbol{\varepsilon}_{n,s}$, where $\overline{\boldsymbol{\lambda}}_n\in\mathbf{R}^{\mathrm{s}}$.

Consider the "sliding" estimation cost function includes at any instant $n+1$ *a priori* information term determined by previously estimate at instant n. Using the representation $\mathbf{w}=\mathbf{\Phi}_{n,s}\overline{\boldsymbol{\lambda}}$, the optimization problem is defined as:

$$J_{n,s}=\left\|\mathbf{y}_{n+1,s}-\mathbf{K}_{n,s}\overline{\boldsymbol{\lambda}}\right\|^2+\gamma^{-1}(\overline{\boldsymbol{\lambda}}-\overline{\boldsymbol{\lambda}}_n)^{\mathrm{T}}\mathbf{K}_{n,s}(\overline{\boldsymbol{\lambda}}-\overline{\boldsymbol{\lambda}}_n)\to\min_{\overline{\boldsymbol{\lambda}}}. \tag{19}$$

Condition of optimality leads to the following normal equations:

$$\left(\overline{\mathbf{K}}_{n,s}^{\mathrm{T}}\overline{\mathbf{K}}_{n,s}+\gamma^{-1}\overline{\mathbf{K}}_{n,s}\right)\overline{\boldsymbol{\lambda}}_n=\overline{\mathbf{K}}_{n,s}^{\mathrm{T}}\mathbf{y}_{n+1,s}+\gamma^{-1}\overline{\mathbf{K}}_{n,s}\overline{\boldsymbol{\lambda}}_n. \tag{20}$$

Using the 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:

$$\overline{\boldsymbol{\lambda}}_{n+1}=\left(\gamma^{-1}\mathbf{I}_s+\overline{\mathbf{K}}_{n,s}\right)^{-1}\left(\gamma^{-1}\overline{\boldsymbol{\lambda}}_n+\mathbf{y}_{n+1}\right) \tag{21}$$

At last, it is necessary to put forward the updating algorithm for inverse regularized sliding kernel matrix $\overline{\mathbf{K}}_{n,s}^{-1}(\gamma)$. Using the approach, proposed by [Vaerenbergh, Javier, Santamar, 2007], 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 "downsizing" matrix $\overline{\mathbf{K}}_{n-1,s-1}(\gamma)$, determined from the sliding kernel matrix representation:

$$\overline{\mathbf{K}}_{n-1,s}(\gamma)=\left(\begin{array}{c:c}\gamma^{-1}+k_{n-s,n-s} & \mathbf{k}_{n-1,s-1}^{\mathrm{T}}(x_{n-s})\\ \hdashline \mathbf{k}_{n-1,s-1}(x_{n-s}) & \overline{\mathbf{K}}_{n-1,s-1}(\gamma)\end{array}\right), \tag{22}$$

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

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

$$\overline{\mathbf{K}}_{n-1,s-1}^{-1}=\mathbf{R}_s\overline{\mathrm{K}}_{n-1,s}^{-1}\mathbf{R}_s^{\mathrm{T}}-(\mathrm{e}_1^{\mathrm{T}}\overline{\mathbf{K}}_{n-1,s}^{-1}\mathrm{e}_1)^{-1}\mathbf{R}_s\overline{\mathbf{K}}_{n-1,s}^{-1}\mathbf{e}_1\mathbf{e}_1^{\mathrm{T}}\overline{\mathbf{K}}_{n-1,s}^{-1}\mathbf{R}_s^{\mathrm{T}}, \tag{23}$$

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

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

$$\overline{\mathbf{K}}_{n,s}(\gamma)=\left(\begin{array}{c:c}\overline{\mathbf{K}}_{n-1,s-1}(\gamma) & \mathbf{k}_{n-1,s-1}(x_n)\\ \hdashline \mathbf{k}_{n-1,s-1}^{\mathrm{T}}(x_n) & \gamma^{-1}+k_{n,n}\end{array}\right), \tag{24}$$

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:

$$\mathbf{K}_{n,s}^{-1}=\left(\begin{array}{c:c}\mathbf{K}_{n-1,s-1}^{-1}(\gamma)+\delta_n^{-1}\mathbf{K}_{n-1,s-1}^{-1}(\gamma)\mathbf{k}_{n-1}(x_n)\mathbf{k}_{n-1}^{\mathrm{T}}(x_n)\mathbf{K}_{n-1,s-1}^{-1}(\gamma) & -\delta_n^{-1}\mathbf{K}_{n-1,s-1}^{-1}(\gamma)\mathbf{k}_{n-1,s-1}^{\mathrm{T}}(x_n)\\ \hdashline -\delta_n^{-1}\mathbf{K}_{n-1,s-1}^{-1}(\gamma)\mathbf{k}_{n-1,s-1}^{\mathrm{T}}(x_n) & \delta_n^{-1}\end{array}\right), \tag{25}$$

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

Finally, expressions (21), (23), (25) produce a recurrent form of SWKM for nonlinear time-series identification.

Conclusion

Recurrent KM approach for nonlinear time-series identification and prediction combining with model reduction technique leads to identification algorithms efficiency improvement. The advantage of such an approach consists not only in computing difficulties reducing and amount of calculation restriction but also in the possibility of on-line operating in non-stationary environments. The key feature of proposed algorithms is that the identified model complexity does not increase as the number of samples increases and time-varying model may be on-line estimated, so recurrent KM algorithms may be successfully used for non-stationary time series identification. Another preference connected with the sufficiently simple possibility of robust modification recurrent KM algorithms design using suitable nonlinear estimation function.

Further inquiry and improvement of proposed approach should be connected with identified model optimization methods development. The most important problem is the model parameters optimization, namely, regularization parameter and kernel tuning parameter. The cross-validation technique is seemed to be the most suitable approach to model parameters optimization. Such issue is closely connected with the general problem of model optimization in compliance with available information via structural risk minimization approach. Dynamic approach to model optimization may be considered as one of the most important directions of further investigation.

Bibliography

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

[Dorffner,1996] G. Dorffner. Neural networks for time series processing // Neural Network World, 1996, 6(4), Pp.447–468.

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

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

[Espinoza, Suykens, De Moor, 2005] M. Espinoza, J.A.K. Suykens, and B. De Moor. Kernel based partially linear models and nonlinear identification // IEEE Transactions on Automatic Control, 2005, 50 (10), Pp. 1602–1606.

[Kivinen, Smola, Williamson, 2004] J. Kivinen, A.J. Smola, R.C. Williamson. Online learning with kernels // IEEE Transactions on Signal Processing, 2004, Vol. 52, Pp. 2165 – 2176.

[Suykens, Van Gestel, De Brabanter, De Moor, Vandewalle,2002] 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.

[Engel, Mannor, Meir, 2004] Y. Engel, S. Mannor, R. Meir. The kernel recursive least-squares algorithm // IEEE Transactions on Signal Processing, 2004, Vol. 52, Pp. 2275 - 2285.

[Vaerenbergh, Javier, Santamar, 2007] S. Van Vaerenbergh, V. Javier, I. Santamar. Nonlinear System Identification using a New Sliding Window Kernel RLS Algorithm // Journal of Communications, 2007, Vol. 2, No. 3.

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

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

[Haykin, 1996] S. Haykin. Adaptive Filter Theory. Prentice Hall, 3rd edition, 1996.

Authors' Information

Leonid Lyubchyk – Head of Department, Department of Computer Mathematics and Mathematical Modeling, National Technical University "Kharkov Polytechnic Institute", Frunze str. 21, Kharkov, 61002, Ukraine; e-mail: lyubchik.leonid@gmail.com

Vladyslav Kolbasin – PhD student, Department of Computer Mathematics and Mathematical Modeling, National Technical University "Kharkov Polytechnic Institute", Frunze str. 21, Kharkov, 61002, Ukraine; e-mail: ka_vlad@list.ru