Two-Level Recursive Identification of Hammerstein System by Interaction Prediction Method

Zygmunt Hasiewicz[1] and Grzegorz Mzyk[2]

[1,2] Wrocław University of Science and Technology, Poland
grzegorz.mzyk@pwr.edu.pl

Abstract. The paper concerns identification of Hammerstein system under nonparametric prior knowledge about the static nonlinear characteristic. The identification task is decomposed by the prediction of the hidden interaction signal. The standard kernel approach is modified to cope with the problem of constant offset between the regression function and the static characteristic in Hammerstein system, which was not solved in the previous papers. The idea is based on alternate updating of the offset and the estimate of the impulse response of the linear block. Both levels of the algorithm are given in the recursive version.

Keywords: System Identification, Hammerstein System, Non-parametric estimation, Decomposition and Coordination.

1 Introduction

The idea of the presented approach is derived from the combined parametric-nonparametric methods of block-oriented nonlinear system identification. The paper is inspired by [5] and [6], where the nonparametric (kernel) regression estimate was used for decomposition of Hammerstein system (Fig. 1) identification task into two independent subproblems. The models of nonlinear and linear subsystems are alternately updated with the use of the predictor of interaction signal v_k. As it is commonly known from literature (see e.g. [3], [11], [10], [9]), the representation of complex systems is not unique owing to inaccessibility of internal signals. For any scale s, the class of Hammerstein systems with the nonlinear characteristics $s\mu()$ and the impulse response $\{\frac{\lambda_r}{s}\}$ is equivalent (indistinguishable) form the input-output point of view. Moreover, the constant nonzero offset appears between the estimated input-output regression function and the true static nonlinear characteristic $\mu()$. In the proposed approach, in the contrary to the papers cited above, and further ideas (see [7] and [8]) where both subsystems are identified separately, it is hopeful that the whole serial model can fit the data without the scale, and the nonlinearity estimate will be free of offset.

2 Statement of the problem

We consider the Hammerstein system, i.e. a tandem consisting of the static nonlinearity $\mu()$ followed by the linear dynamic FIR filter with the impulse response $\{\lambda_r\}_{r=0}^{M}$, driven by the i.i.d. random process u_k and corrupted by the stationary zero-mean white noise z_k of finite variance, independent of the system input

$$y_k = \sum_{r=0}^{M} \lambda_r \mu(u_{k-r}) + z_k,$$

where only input-output data (u_k, y_k) are accessible for measurements – and in particular, the interaction signal v_k cannot be measured.

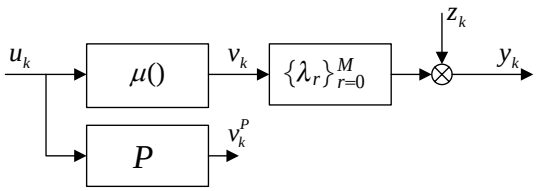

Fig. 1. Hammerstein system with interaction predictor.

It can be easily shown that in such conditions the one dimensional regression function of output on input in Hammerstein system is a scaled and shifted version of the nonlinear characteristic $\mu()$

$$R(u) = E[y_k | u_k = u] = \lambda_0 \mu(u) + d. \tag{1}$$

To decompose the complex system identification problem we propose to introduce the interaction predictor $v_k^P = P(u_k; \{u_k, y_k\}_{k=1}^{N})$ computed on the basis of the set of available input-output measurements $\{u_k, y_k\}_{k=1}^{N}$. Owing to [4] and using the kernel approach ([12]), it is clear that, for a given u, the comprehensible predictor can have the form $P(u; \{u_k, y_k\}_{k=1}^{N}) = \hat{R}(u) - c$, where c may play the role of coordination variable. Such a predictor $P(u; \{u_k, y_k\}_{k=1}^{N})$ is corrected version of kernel regression function estimate $\hat{R}(u)$ (shifted by the constant c). Thus, selection of c will be crucial for the method and c is further indeed treated as coordination variable (tuning factor) which calibrates/tunes the predictor. The use of $y_k - c$ instead of y_k in standard regression function estimate and then application of the model $\mu(u) \simeq P(u)$ instead of $\mu(u) \simeq \hat{R}(u)$, may lead to nonparametric estimate of nonlinear characteristic which is free of bias (offset) d. The problem is (i) how to select the coordination variable c to this end, and (ii) how to easily compute coordinated models of both subsystems. The presented idea is based on progressive identification, in the sense that the corrected model (predictor) of the first subsystem will support identification of the second one, by interaction prediction.

3 Establishing of coordination factor c

Let $\lambda_0 = 1$ and, to begin with, let $\{(u_j^P, y_j^P)\}_{j=1}^{N_P}$ be the initial set of measurement data used for predictor evaluation. The predictor obtained therefrom has the form

$$P\left(u; \{(u_j^P, y_j^P)\}_{j=1}^{N_P}\right) = \sum_{j=1}^{N_P} \frac{K\left(\frac{u_j^P - u}{h_{N_P}}\right)}{\sum_{j=1}^{N_P} K\left(\frac{u_j^P - u}{h_{N_P}}\right)} y_j^P - c = \hat{R}^P(u) - c. \quad (2)$$

For u_k belonging to the input-output pair (u_k, y_k) the proper predicted value v_k^P of the interaction v_k is thus $v_k^P = \hat{R}^P(u_k) - c$. The predicted interaction is next treated as being true, i.e. we apply the interaction prediction principle, yielding the following relation concerning system dynamics $y_k = \sum_{r=0}^{M} \lambda_r v_{k-r}^P + z_k = \left(\overline{v}_k^P\right)^T \lambda + z_k$, where $\overline{v}_k^P = (v_k^P, v_{k-1}^P, \ldots, v_{k-M}^P)^T$, $\lambda = (1, \lambda_1, \ldots, \lambda_M)^T$, letting $\lambda_0 = 1$.

3.1 Local identification of the linear dynamics (1st level task)

Based on predicted interactions, the impulse response λ is here estimated on the basis of a new set of N "measurements" $\{(\overline{v}_{k+l}^P, y_{k+l})\}_{l=0}^{N-1}$ with the use of least squares method. Let $V_N^P = V_N^P(c) = \left(\overline{v}_k^P, \overline{v}_{k+1}^P, \ldots, \overline{v}_{k+(N-1)}^P\right)^T$; $Y_N = \left(y_k, y_{k+1}, \ldots, y_{k+(N-1)}\right)^T$, and $\overline{Y}_N = V_N^P \lambda$, be the linear dynamics model output for a given λ. The goal is thus to solve the following minimization problem: $\left\|Y_N - \overline{Y}_N\right\|_2^2 \to \min_{\lambda}$. As a result we get:

- evaluation of λ:

$$\lambda_N(c) = \left(V_N^{P^T}(c) V_N^P(c)\right)^{-1} V_N^{P^T}(c) Y_N, \quad (3)$$

  being the best estimate of λ for a given coordination factor c, which can be also computed recursively (see Section 5);

- the output of the best model for a given c:

$$\overline{Y}_N(c) = V_N^P(c)\lambda_N(c) = V_N^P(c)\left(V_N^{P^T}(c) V_N^P(c)\right)^{-1} V_N^{P^T}(c) Y_N,$$

  for which it holds that $Y_N - \overline{Y}_N(c) = \left[I - V_N^P(c)\left(V_N^{P^T}(c) V_N^P(c)\right)^{-1} V_N^{P^T}(c)\right] Y_N$.

Our aim in turn is to select c such that the difference between the model output $\overline{Y}_N(c)$ and the measured system output Y_N is minimal (in the least squares sense)and in our approach this is also fundamental from the coordination task point of view.

3.2 Coordination (2nd level task)

Owing to the above, we thus take the following objective function

$$Q_N(c) = \left\|Y_N - \overline{Y}_N(c)\right\|_2^2 \to \min_c,$$

and observe that $\left\|Y_N - \overline{Y}_N(c)\right\|_2^2 = \left(Y_N - \overline{Y}_N(c)\right)^T\left(Y_N - \overline{Y}_N(c)\right) = Y_N^T\left(Y_N - \overline{Y}_N(c)\right)$, since $\overline{Y}_N^T(c)\left(Y_N - \overline{Y}_N(c)\right) = 0$. Consequently, we get

$$Q_N(c) = Y_N^T\left[I - V_N^P(c)\left(V_N^{P^T}(c)V_N^P(c)\right)^{-1}V_N^{P^T}(c)\right]Y_N.$$

Let $\mathbf{1} = [1]$ be a matrix with all entries equal to 1, and let

$$\widehat{\overline{R}}^P(\overline{u}_{k+l}) = \left[\hat{R}^P(u_{k+l}), \hat{R}^P\left(u_{(k+l)-1}\right), \ldots, \hat{R}^P\left(u_{(k+l)-M}\right)\right], \tag{4}$$

for $l = 0,1,\ldots,N-1$. Including that $v_i^P = \hat{R}^P(u_i) - c$, we see that the matrix $V_N^P(c)$ can be rewritten as $V_N^P(c) = P_N - c\mathbf{1}_{N\times(M+1)}$, where

$$P_N \triangleq [\widehat{\overline{R}}^{P^T}(\overline{u}_k), \widehat{\overline{R}}^{P^T}(\overline{u}_{k+1}), \ldots, \widehat{\overline{R}}^{P^T}\left(\overline{u}_{k+(N-1)}\right)]^T.$$

In the objective function $Q_N(c)$ we thus have in particular that

$$W_N^P(c) \triangleq V_N^P(c)\left[V_N^{P^T}(c)V_N^P(c)\right]^{-1}V_N^{P^T}(c) =$$

$$= \left(P_N - c\mathbf{1}_{N\times(M+1)}\right)\left[V_N^{P^T}(c)V_N^P(c)\right]^{-1}\left(P_N - c\mathbf{1}_{N\times(M+1)}\right)^T,$$

and it can be shown in turn that $V_N^{P^T}(c)V_N^P(c) = NT_N$, where

$$T_N = \frac{1}{N}P_N^TP_N - c[\frac{1}{N}P_N^T\mathbf{1}_{N\times(M+1)} + \frac{1}{N}\left(P_N^T\mathbf{1}_{N\times(M+1)}\right)^T] + c^2\mathbf{1}_{(M+1)\times(M+1)},$$

yielding $W_N^P(c) = \left(\frac{1}{\sqrt{N}}P_N - \frac{c}{\sqrt{N}}\mathbf{1}_{N\times(M+1)}\right)T_N^{-1}\left(\frac{1}{\sqrt{N}}P_N - \frac{c}{\sqrt{N}}\mathbf{1}_{N\times(M+1)}\right)^T$, and asymptotically, as N grows large, $W_N^P(c) \simeq P_N[NT_N]^{-1}P_N^T$. Thus, for N large enough, we can put $W_N^P(c) = P_N[NT_N]^{-1}P_N^T = \frac{1}{N}P_NT_N^{-1}P_N^T$ which gives the objective function of the form $Q_N(c) = Y_N^T[I - W_N^P(c)]Y_N = Y_N^T\left[I - \frac{1}{N}P_NT_N^{-1}P_N^T\right]Y_N$. Further, noticing that $T_N^{-1} = (A+B)^{-1}$, where $A = \frac{1}{N}P_N^TP_N$, $B = c^2\mathbf{1}_{(M+1)\times(M+1)} - c\left[\frac{1}{N}P_N^T\mathbf{1}_{N\times(M+1)} + \frac{1}{N}\left(P_N^T\mathbf{1}_{N\times(M+1)}\right)^T\right]$, and using the linear approximation $T_N^{-1} \simeq (I - A^{-1}B)A^{-1}$, we get $\frac{1}{N}P_NT_N^{-1}P_N^T = \frac{1}{N}P_NA^{-1}P_N^T - \frac{1}{N}P_N(A^{-1}B)A^{-1}P_N^T$,
and finally (taking into account that $B = B(c)$) the objective function

$$Q_N(c) = Y_N^TY_N - \frac{1}{N}Y_N^T(P_NA^{-1}P_N^T)Y_N + \frac{1}{N}Y_N^T(P_NA^{-1})B(c)(P_NA^{-1})^TY_N.$$

Hence, $\frac{dQ_N(c)}{dc} = \frac{1}{N} Y_N^T (P_N A^{-1}) \frac{dB(c)}{dc} (P_N A^{-1})^T Y_N$, where

$$\frac{dB(c)}{dc} = 2c\mathbf{1}_{(M+1)\times(M+1)} - \left(\frac{1}{N} P_N^T \mathbf{1}_{N\times(M+1)} + \frac{1}{N}\left(P_N^T \mathbf{1}_{N\times(M+1)}\right)^T\right),$$

and the condition $\frac{dQ_N(c)}{dc} = 0$ gives eventually optimum (due to convexity of $Q_N(c)$) coordination factor

$$c_N = \frac{\left[\left(P_N A^{-1}\right)^T Y_N\right]^T \mathbf{M}_{(M+1)\times(M+1)} \left(P_N A^{-1}\right)^T Y_N}{\left[(P_N A^{-1})^T Y_N\right]^T \{\mathbf{1}_{(M+1)\times(M+1)}\} (P_N A^{-1})^T Y_N},$$

for N large, where $\mathbf{M}_{(M+1)\times(M+1)} = \frac{1}{2N}\left[P_N^T \mathbf{1}_{N\times(M+1)} + \left(P_N^T \mathbf{1}_{N\times(M+1)}\right)^T\right]$.
Noticing that $(P_N A^{-1})^T Y_N = N[P_N (P_N^T P_N)^{-1}]^T Y_N$, and denoting

$$[P_N (P_N^T P_N)^{-1}]^T Y_N \triangleq L_{M+1}, \tag{5}$$

we get $c_N = \frac{L_{M+1}^T \mathbf{M}_{(M+1)\times(M+1)} L_{M+1}}{L_{M+1}^T \{\mathbf{1}_{(M+1)\times(M+1)}\} L_{M+1}}$. For derivation of recursive computation algorithms of matrix $\mathbf{M}$ and vector L with growing N, see Appendix A.

4 Recursive computation of c_N

Using the simplified notation employed in Appendix A, emphasizing the dependence of $\mathbf{M}$ and L on N – the number of data, namely $\mathbf{M}_{(M+1)\times(M+1)} = \mathbf{M}^N$ and $L_{M+1} = L^N$, with which $c_N = \frac{{L^N}^T \mathbf{M}^N L^N}{{L^N}^T \mathbf{1}_{(M+1)\times(M+1)} L^N} \triangleq \frac{\mathcal{L}_N}{\mathcal{M}_N}$ and taking advantage of (18) and (15) in Appendix A, after simple algebra shown in Appendix B, for the numerator $\mathcal{L}_N$ and denominator $\mathcal{M}_N$ we obtain respectively

$$\mathcal{L}_N = \left(\frac{N-1}{N}\right)\mathcal{L}_{N-1} + \mathcal{K}_{\mathcal{L}}^N, \tag{6}$$

and

$$\mathcal{M}_N = \mathcal{M}_{N-1} + \mathcal{K}_{\mathcal{M}}^N, \tag{7}$$

where $\mathcal{K}_{\mathcal{L}}^N = g_N^T \mathbf{M}^N g_N + 2{L^{N-1}}^T \mathbf{M}^N g_N + \frac{1}{N}{L^{N-1}}^T \mathbf{R}^{N-1} L^{N-1}$ and

$$\mathcal{K}_{\mathcal{M}}^N = g_N^T \mathbf{1}_{(M+1)\times(M+1)} g_N + 2{L^{N-1}}^T \mathbf{1}_{(M+1)\times(M+1)} g_N,$$

are proper innovation components with $g_N = G_N \rho_{N-1}$, and G_N and ρ_{N-1} being computed due to (17) and (19) in Appendx A.
Based on (6) and (7), after ordinary calculation, we get further that $c_N = \frac{\mathcal{L}_N}{\mathcal{M}_N} = \varkappa_N \frac{\mathcal{L}_{N-1}}{\mathcal{M}_{N-1}} + t_N$, it is

$$c_N = \varkappa_N c_{N-1} + t_N, \tag{8}$$

or equivalently

$$c_N = c_{N-1} + [t_N - w_N c_{N-1}], \tag{9}$$

where

$$\varkappa_N = \frac{1-\left(\frac{1}{N}\right)}{1+\left(\frac{\mathcal{K}_{\mathcal{M}}^N}{\mathcal{M}_{N-1}}\right)}; \qquad t_N = \frac{\left(\frac{\mathcal{K}_{\mathcal{L}}^N}{\mathcal{M}_{N-1}}\right)}{1+\left(\frac{\mathcal{K}_{\mathcal{M}}^N}{\mathcal{M}_{N-1}}\right)}, \text{ and } w_N \triangleq 1-\varkappa_N = \frac{\left(\frac{1}{N}\right)+\left(\frac{\mathcal{K}_{\mathcal{M}}^N}{\mathcal{M}_{N-1}}\right)}{1+\left(\frac{\mathcal{K}_{\mathcal{M}}^N}{\mathcal{M}_{N-1}}\right)},$$

i.e. a rather simple recursive computation rule for updating the coordination factor c with growing N. Notice that asymptotically, for $N=\infty$, the equation (8) takes the form $c_\infty = \varkappa_\infty c_\infty + t_\infty$,i.e., $(1-\varkappa_\infty)c_\infty = t_\infty$, and hence $c_\infty = \frac{t_\infty}{1-\varkappa_\infty} = \frac{t_\infty}{w_\infty} = \frac{\mathcal{K}_{\mathcal{L}}^\infty}{\mathcal{K}_{\mathcal{M}}^\infty}$, where $\mathcal{K}_{\mathcal{L}}^\infty = g_\infty^T \mathbf{M}^\infty g_\infty + 2{L^\infty}^T \mathbf{M}^\infty g_\infty$, and $\mathcal{K}_{\mathcal{M}}^\infty = g_\infty^T \mathbf{1}_{(M+1)\times(M+1)} g_\infty + 2{L^\infty}^T \mathbf{1}_{(M+1)\times(M+1)} g_\infty$, is the limit value of c in the predictor (2).

5 Recursive identification of the impulse response $\boldsymbol{\lambda}$

Recalling that (see (3) and related denotations) $\lambda_N = \left({V_N^P}^T V_N^P\right)^{-1} {V_N^P}^T Y_N$, and using standard recursive least squares computation scheme we get immediately

$$\lambda_N = \lambda_{N-1} + \Gamma_N \overline{v}_{k+(N-1)}^P \left[y_{k+(N-1)} - \left(\overline{v}_{k+(N-1)}^P\right)^T \lambda_{N-1}\right], \tag{10}$$

where

$$\Gamma_N = \Gamma_{N-1} - \frac{\Gamma_{N-1}\overline{v}_{k+(N-1)}^P \left(\overline{v}_{k+(N-1)}^P\right)^T \Gamma_{N-1}}{1+\left(\overline{v}_{k+(N-1)}^P\right)^T \Gamma_{N-1} \overline{v}_{k+(N-1)}^P}, \tag{11}$$

and $\left(\overline{v}_{k+(N-1)}^P\right)^T = \widehat{\overline{R}}^P\left(\overline{u}_{k+(N-1)}\right) - c_N[1,1,\ldots,1]_{1\times(M+1)}$.

6 The identification algorithm

The above analysis leads eventually to the following multi stage routine for recursive identification of Hammerstein system:

1. Correcting of the coordination variable c ((9))

$$c_N = c_{N-1} + [t_N - w_N c_{N-1}]$$

2. Computing of the corrected non-linearity model (predictor) ((2))

$$P\left(u; \left\{u_j^P, y_j^P\right\}_{j=1}^{N_P}\right) = \hat{R}^P(u) - c_N$$

3. Computing of the corrected interactions (see (4))

$$\left(\overline{v}_{k+(N-1)}^{P}\right)^{T} = \widehat{\overline{R}}^{p}\left(\overline{u}_{k+(N-1)}\right) - c_N[1,1,\ldots,1]_{1\times(M+1)}$$

4. Updating Γ_N ((11)): $\Gamma_N = \Gamma_{N-1} - \Gamma_{N-1}\Psi_{N-1}$
5. Correcting of the estimate of impulse response λ ((10))

$$\lambda_N = \lambda_{N-1} + \Gamma_N \xi_{N-1},$$

where $t_N = \frac{\left(\frac{\mathcal{K}_L^N}{\mathcal{M}_{N-1}}\right)}{1+\left(\frac{\mathcal{K}_M^N}{\mathcal{M}_{N-1}}\right)}$, $w_N = \frac{\left(\frac{1}{N}\right)+\left(\frac{\mathcal{K}_M^N}{\mathcal{M}_{N-1}}\right)}{1+\left(\frac{\mathcal{K}_M^N}{\mathcal{M}_{N-1}}\right)}$, and respectively

$$\Psi_{N-1} = \overline{v}_{k+(N-1)}^{P}\left(\overline{v}_{k+(N-1)}^{P}\right)^{T}\Gamma_{N-1}\left(1+\left(\overline{v}_{k+(N-1)}^{P}\right)^{T}\Gamma_{N-1}\overline{v}_{k+(N-1)}^{P}\right)^{-1},$$

and $\xi_{N-1} = \overline{v}_{k+(N-1)}^{P}[y_{k+(N-1)} - \left(\overline{v}_{k+(N-1)}^{P}\right)^{T}\lambda_{N-1}]$.

7 Numerical implementation in Matlab

The *Matlab* implementation of kernel regression based identification of Hammerstein system, with automatically selected bandwidth parameter, can be found in the *Nonparametric System Identification Toolbox*, proposed by the authors. The actual version of the library and its documentation can be accessed at the WWW page *http://staff.iiar.pwr.wroc.pl/grzegorz.mzyk/KIT*.

8 Summary

It was shown, that the decomposition and coordination concept usually used for solving complex optimization problems can be also applied in block-oriented system identification. Using this strategy, the Hammerstein system identification problem has been decomposed on two independent local subproblems, in which the models of static nonlinearity and linear dynamics are identified separately, and then coordinated. The proposed procedure is based on known and well elaborated estimation algorithms, namely least squares estimation and nonparametric kernel regression function estimation routines. Nonparametric algorithm works without any parametric knowledge of nonlinearity representation. After proper correction, the resulting estimate supports estimation of parameters of linear block and plays the role of coordination variable generator. In result, the model as a whole may be free of the scale and offset even under nonparametric knowledge of the static characteristic, which is in contrast to the existing literature. It was also shown that all stages of the identification procedure can be accomplished recursively, in a rather convenient way.

In the paper, N_p – the numer of data used at the beginning of our routine for predictor set up, was treated as being fixed. However, for better efficiency of the scheme N_p should grow. Owing to the overall context of the paper, it is proper to remark that numerator and denominator of

$$\hat{R}^P(u) = \frac{\frac{1}{N_P}\sum_{j=1}^{N_P} K\left(\frac{u_j^P - u}{h_{N_P}}\right) y_j^P}{\frac{1}{N_P}\sum_{j=1}^{N_P} K\left(\frac{u_j^P - u}{h_{N_P}}\right)} \triangleq \frac{\mathcal{L}_{N_p}^P}{\mathcal{M}_{N_p}^P}$$

in the predictor (2) can also be easily recomputed in a recursive fashion, similarly as in (13)-(14) in Appendix A:

$$\mathcal{L}_{N_p}^P = \frac{N_p - 1}{N_p}\left[\mathcal{L}_{N_p-1}^P + \frac{1}{N_p - 1} K\left(\frac{u_{N_p}^P - u}{h_{N_P}}\right) y_{N_p}^P\right]$$

$$\mathcal{M}_{N_p}^P = \frac{N_p - 1}{N_p}\left[\mathcal{M}_{N_p-1}^P + \frac{1}{N_p - 1} K\left(\frac{u_{N_p}^P - u}{h_{N_P}}\right)\right]$$

and thus gradually updated with growing the number N_p of predictor data.

Finally, it is also worth to notice that in our approach any efficient non-parametric regression function estimate can be used in the role of $\hat{R}^P(u)$ in the predictor (2).

Appendices

Appendix A. Calculation of $\mathbf{M}_{(M+1)\times(M+1)}$ and L_{M+1}

A1. Recursive computing of matrix $\boldsymbol{M}_{(M+1)\times(M+1)}$

Taking into account definitions of P_N and $\mathbf{1}_{N\times(M+1)}$, after simple algebra we obtain

$$\mathbf{M}_{(M+1)\times(M+1)} = \frac{1}{2N}\left[P_N^T \mathbf{1}_{N\times(M+1)} + \left(P_N^T \mathbf{1}_{N\times(M+1)}\right)^T\right] = [m_{ij}]_{i,j=0,1,\ldots,M}$$

where

$$m_{ij} = \frac{1}{2}\left[\underbrace{\frac{1}{N}\sum_{l=0}^{N-1} \hat{R}^P(u_{k+l-i})}_{m_{iN}} + \underbrace{\frac{1}{N}\sum_{l=0}^{N-1} \hat{R}^P\left(u_{k+l-j}\right)}_{m_{jN}}\right], \tag{12}$$

and $m_{ij} = m_{ji}$. Thus, $m_{ij} \triangleq m_{ij,N} = \frac{1}{2}\left[m_{iN} + m_{jN}\right]$.

For empirical means in (12) we simply get recursive versions

$$m_{iN} = \frac{N-1}{N}\left[m_{i,N-1} + \frac{1}{N-1}\hat{R}^P\left(u_{k+(N-1)-i}\right)\right], \tag{13}$$

$$m_{jN} = \frac{N-1}{N}\left[m_{j,N-1} + \frac{1}{N-1}\hat{R}^P\left(u_{k+(N-1)-j}\right)\right], \tag{14}$$

which further leads to $m_{ij,N} = \frac{N-1}{N}\left[m_{ij,N-1} + \frac{1}{N-1}\overline{R}_{ij}^P\right]$, where

$$\overline{R}_{ij}^P = \frac{\hat{R}^P\left(u_{k+(N-1)-i}\right) + \hat{R}^P\left(u_{k+(N-1)-j}\right)}{2}$$

or equivalently $m_{ij,N} = \left(\frac{N-1}{N}\right) m_{ij,N-1} + \frac{1}{N}\overline{R}_{ij}^{P}$. Hence, for the matrix

$$\mathbf{M}_{(M+1)\times(M+1)} \triangleq \mathbf{M}_{(M+1)\times(M+1)}^{N} = \left[m_{ij,N}\right]_{(M+1)\times(M+1)} \triangleq \mathbf{M}^{N}$$

we get

$$\mathbf{M}^{N} = \frac{N-1}{N}\mathbf{M}^{N-1} + \frac{1}{N}\mathbf{R}^{N-1}, \tag{15}$$

where $\mathbf{R}^{N-1} = \left[\overline{R}_{ij}^{P}\right]_{(M+1)\times(M+1)}$.

A2. Recursive computing of vector L_{M+1}
Since (cf. (5))

$$L_{M+1} = (P_N^T P_N)^{-1} P_N^T Y_N, \tag{16}$$

the vector $L_{M+1} = L_{M+1}^{N} \triangleq L^N$ is in fact the least squares model of the linear dynamic object with the input P_N, parameters L and the output Y_N, i.e. the result of the following optimization task $\|Y_N - P_N L\|_2^2 \to \min_L$, where $\|\quad\|_2$ is the Euclidean norm. Taking into account definition of the matrix P_N, and denoting $G_N = (P_N^T P_N)^{-1}$, after standard steps we get recursive version of (16)

$$L^N = L^{N-1} + G_N \widehat{\overline{R}^{P}}^{T}\left(\overline{u}_{k+(N-1)}\right)\left[y_{k+(N-1)} - \widehat{\overline{R}}^{P}\left(\overline{u}_{k+(N-1)}\right)L^{N-1}\right],$$

$$G_N = G_{N-1} - \frac{G_{N-1}\widehat{\overline{R}^{P}}^{T}\left(\overline{u}_{k+(N-1)}\right)\widehat{\overline{R}}^{P}\left(\overline{u}_{k+(N-1)}\right)G_{N-1}}{1+\widehat{\overline{R}}^{P}\left(\overline{u}_{k+(N-1)}\right)G_{N-1}\widehat{\overline{R}^{P}}^{T}\left(\overline{u}_{k+(N-1)}\right)}. \tag{17}$$

Consequently,

$$L^N = L^{N-1} + G_N \rho_{N-1}, \tag{18}$$

where

$$\rho_{N-1} = \widehat{\overline{R}^{P}}^{T}\left(\overline{u}_{k+(N-1)}\right)\left[y_{k+(N-1)} - \widehat{\overline{R}}^{P}\left(\overline{u}_{k+(N-1)}\right)L^{N-1}\right]. \tag{19}$$

Appendix B. Calculating of $\mathcal{L}_N$ and $\mathcal{M}_N$

B1. Recursive computing of the numerator $\mathcal{L}_N$
For $\mathcal{L}_N$, owing to (18), we have

$$\mathcal{L}_N = L^{N^T}\mathbf{M}^N L^N = [L^{N-1} + G_N\rho_{N-1}]^T\mathbf{M}^N[L^{N-1} + G_N\rho_{N-1}]$$

$$= L^{N-1^T}\mathbf{M}^N L^{N-1} + (G_N\rho_{N-1})^T\mathbf{M}^N(G_N\rho_{N-1}) + 2L^{N-1^T}\mathbf{M}^N(G_N\rho_{N-1}),$$

and further, owing to (15), we get

$$L^{N-1^T}\mathbf{M}^N L^{N-1} = \frac{N-1}{N} L^{N-1^T}\mathbf{M}^{N-1} L^{N-1} + \frac{1}{N} L^{N-1^T}\mathbf{R}^{N-1} L^{N-1},$$

which leads to $\mathcal{L}_N = \left(\frac{N-1}{N}\right)\mathcal{L}_{N-1} + \mathcal{K}_{\mathcal{L}}^N$, where

$$\mathcal{K}_{\mathcal{L}}^N = (G_N\rho_{N-1})^T\mathbf{M}^N(G_N\rho_{N-1}) + 2L^{N-1^T}\mathbf{M}^N(G_N\rho_{N-1}) + \left(\frac{1}{N}\right)L^{N-1^T}\mathbf{R}^{N-1}L^{N-1}.$$

B2. Recursive computing of the denominator $\mathcal{M}_N$
For $\mathcal{M}_N$ we have $\mathcal{M}_N = \mathcal{M}_{N-1} + \mathcal{K}_{\mathcal{M}}^N$, where $\mathcal{M}_{N-1} = L^{N-1^T}\mathbf{1}_{(M+1)\times(M+1)}L^{N-1}$, and $\mathcal{K}_{\mathcal{M}}^N = (G_N\rho_{N-1})^T\mathbf{1}_{(M+1)\times(M+1)}(G_N\rho_{N-1}) + 2L^{N-1^T}\mathbf{1}_{(M+1)\times(M+1)}(G_N\rho_{N-1})$.

Acknowledgements

The work was supported by the National Science Centre, Poland, grant No. 2016/21/B/ST7/02284.

References

1. M. Barenthin, M. Enqvist, B. Wahlberg, and H. Hjalmarsson. Gain estimation for Hammerstein systems. IFAC Proceedings Volumes, 39(1):784–789 (2006).
2. W. Findeisen, F.N. Bailey, M. Brdyś, K. Malinowski, P. Tatjewski, and A. Woźniak. Control and Coordination in Hierarchical Systems. J. Wiley, Chichester-New York (1980).
3. F. Giri and E. W. Bai. Block-Oriented Nonlinear System Identification. Lecture Notes in Control and Information Sciences 404. Springer (2010).
4. W. Greblicki and M. Pawlak. Nonparametric System Identification. Cambridge University Press (2008).
5. Z. Hasiewicz and G. Mzyk. Combined parametric-nonparametric identification of Hammerstein systems. IEEE Transactions on Automatic Control, 48(8):1370–1376 (2004).
6. Z. Hasiewicz and G. Mzyk. Hammerstein system identification by non-parametric instrumental variables. International Journal of Control, 82(3):440–455 (2009).
7. G. Mzyk. Nonlinearity recovering in Hammerstein system from short measurement sequence. IEEE Signal Processing Letters, 16(9):762–765 (2009).
8. G. Mzyk. Combined Parametric-Nonparametric Identification of Block-Oriented Systems. Lecture Notes in Control and Information Sciences 454, Springer (2014).
9. G. Mzyk and P. Wachel. Kernel-based identification of Wiener-Hammerstein system. Automatica, 83:275–281 (2017).
10. R. Pintelon and J. Schoukens. System Identification: a Frequency Domain Approach. Wiley-IEEE Press (2004).
11. J. Vöros. Recursive identification of Hammerstein systems with discontinuous nonlinearities containing dead-zones. IEEE Transactions on Automatic Control, 48(12):2203–2206 (2003).
12. M.P. Wand and H.C. Jones. Kernel Smoothing. Chapman and Hall, London (1995).