Discussion on: "Subspace-based Identification Algorithms for Hammerstein and Wiener Models"

W. Greblicki[1] and M. Pawlak[2]

[1]Institute of Engineering Cybernetics, Wroclaw University of Technology, Wroclaw, Poland

[2]Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, Manitoba, R3T5V6, Canada

printed in: European Journal of Control Vol. 11, No. 2, 145-148, 2005

We congratulate the authors for their interesting paper on identification of multi-input multi-output nonlinear cascade systems of the Hammerstein/Wiener type.

The paper by J.C. Gomez and E. Baeyens presents a class of sub-space based algorithms for recovering the characteristics of the multivariate Hammerstein and Wiener systems. The presented contribution are suited for the parametric inference since both the linear and nonlinear subsystems are represented by a finite dimensional parameter space. We will address two related issues which seem to be relevant to the paper. First, we would like to point out possible extension of the authors contribution where some a prior information about the parametric form of the nonlinear characteristics must be taken into consideration. This leads to the constrained version of the authors algorithm.

If, however, the *a priori* information about the characteristics is not available one is forced to apply the nonparametric approach to the identification problem. This issue is examined in the second part of our discussion. We conclude with some general remarks on some further issues related to the paper.

1. Constrained Identification

In many practical applications the non-linear characteristics characterizing the Hammerstein and Wiener systems are known to be non-decreasing, convex, or non-negative. Among these shape constraints the monotonicity seems to be the most common and we will focus on this case in more details. In fact, in all examples given in the paper, the nonlinear characteristics are monotonic. This situation is particularly critical in the Wiener model case when it is common to assume that the nonlinear characteristic $N(\cdot)$ is invertible. No estimation method can guarantee that the estimated characteristic is invertible. This fact has been largely ignored in many papers on the Wiener system identification.

To include the monotone case in the particular identification algorithm one can simply add constraints forcing the solution to meet the monotonicity requirement. This framework can be easily incorporated in the least squares approach to system identification by minimizing a penalized sum of squares. This is the case in the present paper as the authors apply the least squares strategy with the singular value decomposition extension for separation of the parameters of the linear part from the non-linear one.

Here, we show how to extend the presented algorithm to the constrained case when one is looking for an estimate of $N(\cdot)$ which is monotonicity preserving. To fix our idea, without loss of generality, let us consider the single-input-single output version of the Hammerstein system

$$\begin{aligned} x_{k+1} &= Ax_k + bv_k, \\ v_k &= c^T x_k + \epsilon_k, \end{aligned}$$

where b, c are the $d \times 1$ vectors and ϵ_k is the output noise process.

Let also

$$v_k = \sum_{i=1}^{r} \alpha_i g_i(u_k) \tag{1.1}$$

be the parametric representation of the nonlinear characteristic, i.e., the basis functions $g_i(\cdot)$ are selected in advance and the parameter r is assumed to be known. Using the vector notation introduced in the paper we can rewrite the above input–output description in the following equivalent form:

$$\begin{aligned} x_{k+1} &= Ax_k + \Theta U_k, \\ v_k &= c^T x_k + \epsilon_k, \end{aligned}$$

where $\Theta = b\alpha^T$ is the $d \times r$ matrix and U_k is the $r \times 1$ known vector of the transformed input signal u_k via the basis functions $g_i(\cdot), i = 1, \ldots, r$.

The identification procedure proposed in the paper starts with the initial estimates (obtained virtually by any linear system identification method) of A, c, Θ, i.e., we obtain $\hat{A}$, $\hat{c}$, $\hat{\Theta}$. Then the estimates of hidden parameters b, α are derived as the solution of the following least squares problem:

$$(\hat{b}, \hat{\alpha}) = \arg\min_{b,\alpha} ||\hat{\Theta} - b\alpha^T||^2.$$

This problem is solved by the SVD algorithm. All these derivations lead to the following estimate of the nonlinear function

$$\hat{N}(u) = \sum_{i=1}^{r} \hat{\alpha}_i g_i(u).$$

The identical procedure applied to the Wiener system gives the estimate of the inverse function $N(\cdot)$ appearing in this system, i.e., we have

$$\hat{N}^{-1}(u) = \sum_{i=1}^{r} \hat{\alpha}_i g_i(u).$$

It is clear that the monotonicity of $N(\cdot)$ does not imply the monotonicity of $\hat{N}(\cdot)$. Therefore, it is an important issue to construct the constrained solution $\hat{N}(u)$ which is not-decreasing function of u. To do so, let us define the standardized form of the constrained solution

$$N_c(u) = \sum_{i=1}^{r} (\hat{\alpha}_i + \beta_i) g_i(u),$$

where $\{\beta_i\}$ is a new set of unknown parameters. Clearly, if the unconstrained solution $\hat{N}_c(u)$ meets the monotonicity requirement then we should have $\beta_i = 0, i = 1, \ldots, r$.

The monotonicity of $N_c(u)$ gives the following condition for $\{\beta_i\}$

$$\sum_{i=1}^{r} \beta_i g_i^{(i)}(u) \geq -\sum_{i=1}^{r} \hat{\alpha}_i g_i^{(i)}(u),$$

where $g_i^{(i)}(u)$ is the derivative of $g_i(u)$. The above formula can be written in the following vector notation form

$$\beta^T h(u) \geq \alpha(u), \tag{1.2}$$

where $\beta = (\beta_1, \ldots, \beta_r)^T$, $h(u) = (g_1^{(1)}, \ldots, g_r^{(1)})^T$ and $\alpha(u) = -\sum_{i=1}^{r} \hat{\alpha}_i g_i^{(1)}(u)$. Often one needs to do some normalization of the vector $h(u)$ such that $h^T(u)h(u) = 1$. Now, we are in a position to reformulate the procedure proposed in the paper which takes into account the monotonicity constraint. Hence, we seek the solution of the following minimization problem

$$\hat{\beta} = \arg\min_{\beta} ||\hat{\Theta} - \hat{b}(\hat{\alpha} + \beta)^T||^2$$

subject to the constraint in (1.2). Since $\hat{\Theta}$, $\hat{b}$, $\hat{\alpha}$ are already specified the above question is equivalent to the quadratic programming problem with linear constraints. It is worth noting that constraint (1.2) is required to hold for all u. The weaker requirement would ask for (1.2) to be satisfied only at the training input data points u_k, $k = 1, \ldots, N$.

All the aforementioned observation would lead to the following algorithm for finding $\hat{\beta}$:

1. If $a(u) \leq 0$, for all u, i.e., $\sum_{i=1}^{r} \hat{\alpha}_i g_i^{(i)}(u) \geq 0$ then set $\hat{\beta} = 0$. This is the case when unconstrained solution $\hat{N}(u)$ meets the monotonicity constraint.

2. If $a(u) > 0$, for some u then $\hat{\beta} = 0$ is not the solution. In this case define

   $$a(u^*) = \max_u a(u)$$

   and define

   $$\beta^* = a(u^*)h(u^*).$$

   Verify whether β^* satisfies (1.2). This can be done either for all u or for the input data points u_k, $k = 1, \ldots, N$. If (1.2) holds then set $\hat{\beta} = \beta^*$. The case if β^* does not meet (1.2) requires some further investigation and some suggestions can be found in Refs [2,8].

3. Exit the algorithm with the following monotonicity constrained estimate of $N(u)$:

   $$N(u) = \sum_{i=1}^{r} (\hat{\alpha}_i + \hat{\beta}_i) g_i(u).$$

The consistency of the estimate $\hat{N}(u)$ can be established similarly as it has been done in the paper. Yet, another interesting issue (not examined in the paper) is to evaluate the rate of convergence of the proposed estimates. As in any parametric estimation problem one can expect the $O(1/N)$ rate. A precise evaluation of the estimates bias and covariance would be an interesting issue for future research. In particular, the influence of the dimensionality and correlation structure of the input signal on the estimates statistical accuracy remains to be addressed.

2. Non-Parametric Identification

Thus far we have assumed the parametric knowledge of the system nonlinearity defined by formula (1.1). Hence both the basis functions $g_i(u)$, $i = 1, \ldots, r$ and the parameter r are assumed to be known. Often such an assumption is unacceptable and we are forced to turn into non-parametric estimation methods [3]. Such an approach to the identification problem of the Hammerstein and Wiener systems was initiated in [7] and developed in a number of papers afterwards, see Refs [4–7,9,10] and the references cited therein.

The first step into the non-parametric extension of the proposed approach would be to allow the parameter r to be selected adaptively for known class of basis functions. In Ref. [9], the nonlinearity in the Hammerstein system was identified with the help of $g_i(u)$ being the orthogonal Legendre polynomials. Assuming the knowledge of the probability distribution of the input signal the procedure for adaptive selection of r was proposed. The consistency result being the realization of the well known bias − variance trade-off has been obtained under the condition that $r = r(N) \to \infty$ and $r(n)/N \to 0$ as $n \to \infty$. See Ref. [5] for similar developments in the context of recovering the nonlinearity of the Wiener system.

These results seem to be easily incorporated into the framework of the present paper. In fact, the multidimensional nature of the input signal does change the basic algorithms developed in the aforementioned papers. The rates of convergence, however, will strongly depend on the input dimensionality. In fact it is known [3] that the optimal rate of convergence for recovering the d dimensional nonlinear function is $O(N^{-2s/(2s+d)})$, where s is the smoothness (for instance number of existing derivatives) of the nonlinearity. Thus, the dimensionality reduces the accuracy of nonparametric estimates. This is a well known phenomenon called the curse of dimensionality. To alleviate this problem some special low dimensional approximations can be used. Among a number of possibilities the multi-channel structures are particularly attractive, see Ref. [11] for recent developments of the theory of non-parametric identification of multi-channel block oriented systems. As a result of these approximations we can obtain rates being independent on the dimensionality of the input signal. An example of the two-channel Wiener model is presented in Fig. 6 of the present paper. It can be shown that for this model the inverse regression approach introduced in Ref. [4] can be applied and a fully nonparametric estimate of $N^{-1}(\cdot)$ can be derived. The only difficulty which we expect is due to the correlation structure of the inputs signals $(u_1(k), u_2(k))$.

Yet, an another important issue is to how to incorporate the shape constraints (like the monotonicity) into the non-parametric framework. This challenging question is left for future studies.

3. Concluding Remarks

The authors have given a detailed numerical analysis of the proposed subspace algorithm for the parameter estimation of the Hammerstein and Wiener systems. Some preliminary convergence results have also been given. There are further questions to be answered on the statistical accuracy of the estimates. The constrained identification option discussed briefly in Section 1.1 may lead to more accurate estimates; the issue which remains to be addressed in more details. There is a large class of block-oriented models which are useful in many applications. Important examples are: the sandwich model [1] and a class of parallel models [10]. Is it possible to extend the methodology developed in this paper to the case of these important structures?

References

[1] Brillinger DR. The identification of a particular nonlinear time series system. Biometrika 1977; 64: 509–515

[2] Dykstra RL. An algorithm for restricted least squares regression. J Amer Statist Assoc 1983; 77: 621–628

[3] Efromovich S. Nonparametric Curve Estimation. Springer-Verlag, New York, 1999

[4] Greblicki W. Nonparametric identification of Wiener systems. IEEE Trans Inform Theory 1992; 38: 1487–1493

[5] Greblicki W. Nonparametric identification of Wiener systems by orthogonal series. IEEE Trans Automatic Control 1994; 39: 2077–2086

[6] Greblicki W. Nonlinearity recovering in Wiener system driven with correlated signal. IEEE Trans Automatic Control 2004; 49: 1805–1810

[7] Greblicki W, Pawlak M. Identification of discrete Hammerstein systems using kernel regression estimates. IEEE Trans Automatic Control 1986; 31: 74–77

[8] Rice JR. Approximation with convex constraints. SIAM J Appl Math 1963; 11: 15–32

[9] Pawlak M. On the series expansion approach to the identification Hammerstein systems. IEEE Trans on Autom Control 1991; 36: 763–767

[10] Pawlak M, Hasiewicz Z. Nonlinear system identification by the Haar multiresolution analysis. IEEE Trans. Circuits Syst-I: Fund Theory Appl 1998; 45: 945–961

[11] Pawlak M, Hasiewicz Z. Non-parametric identification of nonlinear one and two-channel systems by multiscale expansions. Technical Report, 2005