Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
automatica
Automatica 41 (2005) 1017–1024
www.elsevier.com/locate/automatica

Brief paper

Approximation of the Feasible Parameter Set in worst-case identification of Hammerstein models☆

P. Falugi[a], L. Giarré[b,*], G. Zappa[a]

[a]*Dipartimento di Sistemi e Informatica, Università di Firenze, via di Santa Marta, 50139 Firenze, Italy*
[b]*Dipartimento di Ingegneria dell'Automazione e dei Sistemi, Viale delle Scienze, 90128, Palermo, Italy*

Received 7 January 2004; received in revised form 2 December 2004; accepted 23 December 2004
Available online 17 March 2005

Abstract

The estimation of the Feasible Parameter Set (FPS) for Hammerstein models in a worst-case setting is considered. A bounding procedure is determined both for polytopic and ellipsoidic uncertainties. It consists in the projection of the FPS of the extended parameter vector onto suitable subspaces and in the solution of convex optimization problems which provide Uncertainties Intervals of the model parameters. The bounds obtained are tighter than in the previous approaches.

Keywords: Nonlinear systems identification; Set membership; Hammerstein models; Uncertainty intervals

1. Introduction

Hammerstein models are a special class of *block structured network models* (see Chen (1985) for a complete and detailed characterization): Interconnections of memoryless nonlinear blocks $\mathscr{N}$ and linear dynamics blocks $\mathscr{L}$ have been proven to be successful as simple nonlinear models for a wide number of applications (nonlinear filtering, actuator saturations, audio-visual processing, signal analysis, biologic systems, chemical processes). In particular, the Hammerstein model is the series interconnection of an $\mathscr{N}$ and an $\mathscr{L}$ block. The literature on parametric identification of block oriented models, based on Input/Output noisy measurements, can be essentially divided into three categories:

— *Minimum Prediction Error* (*MPE*) approach, leading to iterative procedures. For $\mathscr{N}\mathscr{L}$ models, for example, the most popular approach is the Narendra–Gallman algorithm (Narendra and Gallman, 1966), an iterative least-squares procedure.
— *Gray box* procedures, where a decoupling of the estimation of the nonlinear and the linear part is obtained through a particular choice of the input signal. For $\mathscr{N}\mathscr{L}$, $\mathscr{L}\mathscr{N}$ and Lur'e models in Pearson and Pottmann(2000), the linear and nonlinear parts are decoupled assuming that the steady-state gain of the linear dynamic model is constrained to be unity, and the steady-state characteristic of the overall model is determined entirely by the static nonlinearity.
— *Two-stage* procedures, where a linear extended parametrization of the nonlinear model is first estimated, and subsequently the estimate is projected onto the nonlinear manifold of model parameters. See Bai (1998) for $\mathscr{N}\mathscr{L}$ systems.

Almost all these contributions assume a statistical description of the noise and are mainly devoted to point estimation

☆ This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor B. Ninness under the direction of Editor T. Soderstrom. Research supported by MURST Grant ex-40% "Robust Techniques for Uncertain Systems Control".

* Corresponding author. Dipartimento di Sistemi e Informatica, Università di Firenze, via di Santa Marta, 50139 Firenze, Italy.
E-mail addresses: falugi@dsi.unifi.it (P. Falugi), giarre@unipa.it (L. Giarré).

0005-1098/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2004.12.010

while little attention is devoted to the computation of confidence regions for the parameter estimates, although they are important for the assessment of the model quality. Conversely, the assumption of Unknown But Bounded (UBB) noise naturally arises the issue of computing the Feasible Parameter Set (FPS) which provides the uncertainties regions for the parameter estimates. Notice that the exact computation of the FPS for block structured models is a difficult task, this set being nonconvex; and even its outer bounding with convex sets has been rarely considered in the literature. In particular in Garulli et al. (2002), the identification of an $\mathcal{N}\mathcal{L}$ model from data generated by diagonal Volterra models is solved. The estimate with minimum ℓ_2 worst-case error is provided by the Chebichev center of the FPS *conditioned* to the *nonlinear* manifold of rank-one matrices and, consequently, the FPS is outer bounded by an ℓ_2 ball. Upper and lower bounds on the parameter Uncertainty Intervals (UI) of $\mathcal{N}\mathcal{L}$ models can be found in Belforte and Gay(2001). In the present work, upper and lower bounds of the FPS are computed for $\mathcal{N}\mathcal{L}$ models affected by UBB noise in ℓ_2 and ℓ_∞ norm. The bounds are much less conservative than in Belforte and Gay(2001).

In order to identify Hammerstein models, the following parametrizations have been adopted:

- Linear block

$$\mathcal{L}(z) = \sum_{i=1}^{M} h_i B_i(z^{-1}), \quad h \triangleq [h_1, \dots, h_M]^{\mathrm{T}}, \tag{1}$$

  where $B_i(z^{-1})$ depends on the chosen rational basis function such as Laguerre, Kautz, orthonormal, etc. For example considering Finite Impulse Response (FIR) parametrization $B_i(z^{-1}) = z^{-i}$;
- Nonlinear block

$$\mathcal{N}(u) = \sum_{i=1}^{N} p_i g_i(u), \quad p \triangleq [p_1, \dots, p_N]^{\mathrm{T}}, \tag{2}$$

  where $g_i(\cdot) : \mathbb{R} \to \mathbb{R}$ are a set of specified basis functions (for instance polynomial or trigonometric functions).

The paper is organized as follows. In Section 2 the mathematical programming problem, providing the parameter UIs, is formulated while in Section 3 its solution is described using suitable relaxation techniques. A simulation section concludes the paper.

2. Problem formulation

Given m noisy Input/Output measurements $\{y_k, u_k; k = 1, \dots, m\}$ on the nonlinear system, we assume that they can be recast in the regression form

$$y_k = \phi_k^{\mathrm{T}} \theta + e_k, \tag{3}$$

where e_k is the noise, θ is the *extended* parameter vector $\theta \triangleq [\theta_1^{\mathrm{T}}, \dots, \theta_N^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{NM}$ and ϕ_k is the corresponding regressor vector.

$$\phi_k \triangleq [\phi_{k1}^{\mathrm{T}}, \dots, \phi_{kN}^{\mathrm{T}}]^{\mathrm{T}},$$
$$\phi_{ki} \triangleq [B_1(z^{-1}) g_i(u_k), \dots, B_M(z^{-1}) g_i(u_k)]^{\mathrm{T}}. \tag{4}$$

Eq. (3) corresponds to an $\mathcal{N}\mathcal{L}$ model if and only if θ admits the decomposition:

$$\theta = p \otimes h \Leftrightarrow \theta_i \triangleq p_i h \in \mathbb{R}^M. \tag{5}$$

Remark 1. The bilinear decomposition (5) applies to other block structured models: for instance in the Lur'e model, one gets

$$y_k = e_k + h^{\mathrm{T}} \phi_{k0} - (p \otimes h)^{\mathrm{T}} \phi_{y_k}, \tag{6}$$

where $\phi_{k0} = [B_1(z^{-1}) u_k, \dots, B_M(z^{-1}) u_k]^{\mathrm{T}}$ and ϕ_{y_k} is obtained from ϕ_k by replacing u with y and e_k is a noise entering after the linear part and before the output y. For extension to the Wiener model see Gómez and Baeyens (2004), where it is assumed that the inner signal connecting the linear and the nonlinear block is affected by noise while the output measurements are noise-free.

From the bilinear decomposition (5) it turns out that additional constraints must be imposed on the parameters p, h in order to get a unique decomposition (i.e. structural identifiability). Moreover, in order to avoid unessential technicalities at this stage, we also impose an experimental identifiability condition on θ, i.e. that the *regressor matrix* $\Phi \triangleq [\phi_1, \phi_2, \dots, \phi_m]^{\mathrm{T}}$ is full column rank. In turn, this corresponds to a suitable persistent excitation for the input u_k.

Assuming that the noise sequence $\{e_k\}$ is UBB, i.e. that $\{e_k\} \in \mathscr{E} \subset \mathbb{R}^{NM}$, we define the *extended FPS*

$$\Omega_\theta \triangleq \{\theta : Y - \Phi\theta \in \mathscr{E}\} \subset \mathbb{R}^{NM} \tag{7}$$

with $Y = [y_1, \dots, y_m]^{\mathrm{T}}$ and $\Phi = [\phi_1, \phi_2, \dots, \phi_m]^{\mathrm{T}}$. Ω_θ is the set of vectors θ not falsified by the measurements, *ignoring* the constraints among the components of θ provided by its bilinear decomposition (5). Ω_θ is compact or convex provided that $\mathscr{E}$ is compact or, respectively, convex.

In particular, we consider the following two cases for $\mathscr{E}$:

(A.1) The noise is bounded in the ℓ_2 norm, i.e.

$$\frac{1}{m} \sum_{k=1}^{m} e_k^2 \leqslant \varepsilon^2. \tag{8}$$

Then

$$\Omega_\theta = \left\{ \theta : \frac{1}{m} \sum_{k=1}^{m} (y_k - \phi_k^{\mathrm{T}} \theta)^2 \leqslant \varepsilon \right\} \tag{9}$$

is an ellipsoid in $\mathbb{R}^{NM}$, i.e.

$$\Omega_\theta = \{\theta : (\theta - \hat{\theta})^{\mathrm{T}} Q(\theta - \hat{\theta}) \leqslant 1\}, \tag{10}$$

where $\hat{\theta} = (\Phi^{\mathrm{T}}\Phi)^{-1}\Phi^{\mathrm{T}}Y$ is the least squares estimate of θ and

$$Q = \Phi^{\mathrm{T}}\Phi\delta^{-1},$$
$$\delta = m\varepsilon - Y^{\mathrm{T}}Y + Y^{\mathrm{T}}\Phi(\Phi^{\mathrm{T}}\Phi)^{-1}\Phi^{\mathrm{T}}Y. \tag{11}$$

(A.2) The noise is bounded in the ℓ_∞ norm, i.e.

$$|e_k| \leqslant \varepsilon, \quad k = 1, \dots, m. \tag{12}$$

Then

$$\Omega_\theta = \{\theta : |y_k - \phi_k^{\mathrm{T}}\theta| \leqslant \varepsilon, \; k = 1 \dots m\} \tag{13}$$

is a bounded convex polytope in $\mathbb{R}^{NM}$.

The aim of this paper is to characterize, as tightly as possible, the FPS of the physical parameters p, h, i.e. the set $\Omega_{ph} \triangleq \{p, h : p \otimes h \in \Omega_\theta\}$. In particular, we are interested in the computation of the UI of each model scalar parameter, i.e. the smallest interval which contains all feasible values of that parameter. This amounts to the solution of the following mathematical programming *problems*:

For $i = 1, \dots, N, \; j = 1, \dots, M$, solve:

$$\begin{array}{ll} \min\limits_{p,h} \; p_i, & \max\limits_{p,h} \; p_i \\ \min\limits_{p,h} \; h_j, & \max\limits_{p,h} \; h_j \\ \text{such that:} & p \otimes h \in \Omega_\theta \end{array} \tag{14}$$

Unfortunately these are nonconvex problems which can be solved by time-consuming branch and bound procedures. In the next section we show how the parameter bounding problems (14) can be simplified, and how in many cases they become convex.

3. Problem relaxation and bounds computation

Let us make the following simplifying assumption:

The constraints are relaxed by replacing the extended FPS Ω_θ with the outer approximation $\Omega_{\theta_1} \times \Omega_{\theta_2} \times \cdots \times \Omega_{\theta_N}$ where $\times$ denotes Cartesian product and Ω_{θ_i} the *projection* of Ω_θ along the coordinates of θ_i.

Now (14) reduces to the *relaxed* problems:

for $i = 2, \dots, N$, compute

$$p_i^- = \min_{p,h} \; p_i, \quad p_i^+ = \max_{p,h} \; p_i, \tag{15}$$

for $i = 1, \dots, M$ compute

$$h_i^- = \min_{p,h} \; h_i, \quad h_i^+ = \max_{p,h} \; h_i, \tag{16}$$

such that

$$p_1 = 1, \quad p_j h \in \Omega_{\theta_j}, \; j = 1, \dots, N, \tag{17}$$

where the structural identifiability constraint $p_1 = 1$ has been specified.

Notice that:

(1) If Ω_θ is a polytope, then Ω_{θ_i} is a polytope as well and can be represented by

$$\Omega_{\theta_i} = \{\theta_i : M_i\theta_i \leqslant m_i\}, \tag{18}$$

where the matrix M_i and the vector m_i can be computed, for instance, by exploiting the Fourier–Motzkin elimination algorithm as in Keerthi and Gilbert(1987).

(2) If Ω_θ is an ellipsoid, then Ω_{θ_i} is an ellipsoid as well:

$$\Omega_{\theta_i} = \{\theta_i : (\theta_i - \hat{\theta}_i)^{\mathrm{T}} Q_i(\theta_i - \hat{\theta}_i) \leqslant 1\}, \tag{19}$$

where the matrix $Q_i \in \mathbb{R}^{M \times M}$ is the Schur complement of a suitable submatrix of Q and therefore is computed analytically.

The geometrical interpretation of the constraints in (17) is the following: the vector $h \in \mathbb{R}^M$ is feasible if $h \in \Omega_{\theta_1}$ and, for $j \geqslant 2$, there exists a scalar parameter p_j such that $p_j h \in \Omega_{\theta_j}$, i.e. if the line passing through h intersects the set Ω_{θ_j}. Moreover, notice that the constraints on the parameters p_j are now decoupled.

We now address in detail the solution to the relaxed problems (17) for the polytopic and ellipsoidal cases.

For *polytopic uncertainties*, exploiting the constraint representation (18) and introducing the *change of coordinates* $\pi_i \triangleq (p_i)^{-1}, \; i = 1, \dots, N$, the problem (15–17) becomes a set of linear programming (LP) problems: for $i = 2, \dots, N$ compute

$$\pi_i^- = \min_{\pi,h} \; \pi_i, \quad \pi_i^+ = \max_{\pi,h} \; \pi_i, \tag{20}$$

for $i = 1, \dots, M$ compute

$$h_i^- = \min_{\pi,h} \; h_i, \quad h_i^+ = \max_{\pi,h} \; h_i, \tag{21}$$

such that

$$\pi \in \mathbb{R}^N, \quad \pi_1 = 1, \tag{22}$$
$$M_j h \leqslant \pi_j m_j \quad \text{if } \pi_j \geqslant 0,$$
$$-M_j h \leqslant -\pi_j m_j \quad \text{if } \pi_j \leqslant 0, \quad j = 1, \dots, N.$$

In fact $p_i^+ = (\pi_i^-)^{-1}$, $p_i^- = (\pi_i^+)^{-1}$; in particular $p_i^+ = 0$ or $p_i^- = 0$ when the solution to (22) is unbounded.

For each orthant in the π space $\mathbb{R}^N$, (20–22) is an LP problem. Usually, it is not necessary to explore all the 2^N orthants. In fact a priori information on the sign of feasible p_j's (and hence of π_j's) allows one to limit to one or very few the number of orthants to be considered.

Remark 2. Notice that one can remove from (22) those constraints, indexed by $j, j \neq 1, i$, such that all entries of m_j are $\geqslant 0$ or $\leqslant 0$, corresponding to polytopes Ω_{θ_j} that contain the origin.

Consider now *ellipsoidal uncertainties*. Exploiting the constraint representation (19), the relaxed problem (16–17) reduces to:

for $i = 2, \dots, N$ compute

$$p_i^- = \min_{p,h} \; p_i, \quad p_i^+ = \max_{p,h} \; p_i, \tag{23}$$

for $i = 1, \dots, M$ compute
$h_i^- = \min_{p,h} \; h_i, \quad h_i^+ = \max_{p,h} \; h_i,$
subject to $p_1 = 1$,

$$(p_j h - \hat{\theta}_j)^{\mathrm{T}} Q_j (p_j h - \hat{\theta}_j) \leqslant 1, \quad j = 1, \dots, N. \tag{24}$$

Contrarily to the polytopic case, in order to achieve a suitable final formulation of the problem the properties of the set of admissible h's will be investigated. In the previous case the discussion has been straightforward due to the possibility of decoupling the variables in the constraints by increasing the dimension of the space.

Proposition 1. *Let $\mathscr{K}_j$ denote the set of vectors h for which there exists a scalar p_j such that $p_j h \in \Omega_{\theta_j}$. Then*

$$\mathscr{K}_j = \{h : h^{\mathrm{T}} M_j h \geqslant 0\}, \tag{25}$$

where

$$M_j \triangleq Q_j \hat{\theta}_j \hat{\theta}_j^{\mathrm{T}} Q_j - \gamma_j Q_j,$$
$$\gamma_j \triangleq \hat{\theta}_j^{\mathrm{T}} Q_j \hat{\theta}_j - 1. \tag{26}$$

Proof. See the Appendix.

Notice that if $\gamma_j < 0$, M_j is positive definite and the constraint $h^{\mathrm{T}} M_j h \geqslant 0$ becomes trivial (see Remark 2). Geometrically $\gamma_j < 0$ means that the ellipsoid Ω_{θ_j} contains the origin and any line, passing through the origin, intersects Ω_{θ_j}. So, without loss of generality, we assume that $\gamma_j \geqslant 0$. In this case, M_j is not sign definite since $\hat{\theta}_j^{\mathrm{T}} M_j \hat{\theta}_j \geqslant 0$ while $h^{\mathrm{T}} M_j h \leqslant 0$ for any h orthogonal to $Q_j \hat{\theta}_j$.

Proposition 2. *$\mathscr{K}_j$ can be decomposed as*

$$\mathscr{K}_j = \mathscr{K}_j^+ \cup \mathscr{K}_j^-, \tag{27}$$

where,

$$\mathscr{K}_j^+ \triangleq \{h : h^{\mathrm{T}} M_j h \geqslant 0, \; h^{\mathrm{T}} Q_j \hat{\theta}_j \geqslant 0\},$$
$$\mathscr{K}_j^- \triangleq -\mathscr{K}_j^+, \tag{28}$$

are convex cones. Moreover $\Omega_{\theta_j} \in \mathscr{K}_j^+$.

Proof. See the Appendix.

Hence, let us denote by

$$\mathscr{C} \triangleq \bigcap_{j \neq 1} \{\Omega_{\theta_1} \cap \mathscr{K}_j\} \tag{29}$$

the set of admissible h. Then the relaxed problem (24) amounts to:

for $i = 2, \dots, N$ compute

$$p_i^- = \min_h \; p_i, \quad p_i^+ = \max_h \; p_i$$
$$\text{subject to } p_i h \in \Omega_{\theta_i}, \quad h \in \mathscr{C} \tag{30}$$

and for $i = 1, \dots, M$ compute

$$h_i^- = \min_h \; h_i, \quad h_i^+ = \max_h \; h_i$$
$$\text{subject to } h \in \mathscr{C}. \tag{31}$$

It is important to check under which conditions the set $\mathscr{C}$, assumed nonempty, is convex. Clearly if, for any $j \neq 1$, Ω_{θ_1} intersects only one between the two convex cones $\mathscr{K}_j^+$, $\mathscr{K}_j^-$, in which $\mathscr{K}_j$ is decomposed, then $\mathscr{C}$, being the intersection of convex sets, is convex. Sufficient conditions are provided by the following two lemmas.

Lemma 1. *Ω_{θ_1} does not intersect both $\mathscr{K}_j^+$ and $\mathscr{K}_j^-$, if, for any $h_1, h_2 \in \Omega_{\theta_1}$ and any $h_3, h_4 \in \Omega_{\theta_j}$, we have $h_1^{\mathrm{T}} h_2 > 0$ and $h_3^{\mathrm{T}} h_4 > 0$.*

Proof. By its geometrical interpretation, the proof is evident, because the condition $h_1^{\mathrm{T}} h_2 > 0$ clearly implies that the tangent cones $\mathscr{K}_1$ and $\mathscr{K}_j$ in the origin are not larger than $90°$. □

Lemma 2. *Ω_{θ_1} does not intersect $\mathscr{K}_j^-$ if*

$$\hat{\theta}_1^{\mathrm{T}} Q_j \hat{\theta}_j > (\hat{\theta}_j^{\mathrm{T}} Q_j Q_1^{-1} Q_j \hat{\theta}_j)^{1/2}. \tag{32}$$

Ω_{θ_1} does not intersect $\mathscr{K}_j^+$ if

$$\hat{\theta}_1^{\mathrm{T}} Q_j \hat{\theta}_j < -(\hat{\theta}_j^{\mathrm{T}} Q_j Q_1^{-1} Q_j \hat{\theta}_j)^{1/2}. \tag{33}$$

Proof (*sketch*). We want to show that, under (32), Ω_{θ_1} belongs to the half-space defined by $\{h : h^{\mathrm{T}} Q_j \hat{\theta}_j > 0\}$. Let us first compute the hyperplane tangent to Ω_{θ_1} in h_0 and normal to $Q_j \hat{\theta}_j$. By differentiating the ellipsoid equation, we have the following relationship between the tangent point h_0 and the normal vector $Q_j \hat{\theta}_j : Q_1 h_0 - Q_1 \hat{\theta}_1 = \lambda Q_j \hat{\theta}_j$. Moreover, $\lambda \in \mathbb{R}$ can be determined by exploiting the ellipsoid equation, since $(h_0 - \hat{\theta}_1)^{\mathrm{T}} Q_1 (h_0 - \hat{\theta}_1) = 1$. Eq. (32) is satisfied if the distance between Ω_{θ_1} and the hyperplane $h^{\mathrm{T}} Q_j \hat{\theta}_j = 0$ is strictly bigger than the projection of $h_0 - \hat{\theta}_1$ in the direction $Q_j \hat{\theta}_j$. Notice that condition (32) guarantees that $h^{\mathrm{T}} Q_j \hat{\theta}_j \geqslant 0 \; \forall h \in \Omega_{\theta_1}$ (the opposite for condition (33)). □

Lemma 2 is particularly useful since it can be easily checked and, in most cases, allows one to select which convex component of $\mathscr{K}_j$ must be included in the definition of the constraining set $\mathscr{C}$.

From the above two lemmas it follows that double intersections may occur only if the diameter of the ellipsoids Ω_{θ_j} is large compared to the distance of its center from the origin, i.e. if the uncertainty on the estimate $\hat{\theta}_j$ is larger than the norm of the estimate itself. It must be underlined that this is a pathological situation corresponding to very low Signal to Noise Ratio (SNR). Actually, in all the numerical experiments carried out this double intersection never occurred. Conversely, when for some j, Ω_{θ_1} intersects both $\mathscr{K}_j^+$ and $\mathscr{K}_j^-$, there is no guarantee that $\mathscr{C}$ is still convex. In any case it can be decomposed in the union of convex sets. More precisely, $\mathscr{C}$ is the union of *at most* 2^l convex sets, where l is the number of indexes j for which a double intersection occurs. Assume now that $\mathscr{C}$ is convex. Then (31) is a convex problem and standard solvers can be used. Moreover, problem (30) amounts to finding those values of p_i for which the scaled ellipsoid $p_i^{-1}\Omega_{\theta_i} \triangleq \{h : \|h - p_i^{-1}\hat{\theta}_i\|^2_{Q_i} \leqslant 1/p_i^2\}$ is tangent to $\mathscr{C}$. This is equivalent to computing the Q_i weighted distance $d(p_i)$ between $p_i^{-1}\hat{\theta}_i$ and $\mathscr{C}$ and looking for those values of p_i for which $d(p_i) = 1/p_i^2$. Given p_i, the computation of the distance is a convex problem, for which again standard solvers can be used; moreover, since the set of admissible values for p_i is an interval, there exist only two solutions $p_i^- \, p_i^+$ to the equation $d(p_i) = 1/p_i^2$. They can be easily found by a bisection technique. If $\mathscr{C}$ is nonconvex, the above procedure must be carried out for each convex component of $\mathscr{C}$.

Defining $\Omega_{p_i} = \{p_i : \Omega_{\theta_i} p_i^{-1} \cap \Omega_{\theta_1} \neq \emptyset\}$, it is easy to show that Ω_{p_i}, if nonempty, is an interval or in the worst-case the union of two intervals.

Remark 3. The tightness of the bounds provided by solving relaxed problems relies on the quality of the approximation of the extended FPS Ω_θ with its projections along the directions spanned by the vectors $\theta_i \in \mathbb{R}^M$, $i=1,\dots,N$. An alternative approximation can be obtained by the projection along the directions spanned by the vectors $\theta_j^* \in \mathbb{R}^N$, $j=1,\dots,M$ provided by the following dual parametrization rearranging the components of the regressor ϕ_k in the new regressor

$$\phi_k^* \triangleq [\phi_{k1}^{*\mathrm{T}}, \dots, \phi_{kM}^{*\mathrm{T}}]^{\mathrm{T}},$$
$$\phi_{kj}^* \triangleq [B_j(z^{-1})g_1(u_k), \dots, B_j(z^{-1})g_N(u_k)]^{\mathrm{T}}. \tag{34}$$

The Input/Output relation (3) can be rewritten as $y_k = \phi_k^{*\mathrm{T}}\theta^* + e_k$ with

$$\theta \triangleq [\theta_1^{*\mathrm{T}}, \dots, \theta_N^{*\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{MN}, \quad \theta_i^* \triangleq h_i p \in \mathbb{R}^N. \tag{35}$$

The vector θ^* will be called the *dual* extended parameter vector of $\mathscr{N}\mathscr{L}$. In this way new bounds are obtained. The latter bounds need to be added with the previous ones as constraints in order to get a closer approximation of the true uncertainty intervals.

This bounding procedure provides the parameter UIs for any input signal which guarantees boundedness of the extended FPS. However, the degree of conservativeness introduced depends on the "shape" of Ω_θ, i.e. on the choice of input signal. In particular, if the input signal is a random steps signal (i.e. a sequence of random amplitude steps of sufficiently long duration such that steady-state behavior is achieved), then the projection of Ω_θ in the directions spanned by the coefficients of p is a tight description for Ω_p and no conservatism is introduced at this stage. Notice that this input selection has been often adopted in the parameter identification of block oriented models since it allows the decoupling in the estimate of the linear and nonlinear part; see Pearson and Pottmann(2000).

The effectiveness of the proposed bounding procedure (BP) obtained by solving problems (30) and (31) is now compared to the bounding algorithm introduced in Belforte and Gay(2001) for polytopic uncertainties and extended here to ellipsoidic uncertainties. It is convenient to define the projections of Ω_θ on each of its components $\Omega_{i,j} \triangleq \{x \in \mathbb{R} : \exists v \in \Omega_\theta : v_{(i-1)M+j} = x\}$ where $1 \leqslant i \leqslant N$ and $1 \leqslant j \leqslant M$. Roughly speaking, Ω_{ij} denotes the uncertainty interval of the jth component of the vector θ_i. If the denominator contains the origin, the ratio of two intervals is the whole real line $\mathbb{R}$.

As a matter of fact, the Belforte and Gay(2001) algorithm, hereafter denoted as BG, amounts to the following claims:

– $h_j \in [h_j^-, h_j^+] = \Omega_{1,j}$, $\quad j = 1, \dots, M$,
– $p_i \in [p_i^-, p_i^+] = [\max_j \min p_i, \min_j \max p_i]$
subject to $h_j \in \Omega_{1,j}$, $p_i h_j \in \Omega_{i,j}$.

The intervals relative to h derived by our method are less conservative. As a matter of fact, they are included in those derived in Belforte and Gay(2001) because $\Omega_{1,1} \times \Omega_{1,2} \times \Omega_{1,3} \times \cdots \times \Omega_{1,M} \supseteq \Omega_{\theta_1}$.

Similarly for p, the BG algorithm can be reformulated as follows:

$p_i \in [\max_j \min p_i, \min_j \max p_i]$
subject to $\Omega_{1,j} \cap \Omega_{i,j} p_i^{-1} \neq \emptyset$.

Our method instead yields

$p_j \in [\min p_j, \max p_j]$
subject to $\Omega_{\theta_1} \cap \Omega_{\theta_j} p_j^{-1} \neq \emptyset$,

which is always a smaller interval since $\Omega_{\theta_1} \cap \Omega_{\theta_j} p_j^{-1} \neq \emptyset \Rightarrow \Omega_{\theta_{1,i}} \cap \Omega_{\theta_{j,i}} p_j^{-1} \neq \emptyset \ \forall i$.

The BG algorithm essentially amounts to solving a relaxed problem where the FPS Ω_θ is outbounded by an axis aligned orthotope. Clearly this is the loosest possible approximation of Ω_θ since any correlation among the components of θ is lost. Hence the bounds provided by the BG algorithm are expected to be more conservative. On the other side the computational load is limited since, even in the polytopic case, each projection is simply obtained by solving two LP problems. Clearly the conservatism of the bounding procedure can be further reduced, exploiting the results of the dual problem illustrated in Remark 3.

3.1. Block oriented model

Note that a model parametrized by the extended vector $\theta \in \mathbb{R}^{NM}$ as in (3) corresponds to the *parallel* of $r\mathcal{NL}$ blocks, if and only if there exist vectors $p^\ell \in \mathbb{R}^N$ and $h^\ell \in \mathbb{R}^M$, $\ell = 1, \dots, r$ such that

$$\theta = \sum_{\ell=1}^{r} p^\ell \otimes h^\ell. \quad (36)$$

The extension of the proposed procedure to the parallel of $r > 1$ pathways ($\mathcal{PNL}_r$) is still under investigation. The main issue is that the problem is not convex, but the use of projections may introduce a high level of simplification in the bounding procedure, and the problems could be decoupled.

4. Numerical examples

The effectiveness of the proposed BP is now illustrated on some numerical examples and compared to the bounding algorithm introduced in Belforte and Gay (2001) for polytopic and ellipsoidic uncertainties. The plants considered in the simulations are Hammerstein models characterized by a nonlinearity of polynomial type $\mathcal{N}(u_k) = \sum_{i=1}^{N} p_i u_k^i$, and by a linear block $\mathcal{L}(z) = \sum_{i=1}^{M} h_i z^{-i}$. In all the identification experiments the noise is bounded either in the ℓ_2 or ℓ_∞ norm with $\varepsilon = 0.5$ while the input signal is a square wave of amplitude ± 1 with random noise superimposed with variance 0.5. This input signal guarantees boundedness of the FPS.

4.1. Example 1

In this first example we consider a very simple plant with $p = [1\ 2\ 3]^{\mathrm{T}} (N = 3)$ and $h = [1\ 2]^{\mathrm{T}}$ $(M = 2)$, corresponding to the extended parameter vector $\theta_{\text{true}} = [1\ 2\ 2\ 4\ 3\ 6]^{\mathrm{T}}$. In an identification experiment, carried out with $m = 100$ measurements using ℓ_2 noise bounds, the extended FPS resulted in an ellipsoid centered in $\hat{\theta} = [0.9901\ 1.9995\ 2.0014\ 4.0011\ 2.9985\ 6.0027]^{\mathrm{T}}$ and whose two dimensional projections Ω_{θ_i}, $i = 1, 2, 3$ are the ellipsoids, centered in $\hat{\theta}_i$ and represented in Fig. 1. Not surprisingly, they are very well aligned in this example where the least squares estimates $\hat{\theta}_i$ are very close to the true values. Therefore Ω_{θ_1} intersects only the convex cones $\mathcal{K}_j^+$, $j = 2, 3$, and the set $\mathscr{C}$ is convex (see the dashed area in Fig. 1).

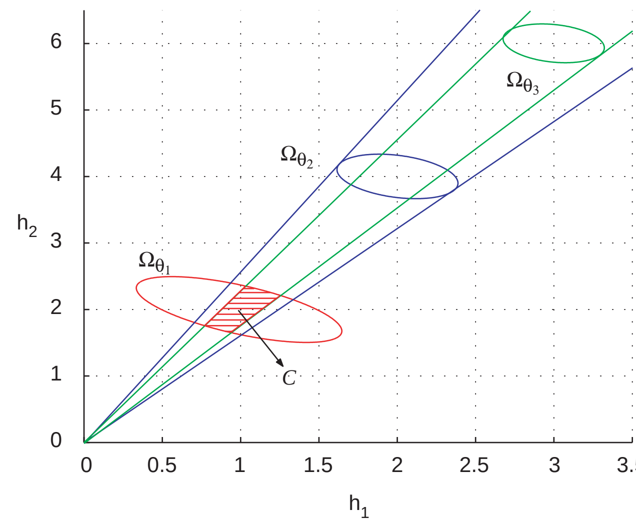

Fig. 1. Projected ellipsoids: Ω_{θ_1}, Ω_{θ_2} and Ω_{θ_3}.

Identification experiments with ℓ_∞ and ℓ_2 noise bounds were carried out and the performance of the BG algorithm and the new BP compared. Results are represented in Tables 1 and 2 where the parameter UI $[p_i^-,\ p_i^+]$, $[h_i^-,\ h_i^+]$

Table 1
Example 1: PUIs: ellipsoidal uncertainty

	BP	BG	d_{BP}	d_{BG}	r
h_1	[0.77 1.25]	[0.33 1.65]	0.48	1.31	63.44%
h_2	[1.65 2.35]	[1.51 2.49]	0.70	0.98	28.97%
p_2	[1.63 2.50]	[1.47 2.88]	0.87	1.41	38.31%
p_3	[2.52 3.66]	[2.29 4.18]	1.15	1.89	39.49%

Table 2
Example 1: PUIs: polytopic uncertainty—exact projections

	BP	BG	d_{BP}	d_{BG}	d_{IA}	r
h_1	[0.98 1.02]	[0.95 1.05]	0.04	0.10	0.02	64.35%
h_2	[1.97 2.03]	[1.96 2.04]	0.06	0.09	0.05	28.08%
p_2	[1.96 2.03]	[1.94 2.05]	0.07	0.11	0.04	34.23%
p_3	[2.95 3.05]	[2.92 3.07]	0.10	0.15	0.08	33.78%

Table 3
Example 2: PUIs: polytopic uncertainty—approximate projections

	BP	BG	d_{BP}	d_{BG}	d_{IA}	r
h_1	[0.98 1.03]	[0.93 1.12]	0.05	0.19	0.01	72.92%
h_2	[1.95 2.04]	[1.93 2.08]	0.09	0.15	0.02	44.52%
h_3	[0.29 0.31]	[0.21 0.40]	0.02	0.19	0.01	88.36%
h_4	[3.91 4.07]	[3.91 4.10]	0.16	0.19	0.04	15.71%
h_5	[0.97 1.02]	[0.91 1.09]	0.05	0.18	0.01	71.11%
h_6	[0.97 1.03]	[0.94 1.11]	0.05	0.17	0.01	69.05%
h_7	[0.48 0.51]	[0.38 0.58]	0.03	0.02	0.01	84.69%
p_2	[2.95 3.07]	[2.92 3.08]	0.12	0.16	0.03	23.64%
p_3	[1.96 2.04]	[1.95 2.05]	0.08	0.11	0.02	24.53%

Table 4
Example 2: PUIs: ellipsoidal uncertainty

	BP	BG	d_{BP}	d_{BG}	r
h_1	[0.91 1.05]	[0.81 1.23]	0.13	0.41	67.15%
h_2	[1.98 2.01]	[1.72 2.34]	0.04	0.62	93.73%
h_3	[0.29 0.31]	[0.18 0.45]	0.02	0.27	93.28%
h_4	[3.94 4.06]	[3.54 4.57]	0.40	1.03	60.72%
h_5	[0.98 1.03]	[0.81 1.22]	0.05	0.41	88.08%
h_6	[0.96 1.04]	[0.81 1.23]	0.08	0.41	81.69%
h_7	[0.46 0.53]	[0.36 0.67]	0.07	0.31	76.62%
p_2	[2.72 3.32]	[2.35 4.07]	0.60	1.72	65.06%
p_3	[1.81 2.34]	[1.56 2.72]	0.53	1.15	54.11%

are reported, together with their width d and the corresponding amplitude reduction percentage

$$r \triangleq 100 \times \frac{d_{BG} - d_{BP}}{d_{BG}}$$

of the new BP with respect to BG.

Moreover, an inner approximation of the true parameter UI was computed, exploiting suitable improved Monte Carlo techniques, and its width d_{IA} is also reported.

4.2. Example 2

The plant considered in this second example is a more complex Hammerstein model characterized by $p=[1\ 3\ 2]^{\mathrm{T}}$ $(N=3)$ and $h=[1\ 2\ 0.3\ 4\ 1\ 1\ 1\ 0.5]^{\mathrm{T}}(M=8)$. In this example, in the case of polytopic uncertainties, an outer bounding approximation of the Fourier–Motzkin algorithm has been used, due to its high computational burden. This approximate projection is characterized by the fact that the projection Ω_{θ_i} has the same number of constraints as the extended FPS Ω_θ. Of course, additional conservativeness is introduced. The identification experiments were carried out with $m=300$ data points. The performance of the BG algorithm and the new BP, for polytopic and ellipsoidal uncertainties, are reported in Table 3 and Table 4, respectively. Even in the polytopic case, where exact projections could not be employed, the improvement provided by the new procedure is clear.

5. Conclusions

A bounding procedure has been proposed for the estimation of the FPS for Hammerstein models in a worst-case setting. It consists in the projection of the FPS $\subset \mathbb{R}^{MN}$ of the extended parameter vector onto suitable M and/or N-dimensional subspaces and in the solution of convex optimization problems which provide the extreme points of the UI of the model parameters. In particular for polytopic FPS, the optimization problem reduces to a linear problem; however, the computation of the exact projection is hard and suboptimal projections are employed. Conversely, for ellipsoidic FPS the computation of the projection is straightforward but the optimization problem requires a bisection procedure which, at each step, solves a convex problem. No specific input signals are required and the obtained bounds are tighter than in other approaches.

Acknowledgements

Giovanni Zappa untimely passed away while this paper was under review. The authors would like to remember him and his enthusiasm and courage during this work in difficult times.

Appendix

Proof of Proposition 1. $h \in \mathscr{K}_j$ if and only if there exists a p_j such that $(p_j h - \hat{\theta}_j)^{\mathrm{T}} Q_j (p_j h - \hat{\theta}_j) = 1$, i.e. if and only if

$$(h^{\mathrm{T}} Q_j h) p_j^2 - 2(h^{\mathrm{T}} Q_j \hat{\theta}_j) p_j + \hat{\theta}_j^{\mathrm{T}} Q_j \hat{\theta}_j - 1 = 0. \quad (37)$$

In turn, this second-order equation in the variable p_j admits real solutions if and only if

$$\Delta_j(h) = (h^{\mathrm{T}} Q_j \hat{\theta}_j)^2 - (\hat{\theta}_j^{\mathrm{T}} Q_j \hat{\theta}_j - 1)(h^{\mathrm{T}} Q_j h) \geqslant 0, \quad (38)$$

i.e. $h^{\mathrm{T}} M_j h \geqslant 0$. □

Proof of Proposition 2. It is straightforward to verify that $\mathscr{K}_j^+$ is a cone, and that $\Omega_{\theta_j} \subset \mathscr{K}_j^+$, since $\Omega_{\theta_j} \subset \mathscr{K}_j$ and its center $\hat{\theta}_j \in \mathscr{K}_j^+$. We show next that $\mathscr{K}_j^+$ is convex. In fact, for $h_1, h_2 \in \mathscr{K}_j^+$, we have

$$(h_1^{\mathrm{T}} Q_j \hat{\theta}_j)^2 \geqslant \gamma_j (h_1^{\mathrm{T}} Q_j h_1),$$
$$(h_2^{\mathrm{T}} Q_j \hat{\theta}_j)^2 \geqslant \gamma_j (h_2^{\mathrm{T}} Q_j h_2). \quad (39)$$

Let $h = \lambda h_1 + (1 - \lambda) h_2$ with $\lambda \in [0, 1]$. Obviously $h^{\mathrm{T}} Q_j \hat{\theta}_j \geqslant 0$. Moreover,

$$h^{\mathrm{T}} M_j h = \lambda^2 h_1^{\mathrm{T}} M_j h_1 + (1-\lambda)^2 h_2^{\mathrm{T}} M_j h_2 + 2\lambda(1-\lambda)(h_1^{\mathrm{T}} M_j h_2). \quad (40)$$

Since $h_1^{\mathrm{T}} M_j h_1 \geqslant 0$ and $h_2^{\mathrm{T}} M_j h_2 \geqslant 0$, we study the term $h_1^{\mathrm{T}} M_j h_2$:

$$h_1^{\mathrm{T}} M_j h_2 = (h_1^{\mathrm{T}} Q_j \hat{\theta}_j)(h_2^{\mathrm{T}} Q_j \hat{\theta}_j) - \gamma_j h_1^{\mathrm{T}} Q_j h_2 \geqslant \gamma_j ((h_1^{\mathrm{T}} Q_j h_1)^{1/2} (h_2^{\mathrm{T}} Q_j h_2)^{1/2} - h_1^{\mathrm{T}} Q_j h_2) \geqslant 0, \quad (41)$$

where the first inequality follows from (39) and the second is the Cauchy inequality. Hence $h \in \mathscr{K}_j^+$. The same reasonings hold for $\mathscr{K}_j^-$. □

References

Bai, E.-W. (1998). An optimal two-stage identification algorithm for Hammerstein–Wiener nonlinear systems. *Automatica*, *34*(3), 333–338.

Belforte, G., & Gay, P. (2001). Discrete time Hammerstein model identification with unknown but bounded errors. *IEE Proceedings of the control theory and applications*, *148*, 523–529.

Chen, H. (1985). Modeling and identification of parallel nonlinear systems: Structural classification and parameter estimation models. *Proceedings of IEEE*, *83*, 39–66.

Garulli, A., Giarrè, L., & Zappa, G. (2002). Identification of approximated Hammerstein models in a worst case setting. *IEEE Transactions on Automatic Control*, *47*(12), 2046–2050.

Gómez, J. C., & Baeyens, E. (2004). Identification of block-oriented nonlinear systems using orthonormal bases. *Journal of Process Control*, *14*, 685–697.

Keerthi, R., & Gilbert, E. G. (1987). Computation of minimum-time feedback control laws for discrete-time systems with state-control constraints. *IEEE Transactions on Automatic Control*, *32*, 432–435.

Narendra, K. S., & 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.

Pearson, R., & Pottmann, M. (2000). Gray-box identification of block oriented nonlinear models. *Journal of Process Control*, *10*, 301–315.

Paola Falugi received the Laurea degree in Electronic Engineering from the University of Florence, Italy in 1997. From 1997 to 1998, she was offered a collaboration position at the University of Siena. She received her PhD degree in the academic year 2001/2002 from the Universitá di Bologna. Since March 2002, she is a Post-doc researcher in the area of Automatic Control at the Dept. of Systems and Computer Science, University of Florence. From February 2001 to July 2001 she was a visiting scholar at the control group, Department of Engineering, Cambridge University (UK). She spent a month as a fellow in the School of Electrical and Electronics Engineering, Nanyang Technological University. Her current research interests are in predictive control and control of constrained nonlinear systems, robust control and system identification.

Laura Giarré was born in Firenze, Italy, July 14, 1960. She received the Laurea degree in Electronic Engineering in 1986 from the Università di Firenze, Itlay and the PhD degree in System Engineering in 1992 from the Università di Bologna, Italy. From 1986 to 1988, she was at the Research Laboratory of the Autostrade SPA as a Research Engineer. From 1988 to 1993, she was with the Università di Firenze, Italy. From 1993 to 1998, she was with the Politecnico di Torino, Italy. In 1998 she joined the Università di Palermo as Associate Professor of Systems Theory. In 1999 she has been visiting professor at the CCEC of the University of California at Santa Barbara. She is co-editor of the book *Multidisciplinary Research in Control* (Springer Verlag, 2002). She is Associate Editor of *Systems and Control Letters*. Her current research interests include system identification and control theory with applications to network systems and congestion control.

Giovanni Zappa was born in Rimini, Italy, September 1, 1947. He received the degree in Physics from the University of Florence in 1972. Since 1974, he worked with the Institute of Electrical Engineering of the University of Florence, holding several teaching and research positions. He was Professor of System Modeling and Identification at the Department of Systems and Computer Science (DSI) of the University of Florence. His research interests included set-membership identification, identification for control, predictive control, adaptive filtering, and parallel signal processing. He passed away on August 23, 2004.