Proceedings of the
46th IEEE Conference on Decision and Control
New Orleans, LA, USA, Dec. 12-14, 2007

Risk Adjusted Identification of a Class of Nonlinear Systems

Mario Sznaier
Electrical and Comp. Engineering Department,
Northeastern University,
Boston, MA 02115.

Constantino Lagoa Wenjing Ma
Department of Electrical Engineering,
The Pennsylvania State University,
University Park, PA 16802.

***Abstract*— This paper addresses the problem of identification of a class of nonlinear systems from a set–membership standpoint. Specifically we consider structures consisting of the interconnection of a Linear Time Invariant plant and a sector bound static nonlinearity. This structure includes as special cases Hammerstein, Wiener and Hammerstein-Wiener systems. Our main result shows that, by pursuing a risk–adjusted approach, the problem can be reduced to a convex LMI optimization form that can be efficiently solved. In addition we provide a convergence analysis that points to an intrinsic limitation of any interpolatory algorithm. These results are illustrated with a simple example.**

I. Introduction

The identification of nonlinear systems has long been seen as an important problem. Different approaches have been taken which rely on different assumptions/approximations of the system to be identified. Many of the previous results rely on a parametric approach where a linear combination of the elements of a basis is used to represent the nonlinear system. Examples of the basis used in the literature include Volterra kernel expansions, neural networks, radial basis function expansions and Fourier series; e.g., see [5], [6], [17], [28], [11], [12], [20]. Also, unstructured methods have been used to tackle the problem on nonlinear system identification; e.g., see [24], [31], [35]. In parallel with these, several authors have studied the problem of identifying nonlinear system in the presence of uncertainty; e.g., [8], [9], [14], [16], [15], [18], [19], [21], [23], [26], [29], [36]. This is the the body of work to which this paper is more closely related to. In particular, the starting point of this paper is the same as [8], [14], [16], [15], [36]; i.e., it is assumed that the nonlinear plant can be represented as a linear plant with a static feedback nonlinearity as an LFT. However, in this previous work, it is assumed that the linear part of the plant is known (apart from a feedback uncertainty bounded in the $\mathcal{H}_\infty$ norm) and several of the "internal" signals can be measured. This greatly limits the applicability of the proposed algorithms in a practical setting.

In this paper, we propose an algorithm for time–domain based identification that avoids these difficulties by pursuing a risk-adjusted approach. Here, in return for an (arbitrarily) small risk of not being able to establish consistency of the data, the problem is reduced to a convex optimization problem. In the second part of the paper we analyze the convergence properties of any interpolatory algorithm, including as a special case the one proposed here, and show that generically there is an intrinsic non–zero identification error. This error reflects the intuitive fact that in general there exists more than one interconnection consistent with the *a priori* information that can explain the observed experimental data. The paper concludes with a simple example illustrating the algorithm and pointing out to directions for future research.

This work was supported by NSF grants ECS–0221562, ITR–0312558, ECS–050166 and CNS–0519897, and AFOSR grant FA9550–05–1–0437.

II. Preliminaries

For ease of reference, next we summarize the notation used in the paper.

$\overline{\sigma}(\mathbf{A})$	maximum singular value of $\mathbf{A}$.
$\mathbf{A} > (\geq) 0$	$\mathbf{A}$ is positive(semi) definite.
$\mathbf{I}, \mathbf{0}$	the identity and zero matrices.
$\mathcal{B}\mathcal{X}(\gamma)$	open γ-ball in a normed space $\mathcal{X}$: $\mathcal{B}\mathcal{X}(\gamma) = \{x \in \mathcal{X} \colon \|x\|_{\mathcal{X}} < \gamma\}$.
$\overline{\mathcal{B}}\mathcal{X}(\gamma)$	closure of $\mathcal{B}\mathcal{X}(\gamma)$.
ℓ_p	Banach space of vector valued real sequences equipped with the norm: $\|x\|_p \doteq \left(\sum_{i=0}^{\infty} \|\mathbf{x}_i\|_p^p\right)^{\frac{1}{p}}$, $p \in [1, \infty]$ and $\|x\|_\infty \doteq \sup_i \|\mathbf{x}_i\|_\infty$.
$\mathcal{P}_n : \ell_p \to R^n$	Truncation operator: $\mathcal{P}_n(x) \doteq \{x_0, \ldots, x_{n-1}\}$
$< ., >$	inner product in ℓ_2.
$\mathcal{H}_\infty$	Space of functions with bounded analytic continuation inside the unit disk, equipped with the norm: $\|G\|_\infty \doteq ess\sup_{\|z\|<1} \overline{\sigma}(G(z))$.
$\mathcal{H}_{\infty,\rho}$	space of transfer functions analytic in $\|z\| \leq \rho$,equipped with the norm $\|G\|_{\infty,\rho} \doteq ess\sup_{\|z\|<\rho} \overline{\sigma}(G(z))$.
$\mathcal{F}_u(M, \Delta)$	Upper LFT of the operators M, Δ: $\mathcal{F}_u(M, \Delta) \doteq M_{22} + M_{21}\Delta(I - M_{11}\Delta)^{-1}M_{12}$
$\mathbf{T}_x$	lower triangular block Toeplitz matrix associated with any finite sequence $\{x_k, k = 0, 1, \cdots, n-1\}$: $\mathbf{T}_x = \begin{bmatrix} x_0 & 0 & \ldots & 0 \\ x_1 & x_0 & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ x_{n-1} & x_{n-2} & \ldots & x_0 \end{bmatrix}$

Next we recall, for ease of reference, a well known interpolation result given a necessary and sufficient condition for the existence of an LTI operator $H \in \mathcal{BH}_{\infty,\rho}$ whose first n Markov parameters (or equivalently, its derivatives at $z = 0$) interpolate a given sequence $\{\mathbf{h}_i\}_{i=0}^{n-1}$. This result has been used both in the context of model (in)validation and robust identification to recast the problem into a convex optimization form (see for instance [7].).

Lemma 1 (Carathéodory-Fejér): Given $K > 0$ and a matrix valued sequence $\{\mathbf{h}_i\}_{i=0}^{n-1}$, there exists a causal, discrete-time, LTI operator $H(z) \in \overline{\mathcal{BH}_{\infty,\rho}}(K)$ such that

$$H(z) = \mathbf{h}_0 + \mathbf{h}z + \mathbf{h}_2 z^2 + \dots \mathbf{h}_{n-1} z^{n-1} + \dots$$

if and only if

$$(\mathbf{R}\mathbf{T}_h^{n_m}\mathbf{R}^{-1})^T(\mathbf{R}\mathbf{T}_h^{n_m}\mathbf{R}^{-1}) \leq K^2\mathbf{I}.$$

where $\mathbf{R} \doteq \mathrm{diag}\{1, \rho, \dots, \rho^{n-1}\}$

Proof: See for instance [25]. ∎

III. Problem Statement

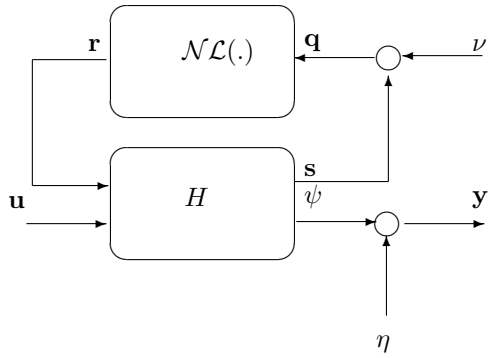

Fig. 1. Nonlinear Identification Setup

Consider the system shown in Figure 1 consisting of the interconnection of a LTI system H and a memoryless, sector bounded nonlinearity $\mathcal{NL}(.)$. The corresponding equations are given by:

$$\begin{bmatrix}\mathbf{q}_k \\ \mathbf{y}_k\end{bmatrix} = (\mathbf{H} * \begin{bmatrix} r \\ u\end{bmatrix})_\mathbf{k} + \begin{bmatrix}\nu_k \\ \eta_k\end{bmatrix} \qquad (1)$$
$$\mathbf{r}_k = \mathcal{NL}(\mathbf{q_k})$$

where $*$ denotes convolution and the signals $\mathbf{u} \in R^{n_u}$ and $\mathbf{y} \in R^{n_y}$ represent the experimental data: a known finite input sequence and its corresponding output sequence, corrupted by unknown but norm–bounded measurement noise η. Note that in general the intermediate signals $\mathbf{s} \in R^{n_p}$ (the output of the LTI system), ν (noise at the input to the nonlinearity) and $\mathbf{r}$, the output of the nonlinearity, are not measurable. Note also that in principle one could also add a third noise source at the output of the nonlinearity. However, doing so would lead to generically hard bilinear problems involving the product of the unknown signal $\mathbf{r}$ and the impulse response of H. Thus, if necessary, noise at the output of the nonlinearity will be handled by incorporating it in the signal ν. Our goal is to, given experimental data consisting of n_m measurements of the input/output pair $\{\mathbf{u}, \mathbf{y}\}$ and some *a priori* information about the plant and noise, establish whether they are consistent, and if so, find a model that interpolates the experimental data within the measurement error level.

In the sequel, we will make the following standard assumptions about the *a priori* information:

A1.- A set membership characterization of the linear portion of the plant is available:

$$H(z) \in \mathcal{S} \subseteq \mathcal{BH}_{\infty,\rho}(K) \qquad (2)$$

that is, we consider exponentially stable plants with a stability margin of $(\rho - 1)$ and a peak response to complex exponential inputs bounded by some known K. The set $\mathcal{S}$ encapsulates any additional a-priori information available about the *structure* of the problem. For instance, in the special case of Wiener systems identification, $\mathcal{S}$ has the form:

$$\mathcal{S}_{wiener} = \left\{ H(z) \in \mathcal{BH}_{\infty,\rho}(K) : H(z) = \begin{pmatrix} 0 & \mathbf{H}_{12}(z) \\ \mathbf{I} & \mathbf{0}\end{pmatrix}\right\}$$

A2.- A bound on the gain of the static nonlinearity is known, e.g.

$$\mathcal{NL}(.) \doteq \mathrm{diag}\left\{\mathcal{NL}_i(.) : ,\ [r_i - \gamma q_i]^T q_i < 0\right\}$$

for some known γ. We will further assume that H_{11}, the $(1,1)$ block of the linear system H, satisfies $\gamma\|H_{11}\|_\infty \leq K_\gamma < 1$ [1]

A3.- The measurement noise satisfies:

$$\begin{aligned}\eta &\in \mathcal{N}_\eta \doteq \{\eta : \|\eta\|_2 \leq \epsilon_\eta\} \\ \nu &\in \mathcal{N}_\nu \doteq \{\nu : \|\nu\|_2 \leq \epsilon_\nu\}\end{aligned}$$

With these assumptions, the problem under consideration can be precisely stated as:

Problem 1: Given the *a priori* information $\mathcal{S}, \mathcal{N} \doteq \mathcal{N}_\eta \times \mathcal{N}_\nu, \gamma$ and the *a posteriori* experimental data $\{\mathbf{y}, \mathbf{u}\}$, determine:

1) if the *a priori* and *a posteriori* information are consistent, i.e., the consistency set

$$\mathcal{T}(\mathbf{y}, n_m, \mathcal{N}) \doteq$$
$$\{H \in \overline{\mathcal{BH}}_{\infty,\rho}(K) :\ y_k = \left[\mathcal{F}_u(H, \mathcal{NL})(\begin{bmatrix}\nu \\ \mathbf{u}\end{bmatrix})\right]_k + \eta_k,$$
$$\text{for some } \mathcal{NL} \text{ and some sequences}$$
$$(\nu_k, \eta_k) \in \mathcal{N}, k = 0, 1, \dots, n_m - 1.\} \qquad (3)$$

is nonempty.

2) If $\mathcal{T} \neq \emptyset$, find a nominal model $\{H, \mathcal{NL}(.)\}$ that interpolates the experimental data[2]

[1] This condition can be relaxed to the existence of some $X > 0$ such that the following LMI is feasible:

$$H^*(e^{j\omega})XH(e^{j\omega}) - \gamma X < 0$$

[2] If $\mathcal{T} = \emptyset$, then the experimental data $\{\mathbf{y}, \mathbf{u}\}$ invalidates the *a priori* assumptions about the class of models and noise, that is, the experimental data cannot be explained by models in these sets.

IV. Risk-adjusted Analysis: Consistency and Identification

Note that Problem 1 is generically non convex, even in the case where the feedback term is a linear operator, due to bilinear terms involving the product of ν and H. Motivated by the earlier work in [32] and [22], in the sequel we will pursue a risk–adjusted convex relaxation to (approximately) check consistency between the *a priori* assumptions and the experimental data. To this effect, begin by noting that $\mathcal{T}(\mathbf{y}, n_m, \mathcal{N}) \neq \emptyset$ if and only if there exists a pair of signals $(\mathbf{r}, \mathbf{q}) \in \ell^2$, $(r_i - \gamma q_i)^T q_i < 0$ such that (1) holds. Moreover, from assumptions [A1] and [A2], it follows (see Figure 1) that

$$\|q\|_2 \leq \frac{1}{(1-K_\gamma)}(\|\nu\|_2 + K\|u\|_2) \doteq K_q \tag{4}$$

The main idea of the proposed algorithm is to establish consistency and identify the linear portion of the interconnection by suitably sampling the set of admissible signals $\{\mathbf{r}, \mathbf{q}\}$. Once consistency is established, the static nonlinearity can be recovered from its graph, using for instance a least-squares fit.

Algorithm 1:

1.- Generate N_s samples, $\{q^i\}_{i=1}^{i=N_s}$, of the set $\mathcal{Q} \doteq \{q\colon \|\mathcal{P}_n(q)\|_2 \leq K_q\}$. Set i=1.

2.- Generate a sequence $\{r_k^i\}$ by sampling the convex set $\|r_k^i\| \leq \gamma\|q_k^i\|$, $k = 1, \ldots, n_m$. Use Lemma 1 to establish whether there exists a candidate plant $H \in \mathcal{S}$ such that (1) holds with $q = q^i$ and $r = r^i$ by solving the convex (in $\mathbf{h}$) problem:

$$\begin{aligned} &(\mathbf{R}\mathbf{T}_h^{n_m}\mathbf{R}^{-1})^T(\mathbf{R}\mathbf{T}_h^{n_m}\mathbf{R}^{-1}) < K^2\mathbf{I} \\ &\gamma^2\left(\mathbf{T}_{h_{11}}^{n_m}\right)^T\mathbf{T}_{h_{11}}^{n_m} < K_\gamma^2\mathbf{I} \\ &(\begin{bmatrix}\mathbf{q}\\ \mathbf{y}\end{bmatrix} - \mathbf{T}_h^{n_m}\begin{bmatrix}\mathbf{r}\\ \mathbf{u}\end{bmatrix}) \in \mathcal{N}, \end{aligned} \tag{5}$$

where $\{h_o, h_1, \ldots, h_{n-1}\}$ are the first n Markov parameters of the linear system $H(z)$ and where $\mathbf{T}_{h_{11}^n}$ is the submatrix associated with the impulse response of its $(1,1)$ block.

3.- If there exists at least one feasible $\mathbf{h}$, stop. Otherwise, set i=i+1 and if $i \leq N_s$ go to step 2.

The algorithm finishes, either by finding $\mathbf{h}$, the first n_m Markov parameters of the the linear portion of the system and one admissible pair of signals (r, q), or after N_s steps. In the latter case, we conclude that –within a given risk– the experimental data invalidates the *a priori* assumptions. On the other hand, if a feasible $\mathbf{h}$ has been found then a model $H(z)$ interpolating the data can be constructed now using the explicit formulas given for instance in [25] or [7], Chapter 2. The identification can be completed now by reconstructing the nonlinearity from its graph (r_k^i, q_k^i) using for instance the *dispersion function method* described in [15], or any other of a number of widely available algorithms for multidimensional data interpolation and curve fitting.

Remark 1: In many cases of interest, the *a priori* information includes a set of bases known to span the nonlinearity, that is: $\mathcal{NL}(q) = \mathbf{B}\Psi(q)$, where $\Psi(q) \in R^{n_\psi \times n_q}$ is a known matrix function and $\mathbf{B} \in R^{n_r \times n_\psi}$ denotes the unknown coefficients. This case can be easily handled by our algorithm by simple modifying step 2 above to sample $\mathbf{B}$ instead of $\mathbf{r}$. Similarly, any structural information available about H (e.g. the special cases of Hammerstein or Wiener systems) can be incorporated as additional constraints when solving the LMI (5).

A. Generating Samples of $\mathbf{q}$

The algorithm above leads to the following question: Which distribution should be used to sample the vector $\mathbf{q} \in Q$? Note that simply using a uniform distribution would result in most of the generated samples being close to the boundary of Q, as the dimension of $\mathbf{q}$ increases. It follows that, since the bound K_q used to define the set Q might be conservative, uniform sampling might lead to misleading results. This effect can be avoided by selecting a probability density function for $\mathbf{q}$ that leads to a desirable distribution of $\|\mathbf{q}\|_2$, for instance uniform. Such a distribution can be obtained by using the following algorithm to generate samples of $\mathbf{q}$: Using a normal distribution with zero mean and whose covariance is the identity matrix, generate a sample ζ with the same dimension of q. Generate also a sample η uniformly distributed in the interval $[0, K_q]$. Then compute the sample of $\mathbf{q}$ as follows

$$q = \frac{\eta}{\|\zeta\|_2}\zeta.$$

It is easily seen that the procedure above leads to samples with both uniformly distributed direction and magnitude.

B. Bound on the Number of Samples

Following ([33], [32]) if the number of samples is large enough the risk of missing a consistent triple (h, r, q) can be made arbitrarily small, as stated next.

Lemma 2: Let (υ, δ) be two positive constants in $(0, 1)$. If

$$N_s \geq \frac{\ln(1/\delta)}{\ln(1/(1-\upsilon))},$$

then the probability of missing a consistent triple (h, r, q) is smaller than υ, and this event occurs with probability greater than $(1-\delta)$, i.e.:

$$\mathbf{Prob}\big\{\mathbf{Prob}\{S(y) \neq \emptyset\} \leq \upsilon\big\} \geq (1-\delta).$$

Proof: Represent 5 as

$$F(\mathbf{h}, \mathbf{q}, \mathbf{r}) < 0$$

Now, given $\mathbf{q}$ and $\mathbf{r}$ let $\alpha(\mathbf{q}, \mathbf{r})$ be the solution of

$$\min_{\alpha, \mathbf{h}} \alpha$$

subject to

$$F(\mathbf{h}, \mathbf{q}, \mathbf{r}) < \alpha I.$$

Note that, given $\mathbf{q}$ and $\mathbf{r}$, the identification problem is feasible if and only if $\alpha(\mathbf{q}, \mathbf{r}) < 0$. Moreover, $\alpha(\mathbf{q}, \mathbf{r})$ is a measurable

function of $\mathbf{q}$ and $\mathbf{r}$. Hence, the feasibility problem can be reformulated as finding $\mathbf{q}^*$ and $\mathbf{r}^*$ such that

$$\alpha(\mathbf{q}^*, \mathbf{r}^*) < 0.$$

The result in the lemma above follows from application of the results in [33] to the function $\alpha(\cdot)$.

■

V. Convergence Analysis

In this section we briefly analyze the convergence properties of the proposed algorithm as the information is completed, that is $\eta, \nu \to 0$ and $n_m \to \infty$. First, note that the proposed algorithm is interpolatory, in the sense that the *true* plant $g^o \doteq \mathcal{F}_u(H^o, \mathcal{NL}^o) \in \mathcal{T}(\mathbf{y}, n_m, \mathcal{N})$. Recall that for any interpolatory algorithm $\mathcal{A}^I$, the *local* and *global* worst-case identification errors, denoted by $e_{id}(y, \mathcal{A}^I)$ and $e_{id}(\mathcal{A}^I)$:

$$e_{id}(y, \mathcal{A}^I) \doteq \sup_{g \in \mathcal{T}(y,n,\mathcal{N})} \|g^o - \mathcal{A}^I(y, n, \mathcal{N})\|$$
$$e_{id}(\mathcal{A}^I) \doteq \sup_{y \in \mathcal{Y}(n,\mathcal{N})} e_{id}(y, \mathcal{A}^I),$$

can be bounded above by the *local* and *global* diameter of information, $\mathcal{D}(y, \mathcal{I})$ and $\mathcal{D}(\mathcal{I})$:

$$\begin{aligned} e_{id}(y, \mathcal{A}^I) &\le d\left(\mathcal{T}(y, n, \mathcal{N})\right) \doteq \mathcal{D}(y, \mathcal{I}) \\ e_{id}(\mathcal{A}^I) &\le \sup_{y \in \mathcal{Y}(n,\mathcal{N})} \mathcal{D}(y, \mathcal{I}) \doteq \mathcal{D}(\mathcal{I}) \end{aligned} \tag{6}$$

where $\mathcal{Y}(n, \mathcal{N})$ stands for the set of all possible experiments compatible with the *a priori* assumptions:

$$\begin{aligned} \mathcal{Y}(n, \mathcal{N}) \doteq \quad & \left\{ y\colon \begin{bmatrix} \mathbf{q} \\ \mathbf{y} \end{bmatrix} = H * \begin{pmatrix} \mathbf{r} \\ \mathbf{u} \end{pmatrix} + \begin{bmatrix} \nu_k \\ \eta_k \end{bmatrix} \right. \\ & \text{for some } H \in \mathcal{S} \text{ and some sequences} \\ & (\nu, \eta) \in \mathcal{N},\ (\mathbf{q}^k, \mathbf{r}^k)\ [\mathbf{r}^{k_i} - \gamma \mathbf{q}^{k_i}]^T \mathbf{q}^{k_i} < 0 \left. \right\} \end{aligned}$$

and the norm of interest is $\|\cdot\|_2$ (see [34], Chapter 4).

Finally, let $\mathcal{T}^*(y)$ denote the consistency set in case of complete and uncorrupted experimental information, i.e.

$$\begin{aligned} \mathcal{T}^*(y) \doteq \quad & \left\{ h \in \mathcal{S}\colon \begin{bmatrix} \mathbf{q} \\ \mathbf{y} \end{bmatrix} = H * \begin{pmatrix} \mathbf{r} \\ \mathbf{u} \end{pmatrix} \right. \\ & \text{for some sequences } (\mathbf{q}_k, \mathbf{r}_k) \\ & [\mathbf{r}^{k_i} - \gamma \mathbf{q}^{k_i}]^T \mathbf{q}^{k_i} < 0 \left. \right\} \end{aligned}$$

and define:

$$e^*(y) \doteq \mathrm{d}(\overline{\mathcal{T}^*(y)}).$$

As we show next, $e^*(y)$ can be seen as an *intrinsic local* worst-case error, in the sense that this is the best that can be achieved by any interpolatory algorithm as the local information is completed.

Theorem 1: Assume that $\mathcal{N}_\eta \doteq \mathcal{B}\ell_2(\epsilon)$ and $\mathcal{N}_\nu \doteq \mathcal{B}\ell_2(\epsilon)$. Then as the information is completed:

$$\lim_{n \to \infty, \epsilon \to 0} d\left[\mathcal{T}(y, n, \epsilon)\right] = e^*(y).$$

Proof: The proof proceeds by showing that the sequence of consistency sets indexed by (n_k, ϵ_k), $\{\overline{\mathcal{T}(y, n_k, \epsilon_k)}\}$, converges to $\overline{\mathcal{T}^*(y)}$. In the sequel, we assume for the sake of notational simplicity that all the signals involved are scalar, but the proof generalizes trivially to the vector case. Begin by noting that if $\{n_k \uparrow, \epsilon_k \downarrow\}$, then the sequence of corresponding consistency sets satisfies:

$$\mathcal{T}(y, n_{k+1}, \epsilon_{k+1}) \subseteq \mathcal{T}(y, n_k, \epsilon_{k+1}) \subseteq \mathcal{T}(y, n_k, \epsilon_k) \tag{7}$$

Therefore its limit $\lim_{k\to\infty} \mathcal{T}(y, n_k, \epsilon_k)$ exists and equals $\cap_{k>0} \overline{\mathcal{T}(y, n_k, \epsilon_k)}$ ([1], page 19).

If the identification problem is well posed, i.e. $\mathcal{T}^*(y) \neq \emptyset$, and given that $h \in \mathcal{T}^*(y) \Leftrightarrow h \in \mathcal{T}(y, n_k, 0) \forall k$, then:

$$\cap_{k>0} \overline{\mathcal{T}(y, n_k, \epsilon_k)} \supseteq \cap_{k>0} \overline{\mathcal{T}(y, n_k, 0)} = \overline{\mathcal{T}^*(y)}.$$

where we have used the fact that $0 \in \mathcal{N}$. In order to show the equality, assume that

$$\exists \text{ some } \tilde{H} \in \cap_{k>0} \overline{\mathcal{T}(y, n_k, \epsilon_k)}, \text{ but } \tilde{H} \notin \overline{\mathcal{T}^*(y)}. \tag{8}$$

Since $\tilde{H} \in \cap_{k>0} \overline{\mathcal{T}(y, n_k, \epsilon_k)}$, it follows that for each k there exists at least one pair of signals (r^k, q^k) and some admissible noise sequences (ν^k, η^k) such that $|r_i^k| \le \gamma |q_k^i|$, $i = 1, \ldots, n_k$ and

$$\begin{bmatrix} q_j^k \\ y_j^k \end{bmatrix} = \left(\tilde{H} \begin{bmatrix} r^k \\ u \end{bmatrix} \right)_j + \begin{bmatrix} \nu_j^k \\ \eta_j^k \end{bmatrix} \quad j = 0, 1, \ldots, n_k$$

Since $\begin{bmatrix} q^k & r^k \end{bmatrix}^T \in \mathcal{B}\ell_2\left((1+\gamma^2)^{0.5} K_q\right)$, from Banach-Alaoglu Theorem it follows that the sequence $\begin{bmatrix} q^k & r^k \end{bmatrix}^T$ contains a convergent subsequence in the weak* topology, that is, there exist some $\tilde{q} \in \mathcal{B}\ell_2(K_q)$, $\tilde{r} \in \mathcal{B}\ell_2(\gamma K_q)$ and a subsequence $\begin{bmatrix} q^{k_i} & r^{k_i} \end{bmatrix}^T$ such that

$$\begin{aligned} &< s_1, q^{k_i} > \to < s_1, \tilde{q} > \\ &< s_2, r^{k_i} > \to < s_2, \tilde{r} > \end{aligned} \quad \text{for all } s_1, s_2 \in \ell^2 \tag{9}$$

Since $\tilde{H} \in \overline{\mathcal{T}(y, n_k, \epsilon_k)} \forall k$, (and hence in $\mathcal{H}_\infty$), it follows that $\tilde{H} * \begin{bmatrix} \tilde{r} \\ u \end{bmatrix} \in \ell^2$ and therefore, given any $\epsilon > 0$, there exists some N such that:

$$\|y - \begin{bmatrix} \tilde{H}_{21} & \tilde{H}_{22} \end{bmatrix} * \begin{bmatrix} \tilde{r} \\ u \end{bmatrix} \|_2 \le \|\mathcal{P}_N(y - \begin{bmatrix} \tilde{H}_{21} & \tilde{H}_{22} \end{bmatrix} * \begin{bmatrix} \tilde{r} \\ u \end{bmatrix})\|_2 + \frac{\epsilon}{3} \tag{10}$$

From equation (9) it follows that, for k_i large enough,

$$\begin{aligned} &\|\mathcal{P}_N(y - \begin{bmatrix} \tilde{H}_{21} & \tilde{H}_{22} \end{bmatrix} * \begin{bmatrix} \tilde{r} \\ u \end{bmatrix})\|_2 \le \\ &\|\mathcal{P}_N(y - \begin{bmatrix} \tilde{H}_{21} & \tilde{H}_{22} \end{bmatrix} * \begin{bmatrix} r^{k_i} \\ u \end{bmatrix})\|_2 + \frac{\epsilon}{3} \\ &\le \frac{2\epsilon}{3} \end{aligned} \tag{11}$$

Combining (10) and (11) we have that, for any $\epsilon > 0$, the signals $(\tilde{q}, \tilde{r})$ satisfy:

$$\| \begin{bmatrix} \tilde{q} \\ y \end{bmatrix} - \tilde{H} * \begin{bmatrix} \tilde{r} \\ u \end{bmatrix} \|_2 \le \epsilon$$

which contradicts the assumption that $\tilde{H} \notin \overline{\mathcal{T}^*(y)}$. ■

Finally, let $\mathcal{Y}^*$ denote the set of all complete and uncorrupted experiments compatible with the *a priori* assumptions $(\mathcal{S}, \mathcal{U})$:

$$\mathcal{Y}^* \doteq \left\{ y\colon y = h * u,\ \ h \in \mathcal{S}, u \in \mathcal{U} \right\},$$

and:

$$\begin{aligned} e_{id}(N,\epsilon) &\doteq \sup_{y\in\mathcal{Y}(N,\epsilon)} \mathrm{d}(\mathcal{T}(y,N,\epsilon)) \\ e^* &\doteq \sup_{y\in\mathcal{Y}^*} \mathrm{d}(\mathcal{T}^*(y)), \end{aligned} \tag{12}$$

where e^* represents an *intrinsic global* worst-case identification errors. The following Theorem extends the notion of convergence introduced above to the global case.

Theorem 2: If $\mathcal{S}$ and $\mathcal{U}$ are sequentially compact and $\mathcal{N} \doteq \mathcal{B}\ell_\infty(\epsilon)$, then:

$$\lim_{N\to\infty,\epsilon\to 0} e_{id}(N,\epsilon) = e^*.$$

Proof: Select $(N_k \uparrow, \epsilon_k \downarrow)$. From (7) combined with the fact that

$$\mathcal{Y}(N_{k+1},\epsilon_{k+1}) \subseteq \mathcal{Y}(N_k,\epsilon_k) \tag{13}$$

it follows that the sequence $\{e_{id}(N_k,\epsilon_k)\}$ is non–increasing, bounded below by 0. Thus, it contains a convergent subsequence([30]. Let $\tilde{e}$ denote its limit, and, by contradiction, assume that:

$$\tilde{e} \doteq \lim_{N_k\to\infty,\epsilon_k\to 0} \sup_{y\in\mathcal{Y}(N,\epsilon)} \mathrm{d}(\mathcal{T}(y,N,\epsilon)) > \sup_{y\in\mathcal{Y}^*} \mathrm{d}(\mathcal{T}^*(y)).$$

Define $\nu \doteq 0.5(\tilde{e} - e^*)$. From (12), it follows that, for each k, there exist an experiment y^k, and two triplets (h^k, u_h^k, η_h^k), $(g^k, u_g^k, \eta_g^k) \in \mathcal{S}\times\mathcal{U}\times\mathcal{N}$ such that:

$$\begin{aligned} \mathbf{y}^k &= \mathbf{T}_{g^k}\mathbf{u}_{g^k} + \boldsymbol{\eta}_{g^k} = \mathbf{T}_{h^k}\mathbf{u}_{h^k} + \boldsymbol{\eta}_{h^k} \\ \|g^k - h^k\| &\geq e_{id}(N_k,\epsilon_k) - \nu. \end{aligned} \tag{14}$$

Since the sets $\mathcal{S}$ and $\mathcal{U}$ are sequentially compact, the sequences $\{h^k\}$, $\{g^k\}$, $\{u_h^k\}$ and $\{u_g^k\}$ all contain subsequences that converge to some limits h^o, g^o, u_h^o, u_g^o satisfying:

$$\begin{aligned} h^o * u_h^o = g^o * u_g^o &\doteq y^o, \quad \text{for some } y^o \\ \|h^o - g^o\| &\geq \tilde{e} - \nu > e^*. \end{aligned} \tag{15}$$

Since by construction $y^o \in \mathcal{Y}^*$ and $h^o, g^o \in \mathcal{T}^*(y^o)$, this contradicts the definition of e^*. ■

VI. Numerical Example

Consider a nonlinear plant of the form (1) consisting of the interconnection of the linear system

$$H(z) = \frac{\begin{bmatrix} 0.2z - 0.1 & 1 \\ 0.3 & z + 0.5 \end{bmatrix}}{z^2 - 0.05z - 0.765}$$

and the following static nonlinear feedback term

$$r = \begin{cases} 0.5q^{\frac{2}{3}} & q \geq 1 \\ 0.5q^{\frac{3}{2}} & 0 \leq q \leq 1 \end{cases}$$

The *a priori* information available for identification purposes is the following: The linear part of the plant belongs to $\mathcal{BH}_{\infty,\rho}(K)$ with $\rho = 1.1$ and $K = 8.20$. The nonlinear static feedback is a sector nonlinearity with $\gamma = 0.5$. Finally, the bounds on the measurement noise and the noise at the input of the nonlinearity are $\epsilon_\eta = 0.24$ and $\epsilon_\nu = 0.16$ respectively. These correspond to a noise level of about 10% of the real values of the signals $\mathbf{s}$ and ψ. The *a posteriori* information consists of the first $n = 15$ samples y of the impulse response of the plant.

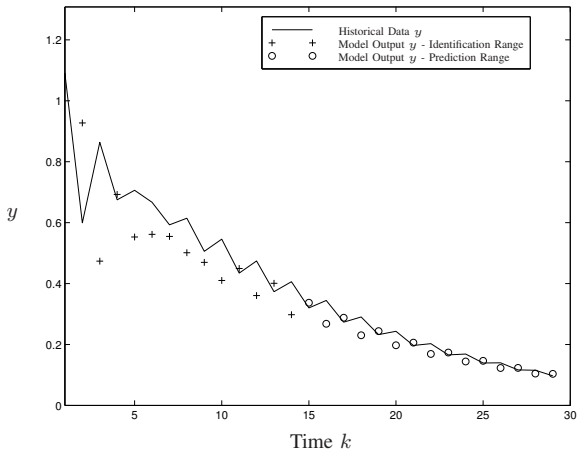

Fig. 2. Simulation Results

Using the proposed algorithm leads and selecting the central interpolant for the linear portion of the plant leads to the following model:

$$H_{id}(z) = \frac{\begin{bmatrix} 0.77 & -1.34z + 0.21 \\ -0.05 & -6.12z + 5.78 \end{bmatrix}}{z - 0.80}$$

The corresponding nonlinear feedback was estimated from the sampled values of $\mathbf{q}$ and $\mathbf{r}$ by using a one dimensional interpolation algorithm. Figure 2 shows that the resulting interconnection can indeed predict the output of the plant (beyond the identification horizon) with an ℓ_2 error no larger than 0.60. It is also worth noting that the plant obtained is not exactly the one which generated the data since, without additional assumptions, there are many plants which are compatible with the available information. However, this example clearly shows that the algorithm presented is able to provide a plant which closely matches the behavior of the real plant.

VII. Conclusion

In this paper we propose an algorithm for deterministic set membership identification of a class of nonlinear system that contains, as special cases, both Hammerstein and Wiener systems, using time–domain data. As shown in the paper, in principle this formulation leads to a non-convex, computationally hard to solve optimization problem. However, by pursuing a risk–adjusted approach, the problem can be relaxed to a convex optimization, at the price of an arbitrarily small probability of mis-identifying the plant. In addition, the convergence results presented in the paper point out to an intrinsic limitation of any interpolatory algorithm when dealing with the interconnection structures considered here: as the information is completed, the consistency set does not, in general, reduce to a singleton. Intuitively, this is due to the fact that the same input/output data can be generated

by more than one admissible interconnection. Thus, in the absence of additional a-priori information, these systems are indistinguishable.

Efforts are currently underway to generalize the techniques proposed here to cases where both the linear dynamics and the nonlinearity are slowly time varying.

REFERENCES

[1] J. P. Aubin and H. Frankowska, Set-Valued Analysis, Systems & Control, Foundations & Applications, Vol. 2, Birkhäuser Boston, 1990.

[2] E.W. Bai, "A blind approach to the Hammerstein-Wiener model identification", *Automatica*, vol. 38, 2002, pp. 967-979.

[3] E.W. Bai "An Optimal Two-Stage Identification Algorithm for Hammerstein-Weiner Nonlinear Systems", *Automatica*, vol. 34, 1998, pp. 333-338.

[4] E.W. Bai, "Frequency Domain Identification of Wiener Models", *Automatica*, vol. 39, 2003, pp 1521-1530.

[5] S.A. Billings, S.Y. Fakhouri, "Identification of a Class of Non-linear Systems Using Correlation Analysis," *Proc. IEE,* v.125, 1978, pp.691–697.

[6] C. Chen, S. Fassois, "Maximum likelihood identification of stochastic Wiener-Hammerstein-type nonlinear systems," *Mech. Syst. Sig. Proc.,* v.6, n.2, 1992, pp.135–153.

[7] J. Chen and G. Gu. *Control–oriented System Identification : an $\mathcal{H}_\infty$ Approach*, John Wiley & Sons, New York, 2000.

[8] M. Claassen, E. Wemhoff, A. Packard and K. Poolla, "On the identifiability of nonlinear maps in a general interconnected system," *39th IEEE Conference on Decision and Control,* 2000, pp. 1248–1253.

[9] L. E. Ghaoui and G. Calafiore, "Worst-case state prediction under structured uncertainty," *1999 American Controls Conference,* San Diego, CA, 1999.

[10] G. Giunta, G. Jacoviti and A. Neri, "Bandpass Nonlinear System Identification by Higher Cross Correlation", IEEE *Transactions on Signal Processing*, vol. 39, 1991, 2092-2095.

[11] W. Greblicki, "Non-parametric orthogonal series identification of Hammerstein systems," *Int. J. Syst. Sci.,* v.20, n.12, 1989, pp.2355-2367.

[12] W. Greblicki, "Nonparametric Approach to Wiener System Identification", IEEE *Transactions on Circuits and Systems*, vol. 44, 1997, pp 538-545.

[13] B.R.J. Haverkamp, C.T. Chou, and U. Verhaegen, "Subspace Identification of Continuous-time Wiener Models", IEEE *Conference of Decision and Control*, vol. 2, 1998, pp 1846-1847.

[14] K. Hsu, M. Claassen, C. Novara, P. Khargonekar, M. Milanese, and K. Poolla, "Nonparametric identifi cation of static nonlinearities in a general interconnected system," *16th IFAC World Congress,* Prague, Czech Republic, 2005.

[15] K. Hsu, T. Vincent, K. Poolla, "Robust Structured Nonlinear System Identification," *45th Conference on Decision and Control,* San Diego, CA, 2006, pp.2518–2522.

[16] K. Hsu, T. Vincent, C. Novara, M. Milanese, and K. Poolla, "Identification of nonlinear maps in interconnected systems," *44th Conference on Descision and Control,* Seville, Spain, 2005.

[17] T.A. Johansen, "Identification of Non-linear Systems using Empirical Data and Prior Knowledge-An Optimization Approach," *Automatica,* v.32, n.3, 1997, pp.337–356.

[18] R. Kosut, M. Lau, and S. Boyd, "Identification of systems with parametric and nonparametric uncertainty," *1990 American Controls Conference,*, Sand Diego, CA, 1990.

[19] J. Krause and P. Khargonekar, "Parameter identification in the presence of nonparametric dynamic uncertainty," *Automatica,* vol. 26, pp. 113 124, 1990.

[20] A. Krzyzak, "Identification of discrete Hammerstein systems by the Fourier series regression estimate," *Int. J. Syst. Sca.,* v.20, n.9, 1989, pp.1729–1744.

[21] L. Ljung, System Identification: Theory for the User, Prentice-Hall Information and System Sciences Series, 1999.

[22] M. C. Mazzaro and M. Sznaier, "A Set-Membership Approach to Blind Identification," *43^{rd} IEEE Confonference on Decision and Control,* Paradise Island, Bahamas, Dec. 2004.

[23] M. Milanese and A. Vicino, "Optimal estimation theory for dynamic systems with set membership uncertaint: An overview," *Automatica,* pp. 9971009, 1991.

[24] K.S. Narendra, P.G. Gallman, "An Iterative Method for the Identification of Nonlinear Systems Using the Hammerstein Model," *IEEE Transactions on Automatic Controntrol,* v.11, n.7, 1966, pp.546–550.

[25] P. A. Parrilo, M. Sznaier, R. S. Sánchez Peña and T. Inanc, "Mixed time/frequency domain based robust identification," *Automatica,* Vol. 34, No. 11, pp. 1375–1389, 1998.

[26] K. Poolla, P. Khargonekar, A. Tikku, and J. Krause, "A time-domain approach to model validation," *IEEE Transactions on Automatic Control,* vol. 39, pp. 951-959, 1994.

[27] R. Raich, G.T. Zhou and M Viberg, "Subspace Based Approaches for Wiener System Identification", IEEE *Transactions on Automatic Control*, vol. 50, 2005, pp 1629-1634.

[28] J. Sjoberg, Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P.Y. Glorennec, H. Hjalmarsson, and A. Juditsdy, "Nonlinear Black-box Modeling in System Identification: a Unified Overview," *Automatica,* v.31, n.12, 1995, pp.1691–1724.

[29] R. Smith and J. Doyle, "Towards a methodology for robust parameter identification," *1990 American Controls Conferece,* San Diego, CA, 1990.

[30] D. L. Stancl and M. L. Stancl, Real Analysis with Point-Set Topology, Monographs and Textbooks in Pure and Applied Mathematics, Vol. 113, Marcel Dekker, Inc., 1987.

[31] P. Stoica, "On the Convergence of an Iterative Algorithm Used for Hammerstein System Identification," *IEEE Transactions on Automatic Control,* v.26, n.4, 1981, pp.967–969.

[32] M. Sznaier, C. M. Lagoa and M.C. Mazzaro, " An Algorithm for Sampling Balls in $\mathcal{H}_\infty$ with Applications to Risk–Adjusted Performance Analysis and Model (In)Validation," *IEEE Transactions on Automatic Control,* March 2005, pp. 410–416.

[33] R. Tempo, E. W. Bai and F. Dabbene (1996). "Probabilistic Robustness Analysis: Explicit Bounds for the Minimum Number of Sampling Points", *IEEE Conference on Decision and Control*, pp. 3418–3423, Kobe, Japan.

[34] J. F. Traub, G. W. Wasilkowski and H. Wozniakowski, Information-Based Complexity, Academic Press, Inc., 1988.

[35] G. Vandersteen and J. Schoukens, "Measurement and Identification of Nonlinear Systems consisting out of Linear Dynamic Blocks and One Static Nonlinearity," *IEEE Instrumentation and Measurement Technology Conference,* v.2, 1997, pp.853-858.

[36] E. Wemhoff, A. Packard, and K. Poolla, "On the identifi cation of nonlinear maps in a general interconnected system," *1999 American Controls Conference,* San Diego, CA, 1999.

[37] D. Westwick and M. Verhaegen, "Identifying MIMO wiener Systems using Subspace Model Identification Methods", *Signal Proces*, vol. 52, 1996, pp 235-258.

[38] T. Wigren, "Convergence Analysis of Recursive Identification Algorithms based on the Nonlinear Wiener Model", IEEE *Transactions on Automatic Control*, vol. 39, 1994, pp 2191-2206.