IRIS AperTO

UNIVERSITÀ DEGLI STUDI DI TORINO

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Optimal Input Design for Set-membership Identification of Hammerstein Models

This is the author's manuscript

Original Citation:

Availability:

This version is available http://hdl.handle.net/2318/120237 since

Published version:

DOI:10.1080/0020717031000067439

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

20 April 2024

INT. J. CONTROL, 2003, VOL. 76, NO. 3, 217–225

Taylor & Francis
Taylor & Francis Group

Optimal input design for set-membership identification of Hammerstein models

GUSTAVO BELFORTE†* and PAOLO GAY‡

In this paper the problem of optimal input design for the identification of Hammerstein models is considered under the assumption that the linear dynamic part of the model is a FIR and that lower and upper bounds are available for the additive measurement errors. The parameters of the Hammerstein model can then be estimated via the identification of a *linearized augmented Hammerstein model*. External approximations of the feasible intervals for the parameters of the original Hammerstein models are then derived (which may correspond to the actual feasible intervals).

This paper deals with the design of input sequences minimizing parameter uncertainty for the *linearized augmented Hammerstein model*. Some new results are also reported about optimal input design for polynomial non-linear blocks, that may be part of Hammerstein models.

1. Introduction

When real systems have non-linear behaviours that cannot just be neglected or approximated by some suitable linearization technique, then non-linearity must be incorporated somehow into their models.

Hammerstein models are a possible answer to such a need. They consist of a non-linear static block followed by a linear dynamic block, as described in figure 1 which corresponds to a discrete-time version of such a model. The inner signal $x(k)$ is usually not available for measurement and only the system input $u(k)$ can be chosen by the designer. The system output $y(k)$ can be measured with some additive measurement noise $e(k)$. Despite their relative simplicity Hammerstein models turn out to be quite adequate to describe some aspects of the behaviour of chemical processes, distillation columns, electric heat exchangers etc. (see, e.g. Korenberg 1973, Kortmann and Unbehauen 1988, Eskinat *et al.* 1991).

The identification of Hammerstein models has been studied since the late sixties (Narendra and Gallman 1966) and a number of identification algorithms have already been proposed (see, e.g. Haber and Unbehauen 1990, Al-Duwaish and Nazmul Karim 1997 and Zi-Qiang 1994). In all of them, to the best of our knowledge, the measurement error $e(k)$ has received a probabilistic description.

The possibility of dealing with unknown but bounded errors has recently been investigated in Belforte and Gay (1999 a) where identification is accomplished via the estimation of the parameters of a *linearized augmented Hammerstein model*. Optimal estimation of these parameters is crucial since they are used to compute external approximations of the feasible intervals for the parameters of the original Hammerstein model (which may correspond to the actual feasible intervals).

In this paper the problem of optimal input design for *linearized augmented Hammerstein model* identification is addressed in a bounded-error context, with the linear dynamic part of the model assumed to be a FIR model. Input sequences that ensure the lowest worst case uncertainty of all the parameter estimates are derived and linked to the optimal identification problem for the non-linear static block standing alone. Such a link provides convenient tools for the computation of optimal input sequences.

Finally some new analytical results are reported about the optimal input design for non-linear blocks in the usual case where they are modelled by polynomial functions. Explicit optimal input sequences are derived.

The results presented in this paper are reminescent of similar results that can be found (e.g. in Kiefer and Wolfowitz 1959, Karlin and Studden 1966 a, Puklsheim and Studden 1993) but it is important to note that these results were not obtained in a bounded-error environment, but in a probabilistic setting. Moreover the cost functions to be minimized differ.

At the end of the paper, a numerical example shows the performance and usefulness of the proposed methodology.

2. Problem formulation

Consider a SISO discrete-time Hammerstein model consisting of a static non-linear block followed by a dynamic linear time-invariant subsystem as shown in figure 1.

The static non-linear part $\phi(\cdot)$ is assumed to be a generalized polynomial so that the (non-measurable) inner signal $x(k)$ can be expressed as

Received 1 December 1999. Accepted 1 October 2002

* Author for correspondence. e-mail: belforte@polito.it
† Dipartimento di Automatica e Informatica, Politecnico di Torino, Italy.
‡ Dipartimento di Economia e Ingegneria Agraria, Forestale e Ambientale, Università degli Studi di Torino, Italy. e-mail: paolo.gay@unito.it

International Journal of Control ISSN 0020–7179 print/ISSN 1366–5820 online © 2003 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/0020717031000067439

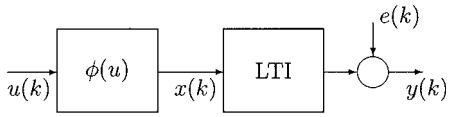

Figure 1. Discrete-time SISO Hammerstein model.

$$x(k) = \phi(u(k)) = \sum_{i=1}^{p} c_i g_i(u(k)) \tag{1}$$

where $g_1(\cdot), g_2(\cdot), \ldots, g_p(\cdot)$ are given linearly independent functions defined on the interval $[u_a, u_b]$ with $g_i(0) = 0$, $i = 1, \ldots, p$, so that the system steady-state output is zero for a zero input u.

The dynamic linear part is assumed to be described by a FIR model

$$y(k) = B(q^{-1})x(k) + e(k) \tag{2}$$

with

$$B(q^{-1}) = b_0 + b_1 q^{-1} + b_2 q^{-2} + \cdots + b_{n_b} q^{-n_b}$$

where q^{-1} is the unit time delay operator, as usual.

The system input $u(k)$ is bounded and can vary only on $[u_a, u_b]$, while the error $e(k)$ is assumed to be unknown but bounded, so that

$$|e(k)| \le E \qquad \forall k \tag{3}$$

where E is a known constant.

Combining (1) and (2), the following input/output relation can be derived

$$\begin{aligned} y(k) = {} & b_0[c_1 g_1(u(k)) + c_2 g_2(u(k)) + \cdots + c_p g_p(u(k))] \\ & + b_1[c_1 g_1(u(k-1)) \\ & + c_2 g_2(u(k-1)) + \cdots + c_p g_p(u(k-1))] \\ & \vdots \\ & + b_{n_b}[c_1 g_1(u(k-n_b)) \\ & + c_2 g_2(u(k-n_b)) + \cdots + c_p g_p(u(k-n_b))] \\ & + e(k) \end{aligned} \tag{4}$$

The parameters b and c are not uniquely related to the total gain of the system and cannot be identified uniquely. To avoid this ambiguity it will be assumed, without loss of generality, that $b_0 = 1$.

The b and c parameters can be rearranged, for notational convenience, into a single vector ξ as

$$\begin{aligned} \xi &= \begin{bmatrix} \xi_1 & \xi_2 & \cdots & \xi_{p+n_b} \end{bmatrix}^{\mathrm{T}} \\ &= \begin{bmatrix} c_1 & c_2 & \cdots & c_p & b_1 & \cdots & b_{n_b} \end{bmatrix}^{\mathrm{T}} \end{aligned}$$

Equation (4) can then be arranged in vector form as

$$y = F(\xi) + e$$

where y is the vector of measurements and e that of errors.

Under the assumption (3), identifying the parameter vector ξ corresponds to finding the set $\mathcal{D}_\xi$ of all ξ consistent with the model (4), the measurements y and the bound E on the absolute value of the errors.

When m measurements are available the corresponding parameter feasible set $\mathcal{D}_\xi$ can be expressed as

$$\mathcal{D}_\xi = \left\{ \xi \in \mathbb{R}^{p+n_b} \colon\ y = F(\xi) + e,\ e \in \Omega_e \right\} \tag{5}$$

where Ω_e is the feasible error set defined as

$$\Omega_e = \{ e \in \mathbb{R}^m \colon\ |e(k)| \le E,\ k = 1, \ldots, m \}$$

Let ξ_i^m and ξ_i^M be the extreme values of the ith component of ξ in $\mathcal{D}_\xi$

$$\xi_i^m = \min_{\xi \in \mathcal{D}_\xi} \xi_i \quad \text{and} \quad \xi_i^M = \max_{\xi \in \mathcal{D}_\xi} \xi_i$$

The parameter uncertainty interval (PUI) of the ith parameter ξ_i is then defined as

$$PUI_{\xi_i} = [\xi_i^m, \xi_i^M]$$

and its width

$$W(PUI_{\xi_i}) = \xi_i^M - \xi_i^m$$

can be introduced as a measure of the reliability of the estimate of ξ_i.

Throughout this paper the size of the set $\mathcal{D}_\xi$ is measured in terms of the corresponding $W(PUI_{\xi_i})$, $i = 1, \ldots, (p + n_b)$.

Define the augmented parameter vector $\theta \in \mathbb{R}^n$, with $n = p(n_b + 1)$, as

$$\theta = [\theta_1 \quad \theta_2 \quad \cdots \quad \theta_n]^{\mathrm{T}}$$

where

$$\left.\begin{array}{cccc} \theta_1 = c_1 & \theta_2 = c_2 & \cdots & \theta_p = c_p \\ \theta_{p+1} = b_1 c_1 & \theta_{p+2} = b_1 c_2 & \cdots & \theta_{2p} = b_1 c_p \\ & \vdots & & \\ \theta_{pn_b+1} = b_{n_b} c_1 & \theta_{pn_b+2} = b_{n_b} c_2 & \cdots & \theta_{p(n_b+1)} = b_{n_b} c_p \end{array}\right\} \tag{6}$$

This vector can be partitioned as

$$\theta = [\theta_{NL}^{\mathrm{T}} \quad \theta_{BL_1}^{\mathrm{T}} \quad \theta_{BL_2}^{\mathrm{T}} \quad \cdots \quad \theta_{BL_{n_b}}^{\mathrm{T}}]^{\mathrm{T}}$$

where $\theta_{NL} = [\theta_1 \quad \cdots \quad \theta_p]^{\mathrm{T}}$ and $\theta_{BL_i} = [\theta_{ip+1} \quad \cdots \quad \theta_{(i+1)p}]^{\mathrm{T}}$, $i = 1, \ldots, n_b$.

Equation (4) can then be rewritten as

$$\begin{aligned} y(k) = {} & \sum_{i=1}^{p} (\theta_i g_i(u(k)) + \theta_{p+i}\, g_i(u(k-1)) \\ & + \theta_{2p+i} g_i(u(k-2)) \\ & + \cdots + \theta_{pn_b+i}\, g_i(u(k-n_b))) + e(k) \end{aligned} \tag{7}$$

In the following, this structure will be referred to as the *linearized augmented Hammerstein model*.

Define $g^{\mathrm{T}}(u(\cdot))$ as

$$g^{\mathrm{T}}(u(\cdot)) = \begin{bmatrix} g_1(u(\cdot)) & g_2(u(\cdot)) & \cdots & g_p(u(\cdot)) \end{bmatrix}$$

so rearrange (7) as

$$y(k) = \begin{bmatrix} g^{\mathrm{T}}(u(k)) & g^{\mathrm{T}}(u(k-1)) & \cdots & g^{\mathrm{T}}(u(k-n_b)) \end{bmatrix}\theta + e(k)$$

When m consecutive output measurements $\{y(k)\}_{k=1}^{k=m}$ are available, together with the corresponding set of $m+n_b$ consecutive inputs $\{u(k)\}_{k=-n_b+1}^{k=m}$, (7) can be expressed more compactly in vector form as

$$y = A(u)\theta + e \tag{8}$$

where the regression matrix is

$$A(u) = \begin{bmatrix} g^{\mathrm{T}}(u(1)) & g^{\mathrm{T}}(u(0)) & \cdots & g^{\mathrm{T}}(u(1-n_b)) \\ g^{\mathrm{T}}(u(2)) & g^{\mathrm{T}}(u(1)) & \cdots & g^{\mathrm{T}}(u(2-n_b)) \\ \vdots & \vdots & \vdots & \\ g^{\mathrm{T}}(u(m)) & g^{\mathrm{T}}(u(m-1)) & \cdots & g^{\mathrm{T}}(u(m-n_b)) \end{bmatrix} \tag{9}$$

where $y = \begin{bmatrix} y(1) & y(2) & \cdots & y(m) \end{bmatrix}^{\mathrm{T}}$ and where $e \in \mathbb{R}^m$ is the corresponding error vector. The parameter feasible set for the augmented parameter vector is defined by

$$\mathcal{D}_\theta = \{\theta \in \mathbb{R}^n \quad y = A(u)\theta + e, \quad e \in \Omega_e\}$$

Provided that $\operatorname{rank} A(u) = \dim\theta$, $\mathcal{D}_\theta$ is a convex polytope described by a subset of the following $2m$ linear inequalities

$$y(k) - E \le a_k^{\mathrm{T}}\theta \le y(k) + E \qquad k = 1, \ldots, m \tag{10}$$

with a_k^{T} the kth row of $A(u)$.

The augmented parameter uncertainty intervals $PUI_{\theta_i} = [\theta_i^m, \theta_i^M]$ and their width $W(PUI_{\theta_i}) = \theta_i^M - \theta_i^m$ are easily computed via the solution of the linear programs

$$\theta_i^m = \min_{\theta \in \mathcal{D}_\theta} \theta_i \quad \text{and} \quad \theta_i^M = \max_{\theta \in \mathcal{D}_\theta} \theta_i$$

Equation (6) implies that the following relations should hold true simultaneously

$$\left.\begin{array}{llll} c_1 = \theta_1 & c_2 = \theta_2 & \cdots & c_p = \theta_p \\ b_1 = \dfrac{\theta_{p+1}}{\theta_1} & b_1 = \dfrac{\theta_{p+2}}{\theta_2} & \cdots & b_1 = \dfrac{\theta_{2p}}{\theta_p} \\ b_2 = \dfrac{\theta_{2p+1}}{\theta_1} & b_2 = \dfrac{\theta_{2p+2}}{\theta_2} & \cdots & b_2 = \dfrac{\theta_{3p}}{\theta_p} \\ & \vdots & & \\ b_{n_b} = \dfrac{\theta_{pn_b+1}}{\theta_1} & b_{n_b} = \dfrac{\theta_{pn_b+2}}{\theta_2} & \cdots & b_{n_b} = \dfrac{\theta_{p(n_b+1)}}{\theta_p} \end{array}\right\} \tag{11}$$

Outer approximations for the PUIs for the parameters of the original Hammerstein model $\tilde{PUI}_{c_i} \supseteq PUI_{c_i}$ and $\tilde{PUI}_{b_i} \supseteq PUI_{b_i}$ can now be derived based on (11).

For the c parameters of the non-linear block the procedure is trivial, as

$$\tilde{PUI}_{c_i} = PUI_{\theta_i} \supseteq PUI_{c_i} \qquad i = 1, \ldots, p \tag{12}$$

Computations are more complex for the b parameters of the FIR model since all equations in (11) should hold true simultaneously. It follows that outer bounds for the b parameters can be obtained from

$$\tilde{PUI}_{b_i} = [\max\{\underline{b}_{i_1}, \quad \underline{b}_{i_2}, \quad \ldots \quad \underline{b}_{i_p}\} \quad \min\{\bar{b}_{i_1}, \quad \bar{b}_{i_2}, \quad \ldots \quad \bar{b}_{i_p}\}] \tag{13}$$

where

$$\underline{b}_{i_j} = \begin{cases} -\infty & \text{if } 0 \in [\theta_j^m, \theta_j^M] \\ \min\left\{\dfrac{\theta_{ip+j}^m}{\theta_j^m}, \dfrac{\theta_{ip+j}^m}{\theta_j^M}, \dfrac{\theta_{ip+j}^M}{\theta_j^m}, \dfrac{\theta_{ip+j}^M}{\theta_j^M}\right\} & \text{otherwise} \end{cases}$$

and

$$\bar{b}_{i_j} = \begin{cases} +\infty & \text{if } 0 \in [\theta_j^m, \theta_j^M] \\ \max\left\{\dfrac{\theta_{ip+j}^m}{\theta_j^m}, \dfrac{\theta_{ip+j}^m}{\theta_j^M}, \dfrac{\theta_{ip+j}^M}{\theta_j^m}, \dfrac{\theta_{ip+j}^M}{\theta_j^M}\right\} & \text{otherwise} \end{cases}$$

Although $\theta_i = c_i$, $i = 1, \ldots, p$, in general PUI_{θ_i} is actually an outer approximation of PUI_{c_i}. The reason for this lies in the fact that, although all the points in $\mathcal{D}_\theta$ represent a feasible augmented Hammerstein model, not all of them correspond to a vector ξ of a Hammerstein model. As already mentioned this occurs only if all equations in (11) are simultaneously satisfied. Unfortunately, the fact that $\theta_i = c_i$, $i = 1, \ldots, p$, does not imply that the components of the vectors in $\mathcal{D}_\theta$, whose ith element is equal to θ_i^m (θ_i^M), will simultaneously satisfy all equations in (11).

The shape, size and location of $\mathcal{D}_\theta$ in parameter space depend on the error bound E (assumed to be known). More importantly they depend on the actual measurements used for deriving $\mathcal{D}_\theta$ (number and information content) and on the error realization. Since the uncertainty in the parameter estimates is related to the size of $\mathcal{D}_\theta$, it also depends on the number and on the informational content of the measurements as well as on the error realization. To get rid of the dependence on the error realization it is common practice to look for the worst parameter feasible set (the largest), in terms of its size, with respect to any possible error realization. In the set-membership literature the size of $\mathcal{D}_\theta$ is often measured in terms of its diameter† in some suitable norm, with l_∞, l_2 and l_1 being the most commonly used. However other

† The diameter $d_{\mathcal{D}}$ of a given set $\mathcal{D}$ is $d_{\mathcal{D}} = \sup_{\theta \in \mathcal{D}, \vartheta \in \mathcal{D}} \|\theta - \vartheta\|$

measures of uncertainty such as the volume of $\mathcal{D}_\theta$ are used as well. When attention is focused on parameter estimation, the size of $\mathcal{D}_\theta$ is frequently evaluated in terms of the widths of the parameter uncertainty intervals $W(PUI_{\theta_i})$, which can be regarded as parameter confidence intervals with confidence coefficient equal to 1. The optimality criterion used in this paper will be the minimization of all the $W(PUI_{\theta_i})$, $i = 1, \ldots, n$, which implies the minimization of the l_∞ diameter of $\mathcal{D}_\theta$. When using this criterion authors often claim that an l_∞ diameter minimization is performed. This is indeed true, however the condition is even stronger as the width of each of the worst-case parameter uncertainty intervals is minimized.

In the bounded-error context, we are interested in finding the (shortest) input string that allows a set of measurements to be obtained that ensures the *smallest worst-case parameter set*, according to some given optimality criterion, here the minimization of all the $W(PUI_{\theta_i})$, $i = 1, \ldots, n$.

With such a criterion the smallest worst-case parameter set is obtained if and only if the identification is performed using a small set of specific measurements denoted as the *optimal measurement set*, possibly together with other possible measurements. The number of measurements in this set lies between the number of parameters (n in the case of $\mathcal{D}_\theta$) and its square (n^2) for any given model and set of possible feasible measurements (Belforte *et al.* 1987). The *smallest worst-case parameter set* can always be derived numerically as described in (Belforte *et al.* 1987). The addition of extra measurements to the *optimal measurement set* may affect the size of the actual set $\mathcal{D}_\theta$ but cannot reduce the smallest worst-case parameter set. This is why we shall not consider such extra measurements. When, for given system and set of feasible measurements, the *optimal measurement set* consists of just as many measurements as there are parameters (n in the case of $\mathcal{D}_\theta$), such a system will be said to correspond to a *minimal worst case*.

Sufficient conditions for a system to be minimal worst case can be found in literature (Belforte *et al.* 1987, Micchelli 1988). It will be the case, for instance, whenever any regression matrix $A \in \mathbb{R}^{m \times n}$ obtained with any possible set of m feasible measurements is *strictly totally positive* (STP) (Karlin and Studden 1966 b), that is whenever all of its minors have positive determinants.

Note however that the ordering of the measurements does not affect $\mathcal{D}_\theta$ but affects A and the sign of the determinants of its minors. For this reason whenever A can be converted, by a suitable row swapping, into an STP matrix we shall take the liberty to refer to it as STP. Note also that for minimal worst-case systems the *optimal measurement set* ensures not only the minimization of all the $W(PUI_{\theta_i})$ $i = 1, \ldots, n$, and thus of the l_∞ diameter, but also of the diameter in any l_h , $1 \leq h \leq \infty$, norm. The worst-case parameter set and related quantities are hereafter denoted by the symbol $\hat{\cdot}$, as in $\hat{\mathcal{D}}_\theta$, $\hat{PUI}_{\theta_i}$, etc. Such quantities are usually computed from (10) forcing $y = 0$.

3. Optimal worst-case input design

In the case considered in this paper, the regression matrix $A(u)$ is only a function of the system inputs so it is possible to look for the shortest input strings that allow the *smallest worst-case parameter set* to be obtained. In the sequel a particular set of input strings that possesses this property and can be derived *a priori* is exhibited and referred to as *optimal input* set.

For the sake of simplicity of the notation and proof, the optimal inputs are derived for the case in which the non-linear block of the Hammerstein FIR model is *minimal worst case* when it is considered in isolation.

Define $\mathcal{L}$ as a set of p distinct levels $l_1, l_2, \ldots, l_p$ for the input

$$\mathcal{L} = \{l_1, l_2, \ldots, l_p : l_i \in [u_a, u_b]\}$$

and $\mathcal{U}(\mathcal{L})$ as the set of $p!$ input vectors obtained by combining the p levels in $\mathcal{L}$

$$\mathcal{U}(\mathcal{L}) = \{u = [u_1, u_2, \ldots, u_p]^{\mathrm{T}} : \\ u_i \in \mathcal{L}, u_i \neq u_j, i \neq j, \forall i, j\}$$

Finally, let $\mathcal{S}(\mathcal{U}(\mathcal{L}))$ be the set of input sequences such that

$$\mathcal{S}(\mathcal{U}(\mathcal{L})) = \left\{ [\underbrace{0 \ldots 0}_{n_b} u_1 \underbrace{0 \ldots 0}_{n_b} u_2 \underbrace{0 \ldots 0}_{n_b} \ldots u_p \underbrace{0 \ldots 0}_{n_b}] : \\ [u_1, \ldots, u_p]^{\mathrm{T}} \in \mathcal{U}(\mathcal{L}) \right\} \quad (14)$$

Sequences of the form (14) are a subset of all possible input sequences. However some of them are *optimal inputs* for the Hammerstein FIR model. The set $\mathcal{S}(\mathcal{U}(\mathcal{L}_{\mathrm{opt}}))$ of all such optimal inputs is obtained from (14) with $\mathcal{L} = \mathcal{L}_{\mathrm{opt}}$ where $\mathcal{L}_{\mathrm{opt}}$ is the set of p distinct levels ensuring the optimal worst case estimation of the non-linear system when considered in isolation. This result is proved in the following theorem.

Theorem 1: *Consider any linear augmented Hammerstein (FIR) model in which the non-linear submodel is minimal worst-case when considered in isolation. Let $\mathcal{L}_{\mathrm{opt}}$ be the set of p distinct levels that ensure the optimal worst-case estimation of the non-linear system when considered in isolation. Then any input sequence $s_{\mathrm{opt}} \in \mathcal{S}(\mathcal{U}(\mathcal{L}_{\mathrm{opt}}))$ minimizes the worst-case widths of the parameter uncertainty intervals for all the parameters θ_i.*

Proof: When using an input vector $s \in \mathcal{S}(\mathcal{U}(\mathcal{L}))$, the regression matrix of relation (9) becomes

$$A(s) = \begin{bmatrix} g^{\mathrm{T}}(u_1) & & & 0 \\ & g^{\mathrm{T}}(u_1) & & \\ & & \ddots & \\ 0 & & & g^{\mathrm{T}}(u_1) \\ \vdots & \vdots & \vdots & \vdots \\ g^{\mathrm{T}}(u_p) & & & 0 \\ & g^{\mathrm{T}}(u_p) & & \\ & & \ddots & \\ 0 & & & g^{\mathrm{T}}(u_p) \end{bmatrix}$$

By suitable row swapping, the relation $y = A(s)\theta + e$ linking measurements, parameters and errors can be rewritten as

$$\bar{y} = \begin{bmatrix} G(\mathcal{L}) & & & 0 \\ & G(\mathcal{L}) & & \\ & & \ddots & \\ 0 & & & G(\mathcal{L}) \end{bmatrix} \begin{bmatrix} \theta_{NL} \\ \theta_{BL_1} \\ \vdots \\ \theta_{BL_{n_b}} \end{bmatrix} + \bar{e} \qquad (15)$$

where

$$G(\mathcal{L}) = \begin{bmatrix} g_1(l_1) & g_2(l_1) & \dots & g_p(l_1) \\ g_1(l_2) & g_2(l_2) & \dots & g_p(l_2) \\ \vdots & \vdots & \vdots & \\ g_1(l_p) & g_2(l_p) & \dots & g_p(l_p) \end{bmatrix} \qquad (16)$$

Partitioning $\bar{y}$ into $n_b + 1$ subvectors of p components each so that $\bar{y} = [\bar{y}^{0\mathrm{T}} \ \bar{y}^{1\mathrm{T}} \ \cdots \ \bar{y}^{n_b\mathrm{T}}]^{\mathrm{T}}$, allows the parameter subvectors θ_{NL} and θ_{BL_i} to be independently identified using only p suitable measurements, namely $\bar{y}^0$ for θ_{NL} and $\bar{y}^i$ for θ_{BL_i}, $i = 1, \dots, n_b$. By hypothesis the standing alone non-linear subsystem is minimal worst-case, therefore the smallest worst-case parameter set is achieved only when all the p distinct input levels $\mathcal{L}_{\mathrm{opt}}$ appear in the input string. Since in this case the input string for independently identifying $\theta_{NL}, \theta_{BL_1}, \dots, \theta_{BL_{n_b}}$ consists of p suitable measurements only, then the worst-case uncertainty is minimal only when the p distinct input levels $\mathcal{L}_{\mathrm{opt}}$ are used so that $G(\mathcal{L}) = G(\mathcal{L}_{\mathrm{opt}})$. Any input sequence $s_{\mathrm{opt}} \in \mathcal{S}(\mathcal{U}(\mathcal{L}_{\mathrm{opt}}))$ ensures $G(\mathcal{L}) = G(\mathcal{L}_{\mathrm{opt}})$ and minimizes therefore the $W(\boldsymbol{P}\hat{\boldsymbol{U}}\boldsymbol{I}_{\theta_i})$, $i = 1, \dots, n$.

Note that the specific structure of the problem evidenced by (15) implies that $W(\boldsymbol{P}\hat{\boldsymbol{U}}\boldsymbol{I}_{\theta_i}) = W(\boldsymbol{P}\hat{\boldsymbol{U}}\boldsymbol{I}_{\theta_{jp+i}})$, $i = 1, \dots, p$, $j = 1, \dots, n_b$.

In the sequel of the proof the worst-case optimality of s_{opt} with respect to any other possible sequence is proved. To this end rewrite (15) as

$$\bar{y} = \begin{bmatrix} G(\mathcal{L}) & A_{12} = 0 \\ A_{21} = 0 & A_{22} \neq 0 \end{bmatrix} \begin{bmatrix} \theta_{NL} \\ \theta_B \end{bmatrix} + \bar{e} \qquad (17)$$

where $\theta_B = [\theta_{BL_1}, \theta_{BL_2}, \dots, \theta_{BL_{n_b}}]$. For a generic input sequence (17) would be replaced by

$$\bar{y} = \begin{bmatrix} A_{11} \neq 0 & A_{12} \neq 0 \\ A_{21} \neq 0 & A_{22} \neq 0 \end{bmatrix} \begin{bmatrix} \theta_{NL} \\ \theta_B \end{bmatrix} + \bar{e} \qquad (18)$$

The worst-case parameter uncertainty widths $W(\boldsymbol{P}\hat{\boldsymbol{U}}\boldsymbol{I}_{\theta_i})$, $i = 1, \dots, p$, would then be no smaller than the parameter uncertainty widths that could be achieved applying the same input sequence but forcing θ_B to zero, since the parameter uncertainty set obtained forcing $\theta_B = 0$ is a subset of the parameter uncertainty set $\hat{\mathcal{D}}_\theta$ with no constraints on θ_B.

We shall therefore force θ_B to zero in the remainder of the proof. The uncertainty in θ_{NL} can then be computed from the a simplified version of (18)

$$\bar{y} = \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix} \theta_{NL} + \bar{e} \qquad (19)$$

However, the worst-case parameter uncertainty widths achievable from (19) are no smaller than those that could be obtained processing

$$\bar{y} = \left[G(\mathcal{L}_{\mathrm{opt}}) \right] \theta_{NL} + \bar{e}$$

which are exactly those obtained from relation (17) where an optimal input sequence is used. This proves that the worst-case parameter uncertainty widths $W(\boldsymbol{P}\hat{\boldsymbol{U}}\boldsymbol{I}_{\theta_i})$, $i = 1, \dots, p$, obtained from (17), where an optimal input sequence is used, are no larger than those obtained from (18), where a generic input sequence is used. By a suitable row and column swapping, (15) can be rearranged as

$$\bar{\bar{y}} = \begin{bmatrix} G(\mathcal{L}) & & & 0 \\ & G(\mathcal{L}) & & \\ & & \ddots & \\ 0 & & & G(\mathcal{L}) \end{bmatrix} \begin{bmatrix} \theta_{BL_i} \\ \theta_{NL} \\ \vdots \end{bmatrix} + \bar{\bar{e}} \qquad (20)$$

where θ_{BL_i} is any of the n_b relevant parameter subvectors. It is then possible to apply to θ_{BL_i} the same reasoning as used for θ_{NL}, which completes the proof. □

Remark 1: The optimal set $\mathcal{L}_{\mathrm{opt}} = \{l_1^{\mathrm{opt}}, l_2^{\mathrm{opt}}, \dots, l_p^{\mathrm{opt}}\}$ of levels required to design the optimal input sequence can be obtained by considering the non-linear block in isolation. In practice, it can be computed using techniques based on linear programming as described in Belforte *et al.* (1987).

Remark 2: Theorem 1 can be extended to the case of not minimal worst-case systems. Optimal worst-case estimation would then be achieved by applying any input sequence $s \in \mathcal{S}(\mathcal{U}(\mathcal{L}_{\mathrm{opt}}))$. The matrices $G(\mathcal{L})$, $G(\mathcal{L}_{\mathrm{opt}})$ and A_{11} would be rectangular.

4. Polynomial basis functions

In order to describe the non-linear block, it is common practice to assume the non-linear function $\phi(\cdot)$ to be a polynomial

$$\phi(u(k)) = c_1 u^{\alpha_1}(k) + c_2 u^{\alpha_2}(k) + \cdots + c_p u^{\alpha_p}(k)$$

where $\alpha_1, \alpha_2, \ldots, \alpha_p$ is any strictly increasing sequence of non-negative integer numbers. This corresponds to choosing the functions $g_1(\cdot), \ldots, g_p(\cdot)$ of (1) as

$$g_i(u(k)) = u^{\alpha_i}(k), \qquad i = 1, \ldots, p \tag{21}$$

Whenever the input signal $u(k)$ belongs to $[u_a, u_b]$ such that $0 \notin (u_a, u_b)$, the following lemma holds.

Lemma 1: *Any non-linear subsystem described by a set of polynomial functions $g_1(\cdot), \ldots, g_p(\cdot)$ as in (21) and studied in isolation, is minimal worst case.*

Proof: Assume p input output data have been collected applying the p distinct levels $l_i \in \mathcal{L}$, $i = 1, \ldots, p$ (without loss of generality these levels are assumed to be ordered so that $l_i > l_j$, $\forall\, i > j$). The model for the non-linear block can be described by

$$x = G(\mathcal{L})c$$

where

$$G(\mathcal{L}) = \begin{bmatrix} l_1^{\alpha_1} & l_1^{\alpha_2} & \cdots & l_1^{\alpha_p} \\ l_2^{\alpha_1} & l_2^{\alpha_2} & \cdots & l_2^{\alpha_p} \\ \vdots & \vdots & & \vdots \\ l_p^{\alpha_1} & l_p^{\alpha_2} & \cdots & l_p^{\alpha_p} \end{bmatrix} \tag{22}$$

is the regression matrix, $c = [c_1, c_2, \ldots, c_p]^{\mathrm{T}}$ is the parameter vector and $x \in \mathbb{R}^p$ is the output of the non-linear block. The matrix $G(\mathcal{L})$ can be obtained by extracting p columns out of the following generalized rectangular Vandermonde matrix

$$V(\mathcal{L}) = \begin{bmatrix} 1 & l_1 & \cdots & l_1^{\alpha_p} \\ 1 & l_2 & \cdots & l_2^{\alpha_p} \\ \vdots & \vdots & & \vdots \\ 1 & l_p & \cdots & l_p^{\alpha_p} \end{bmatrix}$$

Since the Vandermonde matrix $V(\mathcal{L})$ is STP (Karlin and Studden 1966 b), all its minors are positive. Any submatrix of $V(\mathcal{L})$ including $G(\mathcal{L})$ is STP as well. □

For complete polynomials the set of optimal levels $\mathcal{L}_{\text{opt}}$ ensuring the optimal worst-case estimation of the non-linear block considered in isolation can be derived analytically (Belforte *et al.* 1997). It is usually considered desirable that the inner signal x be zero when the input is identically zero which implies that $\alpha_1 > 0$. A common practice is to take

$$g_i(u(k)) = u^i(k), \qquad i = 1, \ldots, p \tag{23}$$

For such polynomials the set of optimal input levels $\mathcal{L}_{\text{opt}}$ in $[0, u_b]$ is provided by the following theorem.

Theorem 2: *For any non-linear block described by a set of functions $g_1(\cdot), \ldots, g_p(\cdot)$ as in (23), the p elements $l_i^{\text{opt}} \in [0, u_b]$ of the optimal set $\mathcal{L}_{\text{opt}}$ are*

$$l_i^{\text{opt}} = \frac{u_b}{1 + \cos(1/2p)\pi}\left(\cos\frac{p-i}{p}\pi + \cos\frac{1}{2p}\pi\right), \qquad i = 1, \ldots, p \tag{24}$$

Proof: Assume again without loss in generality that $l_i > l_j \geq 0$, $\forall i > j$. The regression matrix $G(\mathcal{L})$ for the non-linear block considered in isolation, when p distinct levels $l_i \in \mathcal{L}$, $i = 1, \ldots, p$, have been applied to its input, is STP (see Lemma 1) and therefore minimal worst-case. From the positiveness of its minors, it turns also out that the sign of each element r_{ij} of $G^{-1}(\mathcal{L})$ satisfies

$$\operatorname{sign}(r_{ij}) = (-1)^{i+j} \qquad i, j = 1, \ldots, p \tag{25}$$

as remarked in Belforte *et al.* (1987) and proved in Belforte *et al.* (1984).

From (25), it follows that there are only two vectors $\hat{\theta}_{NL}^{iM}(\mathcal{L})$ and $\hat{\theta}_{NL}^{im}(\mathcal{L})$, $i = 1, \ldots, p$, where the extreme worst-case parameter values $\hat{\theta}_{NL_i}^{m}$, $\hat{\theta}_{NL_i}^{M}$, $i = 1, \ldots, p$, are achieved, instead of $2p$ distinct vectors because

$$\left.\begin{array}{llll} \hat{\theta}_{NL}^{im}(\mathcal{L}) = \hat{\theta}_{NL}^{1M}(\mathcal{L}) & \text{and} & \hat{\theta}_{NL}^{iM}(\mathcal{L}) = \hat{\theta}_{NL}^{1m}(\mathcal{L}) & \text{for i even} \\ \hat{\theta}_{NL}^{im}(\mathcal{L}) = \hat{\theta}_{NL}^{1m}(\mathcal{L}) & \text{and} & \hat{\theta}_{NL}^{iM}(\mathcal{L}) = \hat{\theta}_{NL}^{1M}(\mathcal{L}) & \text{for i odd} \end{array}\right\} \tag{26}$$

Therefore, the two vectors $\hat{\theta}_{NL}^{1M}(\mathcal{L})$ and $\hat{\theta}_{NL}^{1m}(\mathcal{L})$ contain all the information about the size of the worst-case feasible parameter set $\hat{\mathcal{D}}_{NL}(\mathcal{L})$ in terms of parameter uncertainty intervals.

Moreover

$$\hat{\theta}_{NL}^{1M}(\mathcal{L}) = -\hat{\theta}_{NL}^{1m}(\mathcal{L}) = \begin{bmatrix} \hat{\theta}_{NL_1}^{M} \\ \hat{\theta}_{NL_2}^{m} \\ \hat{\theta}_{NL_3}^{M} \\ \vdots \end{bmatrix} = -\begin{bmatrix} \hat{\theta}_{NL_1}^{m} \\ \hat{\theta}_{NL_2}^{M} \\ \hat{\theta}_{NL_3}^{m} \\ \vdots \end{bmatrix}$$

$$= G^{-1}(\mathcal{L}) \begin{bmatrix} +E \\ -E \\ +E \\ \vdots \end{bmatrix}$$

Consider the two polynomials

$$\left.\begin{array}{l} P(l, \mathcal{L}_{\text{opt}}) = \begin{bmatrix} l & l^2 & \cdots & l^p \end{bmatrix} \hat{\theta}_{NL}^{1M}(\mathcal{L}_{\text{opt}}) \\ -P(l, \mathcal{L}_{\text{opt}}) = \begin{bmatrix} l & l^2 & \cdots & l^p \end{bmatrix} \hat{\theta}_{NL}^{1m}(\mathcal{L}_{\text{opt}}) \end{array}\right\} \tag{27}$$

interpolating the points $\pm(-1)^i E$ in the nodes $l_i \in \mathcal{L}_{\text{opt}}$ of (24). Since the first $p-1$ optimal levels $l_1^{\text{opt}}, \ldots, l_{p-1}^{\text{opt}}$ are the zeros of the $p-1$-degree second kind Chebyshev polynomial shifted on the interval $[-u_b(1-\cos(\pi/2p))/(1+\cos(\pi/2p)), u_b]$ and since l_p^{opt} coincides with the upper bound u_b of the input signal u, then l_i^{opt}, $i = 1, \ldots, p$, are the points at which the p-degree first kind Chebyshev polynomial, shifted on the interval $[-u_b(1-\cos(\pi/2p))/(1+\cos(\pi/2p)), u_b]$ achieves its extreme values. It follows that $P(l, \mathcal{L}_{\text{opt}})$ is a Chebyshev polynomial bounded by $\pm E$ in the interval $[-u_b(1-\cos(\pi/2p))/(1+\cos(\pi/2p)), u_b] \supset [0, u_b]$, so $|P(l, \mathcal{L}_{\text{opt}})| \leq E$ for any $l \in [0, u_b]$.

From (27) it follows that

$$-E \leq \begin{bmatrix} l & l^2 & \cdots & l^p \end{bmatrix} \hat{\theta}_{NL}^{1M}(\mathcal{L}_{\text{opt}}) \leq E \qquad \forall l \in [0, u_b] \tag{28}$$

and

$$-E \leq \begin{bmatrix} l & l^2 & \cdots & l^p \end{bmatrix} \hat{\theta}_{NL}^{1m}(\mathcal{L}_{\text{opt}}) \leq E \qquad \forall l \in [0, u_b] \tag{29}$$

Equality holds in (28) and (29) only when $l \in \mathcal{L}_{\text{opt}}$. For any set of levels $\mathcal{L}$ the corresponding set $\hat{\mathcal{D}}_{NL}(\mathcal{L})$ is defined by a set of inequalities of the form

$$-E \leq \begin{bmatrix} l_k & l_k^2 & \cdots & l_k^p \end{bmatrix} \theta_{NL} \leq E \qquad l_k \in \mathcal{L}, \quad k = 1, \ldots, p \tag{30}$$

that derive from relation (10) forcing y to zero in order to deal with the worst-case error. From (28), (29) and (30) it follows that

$$\hat{\theta}_{NL}^{1M}(\mathcal{L}_{\text{opt}}) \in \hat{\mathcal{D}}_{NL}(\mathcal{L}) \quad \text{and} \quad \hat{\theta}_{NL}^{1m}(\mathcal{L}_{\text{opt}}) \in \hat{\mathcal{D}}_{NL}(\mathcal{L}) \qquad \forall \mathcal{L} \tag{31}$$

Equations (26) and (31) then imply that

$$\hat{\theta}_{NL_i}^{M}(\mathcal{L}) \geq \hat{\theta}_{NL_i}^{M}(\mathcal{L}_{\text{opt}}) \quad \text{and} \quad \hat{\theta}_{NL_i}^{m}(\mathcal{L}) \leq \hat{\theta}_{NL_i}^{m}(\mathcal{L}_{\text{opt}}) \qquad i = 1, \ldots, p$$

so

$$\begin{aligned} W(\boldsymbol{P}\hat{\boldsymbol{U}}\boldsymbol{I}_{\theta_{NL_i}}(\mathcal{L})) &= \hat{\theta}_{NL_i}^{M}(\mathcal{L}) - \hat{\theta}_{NL_i}^{m}(\mathcal{L}) \geq W(\boldsymbol{P}\hat{\boldsymbol{U}}\boldsymbol{I}_{\theta_{NL_i}}(\mathcal{L}_{\text{opt}})) \\ &= \hat{\theta}_{NL_i}^{M}(\mathcal{L}_{\text{opt}}) - \hat{\theta}_{NL_i}^{m}(\mathcal{L}_{\text{opt}}) \quad i = 1, \ldots, p \end{aligned} \tag{32}$$

$\mathcal{L}_{\text{opt}}$ is therefore optimal and minimizes the worst-case parameter uncertainty.

Since the vectors $\hat{\theta}_{NL}^{1M}(\mathcal{L})$ and $\hat{\theta}_{NL}^{1m}(\mathcal{L})$ where all the extreme values $\hat{\theta}_{NL_i}^{M}(\mathcal{L})$ and $\hat{\theta}_{NL_i}^{m}(\mathcal{L})$, $i = 1, \ldots, p$, occur are two vertices of $\hat{\mathcal{D}}_{NL}(\mathcal{L})$ where p out of the $2p$ inequalities (30) are saturated, (31) implies that equality holds in (32) if and only if $\hat{\theta}_{NL}^{1M}(\mathcal{L}) = \hat{\theta}_{NL}^{1M}(\mathcal{L}_{\text{opt}})$ and $\hat{\theta}_{NL}^{1m}(\mathcal{L}) = \hat{\theta}_{NL}^{1m}(\mathcal{L}_{\text{opt}})$, which occurs only if $\mathcal{L} = \mathcal{L}_{\text{opt}}$. □

For notational convenience the optimal levels in Theorem 2 have been assumed to belong to the interval $[0, u_b]$, but a set of levels relative to a generic interval $[u_a, u_b]$ can be easily obtained as described in Belforte *et al.* (1997, 2000).

5. Numerical example

Assume that the Hammerstein model has a nonlinear block represented by a fourth-degree polynomial with no constant term

$$x(k) = c_1 u(k) + c_2 u^2(k) + c_3 u^3(k) + c_4 u^4(k) \tag{33}$$

Assume further that the FIR block has four parameters, the first of which b_0 is forced to 1 as described in § 2 and has not to be identified. The nominal parameter values used to generate the simulated data to be used for identification are

$$b_0 = 1 \qquad b = [\,0.5 \quad -0.4 \quad 2\,] \qquad c = [\,1 \quad -3 \quad 2 \quad 0.5\,]$$

The admissible input interval $[u_a, u_b]$ has been assumed to be $[0, 1]$ and randomly generated measurement errors uniformly distributed in the interval $[-E, E]$ for given values of E have been added to the outputs. Since for input strings uniformly generated in $[0, 1]$ the average absolute output value is about 0.3, the error bounds E used in the following are expressed as percentages of this average absolute output value.

To evaluate the performances of the optimal input sequences the following simulations have been performed. The optimal input string (whose length is 20 but which generates 16 measurements only) has been applied to the model generating the data 100 times with $E = 0.1\%$ and 100 times with $E = 1\%$.

For each one of the 100 sets of 16 measurements in each error condition the PUI_{θ_i} and the corresponding $W(PUI_{\theta_i})$, $i = 1, \ldots, 16$, have been computed. The sixteen values of $W(PUI_{\theta_i})$, $i = 1, \ldots, 16$, averaged over the 100 realizations are reported in table 1 for both error levels. The $\boldsymbol{P}\tilde{\boldsymbol{U}}\boldsymbol{I}_{\xi_i}$ and the corresponding $W(\boldsymbol{P}\tilde{\boldsymbol{U}}\boldsymbol{I}_{\xi_i})$, $i = 1, \ldots, 7$, have then been derived according to (11)–(13). The seven values of $W(\boldsymbol{P}\tilde{\boldsymbol{U}}\boldsymbol{I}_{\xi_i})$, $i = 1, \ldots, 7$, averaged over the 100 realizations are reported in table 2 for both error levels.

A second simulation similar to the one already described has then been performed using the two error levels $E = 0.1\%$ and $E = 1\%$ but replacing the optimal input string by randomly generated input strings of 20 elements uniformly distributed in $[0, 1]$. The values of the PUI_{θ_i}, $W(PUI_{\theta_i})$, $i = 1, \ldots, 16$, and $\boldsymbol{P}\tilde{\boldsymbol{U}}\boldsymbol{I}_{\xi_i}$, $W(\boldsymbol{P}\tilde{\boldsymbol{U}}\boldsymbol{I}_{\xi_i})$, $i = 1, \ldots, 7$, have also been computed. There are now error realizations that do not allow finite values of the $W(\boldsymbol{P}\tilde{\boldsymbol{U}}\boldsymbol{I}_{\xi_i})$ to be derived. For the lower error bound ($E = 0.1\%$) the number of cases in which at least one of the $W(\boldsymbol{P}\tilde{\boldsymbol{U}}\boldsymbol{I}_{\xi_i})$, $i = 1, \ldots, 7$, is not finite is

Input	E is equal to 0.1% of the averaged absolute value of y															
Optimal	0.012	0.076	0.130	0.068	0.012	0.075	0.130	0.068	0.012	0.075	0.130	0.068	0.012	0.075	0.130	0.068
Random	0.562	2.629	4.605	2.581	1.203	4.681	6.845	3.425	1.719	6.961	10.207	4.993	0.915	3.598	5.309	2.636
Input	E is equal to 1% of the averaged absolute value of y															
Optimal	0.126	0.763	1.316	0.685	0.126	0.763	1.316	0.685	0.126	0.763	1.316	0.685	0.126	0.763	1.316	0.685
Random	4.667	21.316	36.278	19.611	10.537	41.526	60.914	29.820	7.593	30.448	45.855	22.966	6.152	22.105	33.145	17.344

Table 1. Averaged $W(PUI_{\theta_i}), i = 1, \ldots, 16$ for optimal and random input strings.

Input	E is equal to 0.1% of the averaged absolute value of y						
Optimal	0.0124	0.0756	0.1303	0.0679	0.0187	0.0174	0.0373
Random	0.2936	1.3785	2.3713	1.3282	2.1381	1.3771	2.3821
Input	E is equal to 1% of the averaged absolute value of y						
Optimal	0.1257	0.7628	1.3160	0.6851	0.1890	0.1761	0.3771
Random	1.2102	5.6314	9.4722	5.2179	6.8488	5.2406	8.1064

Table 2. Averaged $W(PUI_{\xi_i}), i = 1, \ldots, 7$, for optimal and random input strings.

Input	E is equal to 10% of the averaged absolute value of y															
Optimal	0.391	2.426	4.228	2.213	0.403	2.487	4.335	2.271	0.443	2.628	4.480	2.316	0.440	2.643	4.530	2.348
Random	1.157	4.939	7.628	3.847	1.162	4.994	7.696	3.865	1.157	4.968	7.657	3.846	1.149	4.926	7.584	3.812
Input	E is equal to 20% of the averaged absolute value of y															
Optimal	0.894	5.359	9.168	4.740	0.851	5.213	9.008	4.687	0.807	4.990	8.717	4.581	0.841	5.088	8.766	4.562
Random	2.432	10.325	15.783	7.895	2.376	10.232	15.799	7.945	2.372	10.122	15.610	7.854	2.404	10.190	15.600	7.810

Table 3. Averaged $W(PUI_{\theta_i}), i = 1, \ldots, 16$, for random and replicated optimal input strings.

equal to 7 while for the higher error bound ($E = 1\%$) this number rises to 48.

The average values of the $W(PUI_{\theta_i})$, $i = 1, \ldots, 16$, and of $W(\tilde{PUI}_{\xi_i})$, $i = 1, \ldots, 7$, obtained omitting these cases are reported in tables 1 and 2 for both error levels. If the error bound E were increased some $W(\tilde{PUI}_{\xi_i})$ could become infinite, even using optimal input strings. Indeed the error levels must be considered in connection with the number of available measurements which, in the present case, is quite low (only 16). When higher measurement error levels must be handled, longer measurement records are needed to allow parameter identification.

When more measurements can be collected than the minimum number ensuring optimal worst-case estimation, the best policy, according to the results described in Belforte and Gay (1999 b), is to replicate the optimal measurements, i.e. to apply a long input sequence obtained repeating several times the optimal input sequence (omitting the first n_b zeros of the sequence that are needed for initialization).

With this in mind, a second simulation study has been performed on the same model with higher error bounds and longer input strings. Again two error levels have been considered, but this time they have been set to $E = 10\%$ and $E = 20\%$. The length of the input strings was set to 84 leading to 80 output measurements (five times more than in the first simulation study).

Again two types of inputs have been compared. The optimal one in which the input sequences are obtained by repeating the optimal string five times (with the omission of the first four zeros). The other inputs consist of randomly generated strings of 84 elements uniformly distributed in $[0, 1]$. For both types of inputs and both error levels 100 measurement sets have been generated from which the $W(PUI_{\theta_i})$, $i = 1, \ldots, 16$, and $W(\tilde{PUI}_{\xi_i})$, $i = 1, \ldots, 7$, have been computed. The average values of these quantities have been reported in tables 3 and 4. Finite parameter uncertainty widths $W(\tilde{PUI}_{\xi_i})$ $i = 1, \ldots, 7$ have always been obtained with the optimal inputs while with the random input there has been one case with error level

Input			E is equal to 10% of the averaged absolute value of y				
Optimal	0.3906	2.4256	4.2279	2.2133	0.6136	0.6279	1.2658
Random	1.1488	4.9060	7.5784	3.8227	3.2333	3.1932	5.9458
Input			E equal to 20% of the averaged absolute value of y				
Optimal	0.8442	5.1109	8.7978	4.5682	2.1973	1.9161	4.1796
Random	2.0513	8.9855	14.0148	7.1002	31.3778	31.1269	44.3552

Table 4. Averaged $W(PUI_{\xi_i}), i = 1, \ldots, 7$, for random and replicated optimal input strings.

$E = 10\%$ and 72 cases with error level $E = 20\%$ in which this was not so.

The results reported in the tables confirm that the optimal input performs much better than the random one. The parameter uncertainty intervals are always much smaller when the optimal input sequence is used. Even more importantly the ξ parameters of the Hammerstein model could always be identified with the optimal input sequences while this was not the case with the random inputs.

Acknowledgements

This work was partially supported by IRITI-CNR and the grants CNR.99.01960.CT07, CNRG002E32, COFIN2000 MM07157922, ASI/I/R/137/01.

References

Al-Duwaish, H., and Nazmul Karim, M., 1997, A new method for the identification of Hammerstein model. *Automatica*, **33**, 1871–1875.

Belforte, G., Bona, B., and Frediani, S., 1984, Optimal sampling schedule with unknown but bounded measurement errors: families of linear models. *Proceedings of the 23rd IEEE Conference on Decision Control*, Las Vegas, NV, USA, pp. 1554–1559.

Belforte, G., Bona, B., and Frediani, S., 1987, Optimal sampling schedule for parameter estimation of linear models with unknown but bounded measurement errors. *IEEE Transactions on Automatic Control*, **32**, 179–182.

Belforte, G., and Gay, P., 1999 *a*, Discrete-time Hammerstein model identification with unknown but bounded errors. *Proc. 38th IEEE Conference on Decision and Control*, Phoenix, AZ, pp. 3202–3203.

Belforte, G., and Gay, P., 1999 *b*, Optimal non-minimal experiment design for static non-linear systems corrupted by unknown but bounded errors. *1999 IFAC World Conference*, Beijing, China, Vol. H, pp. 325–330.

Belforte, G., Gay, P., and Monegato, G., 1997, Optimal sampling schedule for models with l_2 and l_∞ set membership errors. *Proceedings of the 36th Conference on Decision and Control*, San Diego, CA, USA, pp. 744–749.

Belforte, G., Gay, P., and Monegato, G., 2000, Some new properties of Chebyshev polynomials. *Journal of Computational and Applied Mathematics*, **117**, 175–181.

Eskinat, E., Johnson, S. H., and Luyben, W. L., 1991, Use of Hammerstein models in identification of nonlinear systems. *AIChE Journal*, **37**, 255–268.

Haber, R., and Unbehauen, H. 1990, Structure identification of nonlinear dynamic system—a survey on input/output approaches. *Automatica*, **26**, 651–677.

Karlin, S., and Studden, W. J., 1966 *a*, Optimal experimental designs. *Annals of Mathematics and Statistics*, **37**, 783–815.

Karlin, S. J., and Studden, W. J., 1966, *Tchebycheff Systems with Application in Analysis and Statistics* (New York: Interscience Publishers).

Kiefer, J., and Wolfowitz, J., 1959, Optimum design in regression problems. *Annals of Mathematics and Statistics*, **30**, 271–294.

Korenberg, M. J., 1973, Identification of biological cascades of linear and static nonlinear systems. *Proceedings of the 16th Midwest Symposium on Circuit Theory*, Waterloo, Ontario, Canada, pp. 18.2.1–18.2.9.

Kortmann, M., and Unbehauen, H., 1988, Two algorithms for model structure determination of nonlinear dynamic systems with applications to industrial processes. *8th IFAC Symposium on Identification and System Parameter Estimation*, Beijing, China, pp. 939–946.

Micchelli, C. A., 1988, Optimal sampling design for parameter estimation and p-widths under stochastic and deterministic noise. In M. Milanese *et al.* (Eds) *Robustness in Identification and Control* (New York, Plenum Press), pp. 25–40.

Narendra, K. S., and Gallman, P. G., 1966, An iterative method for the identification of nonlinear systems using a Hammerstein model. *IEEE Transactions on Automatic Control*, **11**, 546–550.

Pukelsheim, F., and Studden, W. J., 1993, E-optimal designs for polynomial regression. *Annals of Statistics*, **21**, pp. 402–415

Zi-Qiang, L., 1994, On identification of the controlled plants described by the Hammerstein system. *IEEE Transactions on Automatic Control*, **39**, 569–573.