AN NIIR STRUCTURE USING HL CPWL FUNCTIONS

L. R. CASTRO†, J. L. FIGUEROA‡, and O. E. AGAMENNONI§

†*Dto. de Matemática, Univ. Nac. del Sur, B8000CPB Bahía Blanca, Argentina*
lcastro@ uns.edu.ar
‡*Dto. de Ing. Eléctrica y de Comp. - C.O.N.I.C.E.T., Univ. Nac. del Sur, B8000CPB - Bahía Blanca, Argentina*
figueroa@ uns.edu.ar
§*Dto. de Ing. Eléctrica y de Comp. - C.I.C., Univ. Nac. del Sur, B8000CPB - Bahía Blanca, Argentina*
oagamen@ uns.edu.ar

Abstract— In this paper we present a nonlinear infinite impulse response (NIIR) model structure for black-box identification of nonlinear dynamic systems. The proposed model structure allows the implementation of an identification algorithm in which the degrees of freedom of the Nonlinear Output Error (NOE) model can be easily increased or decreased during the identification process. This property is very attractive to find the appropriate NIIR model, avoiding overfitting. This is done using High Level Canonical Piecewise Linear (HL CPWL) functions with an increasing (decreasing) grid division. Therefore, the algorithm may start using a linear estimation of the model. The parameters of the HL CPWL functions are updated using a simple algorithm based on a modified steepest descent method with an independently adaptive learning rate.

Keywords— Nonlinear identification, NIIR model, PWL functions.

I. INTRODUCTION

The main problem in system identification is to find a good model structure. If it allows to go from a linear model to a nonlinear one during the system identification process, it makes this problem much harder since the set of nonlinear models is richer than the set of linear ones (Sjöberg and Ngia, 1998). If a nonlinear finite impulse (NFIR) structure is used, the model order evaluation problem may be effectively addressed by using regularization theory (Poggio and Girosi, 1990). This is due to the reduction of computational complexity when using NFIR model structures since they allow considering more parameters than needed in the identification algorithm and reducing some of them to zero through the regularization process. If a Wiener like model structure is used, an aggregation approach can be easily implemented as in the Korenberg algorithm (Korenberg and Paarmann, 1991). In the Neural Networks literature there exist growing and pruning methods to deal with the size of a Neural Network during the training process (Haykin, 1994). If NIIR model structures

are used, the problem becomes much more difficult due to the mathematical complexity and the computational cost involved in the identification process.

In this paper we present an NIIR model structure that uses High Level Canonical Piecewise Linear (HL CPWL) functions to develop a nonlinear output error (NOE) identification algorithm. The main feature of this algorithm is its simple mechanism for increasing or decreasing the model approximation capabilities, retaining the approximation achieved when moving from one grid division to another. In this way, it is possible to start the identification with a linear approximation and then increase the model degree of freedom progressively in order to reduce the mismatch up to an acceptable value. On the other hand, a reduced model may be evaluated to alleviate overfitting.

The paper is organized as follows. In Section II, we present the identification algorithm and analyze its advantages and drawbacks; in Section III we develop an example of the proposed methodology and finally, in Section IV we draw some conclusions and comments about future work. In order to be self-contained, in Appendix A we give a brief introduction to HL CPWL functions and their main properties.

II. CPWL IIR IDENTIFICATION

A. From linear to nonlinear

Let us suppose that we want to identify a system given an output vector $\mathbf{y}$ corresponding to an input $\mathbf{u}$. If $\tilde{\mathbf{y}}$ is the estimated vector, let us define

$$\mathbf{u}^k = [\mathbf{u}_1 \ \mathbf{u}_2 \ \ldots \mathbf{u}_k] \tag{1}$$
$$\tilde{\mathbf{y}}^k = [\tilde{\mathbf{y}}_1 \ \tilde{\mathbf{y}}_2 \ \ldots \tilde{\mathbf{y}}_k]. \tag{2}$$

It is well known (see (Sjöberg *et al.*, 1995), for example) that a general black-box model is given by

$$\begin{aligned} \tilde{\mathbf{y}}_k &= f\left(\mathbf{u}^k, \tilde{\mathbf{y}}^{k-1}, \theta\right) \\ &= f\left(\varphi_k, \theta\right), \end{aligned} \tag{3}$$

where $\varphi_k = \varphi\left(\mathbf{u}^k, \tilde{\mathbf{y}}^{k-1}\right)$ is the regression vector and θ is the vector of parameters associated to the function f used to approximate the system's nonlinearity. Therefore, the model is defined once f and the regression vector φ_k are chosen.

Following this idea, we propose a regression vector given by

$$\begin{aligned}
\varphi_k &= [\mathbf{u}_k \ \mathbf{u}_{k-1} \ \ldots \mathbf{u}_{k-M}, \tilde{\mathbf{y}}_{k-1} \ \ldots \tilde{\mathbf{y}}_{k-N}] \\
&= [\mathbf{u}^{k-M}, \tilde{\mathbf{y}}^{k-N}],
\end{aligned} \tag{4}$$

with M, N fixed.

Then our model is defined as follows

$$\begin{aligned}
\tilde{\mathbf{y}}_k &= f_{pwl}\left(\mathbf{u}_k \ldots, \mathbf{u}_{k-M}, \tilde{\mathbf{y}}_{k-1}, \ldots, \tilde{\mathbf{y}}_{k-N}\right) \\
&= f_{pwl}\left([\mathbf{u}^{k-M}, \tilde{\mathbf{y}}^{k-N}]\right),
\end{aligned} \tag{5}$$

where the function f_{pwl} used to approximate the nonlinearity of the model is a HL CPWL function defined, as in Eq. (19), by

$$f_{pwl}\left([\mathbf{u}^{k-M}, \tilde{\mathbf{y}}^{k-N}]\right) = \mathbf{c}\Lambda\left([\mathbf{u}^{k-M}, \tilde{\mathbf{y}}^{k-N}]\right), \tag{6}$$

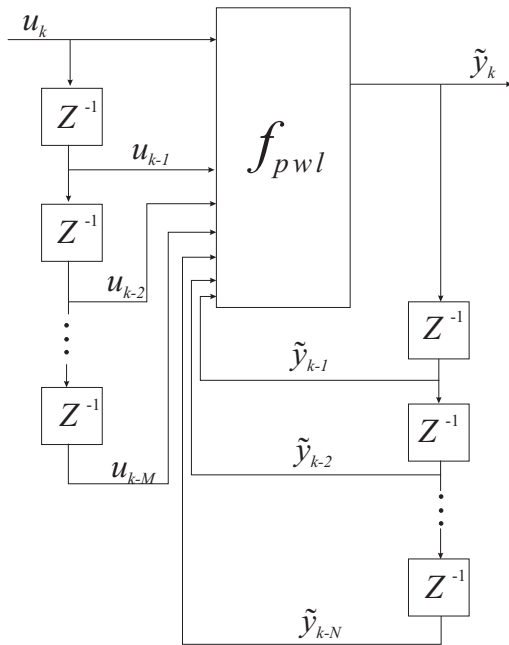and $\tilde{\mathbf{y}}_r, r = 0, \ldots, N$ are initialization values. This model is pictured in Fig. 1.

Figure 1. NIIR HL CPWL model.

The domain of the function f_{pwl} is a compact set $\mathbf{S} \subset \mathbb{R}^m, m = M + N + 1$, defined as follows

$$\mathbf{S} = \{\mathbf{x} \in \mathbb{R}^m : a_i \leq \mathbf{x}_i \leq a_i + \delta \, ndiv, \ i = 1, \ldots, m\}, \tag{7}$$

being δ the fixed grid size and $ndiv$ the number of divisions. So each interval $[a_i, a_i + \delta \, ndiv]$ of the domain $\mathbf{S}$ defined by Eq. (7) is divided into $ndiv$ number of subintervals of equal length δ. As a consequence, when the grid size δ decreases, the number of divisions $ndiv$ increases.

According to Appendix A, the set defined by Eq. (7) is partitioned into polyhedral regions using a simplicial boundary configuration. The f_{pwl} constructed using the methodology described in Appendix A is linear on each simplex and continuous on the adjacent boundaries of the simplices.

In the methodology proposed above, it is possible to start the identification process with a linear approximation to the system. Once the parameters are optimized, the number of divisions $ndiv$ may be increased in order to obtain a better piecewise linear approximation. On the other way, it is possible to go from a fine approximation to a coarser one by decreasing the value of $ndiv$. This modeling facility not only allows to obtain a better quality piecewise linear approximation but also makes it possible to prevent overfitting.

Therefore, when using HL CPWL functions as nonlinear approximators, the parameter $ndiv$ gives a natural ordering of the model since it allows to go from a simple model to a more complex one. The advantages of using this kind of models was pointed out in (Sjöberg and Ngia, 1998, Ch. 1).

In Fig. 2 (a) and (b) the idea of approximating a nonlinear quadratic function using HL CPWL functions with increasing values of $ndiv$ ($ndiv = 2, 4$) is pictured. It can be observed that as long as the value of $ndiv$ increases, the HL CPWL function approximates the nonlinear one more accurately. Also, on the XY plane it can be seen the simplices determined on the region $\mathbf{S}$ by the different number of divisions $ndiv$.
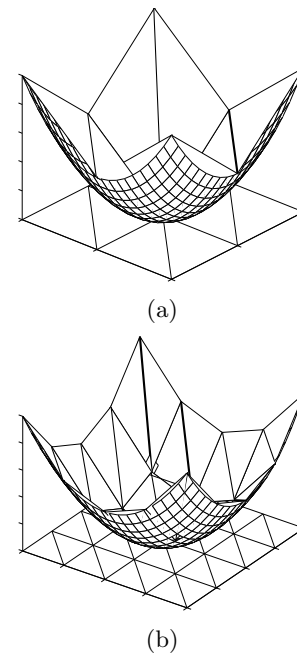

(a)

(b)

Figure 2. HL CPWL approximation for (a) $ndiv = 2$ and (b) $ndiv = 4$.

B. Identification algorithm

Let $(\mathbf{u}_k, \mathbf{y}_k)_{1 \leq k \leq L}$ be the iput/output vectors and $\mathbf{c}^{d-1,*}$ the $(ndiv + 1)^{M+N+1}$-dimensional vector of parameters for a given number of divisions $ndiv = 2^{d-1}, d \in \mathbb{N}$ (if $d = 1$ we have a linear approximation). For $ndiv = 2^d$ we find a new $(ndiv + 1)^{M+N+1}$-

dimensional vector of parameters using a least square approximation technique on the new set of vertices of the region $\mathbf{S}$ and note it $\mathbf{c}^{d,r}, r = 0$.

Now we update the vector of parameters $\mathbf{c}^{d+1,r}, r \geq 1$ using an iterative algorithm that minimizes the square error $E^r, r \geq 1$ between the system $\mathbf{y}$ and the estimate $\tilde{\mathbf{y}}$ at iteration $r, r \geq 1$. The expression of this error in the variables $\mathbf{c}^{d+1,r}$ can be written using Eq. (6), as follows

$$
\begin{aligned}
E^r &= \frac{1}{2}\|\mathbf{y} - \tilde{\mathbf{y}}\|^2 \\
&= \frac{1}{2}\sum_{i=1}^{L}\left[\mathbf{y}_i - \mathbf{c}^{d,r-1}\Lambda\left([\mathbf{u}^{i-M}, \tilde{\mathbf{y}}^{i-N}]\right)\right]^2 . \quad (8)
\end{aligned}
$$

In order to minimize Eq. (8) we use the following modified steepest-descent algorithm. The expression of the components of the gradient vector ∇E^r needed are given by

$$
\begin{aligned}
\nabla E_j^r &= -\sum_{i=1}^{L}\left[\mathbf{y}_i - \mathbf{c}^{d,r-1}\Lambda\left([\mathbf{u}^{i-M}, \tilde{\mathbf{y}}^{i-N}]\right)\right] \cdot \\
&\quad \cdot \left(\Lambda\left([\mathbf{u}^{i-M}, \tilde{\mathbf{y}}^{i-N}]\right)\right)_j, (9)
\end{aligned}
$$

and the vector of parameters $\mathbf{c}^{d,r}$ is updated using the formula

$$
\mathbf{c}^{d,r} \leftarrow \mathbf{c}^{d,r-1} + \Delta\mathbf{c}^{d,r}, \quad (10)
$$

where each component of $\Delta\mathbf{c}^{d,r}$ is adaptively updated using the following algorithm

$$
\Delta\mathbf{c}_j^{d,r} = -\nabla E_j^r \mathbf{lr}_j^r + \Delta\mathbf{c}_j^{d,r-1}\mu, \quad (11)
$$

where $\mathbf{lr}^r$ are modified as described below and the momentum $\mu \in \mathbb{R}$ is fixed.

$$
\mathbf{lr}_j^r = \begin{cases} \mathbf{lr}_j^{r-1}inc & \text{if } \operatorname{sign}\left(\nabla E_j^r\right) = \operatorname{sign}\left(\nabla E_j^{r-1}\right) \\ \mathbf{lr}_j^{r-1}dec & \text{if } \operatorname{sign}\left(\nabla E_j^r\right) \neq \operatorname{sign}\left(\nabla E_j^{r-1}\right), \end{cases} \quad (12)
$$

$inc > 1$ and $dec < 1$ being real, fixed, positive constants.

From the formulation, the local convergence of the method to a minimum immediately follows. The drawback is that the achieved minimum may not be a global minimum but a local one. Also, the high number of parameters generated by the HL CPWL approximation when the number of divisions of the region $\mathbf{S}$ increases, constitutes now a limitation of the method.

In spite of this, the advantages of using HL CPWL functions enumerated below make it worth to define this identification structure.

1. The computation of the gradient is linear in the parameters and straight-forward since the approximation has already been computed in the previous step.

2. The canonical HL CPWL approximation uses the least number of parameters in the sense that any other PWL approximation has greater or equal number of parameters (see (Julián et al., 1999; Julián, 1999)).

3. A very efficient method for computing the HL CPWL approximation (see (Julián et al., 1999; Julián, 1999, Julián et al., 2000)) has been implemented in the MATLAB environment for both HL CPWL and orthonormal HL CPWL functions (Julián, 2000).

III. EXAMPLE

We consider the well known nonlinear system due to Narendra and Parthasarathy (Narendra and Parthasarathy, 1990) given by

$$
\mathbf{y}_k = \frac{1}{1 + \mathbf{y}_{k-1}^2} + \mathbf{u}_k^3, \quad (13)
$$

with $\mathbf{u}$ a random signal with uniform distribution. According to the proposed methodology, the regressor was defined with one input and one delayed output, i.e. $\varphi_k = [\mathbf{u}_k \quad \tilde{\mathbf{y}}_{k-1}]$.

We first generated a linear ARX model of the system given by Eq. (6). In Fig. 3 it is possible to see this linear approximation in a PWL format.
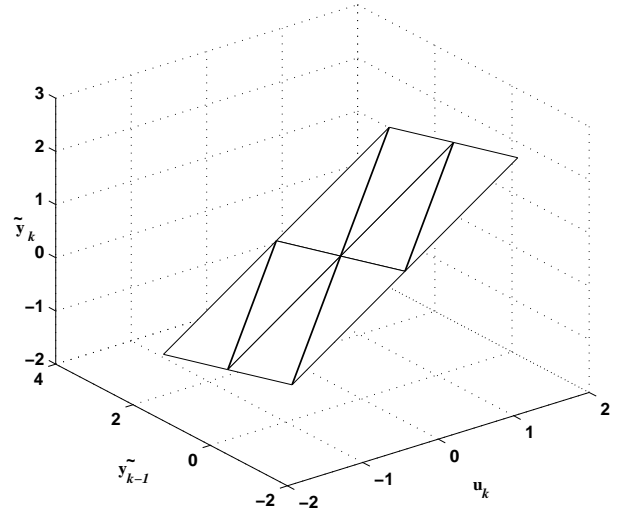

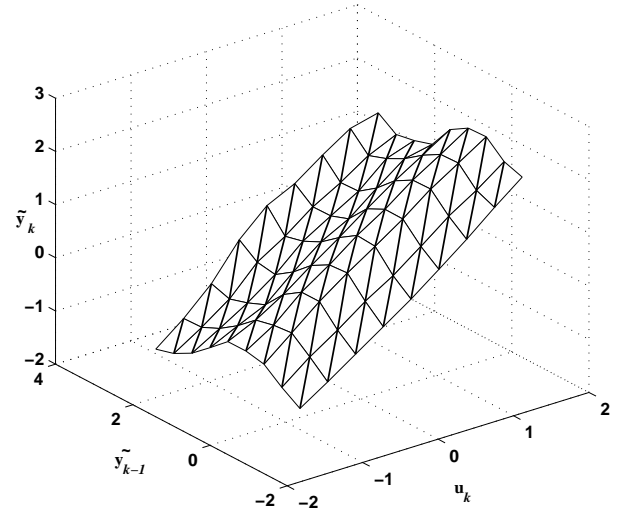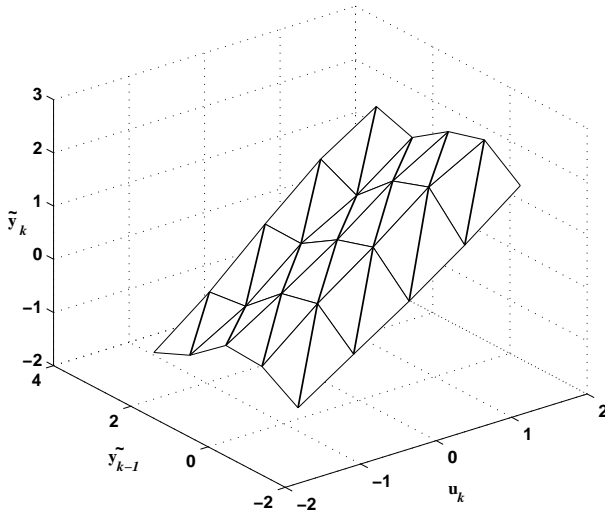Figure 3. Linear approximation using HL CPWL representation.

In order to improve the model performance, we increased the number of divisions to $ndiv = 2$ and optimized the parameters as described in Section IIB. Consequently, a new set of HL CPWL functions was obtained (see Fig. 4). We then repeated the process using $ndiv = 4$. As it can be appreciated in Fig. 5, the approximation rapidly improves.

Finally, the number of division of $\mathbf{S}$ was increased to $ndiv = 8$. The new HL CPWL approximation can be seen in Fig. 6.

As can be appreciated, the approximation to the nonlinear system quickly improves when the number of divisions of the set $\mathbf{S}$ increases. This statement is clearly showed in Fig. 7 and Fig. 8. In Fig. 7 we depicted the parameter optimization RMS error *versus* the number of iterations for each number of divisions. As it can be appreciated, the decreasing rate is high each time the number of divisions is augmented. On the other hand, in Fig. 8 we plotted the approximation and validation errors (*i.e.* the error in data used for approximation and the error in data not used for approximation, respectively) for the ARX and the NIIR HL CPWL models. As it can be clearly seen, there is a significant reduction of both, the approximation and validation errors, as long as the number of divisions increase.

IV. CONCLUSIONS

In this paper a NOE identification algorithm based on HL CPWL functions is presented. The main advantages of the
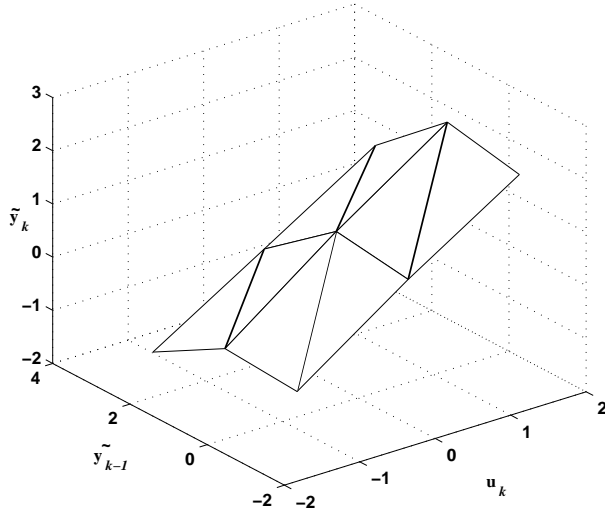
Figure 4. NIIR HL CPWL approximation using $ndiv = 2$.

Figure 5. NIIR HL CPWL approximation using $ndiv = 4$.

Figure 6. NIIR HL CPWL approximation using $ndiv = 8$.
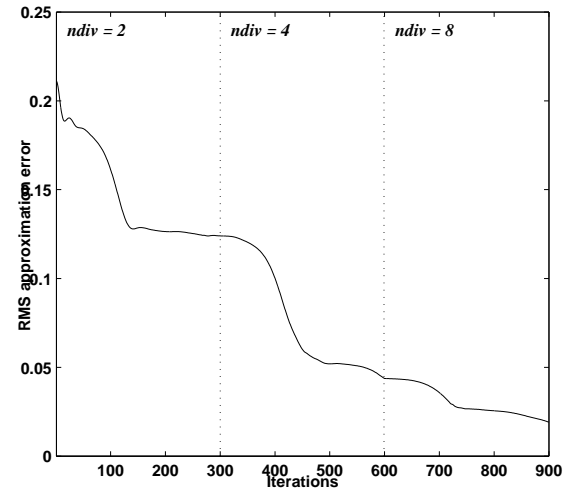
Figure 7. RMSE approximation error for the NIIR HL CPWL models using $ndiv = 2, 4, 8$.

algorithm are the following. We can first mention that this algorithm might be easily implemented in microelectronics due to the efficient computation of the HL CPWL functions and of the gradient. Secondly, we must point out the simplicity of the mechanism for increasing or decreasing the model degree of freedom, retaining the achieved model approximation.

The parameters of the HL CPWL for a given number of divisions could be straightforwardly evaluated from the previous ones. This would avoid using the least square methodology, as mentioned in Section IIB. This is the focus of our future work.

Furthermore, the potentials of our approach have been illustrated with a simulation example.

ACKNOWLEDGMENT

A HL CPWL FUNCTIONS

Definition A.1 *A function $f : \mathbf{S} \subset \mathbb{R}^m \to \mathbb{R}^n$, where $\mathbf{S}$ is a compact set, is PWL if and only if it satisfies the following*

(i) *The domain $\mathbf{S}$ is divided in a number of finite polyhedral regions $R^{(1)}, R^{(2)}, \ldots, R^{(N)}$ such that $\mathbf{S} = \cup_{i=1}^N \overline{R}^{(i)}$, by a finite set of boundaries*

$$H = \{H_i \subset \mathbf{S}, i = 1, 2, \ldots, h\}, \qquad (14)$$

such that each boundary is an $(m-1)$-dimensional hyperplane (or a subset of the hyperplane)

$$H_i = \{\mathbf{x} \in \mathbb{R}^m : \pi_i(\mathbf{x}) = \boldsymbol{\alpha}_i^T \mathbf{x} - \beta_i = 0\}, \qquad (15)$$

where $\boldsymbol{\alpha}_i \in \mathbb{R}^m$ and $\beta_i \in \mathbb{R}$ for $i = 1, 2, \ldots, h$ and cannot be covered by a $(m-2)$-dimensional hyperplane [1].

---

[1] We say that a boundary is *covered* by an hyperplane H if and only if $B \subset H$
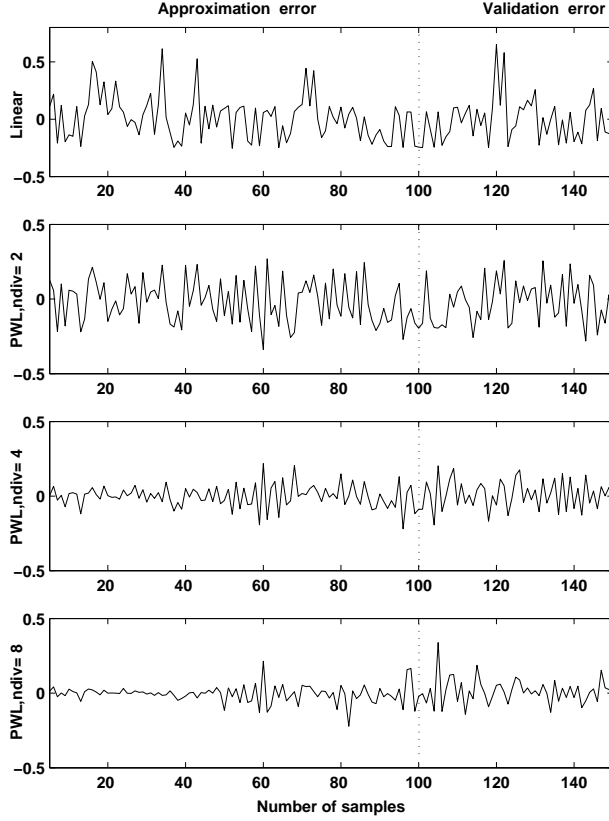
Figure 8. Approximation and validation errors for the NIIR HL CPWL model.

(ii) f is represented by an affine mapping of the form

$$f^{(i)}(\mathbf{x}) = J^{(i)}\mathbf{x} + \mathbf{w}^{(i)}, \qquad (16)$$

for any $\mathbf{x} \in R^{(i)}$; $J^{(i)} \in \mathbb{R}^{n \times m}$ is the Jacobian of the region $R^{(i)}$ and $\mathbf{w}^{(i)} \in \mathbb{R}^n$.

(iii) f is continuous on any boundary between two adjacent regions, i.e.

$$J^{(p)}\mathbf{x} + \mathbf{w}^{(p)} = J^{(q)}\mathbf{x} + \mathbf{w}^{(q)}, \qquad (17)$$

for any $\mathbf{x} \in \overline{R}^{(p)} \cap \overline{R}^{(q)}$.

If $\mathbf{S}$ is defined as in Eq. (7), the space of all continuous PWL mappings defined over the domain $\mathbf{S}$ partitioned with a simplicial boundary configuration H is denoted by $PWL_H[\mathbf{S}]$ and it is a linear vector space with the sum and multiplication of functions by a scalar defined as usual.

A basis for this space, constructed in (Julián *et al.*, 1999) by nesting absolute value functions, can be expressed in vector form as

$$\Lambda = \left[\Lambda^{0^T}, \Lambda^{1^T}, \dots, \Lambda^{m^T}\right]^T, \qquad (18)$$

where Λ^i is the vector containing the generating functions defined in (Julián *et al.*, 1999) with i nesting levels. Accordingly, any $f_p \in PWL_H[\mathbf{S}]$ can be written as

$$f_p(\mathbf{x}) = \mathbf{c}^T \Lambda(\mathbf{x}), \qquad (19)$$

where $\mathbf{c} = \left[\mathbf{c}^{0^T}, \mathbf{c}^{1^T}, \dots, \mathbf{c}^{m^T}\right]^T$, and every vector $\mathbf{c}^i$ is a parameter vector associated with the vector function Λ^i.

Then the HL CPWL functions defined on $\mathbf{S}$ uniformly approximate any continuous function $g : \mathbf{S} \mapsto \mathbb{R}^1$. The HL CPWL approximation to the nonlinear function g is defined (cf. (Julián *et al.*, 1999; Julián, 1999)) as the function $f_p \in PWL_H[\mathbf{S}]$ that satisfies

$$f_{pwl}\left(\mathbf{v}^j\right) = g\left(\mathbf{v}^j\right), \qquad (20)$$

$\mathbf{v}^j$ being the vertices of the simplicial partition H of the domain $\mathbf{S}$. If $g(\cdot)$ is Lipschitz continuous with Lipschitz constant L and the modeling error is defined as

$$\varepsilon = \sup_{\mathbf{x} \in \mathbf{S}} |f_p(\mathbf{x}) - g(\mathbf{x})|, \qquad (21)$$

then we have that

$$\varepsilon \leq \delta L. \qquad (22)$$

In order to obtain an orthonormal basis, it is necessary to define an inner product on $PWL_H[\mathbf{S}]$. If $V_{\mathbf{S}}$ is the set of vertices of $\mathbf{S}$ and f, g belong to $PWL_H[\mathbf{S}]$, then

$$\langle f, g \rangle = \sum_{\mathbf{v}_i \in \mathbf{S}} f(\mathbf{v}_i) g(\mathbf{v}_i), \qquad (23)$$

defines an inner product and so the space $PWL_H[\mathbf{S}]$ becomes a Hilbert space.

The new basis elements are linear combination of (18), that is

$$\Upsilon(\mathbf{x}) = T\Lambda(\mathbf{x}), \qquad (24)$$

and the matrix T may be obtained using the Gram-Schmidt procedure as given in (Julián *et al.*, 2000).

Also, the HL CPWL functions of this class can uniformly approximate any continuous function $g : \mathbf{S} \mapsto \mathbf{R}^1$. For finding the required approximation, we use a routine of (Julián, 2000) that finds a vector of parameters $\mathbf{c}$ that is the solution of the least square problem $\min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\|_2$, being $A = \Upsilon^T(X)$, X the input matrix and $\mathbf{b}$ the output to be approximated in sparse format. In accordance with (Julián *et al.*, 2000), the HL CPWL approximation of the nonlinear function g is defined as the function $f_p \in PWL_H[\mathbf{S}]$ satisfying $f_{pwl} = A\mathbf{c}$.

REFERENCES

Haykin, S. *Neural Networks: a comprehensive foundation*. Macmillan, New York. (1994).

Julián, P. "A toolbox for the piecewise approximation of multidimensional functions". http://www.pedrojulian.com (2000).

Julián, P., A. Desages and M.B. D'Amico. "Orthonormal high level canonical PWL functions with applications to model reduction". *IEEE Trans. on Circ. and Syst.*, **47**, 702-712 (2000).

Julián, P., A. Desages and O. Agamennoni. "High level canonical piecewise representation using a simplicial partition". *IEEE Trans. on Circ. and Syst.*, **44**, 463-480 (1999).

Julián, P.. "A high level canonical piecewise linear representation: theory and applications". PhD thesis, Universidad Nacional del Sur, Bahía Blanca, Argentina. UMI Dissertation Services, Michigan, USA (1999).

Korenberg, M.J. and L.D. Paarmann. "Orthogonal approaches to time-series and system identification". *IEEE Signal Processing Magazine*, 29-43 (1991).

Narendra, K.S. and K. Parthasarathy. "Identification and control of dynamical systems using neural networks". *IEEE Trans. on Neural Networks*, **1**, 4-27 (1990).

Poggio, T. and F. Girosi. "Regularization algorithms for learning that are equivalent to multilayer networks". *Science*, **247**, 978-982 (1990).

Sjöberg, J. and S.H. Ngia. "Neural nets and related model structures for nonlinear system identification". In *Nonlinear modeling: advanced black-box techniques*, J.A.K. Suykens and K. Vandevalle, Eds. Kluwer Academic Publishers, 1-28 (1998).

Sjöberg, J., Q. Zhang, L. Ljung, A. Beneviste, B. Delyon, P. Glorennec, H. Hjalmarsson and A. Juditsky. "Nonlinear blcak-box modeling in system identification: a unified overview". *Automatica*, **35**(12), 1691-1724 (1995).