# Set Membership Identification of Nonlinear Systems

Carlo Novara and Mario Milanese

*Dipartimento di Automatica e Informatica, Politecnico di Torino*
*Corso Duca degli Abruzzi 24 - 10129 Torino - Italy*
*e-mail: cnovara@athena.polito.it, milanese@polito.it*

**Abstract**

In this paper we investigate the problem of finding upper and lower bounds of a real valued function of several variables, on the base of a set of noise corrupted values of the function evaluated at a given set of variables and on some assumptions on function regularity and on noise bounds. Several set membership linear and nonlinear identification problems can be recast into the above problem. Two solutions are proposed. The first one is quite straightforward and leads to the definition of bounds that are the tightest ones but, in high dimensional spaces, computationally expensive. The second solution, relying on approximation properties of neural networks, leads to the evaluation of somewhat more conservative bounds, whose computational complexity is significantly lower than for the optimal bounds. A numerical example, related to the identification and prediction of a Lorenz chaotic system, is presented to show the effectiveness of the proposed approach.

## 1 Introduction

Consider a nonlinear dynamical systems of the form:

$$y\,(t+1) = f\,[w(t)]\,,$$

where $w(t)$ is a regression vector consisting of lagged input and output $y$ and $u$ :

$$\begin{aligned} w(t) \quad = \quad & [y\,(t)\;y\,(t-1)\;...y\,(t-\tau_y) \\ & u\,(t)\;u\,(t-1)\,...u\,(t-\tau_u)]^T, \end{aligned}$$

where $t, \tau_y, \tau_u$ are positive integers.

Consider that the function $f$ 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 a nonlinear model, i.e. finding from measured data a function that approximates $f$, has been widely studied in the literature, (see e.g. [1, 2, 3, 4] and the references therein). In this paper, following the set membership identification philosophy [5, 6, 7], we investigate the problem of finding not a single model but a set of models, described by (possibly tight) upper and lower bounds of $f$:

$$\underline{f}(w) \leq f(w) \leq \overline{f}(w), \forall w.$$

The problem investigated here can indeed be viewed as a multidimensional interpolation problem from noise corrupted data. Interpolation has been widely studied in the numerical analysis literature, in the case of scalar real or complex function with given regularity condition on the derivative (see e.g. [8, 9, 10] and many references in [11]). In most of this literature exact data are considered, but some few papers considered corrupted data, see e.g. [12].

Model set described by such bounds can be used for robust prediction and robust control design, see e.g. [13]. 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. [5, 14, 15, 6, 16, 17, 7, 18, 19, 20] and the references therein, is related to linear systems, while very few papers consider the model set identification of nonlinear systems, see e.g. [21, 22, 23].

In this paper we show how to derive a nonlinear model set described by upper and lower bounds of the unknown function $f$, on the base of available noise corrupted measurements and on some assumptions on function regularity and on noise bounds. Two solutions are proposed. The first one is quite straightforward and leads to the definition of bounds that are the tightest ones but, in high dimensional spaces, computationally expensive. The second solution, relying on strong approximation properties of neural networks, leads to the evaluation of somewhat more conservative bounds, whose computational complexity is significantly lower than for the optimal bounds. A numerical example, related to the identification and prediction of the Lorenz model, is presented to show the effectiveness of the proposed approach.

The proofs of all Propositions are not reported here and can be found in [24].

## 2 Problem formulation

Consider an unknown function:

$$f : W \subseteq \Re^n \to \Re,$$

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

$$\begin{aligned} Z_p \quad & \doteq \quad \{z_k = f\,(w_k + e'_k) + e_k, k = 1, 2, ..., N\}, \\ e_k \quad & \in \quad \Re, e'_k \in \Re^n, \forall k, \end{aligned}$$

evaluated at the set of points

$$W_p \doteq \{w_k \in W, k = 1, 2, ..., N\}.$$

In the paper, we investigate the following problem:

**Problem**
Find upper and lower bounds $\underline{f}$ and $\overline{f}$ of $f$:

$$\underline{f}(w) \leq f(w) \leq \overline{f}(w), \forall w \in W.$$

■

It is clear that if no information is available on function $f$ regularity and on noise size, no finite bounds can be derived. In this paper, the following assumptions are considered:

**Assumptions on $f$:**

$$\begin{aligned} f \in K &\doteq \{f : W \to \Re : \\ f &\in C^1(W), \|\nabla f(w)\| \leq \gamma, \forall w \in W\}. \end{aligned}$$

**Assumptions on noise:**

$$\begin{aligned} \|e'_k\| &\leq \delta_k, k = 1, 2, ..., N, \\ |e_k| &\leq \varepsilon_k, k = 1, 2, ..., N. \end{aligned}$$

■

Here, $\nabla f$ denotes the gradient of $f$ and $\|\cdot\|$ is the Euclidean norm.

A key role in set membership identification is played by the *Feasible Solutions Set*, often indicated as "unfalsified functions set", i.e. the set of all functions consistent with prior information and measured data, defined as

$$FSS \doteq \{f \in K : z_k = f(w_k + e'_k) + e_k, (e'_k, e_k) \in B^e, \forall k\}.$$

where:

$$B^e \doteq \{(e'_k, e_k) : \|e'_k\| \leq \delta_k, |e_k| \leq \varepsilon_k, k = 1, 2, ..., N\}.$$

If prior assumptions on unknown function $f$ and on noise are "true", this set includes $f$, an important property in view of subsequent use for robustness analysis and design. As required in any identification theory, the problem of checking the "truth" of priors arises. Indeed, the only thing that can be actually done is to *check if measured data falsify the priors*. Being the $FSS$ the set of unfalsified systems, this is equivalent to check if $FSS$ is empty. Then, the $FSS$ is assumed to be non-empty. If empty, the prior assumptions on the system and the noise are invalidated by data and have to be suitably modified to give a non-empty $FSS$.

Clearly, the tightest possible bounds, based on such assumptions and on the information provided by available measurements, are of interest. Indeed, the knowledge of the $FSS$ allows to define the following *optimal bounds*:

$$\begin{aligned} \underline{f}^*(w) &\doteq \inf_{f \in FSS} f(w), \\ \overline{f}^*(w) &\doteq \sup_{f \in FSS} f(w). \end{aligned}$$

Such bounds are called optimal being the tightest possible bounds on $f$, based on the available information about this function.

## 3 Evaluation of optimal bounds

At first, we give necessary and sufficient conditions for validating the prior assumptions on $f$ and noise. As just discussed in the problem formulation section, we say that the assumptions are validated if are not falsified by data, i.e. if the measured data can be generated by some function and noise sequence satisfying the assumptions. Thus, to validate the prior assumptions is equivalent to check if $FSS \neq \emptyset$.

Let

$$\begin{aligned} \overline{h}_k &\doteq z_k + \varepsilon_k + \gamma\delta_k, \\ \underline{h}_k &\doteq z_k - \varepsilon_k - \gamma\delta_k, \end{aligned}$$

be the upper and lower bound on $f$ values in $w_k$ and define the functions:

$$f_m(w) \doteq \max_k \left(\underline{h}_k - \gamma \|w - w_k\|\right), \quad (1)$$

$$f_M(w) \doteq \min_k \left(\overline{h}_k + \gamma \|w - w_k\|\right), \quad (2)$$

The following proposition gives necessary and sufficient conditions to check if $FSS \neq \emptyset$.

**Proposition 1**
*$f_M(w_k) \geq \underline{h}_k, \forall k$, is necessary condition for $FSS \neq \emptyset$*
*$f_M(w_k) > \underline{h}_k, \forall k$, is sufficient condition for $FSS \neq \emptyset$.*
■

Note that there is a gap between necessary and sufficient conditions only in case that $f_M(w_j) = \underline{h}_j$ for some $j$. In practice, such cases appear to be quite "rare".

The next result shows that if the prior assumptions are validated, then $f_m$ and $f_M$ are the optimal lower and upper bounds $\underline{f}^*$ and $\overline{f}^*$.

**Proposition 2**
*If $FSS \neq \emptyset$, then, the functions $f_m$ and $f_M$ are:*
*(i) lower and upper bounds of unknown function $f$,*
*(ii) continuous for $\forall w \in W$,*
*(iii) optimal, i.e. $f_m(w) = \underline{f}^*(w)$, $f_M(w) = \overline{f}^*(w)$, $\forall w$.*
■

The computational complexity of evaluating $\overline{f}^*(w)$ or $\underline{f}^*(w)$ for given $w$ is

$$N_s = O(nN).$$

In fact, the computational complexity of $\|w - w_k\|$ is $O(n)$, while the computational complexity of min (or max) is $O(N)$. Unfortunately $N$, to ensure a good exploration of a region $W \subseteq \Re^n$, has often to be exponential in $n$ and so, the complexity becomes

$$N_s = O(n \exp n).$$

This exponential growth, can be avoided adding some further regularity condition on $f$ and using neural networks as approximators. In order to find two neural networks that are lower and upper bound of $f$, let us introduce the hyperbolic Voronoi diagrams.

## 4 Hyperbolic Voronoi diagrams

The hyperbolic Voronoi diagrams are a generalization of standard Voronoi diagrams (see e.g. [25, 26]). Given the points set

$$W_p \doteq \{w_k \in \Re^n, k = 1, 2, ..., N\},$$

they can be defined as follows.

Let $\Delta h_{jk} \in \Re, \gamma \in \Re, w_k, w_j \in \Re^n, j, k = 1, 2, ..., N$; Then we define:

- the $(n-1)$-dimensional hyperbola $H_{kj}$:

$$H_{kj} \doteq \{w \in \Re^n : \|w - w_k\| - \|w - w_j\| = \frac{\Delta h_{jk}}{\gamma}, k \neq j\},$$

- the $n$-dimensional regions $S_{kj}$ containing $w_k$:

$$S_{kj} \doteq \{w \in \Re^n : \|w - w_k\| - \|w - w_j\| < \frac{\Delta h_{jk}}{\gamma}, k \neq j\}$$

- the hyperbolic cell $C_k$ containing $w_k$:

$$C_k \doteq \bigcap_{j \neq k} S_{kj}.$$

The intersections between the surfaces $H_{kj}$ generate other cells of dimension $0 \leq d \leq n-1$ that we call $d$-faces. The cells $C_k$ are called $n$-faces.

We define the *hyperbolic Voronoi diagram* $V$ as the set of all $d$-faces, $0 \leq d \leq n$. Since it depends on set $W_p$, on matrix $\Delta h$ and on $\gamma$, the notation $V(W_p, \Delta h, \gamma)$ is used when necessary.

**Properties**

(i) If $\Delta h_{jk} = 0, \forall j, k$, all hyperboles $H_{kj}$ degenerate into hyperplanes and the definitions become the ones of standard Voronoi diagrams;

(ii) $C_k \neq \emptyset \Longleftrightarrow \gamma > \frac{\Delta h_{kj}}{\|w_k - w_j\|}, \forall j$;

(iii) $C_k \cap C_j = \emptyset, k \neq j$;

(iv) $\bigcup_{k=1}^{N} [C_k] = \Re^n$; where $[C_k]$ is the closure of $C_k$. ■

The following result shows the connection between functions $\underline{f}^*$ and $\overline{f}^*$ defined in (2) and (1), and hyperbolic Voronoi diagrams.

**Proposition 3**

*$\underline{f}^*$ and $\overline{f}^*$ individuate two hyperbolic Voronoi diagrams $V(W_p, \Delta\underline{h}, \gamma)$ and $V(W_p, \Delta\overline{h}, \gamma)$ defined by:*

$$\overline{S}_{jk} \doteq \{w \in \Re^n : \|w - w_j\| - \|w - w_k\| < \frac{\Delta\overline{h}_{kj}}{\gamma}, k \neq j\},$$
$$\underline{S}_{jk} \doteq \{w \in \Re^n : \|w - w_j\| - \|w - w_k\| < \frac{\Delta\underline{h}_{kj}}{\gamma}, k \neq j\},$$

*where $\Delta\overline{h}_{kj} = \overline{h}_k - \overline{h}_j$ and $\Delta\underline{h}_{kj} = \underline{h}_k - \underline{h}_j$.* ■

With the help of hyperbolic Voronoi diagram it is easy to show the following regularity conditions of the optimal bounds.

**Proposition 4**

*$\underline{f}^*$ and $\overline{f}^*$ are Lipschitz-continuous on $\{w \in W : \|w\| < \infty\}$.* ■

## 5 Nominal function and function bounds with neural networks

As previously noted, the computation of $\underline{f}^*$ and $\overline{f}^*$ may be cumbersome for large values of $n$ and $N$. In this section it is shown how, using neural network approximation properties, it is possible to derive bounds which are somewhat more conservative than the optimal ones, but require lower computational burden.

We consider the one hidden layer perceptron $\psi : W \subseteq \Re^n \to \Re$ defined as

$$\psi(w, \Gamma) \doteq \sum_{i=1}^{r} \alpha_i \sigma(\beta_i \cdot w - \lambda_i) + \zeta; \tag{3}$$

where $\Gamma \doteq \{\alpha_i, \lambda_i, \zeta \in \Re, \beta_i \in \Re^n, i = 1, 2, ..., r\}$ is a set of parameters and the activation function $\sigma : \Re \to \Re$ is a sigmoidal function like $\tanh(x)$ or $1/(1 + \exp(-x))$.

The one hidden layer perceptron (3) can approximate functions with an approximation degree that is stated by the following result (see e.g. [27, 28, 29, 30]).

**Theorem 5 (Barron, 1993)**

*If a function $f : W \subseteq \Re^n \to \Re$ satisfies the condition*

$$C_f \doteq \int_{\Re^n} \|\omega\| \left|\widetilde{f}(\omega)\right| d\omega < \infty, \tag{4}$$

*where $\|\cdot\|$ is the euclidean norm and $\widetilde{f}$ is the Fourier transform of $f$, then a set of parameters $\Gamma$ exists such that:*

$$\int [f(w) - \psi(w, \Gamma)]^2 \mu(dx) \leq \frac{D^2 C_f^2}{r};$$

*where $\mu$ is an arbitrary probability measure defined on the $n$-dimensional sphere of diameter $D$ in which $f$ is approximated*

. ■

**Remarks**

(i) Condition (4) implies that $\|\nabla f(w)\| \leq C_f, \forall w \in W$.

(ii) The approximation error, defined as $E \doteq \sqrt{\int [f(w) - \psi(w, \Gamma)]^2 \mu(dx)}$, doesn't depend on dimension $n$ of the space on which $f$ and $\psi$ are defined. This implies that calculation of $\psi$ for given $w$, with $r$ fixed, has complexity

$$N_s = O(rn).$$

■

In order to retain such interesting approximation propriety, from now it is assumed that $f$ satisfies (4).

Now consider a neural network

$$\overline{\psi}(w) \doteq \sum_{i=1}^{r} \overline{\alpha}_i \sigma\left(\overline{\beta}_i \cdot w - \overline{\lambda}_i\right) + \overline{\zeta},$$

trained on the set: $\{(w_k, z_k + c(w_k)), k = 1, 2, ..., N\}$, where the correction function $c : W \to \Re_+$ satisfies hypothesis (4) and is chosen large enough to ensure that $f(w) \leq \overline{\psi}(w), \forall w \in W$.

In the following we give a criterion to check if $c$ is large enough. This criterion needs the computation of all

0-faces of a hyperbolic Voronoi diagram suitably defined on the base of the hyperbolic Voronoi diagram $V\left(W_p, \Delta\overline{h}, \gamma\right)$ and some hyperplanes associated to the terms of (3).

Let

$$\overline{\psi}_c(w) \doteq \sum_{i=1}^{r} \overline{\alpha}_i \overline{\sigma}_i\left(\overline{\beta}_i \cdot w - \overline{\lambda}_i\right) + \overline{\zeta},$$

be the function with the same parameters of $\overline{\psi}$ but with a different activation function $\overline{\sigma}_i$ :

$$\text{if } \overline{\alpha}_i > 0, \qquad \overline{\sigma}_i(h) \doteq \begin{cases} -1, & h < -1/\overline{\alpha}_i, \\ h, & -1/\overline{\alpha}_i \leq h \leq 0, \\ \sigma(h), & h > 0; \end{cases}$$

$$\text{if } \overline{\alpha}_i < 0, \qquad \overline{\sigma}_i(h) \doteq \begin{cases} \sigma(h), & h < 0, \\ h, & 0 \leq h \leq -1/\overline{\alpha}_i, \\ 1 & h > -1/\overline{\alpha}_i. \end{cases}$$

The function $\overline{\psi}_c$ is a lower bound of $\overline{\psi}$ and is concave for $\overline{\beta}_i \cdot w - \overline{\lambda}_i < -1/\overline{\alpha}_i$, $\overline{\beta}_i \cdot w - \overline{\lambda}_i = -1/\overline{\alpha}_i$ and $\overline{\beta}_i \cdot w - \overline{\lambda}_i > -1/\overline{\alpha}_n$.

Now consider the hyperbolic Voronoi diagram $\overline{T}$ individuated
by the $d$-faces, $0 \leq d \leq n$, generated by intersection between the hyperbola $\overline{H}_{kj}$, $k, j = 1, 2, ..., N$, and the hyperplanes $\left\{w : \overline{\beta}_i \cdot w - \overline{\lambda}_i = -1/\overline{\alpha}_i, i = 1, ..., r\right\}$ and consider the set

$$\overline{Q}^0 \doteq \overline{T}^0 \cap W_p,$$

where $\overline{T}^0$ is the collection of all 0-faces of $\overline{T}$.

Defining $W_c \doteq \mathrm{conv}(W_p)$ as the convex hull of $W_p$ it is possible to check if $c$ is large enough by checking if the hypothesis of following proposition is verified. In order to simplify some definitions and notations, it is assumed $W_c \subseteq W$.

**Proposition 6**

*If* $\overline{f}^*(w) \leq \overline{\psi}_c(w), \forall w \in \overline{Q}^0$, *then:*
$f(w) \leq \overline{\psi}(w), \forall w \in W_c$.

■

In a similar way, given a neural networks

$$\underline{\psi}(w) = \sum_{i=1}^{r} \underline{\alpha}_i \sigma\left(\underline{\beta}_i \cdot w - \underline{\gamma}_i\right) + \underline{\zeta},$$

trained on the set $\{(w_k, z_k - c(w_k)), k = 1, 2, ..., N\}$, a function $\underline{\psi}_c(w) \geq \underline{\psi}(w), w \in \Re^n$, that is convex for $\underline{\beta}_i \cdot w - \underline{\gamma}_i < 1/\underline{\alpha}_i$, $\underline{\beta}_i \cdot w - \underline{\gamma}_i = 1/\underline{\alpha}_i$ and $\underline{\beta}_i \cdot w - \underline{\gamma}_i > 1/\underline{\alpha}_i$, is defined:

$$\underline{\psi}_c(w) = \sum_{i=1}^{r} \underline{\alpha}_i \underline{\sigma}_i\left(\underline{\beta}_i \cdot w - \underline{\gamma}_i\right) + \underline{\zeta},$$

where:

$$\text{if } \overline{\alpha}_i > 0, \qquad \underline{\sigma}_i(h) = \begin{cases} \sigma(h), & h < 0, \\ h, & 0 \leq h \leq 1/\underline{\alpha}_i, \\ 1, & h > 1/\underline{\alpha}_i, \end{cases}$$

$$\text{if } \overline{\alpha}_i < 0, \qquad \underline{\sigma}_i(h) \doteq \begin{cases} -1, & h < 1/\overline{\alpha}_i, \\ h, & 1/\overline{\alpha}_i \leq h \leq 0, \\ \sigma(h) & h > 0. \end{cases}$$

A set of points $\underline{Q}^0$ is constructed analogously to $\overline{Q}^0$. Then it's possible to check if $c$ is large enough to ensure $f(w) \geq \underline{\psi}(w), \forall w \in W_c$ by using the following proposition.

**Proposition 7**

*If* $\underline{f}^*(w) \geq \underline{\psi}_c(w), \forall w \in \underline{Q}^0$, *then:*
$f(w) \geq \underline{\psi}(w), \forall w \in W_c$.

■

**Remarks**

(i) Simple choices for $c(w)$ can be:

$$\begin{aligned} c(w_k) &= Cz_k, \\ c(w_k) &= C\varepsilon_k, \\ c(w_k) &= C. \end{aligned}$$

With such choices it is possible to obtain verification of proposition (6) and (7) only tuning the single quantity $C \in \Re$.

(ii) Once the optimal or neural bounds are found, a nominal function can be defined as the mean of the two bounds. If the nominal function is defined as the mean of the optimal bounds it is called *central function.*

## 6 Example

Using the Lorenz model demo of Matlab Differential Equation Editor, a time series of 600 data has been generated. The Lorenz model (see e.g. [31]) is defined by the following set of differential equations:

$$\begin{aligned} \dot{x}_1 &= -\sigma x_1 + \sigma x_2, \\ \dot{x}_2 &= -x_1 x_3 + \rho x_1 - x_2, \\ \dot{x}_3 &= x_1 x_2 - \beta x_3 + x_1, \\ y &= x_1, \end{aligned}$$

We used the parameters values $\beta = 2.6667$, $\rho = 28$, $\sigma = 10$ and a sampling time of 0.025. Note that for such parameters the system displays a chaotic behavior.

The corrupted data have been obtained as $z(t) = y(t) + e(t)$, where $e(t)$ is a i.i.d. random uniform noise such that $|e(t)| \leq 0.05 y(t), \forall t$. The data set has been divided into an estimation set, composed of the first 460 data, and a validation set, composed of the remaining ones.

A nonlinear autoregressive model (NAR) of 3-th order

$$\begin{aligned} y(t+1) &= f[w(t)], \\ w(t) &= [y(t) \;\; y(t-1) \;\; y(t-2)]^T, \end{aligned}$$

has been considered.

The following values has been assumed for noise bounds and function $f$ regularity: $\varepsilon(t) = 0.05\,|y(t)|$, $\delta(t) = 0.05\,\|w(t)\|$, $\gamma = 2.5$. Note that the assumed value of $\gamma$ is quite "cautious". In fact, the minimum value of $\gamma$ for which the sufficient condition of Proposition 1 is verified, has been estimated and is 1.8.

Then, using Proposition 2, the optimal bounds $\overline{f}^*(w)$, $\underline{f}^*(w)$ and the central function $\widehat{f}(w) = \left[\overline{f}^*(w) + \underline{f}^*(w)\right]/2$ have been identified on the estimation data set. Note that in this example, when $n = 3$

and $N = 460$, the computational complexity of evaluating the optimal bounds is quite acceptable. In fact the computing time of $\overline{f}^*(w)$ and $\underline{f}^*(w)$ for given $w$ is 0.06 sec. on a 650 Mhz PC. The central function $\widehat{f}(w)$ has been used as 1-step ahead predictor for $y(t+1)$ :

$$\begin{aligned} \widehat{y}(t+1) &= \widehat{f}[\widehat{w}(t)], \\ \widehat{w}(t) &= [z(t)\ \ z(t-1)\ \ z(t-2)]^T, \end{aligned}$$

and $\overline{f}^*(w)$ and $\underline{f}^*(w)$ provides the uncertainty range for such predictor. The results on the validation data set are reported in figure 1. Moreover, a $t_a$ steps ahead predictor for $y(t+t_a)$ has been obtained as

$$\begin{aligned} \widehat{y}(t+t_a) &= \widehat{f}[\widehat{w}(t+t_a-1)], \\ \widehat{w}(t+t_a-1) &= [\widehat{y}(t+t_a-1) \\ &\qquad \widehat{y}(t+t_a-2)\ \ \widehat{y}(t+t_a-3)]^T. \end{aligned}$$

The results on the validation data for $t_a = 3$ and $t_a = 10$ are reported in figures 2 and 3, respectively. The Root Mean Squares Error (RMSE) resulted to be 0.9847 and 1.2587 respectively.

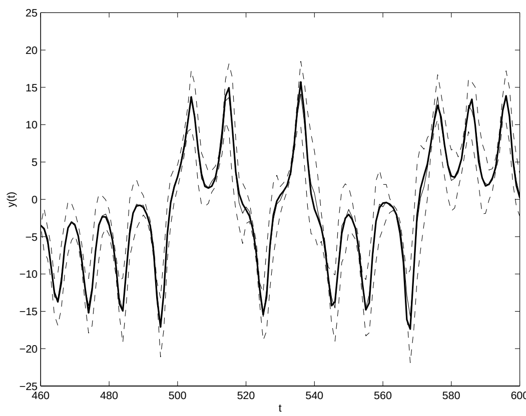


**Figure 1:** Lorenz model (bold line), 1 steps ahead prediction of central NAR model (thin line) and uncertainty bounds (dashed line).

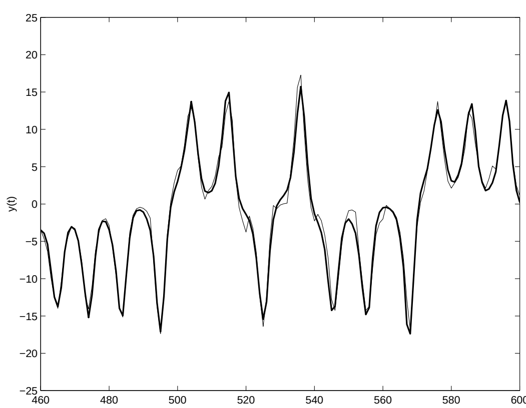


**Figure 2:** Lorenz model (bold line), 3 steps ahead prediction of central NAR model (thin line); RMSE=0.9847.

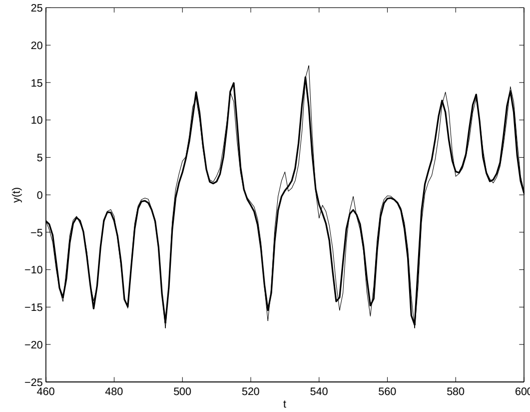


**Figure 3:** Lorenz model (bold line), 10 steps ahead prediction of central NAR model (thin line); RMSE=1.2587.

For comparison, linear AR models

$$y(t+1) = a_1 y(t) + \ldots + a_\tau y(t-\tau),$$

with $\tau = 2, 3, 4, 6, 8, 10, 12, 15, 20$, have been identified. The 4-th order model ($\tau = 3$) showed to have the best 3 and 10 steps ahead prediction, with a RMSE of 6.3049 and 8.0757, respectively. The results on the validation set are reported in figures 4 and 5, respectively.

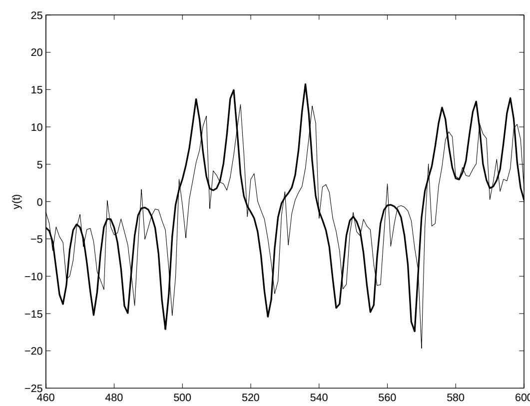


**Figure 4:** Lorenz model (bold line), 3 steps ahead prediction of 4-th order AR model (thin line); RMSE=6.3049.

## 7 Conclusions

In this paper, the problem of identifying nonlinear regression systems is considered. Following the Set Membership Identification paradigm, the problem of finding not only a model but also evaluating its accuracy is solved by evaluating upper and lower bounds of the unknown regression function under some regularity conditions. Two solutions are proposed. The first one is quite straightforward and leads to the definition of bounds that are the tightest ones but, in high dimensional spaces, computationally expensive. The second solution, relying on strong approximation properties of neural networks, leads to the evaluation of somewhat more conservative bounds, whose computational complexity is significantly lower than for the optimal bounds.

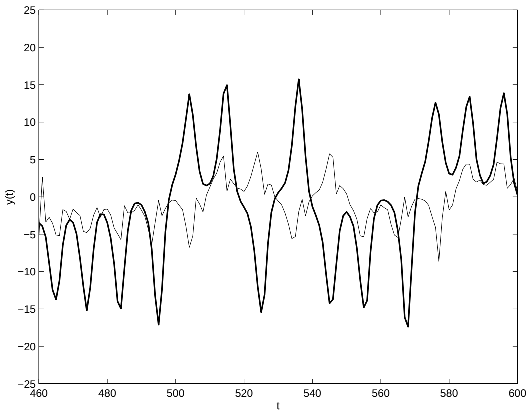


**Figure 5:** Lorenz model (bold line), 10 steps ahead prediction of 4-th order AR model (thin line); RMSE=8.0757.

Model set described by the defined upper and lower bounds can be used for robust prediction and are also in the suitable form required by the robust control design techniques for nonlinear systems described in [13].

It can be observed that a global bound on the gradient of the unknown regression function to be identified is used here. Following similar lines of reasoning of the gain scheduling (see e.g. [32]) or the Just-in-Time modeling approaches (see e.g. [33, 34]), local bounds on the gradient could be considered, possibly leading to significant improvements in terms of computational complexity and model set accuracy. It can be also observed that, in this paper, identification of nonlinear regressive systems is considered, but several other linear and nonlinear system identification problems, such as $H^\infty$ identification of linear systems, identification of the nonlinear part of Wiener and Hammerstein models can be approached in a similar way. These problems are currently under investigation.

## References

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

[2] A. Juditsky, H. Hjalmarsson, A. Benveniste, B.Delyon, L. Ljung, J. Sjöberg, and Q. Zhang, "Nonlinear black-box modeling in system identification: Mathematical foundations," *Automatica*, vol. 31, pp. 1725–1750, 1995.

[3] K. S. Narenda and S. Mukhopadhyay, "Neural networks for system identification," in *Sysid '97*, vol. 2, pp. 763–770, 1997.

[4] R. Isermann, S. Ernst, and O. Nelles, "Identification with dynamic neural networks -architectures, comparisons, applications-," in *Sysid '97*, vol. 3, pp. 997–1022, 1997.

[5] M. Milanese, R. Tempo, and A. Vicino, eds., *Robustness in Identification and Control*. New York: Plenum Press, 1989.

[6] R. S. Smith and M. Dahleh, eds., *The Modeling of Uncertainty in Control Systems*. London: Springer-Verlag, 1994.

[7] M. Milanese, J. Norton, H. Piet-Lahanier, and É. Walter, eds., *Bounding Approaches to System Identification*. New York: Plenum Press, 1996.

[8] B. D. Bojanov, "Best methods of interpolation for certain classes of differentiable functions," *Math. Notes*, vol. 17, pp. 301–309, 1975.

[9] P. W. Gaffney and M. J. D. Powell, *Optimal Interpolation*, vol. 506. Berlin and New York: Springer Verlag, 1976.

[10] M. Golomb, *Interpolation operators as optimal recovery schemes for classes of analytic functions*, vol. 506. New York: Plenum Press, 1977.

[11] J. F. Traub and H. Woźniakowski, *A General Theory of Optimal Algorithms*. Academic Press, Inc., 1980.

[12] A. G. Marchuk and K. Osipenko, "Best approximation theory of functions specified with an error at a finite number of points," *Math. Notes*, vol. 17, pp. 207–212, 1975.

[13] Z. Qu, *Robust Control of Nonlinear Uncertain Systems*. Wiley series in nonlinear science, 1998.

[14] M. Milanese and A. Vicino, "Optimal estimation theory for dynamic systems with set membership uncertainty: an overview," *Automatica*, vol. 27, no. 6, pp. 997–1009, 1991.

[15] M. Milanese and A. Vicino, "Information-based complexity and nonparametric worst-case system identification," *Journal of Complexity*, vol. 9, pp. 427–446, 1993.

[16] B. Ninness and G. C. Goodwin, "Estimation of model quality," *Automatica*, vol. 31, no. 12, pp. 1771–1797, 1995.

[17] P. M. Mäkilä, J. R. Partington, and T. K. Gustafsson, "Worst-case control-relevant identification," *Automatica*, vol. 31, no. 12, pp. 1799–1819, 1995.

[18] J. R. Partington, *Interpolation, Identification and Sampling*, vol. 17. New York: Clarendon Press - Oxford, 1997.

[19] L. Giarré, M. Milanese, and M. Taragna, "$H_\infty$ identification and model quality evaluation," *IEEE Transactions on Automatic Control*, vol. AC-42, no. 2, pp. 188–199, 1997.

[20] A. Garulli, A. Tesi, and A. Vicino, eds., *Robustness in Identification and Control*, vol. 245 of *Lecture Notes in Control and Information Sciences*. Godalming, UK: Springer-Verlag, 1999.

[21] M. Krüger, E. Wemhoff, A. Pakard, and K. Poolla, "Semi-parametric methods for system identification," *Lecture Notes in Control and Information Sciences 245, Springer*, pp. 47–61, 1999.

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

[23] A. Alessandri, M. Baglietto, T. Parisini, and R. Zoppoli, "A neural state estimator with bounded errors for nonlinear systems," *IEEE Transaction on Automatic Control*, vol. 44, pp. 2028–2042, 1999.

[24] C. Novara and M. Milanese, "Set membership identification of nonlinear systems," *Politecnico di Torino Internal Report*, 2000.

[25] H. Edelsbrunner, *Algorithms in Combinatorial Geometry*. Berlin: Springer-Verlag, 1987.

[26] F. P. Preparata and M. J. Shamos, *Computational Geometry*. New York: Springer-Verlag, 1985.

[27] A. R. Barron, "Universal approximation bounds for superposition of a sigmoidal function," *IEEE Transaction on Information Theory*, vol. 39, pp. 930–945, 1993.

[28] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer Verlag, 1995.

[29] A. Pinkus, "Approximation theory of the mlp model in neural networks," *Acta Numerica, Cambridge University Press, 1999*, pp. 143–195, 1999.

[30] K. Hornik, M. Stinchcombe, H.White, and P. Auer, "Degree of approximation results for feedforward networks approximating unknown mappings and their derivatives," *Neural Computation*, vol. 6, pp. 1262–1275, 1994.

[31] H. G. Schuster, *Deterministic chaos: an introduction*. Physik Verlag, 1984.

[32] K. J. A. strom and B. Wittenmark, *Adaptive control*. Addison Wesley, 1989.

[33] A. Stenman, F. Gustafsson, and Ljung, "Just in time models for dynamical systems," in *35th Conference on Decision and Control*, (Kobe, Japan), pp. 1115–1120, 1996.

[34] Q. Zheng and H. Kimura, "Just in time modeling for function prediction and its applications," in *3rd Asian Control Conference*, (Shanghai), 2000.