NONPARAMETRIC IDENTIFICATION OF TWO-CHANNEL NONLINEAR SYSTEMS

W. Greblicki* and M. Pawlak**

*Department of Electrical Engineering, The University of Manitoba, Winnipeg, Canada, R3T 2N2
On leave from Institute of Engineering Cybernetics, Technical University of Wroclaw, Wroclaw, Poland

** Department of Electrical Engineering, The University of Manitoba, Winnipeg, Canada, R3T 2N2

Abstract

In this paper, a discrete-time two-channel non-linear system is identified. Each branch of the system has the form of the Hammerstein model, i.e., a nonlinear gain function followed by a dynamic linear system. The dynamic subsystems are recovered using the standard correlation method. The main results are concerned with the estimation of the nonlinear memoryless subsystems. The class of nonlinearities considered in the paper, consists of those Borel functions that do not increase faster than linear functions. The identification algorithm is a nonparametric kernel estimate of the regression function. The statistically dependent, as well as independent random signal inputs are assumed. For the first case, the algorithm achieves a rate of convergence of the order $O(n^{-1/4})$ while the latter one, $O(n^{-1/3})$ is achieved in probability, where n is the sample size.

1. Introduction

In this paper we identify a nonlinear discrete-time system shown in Fig. 1. This is a two-channel system in which each branch is in the form of the Hammerstein model, i.e., the first subsystem is nonlinear and memoryless while the second is linear and dynamic. The nonlinear characteristics are denoted by $m_i, i = 1, 2$ and the weighting functions by $\{g_i\}$, $\{h_i\}$. The inputs U_n and V_n are stationary white noises with joint density function $f(u, v)$. The output Y_n is distributed by a zero mean stationary white noise ε_n. Moreover η_n and ξ_n are also zero mean white noises.

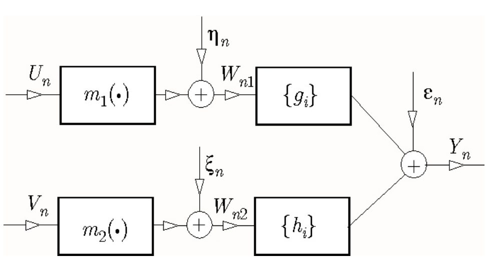

Fig. 1. The identified nonlinear system.

The systems of such form are encountered, for example, in digital data transmission (see, for example, Maqusi [9]) as well as in image processing (12). The problem of the paper is to identify both the subsystems in each branch from the input-output observations of the whole system. The signals $W_{n,1}$ and $W_{n,2}$ interconnecting both the subsystems are inaccessible to measurements.

Such a problem, for a single Hammerstein system, has been investigated by a number of authors, see the original work of Narendra and Gallman [11] as well as later papers of Billings and Fakhouri [1] and Kung and Shih [8]. In [2] the two-channel system was discussed. So far, all authors studying the problem have assumed that nonlinearities are of a polynomial form. Therefore the problem investigated by them is parametric for they estimate only a finite number of coefficients.

Recently, observing that the nonlinear characteristics can be expressed as a regression function, Greblicki and Pawlak [5, 6] proposed a new approach based on nonparametric estimation of regression function. Their idea made possible to introduce by them, an algorithm not only completely independent of procedures which can be used for identification of the linear subsystem, but, first of all, recovers nonlinear characteristics even if they are not continuous.

In this paper we extend their approach for the case of two-channel nonlinear systems. Specifically, the rate of convergence of proposed methods is given. We consider the case of dependent as well as independent input random variables (Un, Vn). For the latter case we are enable to estimate separately each nonlinear characteristic.

2. Identification Algorithms

We assume that the input signals (Un, Vn), $n = ..., -1, 0, 1, ...$ are the stationary white noises and that

$$EU_n^2 + EV_n^2 < \infty. \quad (1)$$

The nonlinear characteristics of the system are Borel functions and satisfy the following condition

$$|m_i(u)| \le c_{1i} + c_{2i}|u|, \quad (2)$$

c_{1i} and c_{2i} some constants, $i = 1, 2$.

In turn, the linear part of the system is described by the following state equations:

$$\left.\begin{aligned} X_{n+1,i} = \;& A_i X_{n,i} + b_i W_{n,i} \\ Y_n = \;& c_1^T X_{n,1} + c_2^T X_{n,2} \\ & + W_{n,1} + W_{n,2} + \varepsilon_n, \end{aligned}\right\} \tag{3}$$

$i = 1, 2$, where $n = ..., -1, 0, 1, ...$, $W_{n,1} = m_1(U_n) + \eta_n$ and $W_{n,2} = m_2(U_n) + \xi_n$, $X_{n,1}$, $X_{n,2}$ are state vectors of the first and second channel, respectively, whereas b_i, c_i, $i = 1, 2$, are some vectors. A_1, A_2 are matrices assumed to be asymptotically stable. We assume, moreover, that $\{(U_n, V_n)\}$ and $\{\eta_n\}$ as well as $\{\xi_n\}$, and $\{\varepsilon_n\}$ are mutually independent.

From (3) it follows that

$$Y_n = \sum_{j=0}^{\infty} g_j W_{n-j,1} + \sum_{j=0}^{\infty} h_j W_{n-j,2} + \varepsilon_n, \tag{4}$$

where $g_j = c_1^T A_1^{j-1} b_1$, $j = 1, 2, \ldots$, $g_0 = 1$ and $h_j = c_2^T A_2^{j-1} b_2$, $j = 1, 2, \ldots$, $h_0 = 1$.

It is clear that (1) and (2) imply $EW_{n,i}^2 < \infty$, $i = 1, 2$. From this and asymptotical stability of A_i, $i = 1, 2$, it follows that Y_n is random variable. Moreover $\{Y_n\}$, $n = \ldots, -1, 0, 1, \ldots$ is a strictly stationary and ergodic process, Hannan [6].

It means that the restrictions (1) and (2) are totally independent of the identification method presented in the paper. They simply make the problem well posed in the sense that the output of the system is a random variable. The class of functions satisfying (2) is so wide that it cannot be parametrized. Due to that, the problem of recovering the characteristics is nonparametric.

Using the idea of Greblicki and Pawlak [5] we first compute the regression function of Y_n on the inputs variables (U_n, V_n). From (4) it results that

$$\begin{aligned} E\{Y_n | U_n = u, V_n = v\} &= m_1(u) + m_2(v) + \gamma \\ &= \psi(u, v), \text{ say,} \end{aligned} \tag{5}$$

where $\gamma = E\{m_1(U)\} \sum_{i=1}^{\infty} g_i + E\{m_2(U)\} \sum_{i=1}^{\infty} h_i$. Therefore, we can estimate merely a linear combination of m_1 and m_2. Estimating m_1 and m_2 separately is, generally, beyond our possibilities. Such a result is to be expected since no restrictions have been placed on the functions m_1, m_2, on the distribution of the input variables as well as on the dynamical subsystems.

In order to estimate $m_1(u) + m_2(v)$ (which is still applicable) we first do the regression function in (5). For this we apply

$$\hat{\psi}(u, v) = \frac{\displaystyle\sum_{i=0}^{n-1} Y_i K\left(\frac{u - U_i}{r(n)}\right) K\left(\frac{v - V_i}{r(n)}\right)}{\displaystyle\sum_{i=0}^{n-1} K\left(\frac{u - U_i}{r(n)}\right) K\left(\frac{v - V_i}{r(n)}\right)}, \tag{6}$$

where K is the bounded and nonnegative Borel kernel and $\{r(n)\}$ is a sequence of positive numbers. $\{U_i, V_i; Y_i\}$, $i = 0, 1, \ldots, n$ is a given sample of the input and output signals of the whole system.

Let us assume that

$$m_1(u = 0) = m_2(u = 0) = 0. \tag{7}$$

that is satisfied in most applications. Due to (7) we can estimate $m_1(u) + m_2(v)$ with $\hat{m}(u, v)$ which is defined in the following way:

$$\hat{m}(u, v) = \hat{\psi}(u, v) - \hat{\psi}(u = 0, v = 0).$$

Estimate (6) has been studied in statistical literature since the original works of Watson [14] and Nadaraya [10] or the more recent results of Devroye [3] as well As Greblicki, Krzyżak and Pawlak [4]. All of those authors studied, however, a simpler case since they assumed that pairs $(U_i, V_i; Y_i)$'s are mutually independent. In (6), $\{Y_n\}$ being the output of a dynamic system, is a sequence of dependent random variables.

Under some smoothing conditions on the nonlinear characteristics we show that $\hat{m}(u, v)$ converges to $m_1(u) + m_2(v)$ as rapidly as $O(n^{-1/4})$ in probability.

Now we make use of the additional restriction of the paper i.e.,the assumption that

$$\left.\begin{array}{c} U_n \text{ and } V_n \text{ are mutually independent} \\ \text{random variables.} \end{array}\right\} \tag{8}$$

This restriction enables us to estimate each nonlinear characteristics separately. Indeed, due to (8) and (4) we have

$$E\{Y_n | U_n = u\} = m_1(u) + \gamma_1$$

and

$$E\{Y_n | V_n = v\} = m_2(v) + \gamma_2,$$

where

$$\gamma_1 = E\{m_1(U)\} \sum_{i=1}^{\infty} g_i + E\{m_2(V)\} \sum_{i=0}^{\infty} h_i,$$

and

$$\gamma_2 = E\{m_1(U)\} \sum_{i=0}^{\infty} g_i + E\{m_2(V)\} \sum_{i=1}^{\infty} h_i.$$

Thus, because of (7), estimates of $m_1(u)$ and $m_2(v)$ are given by

$$\hat{m}_1(u) = \hat{\mu}(u) - \hat{\mu}(u = 0)$$

and

$$\hat{m}_2(u) = \hat{v}(u) - \hat{v}(u = 0),$$

where

$$\hat{\mu}(u) = \frac{\sum_{i=0}^{n-1} Y_i K\left(\frac{u - U_i}{r(n)}\right)}{\sum_{i=0}^{n-1} K\left(\frac{u - U_i}{r(n)}\right)}, \tag{9}$$

and

$$\hat{v}(u) = \frac{\sum_{i=0}^{n-1} Y_i K\left(\frac{v - V_i}{r(n)}\right)}{\sum_{i=0}^{n-1} K\left(\frac{v - V_i}{r(n)}\right)}. \tag{10}$$

Thus (9) and (10) define effective estimates of the nonlinear characteristics in each branch of the system. They are extremely easy in calculations and are completely independent of procedures which can be used for identification of the linear subsystems. In the paper we show that $\hat{m}_1(u)$ and $\hat{m}_2(u)$ converge to $m_1(u)$ and $m_2(v)$, respectively as rapidly as $O(n^{-1/3})$ in probability. Additionally, the next chapter will be devoted to the dynamic subsystems identification.

3. Linear Subsystems Identification

For identification of the linear subsystems, we shall employ the standard correlation approach. It is assumed that condition (8) is in force. From (4) we see that

$$\text{cov}(Y_n, U_n) = \alpha, \tag{11}$$

$$\text{cov}(Y_{n+i}, U_n) = \alpha g_i, i = 1, 2, \ldots, \tag{12}$$

$$\text{cov}(Y_n, V_n) = \beta, \tag{13}$$

$$\text{cov}(Y_{n+i}, V_n) = \alpha h_i, i = 1, 2, \ldots, \tag{14}$$

where $\alpha =$cov$(m_1(U), U)$ and $\beta =$cov$(m_2(V), V)$. Let us observe that (1) and (2) assure α and β to exist. Due to (11)–(14), if only $\alpha, \beta > 0$, we can define $\hat{g}_i = \hat{\rho}_i/\hat{\rho}_0$ and $\hat{h}_i = \hat{\delta}_i/\hat{\delta}_0$ as estimates of g_i and h_i, respectively. Here

$$\hat{\rho}_i = \frac{1}{n}\sum_{j=0}^{n-1-i}\left(Y_{j+i} - \bar{Y}\right)\left(U_j - \bar{U}\right),$$

$$\bar{Y} = \frac{1}{n}\sum_{j=0}^{n-1-i} Y_{j+i}, \qquad \bar{U} = \frac{1}{n}\sum_{j=0}^{n-1} U_j.$$

$\hat{\delta}_i$ is defined in the same way replacing U_i with V_i. Recalling that $\{Y_n\}$ is strictly stationary and ergodic process, one can easily show that $\hat{\rho}_i \stackrel{n}{\to} \alpha g_i$ and $\hat{\delta}_i \stackrel{n}{\to} \beta h_i$, $i = 0, 1, 2, ...,$ in probability. Consequently we have $\hat{g}_i \stackrel{n}{\to} g_i$ and $\hat{h}_i \stackrel{n}{\to} h_i$, in probability. In this way we have proved the following theorem.

Theorem 1 Let matrices A_1, A_2 be asymptotically stable. Let (1) and (2) hold. If $\alpha, \beta \neq 0$, then

$$\hat{g}_i \stackrel{n}{\to} g_i \text{ in probability}$$

and

$$\hat{h}_i \stackrel{n}{\to} h_i \text{ in probability.}$$

4. Convergence of the Identification Algorithm

In this section we prove that estimates $\hat{m}(u, v)$, $\hat{m}_1(u)$ and $\hat{m}_2(v)$ converge to $m_1(u) + m_2(v)$, $m_1(u)$ and $m_2(v)$, respectively with certain rate.

Theorem 2 Let matrices A_1, A_2 be asymptotically stable and let (1), (2) and (7) hold. Suppose that the bounded nonnegative Borel kernel has a compact support. Let $m_1(u)$ and $m_2(v)$ have finite derivatives at u and v, respectively, and let $f(u, v) > 0$ for all $(u, v) \in R^2$. If

$$r(n) \backsim n^{-1/4} \tag{15}$$

then

$$|\hat{m}(u, v) - [m_1(u) + m_2(v)]| = O(n^{-1/4})$$

in probability.

The next theorem is concerned with the estimates $\hat{m}_1(u)$ and $\hat{m}_2(v)$ and may be proved in the same way as Theorem 2.

Theorem 3 Let all the assumptions of Theorem 2 he satisfied. Suppose that (8) holds. If

$$r(n)n \backsim 1/3,$$

then

$$|\hat{m}_1(u) - m_1(u)| = O(n^{-1/3}) \tag{16}$$

and

$$|\hat{m}_2(u) - m_2(v)| = O(n^{-1/3})$$

in probability.

Remark 1 It is worthy to note that no smoothing conditions have been placed on the density of input variable. The consistency simply holds in every differentiability point of $m_1(u)$ and $m_2(v)$.

Proof of Theorem 2. For convenience, by K_{i1}, K_{i2} we denote

$$K\left(\frac{u-U_i}{r(n)}\right) \text{ and } K\left(\frac{v-V_i}{r(n)}\right),$$

respectively.

It is clear that it suffices to verify that $\left|\hat{\psi}(u,v)-\psi(u,v)\right| = O(n^{-1/4})$, in probability. Denoting

$$\hat{e}(u,v)=\frac{1}{nr^2(n)}\sum_{i=0}^{n-1}Y_iK_{i1}K_{i2}$$

and

$$\hat{f}(u,v)=\frac{1}{nr^2(n)}\sum_{i=0}^{n-1}K_{i1}K_{i2},$$

we have

$$\hat{\psi}(u,v)=\frac{\hat{e}(u,v)}{\hat{f}(u,v)}.$$

Clearly

$$\begin{aligned}&P\left\{\left|\hat{\psi}(u,v)-\psi(u,v)\right|>\varepsilon\right\}\\ &\le P\left\{\left|\hat{a}(u,v)-\psi(u,v)\right|>t\right\}\\ &+P\left\{\hat{b}(u,v)-E\hat{b}(u,v)>t\right\}\end{aligned} \quad (17)$$

where

$$\hat{a}(u,v)=\frac{\hat{e}(u,v)}{E\hat{f}(u,v)},$$

$$\hat{b}(u,v)=\frac{\hat{f}(u,v)}{E\hat{f}(u,v)}$$

and

$$t=\frac{\varepsilon}{\varepsilon+|\psi(u,v)|+1}.$$

By Chevyshev's inequality, the first probability on the right side of (17) is bounded from above by $\mathrm{var}(\hat{a}(u,v))+\left[E\hat{a}(u,v)-\psi(u,v)\right]^2/t^2$. Using arguments as in [5], after tedious algebra, one can verify that

$$\mathrm{var}(\hat{a}(u,v))=O\left(\frac{1}{nr^2(n)}\right). \quad (18)$$

Moreover, by (4) we get

$$\begin{aligned}&E\left|\hat{a}(u,v)-\psi(u,v)\right|\\ &=\left|\frac{E\left\{m_1(U_0)+m_2(V_0))K_{01}K_{02}\right\}}{EK_{01}K_{02}}-[m_1(u)+m_2(v)]\right|\\ &\le\left|\frac{\int|m_1(z)-m_1(u)|K\left(\frac{z-u}{r(n)}\right)K\left(\frac{t-v}{r(n)}\right)dzdt}{\int K\left(\frac{z-u}{r(n)}\right)K\left(\frac{t-v}{r(n)}\right)dzdt}\right|\\ &+\left|\frac{\int|m_2(z)-m_2(v)|K\left(\frac{z-u}{r(n)}\right)K\left(\frac{t-v}{r(n)}\right)dzdt}{\int K\left(\frac{z-u}{r(n)}\right)K\left(\frac{t-v}{r(n)}\right)dzdt}\right|.\end{aligned}$$

The above quantity is certainly not greater than

$$\begin{aligned}&c_1\sup_{z\in u+r(n)A}|z-u|+c_2\sup_{z\in v+r(n)A}|z-v|\\ &\le(c_1+c_2)|A|r(n),\end{aligned}$$

where $c_1 = |m_1\prime(u)|$, $c_2 = |m_2\prime(v)|$ and $|A|$ is the Lebesgue's measure of the support A of the kernel function. Since $\mathrm{var}(\hat{b}(u,v)) = (1/nr^2(n))$, (18) and (15) we have completed the proof. □

Remark 2 We note for us that the regression function in (5) has the additive form. It is natural to conjecture that for such a case $m(u,v)$ should achieve the same rate as in (16), i.e. $O(n^{-1/3})$ instead of $O(n^{-1/4})$. The solution of this problem, however, is not obvious even if pairs $(U_n,V_n;Y_n)$ are mutually independent i.e. if the dynamic subsystems degenerate to a memory linear amplifiers, Stone [13].

5. Concluding Remarks

In this paper we have investigated the identification of two-channel nonlinear systems. It is evident that the methods may be generalized to m-channel systems. Then, the rate of convergence in Theorem 2 would be of order $O(n^{-1/(2+m)})$ while, due to (8), the rate in (16) would be unaltered.

Aknowledgement

This work was supported by the NSERC Grant A8131.

References

[1] S.A. Billings and S.Y. Fakhouri, "Non-linear system identification using the Hammerstein model", *Int. J. Systems Sci.*, vol. 10, pp. 567-578, 1979.

[2] R.G. Corlis and R. Luus, "Use of residuals in the identification and control of two-input, single-output systems," *Ind. Eng. Chem. Fundamentals*, vol. 8, pp. 246-253, 1969.

[3] L. Devroye, "On the almost everywhere convergence of nonparametric regression function estimates", *Ann. Statist.*, vol. 9, pp. 1310-1319, 1981.

[4] W. Greblicki, A. Krzyżak and M. Pawlak, "Distribution-free pointwise consistency of kernel regression estimate", *Ann. Statist.*, vol. 12, pp. 1570-1775-1984.

[5] W. Greblicki and M. Pawlak, "Identification of discrete Hammerstein systems using kernel regression estimates", *IEEE Trans. Automat. Control*, vol. AC-11, pp. 74-77, 1986.

[6] W. Greblicki and 11. Pawlak, "Hammerstein system identification by nonparametric regression estimation", *Int. J. Control*, 1986, to appear.

[7] E.J. Hannan, "*Multiple Time Series*", New York, Wiley, 1970.

[8] P.C. Kung and D.H. Shih, "Analysis and identification of Hammerstein model non-linear delay systems using block-pulse function expansions", *Int. J. Control*, vol. 43, pp. 139-147, 1986.

[9] M. Maqusi, "Performance of baseband digital date transmission in nonlinear channels with memory", *IEEE Trans. Commun.*, vol. COM-33, pp. 715-719, 1985.

[10] B.A. Nadaraya, "On estimating regression", *Theor. Probability Appl.*, vol. 9, 141-142, 1964.

[11] M.S. Narendra and P.G. Gallman, "An iterative method for the identification of nonlinear systems using the Hammerstein model", *IEEE Trans. Automat. Control*, vol. AC-11, pp.546-550, 1966.

[12] A.A. Sawchuk and T.C. Strand, "Fourier optics in nonlinear image processing", In *Applications of Optical Fourier Transforms*, H. Stark, Ed., pp. 371-429, New York, Academic Press, 1982.

[13] C.J. Stone, "Additive regression and other nonparametric models", *Ann. Statist*, vol. 13, pp. 689-705, 1985.

[14] G.S. Watson, "Smooth regression analysis", *Sankhyā*, S.A., vol. 26, pp. 335-372, 1964.