15th Triennial World Congress, Barcelona, Spain

SET MEMBERSHIP ESTIMATION OF NONLINEAR REGRESSIONS

Mario Milanese and Carlo Novara

Dipartimento di Automatica e Informatica
Politecnico di Torino
Email: milanese@polito.it, novara@polito.it

Abstract: In this paper we propose a method, based on a Set Membership approach, for the estimation of nonlinear regressions models. At the contrary of most of the existing identification approaches, the method presented in this paper does not need any assumption about the functional form of the model to be identified, but uses only some prior information on its regularity and on the size of noise corrupting the measurements. The aim is to evaluate not only a nominal model but a model set, describing the inherent uncertainty of the regression function coming from finite and noise corrupted data. This is obtained by computing the optimal bounds on the regression function , i.e. its tightest lower and upper bounds compatible with measured data and with the given assumptions on the regression function and on noise. Moreover, necessary and sufficient conditions are given for validating the prior assumptions. The effectiveness of the method is tested on a water heater identification problem, where the obtained models are compared in simulation with other nonlinear models obtained by neural networks, Just In Time and Fuzzy approaches.

Keywords: Nonlinear Identification, Uncertainty, Set Membership Identification.

1. INTRODUCTION

Consider a nonlinear dynamical systems of the form:

$$y_{t+1} = f^o\left(\varphi_t\right)$$

where φ_t is a regression vector consisting of lagged input and output u and y:

$$\varphi_t = \left[y_t \, y_{t-1} \ldots y_{t-\tau_y} u_t \, u_{t-1} \ldots u_{t-\tau_u}\right]^T$$

where t, τ_y, τ_u are positive integers.

Consider that the function f^o is not known, but a set of output noise corrupted measurements are available:

$$z_t = y_t + e_t, \; t = 1, .., N$$

where e_t is measurement noise.

The problem of identifying nonlinear regression models, i.e. finding from measured data a function that approximates f^o, has been widely studied in the literature, (see e.g. (Sjöberg *et al.* 1995, Narenda and Mukhopadhyay 1997, Muller *et al.* 1999) and the references therein). Most of the existing identification methods need the choice of a functional form of regression function and this choice is usually the result of heuristic searches. These searches may be quite time consuming, and lead only to approximate model structures, whose errors may be responsible of bad performances in the intended uses of the model, e.g. prediction, control design, etc. Moreover, statistical assumptions on noise are typically taken such as stationarity, ergodicity, uncorrelation, type of distribution, etc., whose validity may be difficult to be reliably tested in many applications.

In this paper, a method is presented which requires only information on the regularity of the regression function and on the size of noise. Following the set membership identification philosophy (Milanese *et al.* 1996, Garulli *et al.* 1999), we investigate the problem of finding not a single model but a set of models, described by (possibly tight) upper and lower bounds of f^o:

$$\underline{f}(\varphi) \leq f^o(\varphi) \leq \overline{f}(\varphi), \forall \varphi$$

The problem investigated here can indeed be viewed as a multidimensional interpolation problem from noise corrupted data, which has been

[1] This research was supported in part by funds of Ministero dell'Università e della Ricerca Scientifica e Tecnologica under the project "Robustness techniques for control of uncertain systems".

mainly studied in the case of scalar real or complex function, see e.g. (Gaffney and Powell 1976, Golomb 1977) and many references in (Traub and Woźniakowski 1980)).

Model set described by such bounds can be used for robust prediction, see e.g. (Novara and Milanese 2001), and robust control design, see e.g. (Qu 1998). Robustness had become in past two decades a central issue of system and control theory. This need is motivated by the fact that any model may be only an approximation of the system to be analyzed or controlled. Then the system is not supposed exactly described by the model, but only belonging to a set of models obtained by perturbations of the model, whose size measures the model uncertainty. Such a set, often indicated as *uncertainty model set* or model set for short, has to be identified from available information and data in a suitable way to be used for analysis and design purposes. There is now a large literature on such topics. Most of the literature on model set identification, see e.g. (Milanese and Vicino 1991, Ninness and Goodwin 1995, Milanese *et al.* 1996, Garulli *et al.* 1999) and the references therein, is related to linear systems, while very few papers consider the model set identification of nonlinear systems, see e.g. (Smith and Dullerud 1999, Alessandri *et al.* 1999).

In this paper we show how to derive a nonlinear model set described by upper and lower bounds of the unknown function f^o, on the base of available noise corrupted measurements and on some assumptions on function regularity and on noise bounds. An example, related to the model identification for a heater, is presented to show the effectiveness of the proposed approach. Applications to nonlinear time series prediction can be found in (Novara and Milanese 2001, Novara 2002).

2. PROBLEM FORMULATION

Consider an unknown function:

$$f^o(\varphi) : \Phi \subseteq \Re^n \to \Re$$

and suppose that a set of noise corrupted values of f^o is given:

$$Z_p \doteq \{z_k = f^o\left(\varphi_k + e'_k\right) + e_k, k = 1, 2, ..., N\}$$

evaluated at the set of points:

$$\Phi_p \doteq \{\varphi_k \in \Phi, k = 1, 2, ..., N\}$$

In the paper, we investigate the following problem:

Problem:

Find lower and upper bounds $\underline{f}$ and $\overline{f}$ of f^o:

$$\underline{f}(\varphi) \leq f^o(\varphi) \leq \overline{f}(\varphi), \forall\varphi \in \Phi$$

■

It is clear that if no information is available on function f^o regularity and on the size of noises e'_{ki} and e_k, no finite bounds can be derived. In this paper, the following assumptions are considered:

Assumptions on $f^o(\varphi)$:

$f^o \in K_2^\nu \doteq \{f \in C^1, \|\nabla f\left(\varphi\right)\|_2^\nu \leq \gamma, \forall\varphi \in \Phi\}$

Assumptions on noises:

$$\begin{array}{l} |e'_{ki}| \leq \varepsilon'_{ki},\ i = 1, 2, ..., n,\ k = 1, 2, ..., N \\ |e_k| \leq \varepsilon_k,\ k = 1, 2, ..., N \end{array}$$

■

Here, $||\cdot||_2^\nu$ denotes the weighted euclidean norm given by:

$$\|x\|_2^\nu \doteq \sqrt{\sum_{i=1}^n \nu_i x_i^2},\ \nu_i > 0$$

A weighted norm is used in order to adapt to the properties of data, by properly choosing weights ν_i. Such scaling is very important when the gradient components have quite different magnitudes. A suitable choice for the weights ν_i can be made by deriving (e.g. from a neural approximation of f^o or directly from data) some estimates of the quantities $\mu_i = \max_{\varphi\in\Phi}\left|\frac{\partial f^o(\varphi)}{\partial\varphi_i}\right|,\ i = 1, 2, ..., n.$ Here, φ_i denotes the i-th component of vector $\varphi = [\varphi_1, \varphi_2, ..., \varphi_n]$. These estimates support the evidence that:

$f^o \in K_\infty^\mu \doteq \{f \in C^1, \|\nabla f\left(\varphi\right)\|_\infty^\mu \leq \gamma, \forall\varphi \in \Phi\}$

where $||\cdot||_\infty^\mu$ denotes the weighted ℓ_∞ norm given by:

$$\|x\|_\infty^\mu \doteq \max_{i=1,\ldots,n} |x_i|\,\mu_i^{-1},\ \mu_i > 0$$

Then, it is desired to approximate the set K_∞^μ with a set K_2^ν by suitably choosing ν. Inner, e.g. $K_2^\nu \subseteq K_\infty^\mu$, or outer, e.g. $K_2^\nu \supseteq K_\infty^\mu$, approximations can be looked for. This is equivalent to look, in the n-dimensional gradient space, for inner or outer approximations of the weighted ℓ_∞ ball $B_\infty^\mu \doteq \{x \in R^n : \|x\|_\infty^\mu \leq 1\}$ with a weighted ℓ_2 ball $B_2^\nu \doteq \{x \in R^n : \|x\|_2^\nu \leq 1\}$. By taking the ratio of the volumes of the two sets as measure of approximation goodness, maximal volume inner approximation and minimal volume outer approximation are optimal. The following lemma shows how these optimal solutions can be obtained.

Lemma 1.

i) The optimal (minimal volume) outer approximation is $B_2^{\overline{\nu}}$, where:

$$\overline{\nu}_i = (n\mu_i^2)^{-1},\ i = 1, ..., n$$

ii) The optimal (maximal volume) inner approximation is $B_2^{\underline{\nu}}$, where:

$$\underline{\nu}_i = (\mu_i^2)^{-1}\ i = 1, ..., n$$

Proof.

i) The sets B_∞^μ and B_2^ν are axis aligned box and ellipsoid, respectively, centered in the origin. The minimal volume ellipsoid containing a box have to be tangent to the vertices of the box. Such tangent ellipsoids are such that $\sqrt{\sum_{i=1}^n \nu_i\mu_i^2} = 1,\ \nu_i > 0$. Recalling that the volume of an ellipsoid is proportional to the product of the lengths of his principal axes, the minimal volume outer approximation $B_2^{\overline{\nu}}$ can be obtained by solving the problem:

$$\max_{\substack{\nu_i > 0 \\ \sqrt{\sum_{i=1}^{n} \nu_i \mu_i^2} = 1}} \prod_{i=1}^{n} \nu_i \tag{1}$$

By using a Lagrangian technique, it is immediate to verify that $\overline{\nu}_i$ is the unique solution of the necessary and sufficient conditions for being a solution of (1).
ii) The proof is trivial. ■

Once the weights ν_i have been chosen, in order to simplify the notation, it is convenient to take the following linear transformation $T(\varphi) \doteq [T(\varphi_1) \; T(\varphi_2) \; \ldots \; T(\varphi_n)]^T : \Phi \to W$:

$$w_i = T(\varphi_i) \doteq \frac{1}{\sqrt{\nu_i}} \varphi_i, \; i = 1, 2, ..., n$$

The evaluation set Φ_p is also transformed in the set W_p defined by:

$$W_p \doteq \{w_k = T(\varphi_k), k = 1, 2, ..., N\}$$

In the following of the paper, the notation $f^o(w)$ will be used for the transformed function instead of $f^o\left[T^{-1}(w)\right]$.
In the space of transformed variables w, the prior assumptions on f^o and on noises can be written in the following simpler form, involving an unweighted ℓ_2 bound on ∇f^o:

Assumptions on $f^o(w)$:

$f^o \in K_2 \doteq \{f \in C^1(W), \|\nabla f(w)\| \leq \gamma, \forall w \in W\}$

Assumptions on noises:

$$|e_{ki}^w| \leq \delta'_{ki}, \; i = 1, 2, ..., n, \; k = 1, 2, ..., N$$
$$|e_k| \leq \varepsilon_k, \; k = 1, 2, ..., N$$

■

Here $\|\cdot\| = \|\cdot\|_2^{\nu}$ with $\nu = (1, 1, ..., 1)$, $e_{ki}^w \doteq T(e'_{ki})$ and $\delta'_{ki} \doteq T(\varepsilon'_{ki})$.
A key role in set membership identification is played by the *Feasible Systems Set*, often indicated as "unfalsified functions set", i.e. the set of all functions consistent with prior information and measured data.

Definition 1. Feasible Systems Set

$$FSS \doteq \{f \in K_2 : z_k = f(w_k + e_k^w) + e_k, \; (e_k^w, e_k) \in B^e, \forall k\}$$

where: $B^e \doteq \{(e_k^w, e_k) : |e_{ki}^w| \leq \delta'_{ki}, |e_k| \leq \varepsilon_k, i = 1, 2, ...n, k = 1, 2, ..., N\}$. ■

The Feasible Systems Set FSS summarizes all the available information (measured data and prior information on f^o and noise e and e'). If prior assumptions are "true", then $f^o \in FSS$, an important property in view of subsequent use for prediction or control. As required in any identification theory, the problem of checking the validity of prior assumptions arises. Indeed, the only thing that can be actually done is to check if prior assumptions are invalidated by data, evaluating if no unfalsified system exists, i.e. if FSS is empty. However, it is usual to introduce the concept of prior assumption validation as follows.

Definition 2. Validation of prior assumptions
Prior assumptions are considered validated if:

$$FSS \neq \emptyset$$

■

The tightest possible bounds, based on such assumptions and on the information provided by available measurements are:

$$\underline{f}^*(w) \doteq \inf_{f \in FSS} f(w) \leq f^o(w) \leq$$
$$\leq \sup_{f \in FSS} f(w) \doteq \overline{f}^*(w), \; \forall w \in W$$

giving rise to the following definition:

Definition 3. Optimal bounds and central function

i) $\underline{f}^*(w)$ *and* $\overline{f}^*(w)$ *are called optimal lower and upper bound.*

ii) The mean of the optimal bounds:

$$f^c(w) \doteq \frac{1}{2}\left[\underline{f}^*(w) + \overline{f}^*(w)\right]$$

is called central function. ■

The two problems of giving conditions for prior assumptions validation and evaluation of optimal bounds are investigated in the next section.

3. PRIOR ASSUMPTIONS VALIDATION AND OPTIMAL BOUNDS EVALUATION

In this section, necessary and sufficient conditions for checking the assumptions validity are given.
Let us introduce the quantities:

$$f_U(w) \doteq \min_{k=1,\ldots,N} \left(\overline{h}_k + \gamma \|w - w_k\|\right)$$
$$f_L(w) \doteq \max_{k=1,\ldots,N} \left(\underline{h}_k - \gamma \|w - w_k\|\right)$$

where: $\overline{h}_k \doteq z_k + \varepsilon_k + \gamma\delta_k$, $\underline{h}_k \doteq z_k - \varepsilon_k - \gamma\delta_k$, $\delta_k \doteq \sqrt{\sum_{i=1}^{n} \delta_{ki}'^2}$.

Theorem 2.

i) $f_U(w_k) \geq \underline{h}_k, \; k = 1, 2, ..., N,$
is necessary condition for prior assumptions to be validated.
ii) $f_U(w_k) > \underline{h}_k, \; k = 1, 2, ..., N,$
is sufficient condition for prior assumptions to be validated.

Proof.

i) First we show that $f_U(w)$ and $f_L(w)$ are the solutions of the following optimization problems:

$$f_U(w) = \sup_{f \in \overline{F}} f(w)$$
$$f_L(w) = \inf_{f \in \underline{F}} f(w)$$

$$\overline{F} \doteq \{f \in K_2 : z_k \leq f(w_k) + \varepsilon_k + \gamma\delta_k, k = 1, .., N\}$$
$$\underline{F} \doteq \{f \in K_2 : z_k \geq f(w_k) - \varepsilon_k - \gamma\delta_k, k = 1, .., N\}$$

Suppose that w_k is the noise-corrupted value of $\widehat{w}_k = w_k + e_k^w$. For mean value theorem, for each $w \in W$ and each $f \in \overline{F}$ a $w' \in W$ exists such that:

$$f(w) = f(\widehat{w}_k) + \nabla f(w') \cdot (w - \widehat{w}_k)$$

Being $\nabla f(w') \cdot (w - \widehat{w}_k) \leq \|\nabla f(w')\| \|w - \widehat{w}_k\|$, $f(\widehat{w}_k) \leq z_k + \varepsilon_k$ and $\|\nabla f(w')\| \leq \gamma, \forall w' \in W$ and using the inequalities:

$$\begin{aligned} \|w - \widehat{w}_k\| &\leq \|w - w_k\| + \|\widehat{w}_k - w_k\| \leq \\ &\leq \|w - w_k\| + \delta_k \end{aligned}$$

we obtain:

$$f(w) \leq z_k + \varepsilon_k + \gamma\delta_k + \gamma \|w - w_k\|, \forall w \in W$$

This does hold for $\forall k$, then f_U is an upper bound of f. Since every value of f below f_U is allowed by mean value theorem and since f is a generic function belonging to $\overline{F}$, then $f_U(w) = \sup_{f \in \overline{F}} f(w)$. In the same way it results that $f_L(w) = \inf_{f \in \underline{F}} f(w)$.
Now suppose that a k exists for which $f_U(w_k) < \underline{h}_k$. Being $f_L(w_k) = \underline{h}_k$, by definition, it follows $f_U(w_k) < f_L(w_k)$. This implies that $\underline{F} \cap \overline{F} = \emptyset$. Since $FSS = \underline{F} \cap \overline{F}$ it results that $FSS = \emptyset$.
ii) Now suppose that $f_U(w_k) > \underline{h}_k, \forall k$. Then, by defining $k_1 = \arg\min_k \left(\overline{h}_k + \gamma \|w - w_k\|\right)$ and $k_2 = \arg\max_k \left(\underline{h}_k - \gamma \|w - w_k\|\right)$ we get the following inequalities:

$$\begin{aligned} & f_U(w) - f_L(w) = \\ & = \overline{h}_{k_1} - \underline{h}_{k_2} + \gamma \left(\|w - w_{k_1}\| + \|w - w_{k_2}\|\right) \geq \\ & \geq \overline{h}_{k_1} - \underline{h}_{k_2} + \gamma \|w_{k_1} - w_{k_2}\| \geq \\ & \geq f_U(w_{k_2}) - \underline{h}_{k_2} > 0, \ \forall w \in W \end{aligned}$$

Being $f_U(w) - f_L(w) > 0, \forall w \in W$, it is possible to find a function belonging to FSS. To this end, it suffices to find $g(w) \in K_2$ such that:

$$f_L(w) \leq g(w) \leq f_U(w), \forall w \in W \qquad (2)$$

since this implies that the condition $z_k = g(w_k + e_k^w) + e_k$, $(e_k^w, e_k) \in B^e$, $k = 1, 2, ..., N$, is fulfilled. Such function can be constructed by considering the function:

$$f_C(w) = \frac{1}{2}\left[f_U(w) + f_L(w)\right]$$

which clearly satisfies (2) but $f_C(w) \notin FSS$. In fact, $f_C(w) \in C^1$ and $\|\nabla f_C(w)\| \leq \gamma$ almost everywhere on W except on a set of points of zero measure where $\nabla f_C(w)$ is discontinuous (see (Novara and Milanese 2000)). This means that $f_C(w) \notin K_2$, only because of these gradient discontinuity points. However, being $f_L(w) < f_U(w)$, it is possible to construct a function $g(w)$ satisfying the following conditions:
- for w belonging to a neighborhood of each discontinuity point of $\nabla f_C(w)$: (1) $g(w) \in C^1$ (2) $\|\nabla g(w)\| \leq \gamma$ (3) $\underline{f}(w) \leq g(w) \leq \overline{f}(w)$;
- for w out of the neighborhoods: $g(w) = f_C(w)$;
Thus, $g(w)$ fulfills all conditions defining FSS, showing that $FSS \neq \emptyset$. ■

The previous proposition can be used for choosing values of $\varepsilon = [\varepsilon_1 \ ... \ \varepsilon_N]$, $\delta = [\delta_1 \ ... \ \delta_N]$ and γ assuring that prior assumptions are not invalidated by data. The space $(\varepsilon, \delta, \gamma)$ is divided in two regions. One region corresponds to values of ε, δ and γ falsified by data ($FSS = \emptyset$), the other corresponds to values of ε, δ and γ validated by data ($FSS \neq \emptyset$). Clearly, ε, δ and γ must be chosen in the validated parameters region. In the space $(\varepsilon, \delta, \gamma)$, the function:

$$\gamma^*(\varepsilon, \delta) \doteq \inf_{FSS \neq \emptyset} \gamma$$

individuates a surface that separates falsified values of ε, δ and γ from validated ones, The surface $\gamma^*(\varepsilon, \delta)$ can be evaluated by considering several values of ε and δ and obtaining the corresponding values of $\gamma^*(\varepsilon, \delta)$ by means of theorem 2.
The next result shows that, if prior assumption are validated, then $f_U(w)$ and $f_L(w)$ are optimal bounds.

Theorem 3.

If $FSS \neq \emptyset$, then:

$$\overline{f}^*(w) = f_U(w)$$
$$\underline{f}^*(w) = f_L(w)$$

Proof.
In the proof of Theorem 2, it has been shown that: $f_U(w) = \sup_{f \in \overline{F}} f(w)$, $f_L(w) = \inf_{f \in \underline{F}} f(w)$ with $\overline{F} \doteq \{f \in K_2 : z_k \leq f(w_k) + \varepsilon_k + \gamma\delta_k, k = 1, 2, ..., N\}$ and $\underline{F} \doteq \{f \in K_2 : z_k \geq f(w_k) - \varepsilon_k - \gamma\delta_k, k = 1, 2, ..., N\}$.On the other hand, it is immediate to verify that:

$$\sup_{f \in \overline{F}} f(w) = \sup_{f \in FSS} f(w)$$
$$\inf_{f \in \underline{F}} f(w) = \inf_{f \in FSS} f(w)$$

and the claim follows. ■

The computational complexity of evaluating the optimal bounds is $O(nN)$. In many practical applications, series of less than few thousands data are available. For such cases, the computing times are quite acceptable, e.g. less than few second on a personal computer for each step ahead prediction. For larger set of data, the computational complexity may be reduced using the approach based on neural networks approximation and hyperbolic Voronoi diagrams proposed in (Novara and Milanese 2000, Novara 2002) . ■

Up to now, a global bound on $\|\nabla f^o(w)\|$ over all W has been considered. However, according to what done in other contexts (see e.g. (Stenman *et al.* 1996, Zheng and Kimura 2001)), a local approach can be taken in order to obtain improvements in identification accuracy. A very simple approach allowing to use local assumptions on $\nabla f^o(w)$, is based on the evaluation of a function f^a approximating f^o (e.g. the central function f^c obtained by using a global bound or a neural networks function) and on the application of the method described in this paper to the residue function, defined as:

$$\Delta(w) \doteq f^o(w) - f^a(w)$$

starting from the set of values:

$$\Delta z_k = z_k - f^a(w_k), \ k = 1, 2, ..., N$$

See (Novara 2002) for more details about this local approach. ∎

4. EXAMPLE

In this example we investigate the water heater identification problem considered also in (Stenman *et al.* 1996). The system is constituted by a volume of water heated by a resistor element. The heating process can be described by an output variable, i.e. the temperature T_t of the water, and by an input variable, i.e. the voltage u_t that controls the resistor by means of a thyristor. It is expected that the main nonlinearities is due to nonlinear characteristic of the thyristor.

The data set is the one used also in (Stenman *et al.* 1996) and is given by a series of 3000 samples of T_t and u_t recorded every 3 seconds. The data set is divided into an estimation set, composed by the first 2000 data, and a validation set, composed by the remaining 1000 data. The estimation set was used to identify two Nonlinear Set Membership models and a neural networks model, the validation set was used to test the identified models in simulation and to compare the simulation performances with those presented in (Stenman *et al.* 1996), where a just in time model (JIT) and a fuzzy model are considered.

The following regression has been considered in all these methods:

$$y_{t+1} = f(\varphi_t)$$
$$\varphi_t = [T_t \ T_{t-1} \ u_{t-3} \ u_{t-4}]^T$$

where the delay of the inputs is suggested by the delay of 12 to 15 seconds between input and output that can be observed on the data set, as explained in (Stenman *et al.* 1996).

Nonlinear Set Membership model NSMG

The NSMG model is obtained by taking:

$$f(\varphi) = f_G^c(w)$$
$$w_i = T(\varphi_i) = \frac{1}{\sqrt{\nu_i}}\varphi_i, \ i = 1, 2, ..., n$$
$$\nu = [2.367 \quad 6.925 \quad 0.014 \quad 0.009]$$

where f_G^c is the central function evaluated on the base of the following assumptions:

$$\varepsilon_t = 0.5, \ \varepsilon'_{t1,2} = 0.5, \ \varepsilon'_{t3,4} = 0.01, \ \forall t$$
$$\gamma = 1.49$$

The weights ν_i have been obtained by computing approximate bounds on the partial derivatives of a neural network $\psi_o(\varphi)$ trained on the estimation set: $\mu_i = \max_{k=1,...,2000} \left| \frac{\partial \psi_o(\varphi)}{\partial \varphi_i} \right|_{\varphi=\varphi_k}$ and by choosing $\nu_i = 1/\mu_i^2$. The selected γ appears to be quite "cautious", since the minimum value of the bound on $\|\nabla f^o(\varphi)\|_2^\nu$ for which the sufficient condition of Theorem 2 is satisfied, resulted to be $\gamma^* = 1.1$.

Neural Network model NN

The NN model is obtained by taking:

$$f(\varphi) = \psi(w)$$

where the function ψ is a one hidden layer neural network (see e.g. (Hertz *et al.* 1991, Vapnik 1995)) composed by r neurons:

$$\psi(\varphi) = \sum_{i=1}^{r} \alpha_i \sigma(\beta_i w - \lambda_i) + \zeta \qquad (3)$$

Here $\alpha_i, \lambda_i, \zeta \in \Re$, $\beta_i \in \Re^n$, are parameters and $\sigma(x) = 2/(1 + e^{-2x}) - 1$ is a sigmoidal function. Several neural networks of the form (3) with different values of r (from $r = 3$ to $r = 20$) have been trained on the estimation set. A neural network with $r = 8$, showing good performances in simulation, has been chosen for the model NN.

Nonlinear Set Membership model NSML

The NSML model was obtained by means of the local approach mentioned in the previous section with $f^a(w) = \psi(w)$ and by taking:

$$f(\varphi) = \psi(w) + f_L^c(w)$$

where $f_L^c = \frac{1}{2}\left[\overline{\Delta}^*(w) + \underline{\Delta}^*(w)\right]$. The bounds $\overline{\Delta}^*(w)$ and $\underline{\Delta}^*(w)$ on the residue function $\Delta(w) \doteq f^o(w) - \psi(w)$ have been evaluated on the base of the following assumptions:

$$\varepsilon_t = 0.5, \ \varepsilon'_{t1,2} = 0.5, \ \varepsilon'_{t3,4} = 0.01, \ \forall t$$
$$\gamma = 0.26$$

The selected γ appears to be quite "cautious", since the minimum value of the bound on $\|\nabla f^o(\varphi)\|_2^\nu$ for which the sufficient condition of Theorem 2 is satisfied, resulted to be $\gamma^* = 0.15$.

Simulation performances

In table 1 the root mean squared errors (RMSE) obtained by the mentioned methods in simulation of the validation data set are reported.

Model	NSMG	NSML	NN	JIT	Fuzzy
RMSE	0.974	0.789	0.798	0.886	1.020

Table 5. Simulation errors.

5. CONCLUSIONS

In this paper, following the Set Membership Identification paradigm, the problem of estimating regressions model sets is solved by evaluating upper and lower bounds of the unknown regression function under some regularity conditions. The interest of the method lies on the fact that it does not need the choice of a functional form of the regression function as required by most of existing nonlinear identification methods. This choice, that is usually the result of heuristic searches, may be quite time consuming and leads only to approximate model structures, whose errors may be responsible of bad propagation of prediction errors. Model set described by the derived upper and lower bounds can be used for robust prediction (Novara and Milanese 2001) and for robust control design techniques for nonlinear systems (Qu 1998).

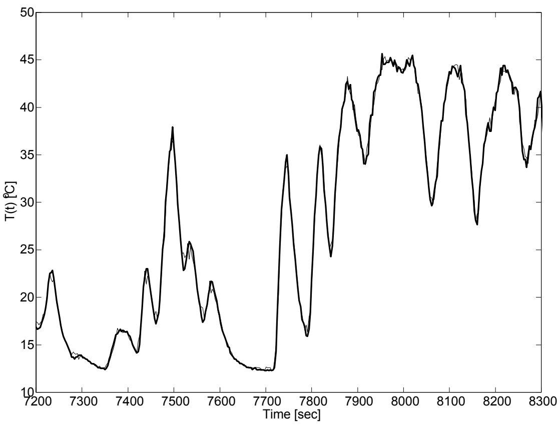

Fig. 1. Validation set: measured temperature (bold line) and NSML simulation (solid line).

Acknowledgments: The authors thank Professors Lennart Ljung and Anders Stenman for having provided the heater data set.

6. REFERENCES

Alessandri, A., M. Baglietto, T. Parisini and R. Zoppoli (1999). A neural state estimator with bounded errors for nonlinear systems. *IEEE Transaction on Automatic Control* **44**, 2028–2042.

Gaffney, P. W. and M. J. D. Powell (1976). *Optimal Interpolation*. Vol. 506. Springer Verlag. Berlin and New York.

Garulli, A., Tesi, A. and Vicino, A., Eds.) (1999). *Robustness in Identification and Control*. Vol. 245 of *Lecture Notes in Control and Information Sciences*. Springer-Verlag. Godalming, UK.

Golomb, M. (1977). *Interpolation operators as optimal recovery schemes for classes of analytic functions*. Vol. 506. Plenum Press. New York.

Hertz, J., A. Krogh and G. Palmer (1991). *Introduction to the theory of neural computation*. Addison - Wesley, Santa Fe Institute studies in the sciences of complexity. Reading (Mass.).

Milanese, M. and A. Vicino (1991). Optimal estimation theory for dynamic systems with set membership uncertainty: an overview. *Automatica* **27**(6), 997–1009.

Milanese, M., Norton, J., Piet-Lahanier, H. and Walter, É., Eds.) (1996). *Bounding Approaches to System Identification*. Plenum Press. New York.

Muller, K. R., A. Smola, G. Rätsch, B. Schölkopf, J. Kohlmorgen and V. Vapnik (1999). Predicting time series with support vector machines. In: *Advances in Kernel Methods — Support Vector Learning*. Cambridge, MA. pp. 243–254.

Narenda, K. S. and S. Mukhopadhyay (1997). Neural networks for system identification. In: *Sysid '97*. Vol. 2. pp. 763–770.

Ninness, B. and G. C. Goodwin (1995). Estimation of model quality. *Automatica* **31**(12), 1771–1797.

Novara, C. (2002). *Set Membership Identification of Nonlinear Systems*. PhD Thesis, Politecnico di Torino. Torino, Italy.

Novara, C. and M. Milanese (2000). Set membership identification of nonlinear systems. In: *Proc. of the 39th IEEE Conference on Decision and Control*. Sydney, AU. pp. 2831–2836.

Novara, C. and M. Milanese (2001). Set membership prediction of nonlinear time series. In: *Proc. of the 40th IEEE Conference on Decision and Control*. Orlando, FL.

Qu, Z. (1998). *Robust Control of Nonlinear Uncertain Systems*. Wiley series in nonlinear science.

Sjöberg, J., Q. Zhang, L. Ljung, A. Benveniste, B.Delyon, P. Glorennec, H. Hjalmarsson and A. Juditsky (1995). Nonlinear black-box modeling in system identification: a unified overview. *Automatica* **31**, 1691–1723.

Smith, R. and G. Dullerud (1999). Modeling and validation of nonlinear feedback systems. *Lecture Notes in Control and Information Sciences 245, Springer* pp. 87–101.

Stenman, A., F. Gustafsson and Ljung (1996). Just in time models for dynamical systems. In: *Proc. of the 35th IEEE Conference on Decision and Control*. Kobe, Japan. pp. 1115–1120.

Traub, J. F. and H. Woźniakowski (1980). *A General Theory of Optimal Algorithms*. Academic Press, Inc.

Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer Verlag.

Zheng, Q. and H. Kimura (2001). Just in time modeling for function prediction and its applications. *Asian Journal of Control, Vol. 3, No. 1*, pp. 35–44.