Efficient identification of Hammerstein systems by two-level optimization with decomposition

Grzegorz Mzyk

Wrocław University of Science and Technology, Poland

e-mail: grzegorz.mzyk@pwr.edu.pl

Abstract—**The paper considers popular problem of Hammerstein system identification. It is inspired by the real problem concerning modeling of differential scanning calorimetry for chalcogenide glass properties examination. In spite of variety of identification methods proposed in the literature, none of them can be applied directly, due to specific practical limitations. The most popular approaches, e.g. overparametrization approach, or nonparametric regression estimation, require relatively large number of data or lead to very complicated numerical tasks. The proposed algorithm consists of two steps. Firstly, the impulse response of the linear block is identified by the standard least squares method, assuming i.i.d. input excitation. Next, the coefficients of orthogonal expansion of nonlinear characteristic are estimated independently by iterative optimization, provided that the criterion function is convex. Results of simulation examples give promising results, i.e., satisfactory accuracy and relatively fast computations.**

I. Introduction

In the paper, the nonlinear dynamic system modeling problem is considered. If the linear model is not sufficient, the block-oriented structure of the model is commonly applied, including static nonlinear blocks, connected with linear dynamics. The most popular and the easiest from the system identification point of view is Hammerstein structure (see Fig. 1), consisting of static nonlinear element connected in series with linear dynamic block. Owing to numerous potential application the problem has been intensively elaborated since 1960's (see, *e.g.*, [19], [4] , [3], [22], and [2]). The ideas can be roughly divided into several categories:

- the key term separation methods [23];
- the hierarchical identification methods [5];
- the projection or gradient methods (stochastic gradient, heuristic, and genetic algorithms) [6];
- the iterative methods [19], [4], [14];
- the overparametrization approach [1], [7];
- nonparametric (kernel and orthogonal), regression-based identification [10], [8], [21];
- combined methods [11], [12], [9], [15], [17], [16], [18].

Each method is characterized by the specific set of prior assumptions imposed on the input, the noise, and the class of admitted nonlinear characteristics. In the real problems (see *e.g.* [13]), they cannot be verified and the selection of proper method can be problematic. In particular, the overparametrization approach leads to very attractive linear-in-parameters representation of Hammerstein system, but it generates the multivariable algebraic problem. In the nonparametric kernel regression methods, the prior knowledge can be poor, but its asymptotic properties reveal for relatively large amount of measurement data. The purpose of the paper is to propose numerically attractive and accurate identification method for i.i.d. random input.

The paper is organized as follows. In Section II, the problem is formulated in detail. Then, in Section III, two popular approaches known from literature are shortly described, and their drawbacks are indicated. Next, in Section IV, our algorithm is proposed and analyzed. Finally, the results of simulation examples are given in Section V.

II. Statement of the problem

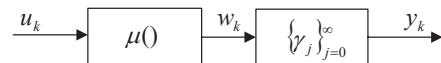The Hammerstein system, shown in Fig. 1 is described by

Fig. 1. Hammerstein system

the following equations

$$y_k = \sum_{j=0}^{\infty} \gamma_j \mu\left(u_{k-j}\right), \qquad (1)$$

$$\mu\left(u\right) = \sum_{i=1}^{\infty} \alpha_i \varphi_i\left(u\right).$$

We assume that:

A1. The nonlinear characteristic of static block, $\mu\left(u\right)$, is square integrable, *i.e.*, $\mu\left(u\right) \subset L_2$. The functions $\left\{\varphi_i\left(\right)\right\}_{i=0}^{\infty}$ constitutes complete orthonormal basis in L_2, i.e.

$$E\left\{\varphi_{i_1}\left(u\right)\varphi_{i_2}\left(u\right)\right\} = \left\{ \begin{array}{l} 1, \text{ as } i_1 = i_2 \\ 0, \text{ elsewhere} \end{array} \right. .$$

A2. The linear dynamic block is asymptotically stable, *i.e.*, $\sum_{j=0}^{\infty} \left|\gamma_j\right| < \infty$. To guarantee uniqueness of representation, without any loss of generality, we also take technical assumption that the steady-state gain of the linear component is one ($\sum_{j=0}^{\infty} \gamma_j = 1$). For detail discussion see [11].

A3. The input $\{u_k\}$ is an *i.i.d.* random process, with the probability density function $f\left(u\right)$ positive and continuous on compact support, *i.e.*, $f\left(u\right) > 0$, as $u \in [u_{\min}, u_{\max}]$.

The goal is to build the finite approximation model of the system

$$\overline{y}_k = \sum_{j=0}^{p} g_j m\left(u_{k-j}\right), \qquad (2)$$

$$m\left(u\right) = \sum_{i=1}^{s} a_i \varphi_i\left(u\right),$$

using the input-output measurement pairs $\{(u_k, y_k)\}_{k=1}^{N}$, such that

$$Q\left(g, a\right) = E\left(\overline{y}_k - y_k\right)^2 \to \min_{g,a}, \qquad (3)$$

where

$$g = (g_0, g_1, ..., g_p)^T \quad \text{and} \quad a = (a_1, a_2, ..., a_s)^T$$

are model parameters. For simplicity of presentation we assume that the linear block can be satisfactory approximated by the FIR model $\{\gamma_j\}_{j=0}^{p}$, and the terms $\gamma_{p+1}, \gamma_{p+2}, ...$ can be neglected, *i.e.*, we assume $\gamma_j = 0$ as $j > p$.

III. STATE OF THE ART

In this section we shortly describe two existing methods for Hammerstein system identification and indicate its limitations and drawbacks from the practical point of view.

A. Overparametrization approach

Introducing the vector of aggregated parameters

$$\theta = \left(\gamma_0 \alpha_1, .., \gamma_0 \alpha_1, ..., \gamma_p \alpha_1, .., \gamma_p \alpha_s\right)^T,$$

including all mixed products of γ_j's and α_i's, one can represent the Hammerstein system in by the linear equation with respect to θ, *i.e.*,

$$y_k = \phi_k^T \theta + \epsilon,$$

where ϵ is arbitrarily small approximation error, connected with the neglected tail $\{\alpha_i\}_{s+1}^{\infty}$. Owing to this, θ can be estimated as follows

$$\widehat{\theta} = \left(\Phi^T \Phi\right)^{-1} \Phi^T Y, \qquad (4)$$

where

$$\Phi = \left(\phi_1^T, \phi_2^T, ..., \phi_N^T\right)^T, \quad Y = (y_1, y_2, ..., y_N)^T,$$

and

$$\phi_k = \left(\varphi_1\left(u_k\right), .., \varphi_s\left(u_k\right), ..., \varphi_1\left(u_{k-p}\right), .., \varphi_s\left(u_{k-p}\right)\right)^T$$

are mapped regressor vectors. The vectors $\gamma = \left(\gamma_0, \gamma_1, ..., \gamma_p\right)^T$ and $\alpha = (\alpha_1, \alpha_2, ..., \alpha_s)^T$ can be extracted from the estimate of θ with the use of singular value decomposition technique ([1]). Although the algorithm is based on the standard linear least squares procedure, the dimensionality of θ can exclude the idea in many practical applications. The number of measurements N must be necessary greater than the number of aggregated parameters $n = \dim \theta = (p+1)s$, and at least n of ϕ_k vectors must be linearly independent for the estimate to be well defined. Moreover, if n is close to N, then the estimate is very sensitive on the potential output noise.

B. Nonparametric kernel regression estimation

In the standard regression-based nonparametric methods (kernel or orthogonal, see *e.g.* [10]) the system output is treated as the sum of the two components

$$y_k = \gamma_0 \mu\left(u_k\right) + \sum_{j=1}^{\infty} \gamma_j \mu\left(u_{k-j}\right).$$

First of them, $\gamma_0 \mu\left(u_k\right)$, is treated as informative term, and the remaining part, the tail $\sum_{j=1}^{\infty} \gamma_j \mu\left(u_{k-j}\right)$, as the "system" noise. The observation that the simple input-output static regression

$$R(u) \triangleq E\left\{y_k | u_k = u\right\} = \gamma_0 \mu(u) + c$$

$$c = \sum_{j=1}^{p} \gamma_j E\mu(u_1)$$

is equivalent to the system nonlinearity (up to some scale and offset), allows for its estimation by, *e.g.*, kernel method

$$\widehat{\mu}\left(u\right) = \frac{\sum_{k=1}^{N} y_k K\left(\frac{u_k - u}{h}\right)}{\sum_{k=1}^{N} y_k K\left(\frac{u_k - u}{h}\right)} \qquad (5)$$

under very poor prior knowledge. The parametric form of $\mu\left(\right)$ and the difference equation describing linear filter need not to be known. Nevertheless, asymptotic properties of the estimates works for relatively large number of measurements. For moderate N, the "noise" term $\sum_{j=1}^{\infty} \gamma_j \mu\left(u_{k-j}\right)$ can dominate the usable one, *i.e.* $\gamma_0 \mu\left(u_k\right)$, and produce high variance error. This problem is widely discussed in [15].

IV. THE PROPOSED ALGORITHM

The proposed algorithm has two steps. It starts from estimation of impulse response of the linear block by the least squares method. Then, in second step, the nonlinear block is decomposed on s parallel channels, and each channel is identified independently of the others, by iterative optimization.

A. Identification of linear dynamic block

The impulse response $\gamma = \left(\gamma_0, \gamma_1, ..., \gamma_p\right)^T$ is estimated by the least squares method

$$\widehat{\gamma} = \left(\Xi_N^T \Xi_N\right)^{-1} \Xi_N^T Y_N, \qquad (6)$$

where

$$\Xi_N = \left(\vartheta_1^T, \vartheta_2^T, ..., \vartheta_N^T\right)^T,$$

$$\vartheta_k = \left(u_k^*, u_{k-1}^*, ..., u_{k-p}^*\right)^T,$$

$$u_k^* = u_k - Eu_k,$$

$$Y_N = (y_1^*, y_2^*, ..., y_N^*)^T,$$

$$y_k^* = y_k - Ey_k.$$

The estimate $\widehat{\gamma}$ should be then normalized, such that $\sum_{j=0}^{p} \widehat{\gamma}_j = 1$ (see Assumption A2). The convergence $\widehat{\gamma} \to \gamma$, as $N \to \infty$, is a simple consequence of the fact that $E\left(y_{k+j}^* u_k^*\right)$ is proportional to γ_j.

B. Identification of static nonlinear characteristics

The model (2) of the Hammerstein system is shown in Fig. 2. It consists of s parallel channels with coefficients a_i, $i = 1, 2, ..., s$. As it will be shown below, the channels can be identified independently owing to orthogonality of basis functions φ_i's with respect to the input probability density function $f(u)$. The algorithm looks for the best model
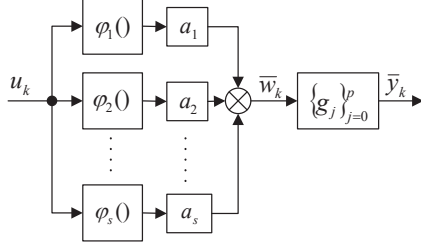

Fig. 2.   The parametric model

parameters g and a to minimize the following criterion

$$Q(g, a) = E\left\{\overline{y}_k(g, a) - y_k\right\}^2 \to \min_{g, a}.$$

Let us introduce the symbol

$$\delta_i \triangleq a_i - \alpha_i \tag{7}$$

for the error of individual coefficient α_i of the nonlinearity representation, and let

$$x_j \triangleq \sum_{i=1}^{s} \left(g_j a_i - \gamma_j \alpha_i\right) \varphi_i\left(u_{k-j}\right) \tag{8}$$

be the output error connected with δ_i. One can write that

$$
\begin{aligned}
x_j &= \sum_{i=1}^{s} \left[g_j\left(\alpha_i + \delta_i\right) - \gamma_j \alpha_i\right] \varphi_i\left(u_{k-j}\right) = \\
&= \sum_{i=1}^{s} \left[\left(g_j - \gamma_j\right)\alpha_i + g_j \delta_i\right]\varphi_i\left(u_{k-j}\right),
\end{aligned}
$$

and hence

$$
\begin{aligned}
\overline{y}_k - y_k &= \sum_{j=0}^{p} x_j = \sum_{j=0}^{p}\sum_{i=1}^{s}\left(g_j - \gamma_j\right)\alpha_i\varphi_i\left(u_{k-j}\right) + \\
&\quad + \sum_{j=0}^{p}\sum_{i=1}^{s}g_j\delta_i\varphi_i\left(u_{k-j}\right).
\end{aligned}
$$

For $g_j = \gamma_j$ we obtain that

$$\overline{y}_k - y_k = \sum_{i=1}^{s}\sum_{j=0}^{p}\gamma_j\delta_i\varphi_i\left(u_{k-j}\right). \tag{9}$$

Now, we focus on estimation of α_{i_0} and rewrite (9) as follows

$$\overline{y}_k - y_k = \sum_{j=0}^{p}\gamma_j\delta_{i_0}\varphi_{i_0}\left(u_{k-j}\right) + c_{i_0},$$

where

$$c_{i_0} = \sum_{i=1, i\neq i_0}^{s}\sum_{j=0}^{p}\gamma_j\delta_i\varphi_i\left(u_{k-j}\right).$$

Hence

$$
\begin{aligned}
E\left(\overline{y}_k - y_k\right)^2 &= E\left(\sum_{j=0}^{p}\gamma_j\delta_{i_0}\varphi_{i_0}\left(u_{k-j}\right)\right)^2 + Ec_{i_0}^2 + \\
&\quad + 2E\left\{\left(\sum_{j=0}^{p}\gamma_j\delta_{i_0}\varphi_{i_0}\left(u_{k-j}\right)\right)c_{i_0}\right\}.
\end{aligned}
$$

For fixed $a_1, ..., a_{i_0-1}, a_{i_0+1}, ..., a_s$, it can easily be shown that

$$Ec_{i_0}^2 = \beta_0,$$

$$2E\left\{\left(\sum_{j=0}^{p}\gamma_j\delta_{i_0}\varphi_{i_0}\left(u_{k-j}\right)\right)c_{i_0}\right\} = \beta_1\delta_{i_0},$$

$$E\left(\sum_{j=0}^{p}\gamma_j\delta_{i_0}\varphi_{i_0}\left(u_{k-j}\right)\right)^2 = \beta_2\delta_{i_0}^2,$$

where β_0, β_1, and β_2 are some unknown constants, *i.e.*,

$$Q\left(a_{i_0}\right) = \beta_2\delta_{i_0}^2 + \beta_1\delta_{i_0} + \beta_0,$$

and moreover

$$\frac{\partial^2 Q\left(a_{i_0}\right)}{\partial \delta_{i_0}^2} = 2\beta_2 > 0,$$

since

$$E\left(\sum_{j=0}^{p}\gamma_j\delta_{i_0}\varphi_{i_0}\left(u_{k-j}\right)\right)^2 > 0.$$

Consequently, $Q\left(a_{i_0}\right)$ is convex with respect to $a_{i_0} = \delta_{i_0} + \alpha_{i_0}$, independently of the remaining parameters. In the computer implementation the criterion $Q\left(a_{i_0}\right)$ is replaced with its empirical version

$$\widetilde{Q}\left(a_{i_0}\right) = \frac{1}{N}\sum_{k=1}^{N}\left(\overline{y}_k - y_k\right)^2. \tag{10}$$

The procedure is as follows.

Initialization. Set κ_0, λ_0, and ε_0, such that $\kappa_0 < \alpha_{i_0} < \lambda_0$, and $0 < \varepsilon_0 < \frac{\kappa_0 + \lambda_0}{2}$.

Step n-th. If $\widetilde{Q}\left(\frac{\kappa_n + \lambda_n}{2} - \varepsilon_n\right) \geq \widetilde{Q}\left(\frac{\kappa_n + \lambda_n}{2} - \varepsilon_n\right)$ then set $\kappa_{n+1} := \frac{\kappa_n + \lambda_n}{2} - \varepsilon_n$, and $\lambda_{n+1} := \lambda_n$.

If $\widetilde{Q}\left(\frac{\kappa_n + \lambda_n}{2} - \varepsilon_n\right) < \widetilde{Q}\left(\frac{\kappa_n + \lambda_n}{2} - \varepsilon_n\right)$ then set $\kappa_{n+1} := \kappa_n$, and $\lambda_{n+1} := \frac{\kappa_n + \lambda_n}{2} + \varepsilon_n$.

Set $\varepsilon_{n+1} := \varepsilon_n/2$.

Stop condition. Stop, if $|\lambda_n - \kappa_n|$ is appropriately small.

V. SIMULATION EXAMPLE

To illustrate advantages of the approach we show simple experiment with the simulated Hammerstein system. The system with nonlinear characteristic $\mu(u) = 1 - |u|$ followed by the FIR filter with the 10-element impulse response $\gamma = (0.1, 0.1, ..., 0.1)$ was excited by the uniformly distributed input process $u_k \sim U[-1, 1]$. The output was disturbed by the

TABLE I
MISE ERROR FOR VARIOUS METHODS

	$N = 50$	$N = 100$
kernel estimate	0,2881	0,0190
LS+SVD	cannot be computed	0,1120
proposed method	0,0007	0,0001

noise process $z_k \sim U[-1, 1]$. For $N = 50$ and $N = 100$ input-output data, the following Mean Integrated Squared Error of the nonlinear characteristic

$$MISE(\widehat{\mu}(u)) = E \int_{-1}^{1} (\widehat{\mu}(u) - \mu(u))^2$$

was computed numerically. In the parametric methods we set respectively $s = 10$ and $p = 9$ with cosine orthonormal basis $\left\{ \frac{1}{\sqrt{2}}, \cos \pi x, \cos 2\pi x, ... \right\}$. The results are shown in Tab. I.

VI. CONCLUSIONS

For simplicity of presentation, in the paper we assumed uncorrelated input process, and FIR representation of the linear block. The idea can be naturally generalized for ARMA IIR models (see, *e.g.*, [20] for the methods based on the best linear approximation). As regards the colored input, the instrumental variables technique can be applied instead of least squares, to cope with the problem (see, *e.g.*, [17]). The main advantage of the idea is decomposition of the optimization task into two independent subproblems. The impulse response $\{\gamma_j\}_{j=0}^{p}$ of the linear dynamic block is estimated completely independently of the coefficients of static nonlinear characteristic, using standard and well elaborated linear least squares procedure. In the second step, thanks to mutual orthogonality of basis functions $\varphi_i()$'s, all parameters α_i's are estimated independently by simple iterative optimization of convex function. We emphasize that if the input is not uniformly distributed, it is still possible to design appropriate basis $\{\varphi_i()\}$, orthogonal with respect to the input probability density function ([21]). Comparing to the overparametrization approach, we avoid construction of multidimensional vectors, which significantly widens the scope of potential applications, particularly for short data sequences. Moreover, in the contrary to nonparametric orthogonal expansion algorithms ([21]), estimation of coefficients of the nonlinear characteristic is supported by the model of linear dynamics, obtained in the first step of the procedure. Results of the simulation examples show that the comparable accuracy can by achieved for smaller number of data. Decomposition of the procedure on independent channels allows for application of parallel computation implementation. All derivations concerning identification of the nonlinear element were made under assumption that the accurate impulse response $\{\gamma_j\}_{j=0}^{p}$ is known. The formal analysis of the estimates $\widehat{\alpha}_i$'s, when the true $\{\gamma_j\}_{j=0}^{p}$ are replaced with its estimates $\{\widehat{\gamma}_j\}_{j=0}^{p}$, remains open for future research.

REFERENCES

[1] E.W. Bai. An optimal two-stage identification algorithm for Hammerstein-Wiener nonlinear systems. *Automatica*, 34(3):333–338, 1998.
[2] E.W. Bai and D. Li. Convergence of the iterative Hammerstein system identification algorithm. *IEEE Transactions on Automatic Control*, 49(11):1929–1940, 2004.
[3] S.A. Billings and S.Y. Fakhouri. Identification of systems containing linear dynamic and static nonlinear elements. *Automatica*, 18(1):15–26, 1982.
[4] F. Chang and R. Luus. A noniterative method for identification using Hammerstein model. *IEEE Transactions on Automatic Control*, 16(5):464–468, 1971.
[5] F. Ding and T. Chen. Hierarchical least squares identification methods for multivariable systems. *IEEE Transactions on Automatic Control*, 50(3):397–402, 2005.
[6] F. Ding, X.P. Liu, and G. Liu. Identification methods for Hammerstein nonlinear systems. *Digital Signal Processing*, 21(2):215–238, 2011.
[7] F. Giri and E. W. Bai. *Block-Oriented Nonlinear System Identification*. Lecture Notes in Control and Information Sciences 404. Springer, 2010.
[8] W. Greblicki. Nonparametric orthogonal series identification of Hammerstein systems. *International Journal of System Science*, 20:2355–2367, 1989.
[9] W. Greblicki and G. Mzyk. Semiparametric approach to Hammerstein system identification. In *Proceedings of the 15th IFAC Symposium on System Identification, Saint-Malo, France*, pages 1680–1685, 2009.
[10] W. Greblicki and M. Pawlak. Identification of discrete Hammerstein systems using kernel regression estimates. *IEEE Transactions on Automatic Control*, 31:74–77, 1986.
[11] Z. Hasiewicz and G. Mzyk. Combined parametric-nonparametric identification of Hammerstein systems. *IEEE Transactions on Automatic Control*, 48(8):1370–1376, 2004.
[12] Z. Hasiewicz and G. Mzyk. Hammerstein system identification by nonparametric instrumental variables. *International Journal of Control*, 82(3):440–455, 2009.
[13] B. Kozdraś and G. Mzyk. Identification of the heating process in Differential Scanning Calorimetry with the use of Hammerstein model. *submitted to Control Engineering Practice*, 2016.
[14] G. Li, C. Wen, W.X. Zheng, and G. Zhao. Iterative identification of block-oriented nonlinear systems based on biconvex optimization. *Systems & Control Letters*, 79:68–75, 2015.
[15] G. Mzyk. Nonlinearity recovering in Hammerstein system from short measurement sequence. *IEEE Signal Processing Letters*, 16(9):762–765, 2009.
[16] G. Mzyk. *Combined Parametric-Nonparametric Identification of Block-Oriented Systems*. Lecture Notes in Control and Information Sciences 454. Springer, 2014.
[17] G. Mzyk. Instrumental variables for nonlinearity recovering in block-oriented systems driven by correlated signals. *International Journal of Systems Science*, 46(2):246–254, 2015.
[18] G. Mzyk and P. Wachel. Kernel-based identification of Wiener-Hammerstein system. *Automatica*, 83:275–281, 2017.
[19] K.S. Narendra and P. G. Gallman. An iterative method for the identification of nonlinear systems using the Hammerstein model. *IEEE Transactions on Automatic Control*, 11:546–550, 1966.
[20] R. Pintelon and J. Schoukens. *System Identification: a Frequency Domain Approach*. Wiley-IEEE Press, 2004.
[21] P. Śliwiński. *Nonlinear System Identification by Haar Wavelets*. Lecture Notes in Statistics 210. Springer, 2013.
[22] P. Stoica and T. Söderström. Instrumental-variable methods for identification of Hammerstein systems. *International Journal of Control*, 35(3):459–476, 1982.
[23] J. Vörös. Recursive identification of Hammerstein systems with discontinuous nonlinearities containing dead-zones. *IEEE Transactions on Automatic Control*, 48(12):2203–2206, 2003.