On-line nonlinear identification using PWL mappings with applications to Wiener modelling

P. JULIAN *, O. AGAMENNONI†
Departamento de Ingenieria Eléctrica
Universidad Nacional del Sur
Av. Alem 1253, (8000) Bahia Blanca
ARGENTINA
e-mail: pjulian@criba.edu.ar

AND EDGARD N. SANCHEZ
CINVESTAV, Unidad Guadalajara
Apartado Postal 31-438, Plaza La Luna
Guadalajara, Jalisco C.P. 44550
MEXICO

Abstract: In this paper a nonlinear model structure oriented to online identification of fadding memory systems is presented. It is based on the Wiener structure and it is composed by two cascade blocks: a linear dynamic system followed by a nonlinear static system, which is designed using high level piecewise linear functions. The proposed resulting structure has some properties which are specially atractive for online inplementation.

Keywords: Piecewise-linear approximations, Wiener modelling.

1 Introduction

One of the main issues in system identification is the model structure, specially when nonlinear systems are considered. The model structure has a direct incidence in solving the bias-variance traded-off, *i.e.* enough degree of flexibility to adapt to any nonlinear function while overfitting is prevented. Two basic approaches are used to solve the bias-variance trade-off. The first one is based on a trial and error search for the model structure that offers the best performance with the testing data set. The second approach introduces a "regularization" mechanism that prevents a subset of the parameters to be adapted from the data. This permits to generate a model with different degrees of flexibility according to the mechanism used. A good discussion and a set of references on this subject may be found in [1].

Another important issue appears when on-line identification mechanisms are considered, for example, in adaptive control structure setups. In this case, the identification algorithm has to deal with system parameters and drift variations when moving from one operation region to another. It is usually assumed that the parameter variation of the system has a slower time scale than the dominant time constant of the system. Then, if the system is moving to a new region of the domain space, the model structure has to incorporate the new information without affecting the previous one. This is the stability-plasticity trade-off that is well known specially in the context of neural networks [2].

In this paper a nonlinear model structure oriented to on-line identification of fading memory systems is presented. It is based on the Wiener structure [3] and it is composed by two cascade blocks: a linear dynamic system followed by a nonlinear static system, as follows:

$$\begin{aligned} \dot{\mathbf{x}}(t) &= A\mathbf{x}(t) + Bu(t) \\ y(t) &= g(\mathbf{x}(t)) \end{aligned} \tag{1.1}$$

with $u(t)$ and $y(t)$ the input and output of the system respectively, $\mathbf{x}(t)$ the state vector and A the system matrix. Within this context, Laguerre or Kautz orthogonal basis can be considered depending on the dynamic nature of the system. (To model a resonant system a Kautz representation is shorter than the Laguerre representation since the elements of the basis function may be tuned to oscillate at or near the resonant frequency of the system.) If a Laguerre basis is used [4], the state vector $\mathbf{x}(t)$ can be considered as a vector of different filtered versions of the

*P. Julián is with CONICET (Comité Nacional de Investigaciones Científicas y Técnicas)

†O. Agamennoni is with CIC (Comité de Investigaciones de la Pcia. de Bs. As.)

input $u(t)$

$$x_i(s) = \sqrt{2p}\frac{(s-p)^{i-1}}{(s+p)^i}u(s)$$

The nonlinear static mapping $g(.)$ of the original Wiener series [3], was performed by Hermite polynomials. In the present paper $g(.)$ is realized using canonical piecewise linear functions [6]. As shown in [7], [5] and [9], these functions can approximate uniformly any continuous function defined on a compact domain. Another important aspect is that they have a special structure which is linear by regions and permits to express a nonlinear problem as several linear problems. In addition, the coefficients of the expression can be obtained efficiently from the resolution of a linear system characterized by a lower triangular matrix [5]. As result, the structure which is proposed has some properties which are specially attractive for on-line implementations.

1.1 Basic Definitions

Definition 1 *A function* $f : \mathbf{D} \subset \mathbf{R}^n \mapsto \mathbf{R}^m$, $\mathbf{D}$ *compact is a PWL function iff:*
1) $\mathbf{D}$ *is partitioned into a finite number of polyhedral regions* $R^{(1)}$, $R^{(2)}$, ..., $R^{(N)}$ *(so that* $\mathbf{D} = \bigcup_{i=1}^{N} \bar{R}^{(i)}$*), by a finite set of boundaries* H_i, $i = 1, 2, ..., h$, *where each boundary is an* $n-1$ *dimensional hyperplane (or a subset of),*

$$H_i = \left\{ \mathbf{x} : \pi_i(\mathbf{x}) := \alpha_i^T \mathbf{x} - \beta_i = 0 \right\}, \qquad (1.2)$$

and can not be covered [1] *by an hyperplane* $(n-2)$ *dimensional, where* $\alpha_i \in \mathbf{R}^n$ *and* $\beta_i \in \mathbf{R}^1$, $\forall i$.
2) f *is expressed as*

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

for any $\mathbf{x} \in R^{(i)}$, *where* $J^{(i)} \in \mathbf{R}^{m \times n}$ *is the jacobian matrix of region* $R^{(i)}$, *and* $w^{(i)} \in \mathbf{R}^m$.
3) f *is continuous on every boundary of adjacent regions, i.e.,*

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

$\forall \mathbf{x} \in \bar{R}^{(p)} \cap \bar{R}^{(q)}$.

Definition 2 *Let* $\mathbf{x}_0$, $\mathbf{x}_1$, ...,$\mathbf{x}_n$ *be* $n+1$ *points in* $\mathbf{R}^n$. *A* simplex *is defined as the region* $\{\mathbf{x} : \mathbf{x} = \sum_{i=0}^{n} \mu_i \mathbf{x}_i\}$, *where* $\mu_i \in [0, 1]$, $i \in \{1, n\}$ *and* $\sum_{i=0}^{n} \mu_i = 1$.

[1] A boundary B is said to be covered by an hyperplane H iff $B \subset H$.

A simplex is said to be proper if and only if it satisfies

$$rank\left(\begin{bmatrix} x_0 & x_1 & \cdots & x_n \\ 1 & 1 & \cdots & 1 \end{bmatrix}\right) = n+1. \qquad (1.3)$$

Geometrically, this means that a proper simplex can not be contained in an $(n-1)$ dimensional hyperplane. The set of vertices of a simplex $R^{(i)}$ will be noted as $vert\left(R^{(i)}\right)$.

2 Preliminaries

Without loss of generality, the input state vector will be assumed to lie in a compact region $D = \{x \in \mathbf{R}^n : \|\mathbf{x}\|_\infty \leq d\}$. As it was exposed in [9] and [5], a simplicial boundary configuration with grid step δ, namely H, will be used to partition this domain into proper simplices of $n+1$ vertices. In this way, it follows that the set $PWL_H[D]$ of all continuous PWL mappings $f : D \mapsto \mathbf{R}^1$, defined on D with the boundary configuration H, is a linear vector space, if addition and multiplication by a scalar $r \in \mathbf{R}$ are defined as:

$$\begin{array}{l} a)\,(f+g)(\mathbf{z}) = f(z) + g(z), \forall z \in D \\ b)\,(r \cdot f)(\mathbf{z}) = r \cdot f(z), \forall z \in D \end{array} \qquad (2.1)$$

where $f, g \in PWL_H[D]$. In addition, every function $f \in PWL_H[D]$ is uniquely determined by the values of f at the vertices of D, i.e., f is uniquely defined by the set of values

$$\left\{ f(v_i), v_i \in vert\left(R^{(i)}\right), \forall R^{(i)} \subset \mathbf{D} \right\} \qquad (2.2)$$

In [9] an equivalent representation was introduced, using a set of basis functions, that permits to represent any $f \in PWL_H[D]$ in a compact form $f = c^T \Lambda(x)$, where c is a vector of parameters and Λ is a vector composed of high level PWL functions with different absolute value nestings, ranging from 1 to n. This equivalence implies that there is a one to one relationship between the vector $\tilde{c}$ and the set of values of f at the vertices of the partition. Moreover, with the introduction of an adequate inner product in the space $PWL_H[D]$ it is possible to obtain a set of orthonormal basis (see [10]), so that $f = \tilde{c}^T \Lambda(x)$ and in this case each element of $\tilde{c}$ is the value of function f at some vertex of D. As it was exposed in [9], from an application standpoint it is convenient to formulate the PWL function using the functional form. However, for the purposes of this paper it is more illustrative to state the identification algorithm in terms of the values (2.2).

3 Identification Strategy

The objective of the identification is the computation of the nonlinear function $g(\cdot)$ in (1.1), from a sequence of samplings of the signals $x(t)$ and $y(t)$. The structure chosen to represent the function is a PWL mapping with a simplicial partition according to [9]. Before introducing the main idea, the following concepts are necessary:

Definition 3 *Let $P = \{x_0, ..., x_q\}$, $x_i \in \mathbf{R}^n$ be a set of samples. If there exists a subset $P' = \{x_{i_0}, ..., x_{i_n}\}$ of $n+1$ points of P which determine a proper simplex, then the points of P' will be referred to as "valid" samples of the set P.*

Lemma 1
The samples $x_0, x_1, ..., x_{n+1} \in \mathbf{R}^n$, are valid if and only if $rank([x_1 - x_0, x_2 - x_0, ..., x_n - x_0]) = n$.

See Appendix
First, consider the identification of a PWL function, namely g_p. In this case, for any given region $R^{(i)}$ of the domain space, g_p has a linear affine expression $g_p(x) = J^{(i)}x + w^{(i)}$, where $J^{(i)} \in \mathbf{R}^{1\times q}$, $w^{(i)} \in \mathbf{R}^1$, $\forall i$. Due to its local linear affine form, only $n+1$ "valid" points in $\mathbf{R}^n$ and their corresponding function values are necessary to fully characterize $J^{(i)}$ and $w^{(i)}$. In consequence, any further sample in the same region, in excess of the first $n+1$ "valid" samples, do not provide additional information. Accordingly, different choices of the initial set of $n+1$ valid samples will give as result the same values for the jacobian and offset vector of $R^{(i)}$.
However, the measured data corresponds to a nonlinear function g, and in addition measurement noise is always present. As a consequence of this, the values obtained for $J^{(i)}$ and $w^{(i)}$, will vary in general with the choice of the set of valid samples $x_j \in R^{(i)}$.
The strategy to incorporate in the model all the information provided by several different estimations of $J^{(i)}$ and $w^{(i)}$ (or equivalently, several choices of sets of $n+1$ valid samples) is the following; with every estimate obtained, namely $\hat{J}^{(i)}$ and $\hat{w}^{(i)}$, an upper g_U and a lower g_L PWL functions are determined in such a way that

$$g_L(x_j) \leq \hat{J}^{(i)}x_j + \hat{w}^{(i)} \leq g_U(x_j)$$

holds for every valid $x_j \in X$, and for every estimate $\left(\hat{J}^{(i)}, \hat{w}^{(i)}\right)$ obtained. This procedure guarantees a robust approximation in the sense that both PWL functions will "contain" the measured values of $y(t)$. Finally, a nominal function can be determined as $g_N = (g_U + g_L)/2$.
Clearly, two different algorithms are needed to work in parallel. One, to classify the samples in order to select only valid samples, and the other to develop both PWL functions g_U and g_L using the valid data. Next, both of them are explained in detail.

3.1 Selection of valid samples

This section centers in finding on-line sets of $n+1$ valid samples, for a given sequence of samplings of the input state vector $X = \{x(t_0), x(t_1), ..., x(t_m)\}$. The selection procedure is as follows: suppose that k valid samples $x_0, x_1, ..., x_{k-1}$ have been obtained. This implies that $rank\left(\left[x'_1, x'_2, ..., x'_{k-1}\right]\right) = k-1$, where $x'_i = x_i - x_0$. Then, using the Gram-Schmit (GS) procedure with the $k-1$ vectors x'_1, ...,x'_{k-1} an orthonormal basis e'_1, ...,e'_{k-1} of $\mathbf{R}^n$ is determined.
Next, a candidate sample $\tilde{x}_k \in X$ is chosen. To qualify as a valid sample, it must satisfy that $rank\left(\left[x'_1, x'_2, ..., x'_{k-1}, \tilde{x}'_k\right]\right) = k$. To check if this condition holds, we calculate its associated orthonormal vector $\tilde{e}_k$ using the GS procedure with the vectors already found x'_1, ...,x'_{k-1} and the new vector $\tilde{x}'_k$. If $\tilde{e}_k \neq 0$, we choose $x_k = \tilde{x}_k$, $e_k = \tilde{e}_k$ because in this case the vectors $x'_1, x'_2, ..., x'_k$ are linearly independent and then $rank([x'_1, x'_2, ..., x'_k]) = k$. On the contrary, if $\tilde{e}_k = 0$, $\tilde{x}'_k$ is a linear combination of the vectors $x'_1, x'_2, ..., x'_{k-1}$ and $rank\left(\left[x'_1, x'_2, ..., x'_{k-1}, \tilde{x}'_k\right]\right) = k-1$. In this case, a new $\tilde{x}_k \in X$ is chosen.
This procedure is repeated, following the steps listed below, until
$n+1$ valid samples $x_0(1), x_1(1), ..., x_n(1)$ are obtained.
step i) Initialize $i = 0$, $x_0(1) = x(t_i)$ and a tolerance factor $\varepsilon > 0$.
step ii) $(i = i+1)$ Define

$$e_1 = \frac{x'(t_i)}{\|x'(t_i)\|} = \frac{x(t_i) - x_0(1)}{\|x(t_i) - x_0(1)\|}$$

Repeat this step until $|\langle x'(t_i), e_1\rangle| > \varepsilon$, then set $x_1(1) = x(t_i)$.
step j) $(i = i+1)$ Define

$$e_j = \frac{x'(t_i) - \sum_{q=0}^{j-1} \langle x'(t_i), e_q\rangle e_q}{\left\|x'(t_i) - \sum_{q=0}^{j-1} \langle x'(t_i), e_q\rangle e_q\right\|}$$

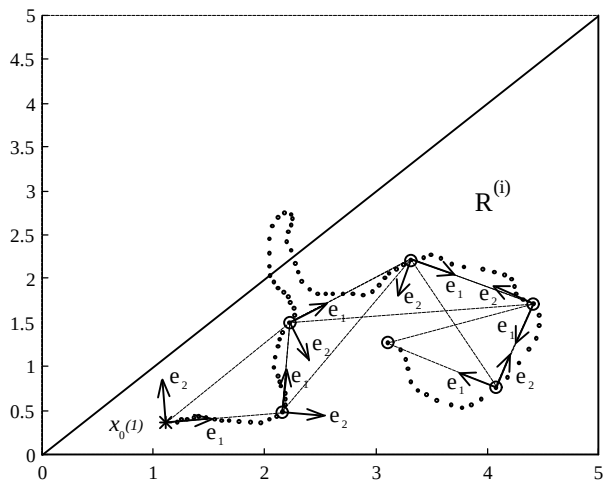

Figure 1: Selection of valid samples

Repeat this step until $|\langle x'(t_i), e_j\rangle| > \varepsilon$, then set $x_j(1) = x(t_i)$.
After n steps, a set $S(1)$ of $n+1$ initial valid points $S(1) = \{x_0(1), x_1(1), ..., x_n(1)\}$ is determined.
Next, it is necessary to process the incoming samples. When a new valid sample $\hat{x}$ arrives, the valid sample $x_0(k)$ is discarded. Then, the first $n-1$ valid samples are obtained as

$$\begin{cases} x_0(k+1) = x_1(k) \\ x_1(k+1) = x_2(k) \\ \vdots \\ x_{n-1}(k+1) = x_n(k) \end{cases}$$

and the new sample $\hat{x}$ is incorporated as

$$x_n(k+1) = \hat{x}$$

The validity of $\hat{x}$ is checked using the condition $|\langle \hat{x}', e_n\rangle| > \varepsilon$, where $e_n = v/\|v\|$, $v = \hat{x}' - \sum_{q=0}^{n-1} \langle \hat{x}', e_q\rangle e_q$. This is illustrated in fig. 1 for a domain in $\mathbf{R}^2$.
In this way a set of valid samples

$$S(k) = \{x_0(k), ..., x_n(k)\} \tag{3.1}$$

is always available.validsamples
If while working in simplex $R^{(i)}$, the new valid sample $\hat{x}$ is found in another simplex $R^{(j)}$, the algorithm is re-initialized. This is done by setting $x_0(1) = \hat{x}$ and using $R^{(j)}$ as the new working simplex .

3.2 On line adaptation of g_U and g_L

This algorithm uses as input the set (3.1) and its associated set of measured values $Y(k) = \{y_0(k), ..., y_n(k)\}$, where $y_j(k) = g(x_j(k))$, $j = 0, 1, ..., n$.
i) First, the region $R^{(i)}$ where the set $S(k)$ belongs to, is identified.
ii) Using the valid samples, the equation of the PWL function, corresponding to the region $R^{(i)}$, at step k, given by $\left(J^{(i)}(k), w^{(i)}(k)\right)$ is determined solving the linear system

$$\begin{bmatrix} y_0(k) \\ y_1(k) \\ \vdots \\ y_n(k) \end{bmatrix} = M(k)^T \begin{bmatrix} J^{(i)^T}(k) \\ w^{(i)}(k) \end{bmatrix} \tag{3.2}$$

where $M(k) = \begin{bmatrix} x_0(k) & x_1(k) & \cdots & x_n(k) \\ 1 & 1 & \cdots & 1 \end{bmatrix}$. As the next lemma states, the solution of (3.2) can be done without matrix inversions, using back substitution.

Lemma 2 *The equation system (3.2) is equivalent to the lower triangular system*

$$\begin{bmatrix} y'_1(k) \\ y'_2(k) \\ \vdots \\ y'_n(k) \end{bmatrix} = \begin{bmatrix} a_{1,1} & 0 & \cdots & 0 \\ a_{2,1} & a_{2,2} & \ddots & \vdots \\ & \ddots & \ddots & 0 \\ a_{n,1} & & a_{n,n-1} & a_{n,n} \end{bmatrix} J^{(i)^T}(k)$$
$$w^{(i)} = y_0(k) - J^{(i)}(k)\, x_0(k)$$

where $a_{i,j} = \langle x'_i(k), e_j(k)\rangle$, $\forall i > j$ *and* $y'_j(k) = y_j(k) - y_0(k)$.

Remark: Note that this equation system has always solution because $a_{i,i} = \langle x'_i(k), e_j(k)\rangle > \varepsilon \neq 0$.
iii) With the values obtained for the pair $\left(J^{(i)}(k), w^{(i)}(k)\right)$, the values of the estimated non-linear function $\hat{g}$ at the vertices $v_0, v_1, ..., v_n \in vert\left(R^{(i)}\right)$ are found:

$$\hat{g}(v_q(k)) = J^{(i)}(k)\, v_q(k) + w^{(i)}(k), q = 0, 1, ..., n.$$

iv) The values of g_U and g_L at the vertices $v \in vert\left(R^{(i)}\right)$ are adjusted according to:

$$g_U|_{v_q}(k+1) = \max\left\{g_U|_{v_q}(k), g(v_q(k))\right\},$$
$$g_L|_{v_q}(k+1) = \min\left\{g_L|_{v_q}(k), g(v_q(k))\right\}$$

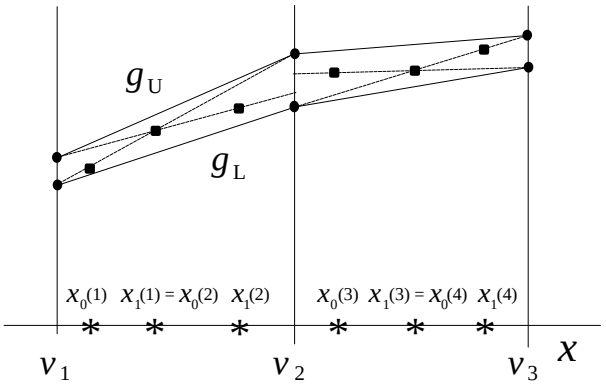

Figure 2: Adaptation of g_U and g_L

$q = 0, 1, ..., n$.
ajusto An example in $\mathbf{R}^1$ with two simplices and three samples on each region is shown in fig. 2.
Note here that in each step only $2(n+1)$ parameters are adjusted.

3.3 Numerical Remarks

The constructive algorithm proposed guarantees that only $n+1$ valid samples are necessary to estimate the function in region $R^{(i)}$. This is due to the fact that the algorithm fully exploits the PWL structure of the function and the structure of the domain. Within the framework presented, less than $n+1$ valid samples collected in one region can be considered as not enough information to characterize the function. Accordingly, in applications it is convenient to use an algorithm in parallel to indicate which regions of the domain have not been identified to avoid their use.

4 Conclusions

An algorithm has been proposed for the on-line identification on nonlinear functions using PWL functions. It fully exploits the local linear affine form which PWL functions have for a given region of the domain. It is mainly intended to be used in Wiener model structures, in connection with the identification of dynamic nonlinear systems. Further research includes an intensive comparison with another identification strategies like those proposed in [11] and [12].

5 Appendix

Proof of lemma 1 : If the coordinates change $x'_i = x_i - x_0$ is considered, condition (1.3) can be written as $rank(\begin{bmatrix} 0 & x'_1 & \cdots & x'_n \\ 1 & 1 & \cdots & 1 \end{bmatrix}) = n+1$, which clearly is equivalent to satisfy $rank(\begin{bmatrix} x'_1 & x'_2 & \cdots & x'_n \end{bmatrix}) = n$, which is the condition stated by the lemma.
Proof of lemma 2 : First, consider the expression of a generic $y_j(k)$ written as

$$y_j(k) = J^{(i)}(k)(x_j(k) - x_0(k)) + \hat{w}^{(i)}(k) \quad (5.1)$$

where $\hat{w}^{(i)}(k) = w^{(i)}(k) + J^{(i)}(k)\, x_0(k)$. If $x_j(k) = x_0(k)$, then $y_0(k) = \hat{w}^{(i)}(k)$, so that (5.1) is

$$y_j(k) - y_0(k) = J^{(i)}(k)\, x'_j(k) \quad (5.2)$$

In addition, $w^{(i)}(k) = y_0(k) - J^{(i)}(k)\, x_0(k)$ and the second equation of (3.2) holds. Second, note that $x_j(k) - x_0(k) = x'_j(k)$ can be written as

$$x'_j(k) = \sum_{p=0}^{j} \left\langle x'_j(k), e_p \right\rangle e_p \quad (5.3)$$

Replacing (5.3) into (5.2) gives

$$y_j(k) = \sum_{p=0}^{j} \left\langle x'_j(k), e_p \right\rangle J^{(i)}(k)\, e_p + y_0(k) = \\ = \sum_{p=0}^{j} \left\langle x'_j(k), e_p \right\rangle J_p^{(i)}(k) + y_0(k)$$

and the lemma holds after noting that $a_{j,p} = \left\langle x'_j(k), e_p \right\rangle$.

References

[1] J. Suykens and J. Vandewalle, *Nonlinear Modeling: Avanced Black-Box Techniques.* Kluwer Academic Publishers, 1998.

[2] A. Nigrin, *Neural Networks for pattern recognition.* The MIT Press, 1993.

[3] M. Schetzen, *The Volterra and Wiener theories of nonlinear systems.* John Wiley & Sons, Chap. 2-6, 1980.

[4] A. Elshafei, G. Dumont, and E. Elnaggar, "Adaptive gpc based on laguerre-filters modelling," *Automatica*, vol. 30, pp. 1913–1920, 1994.

[5] P. Julián, A. Desages, and O. Agamennoni, "High level canonical piecewise linear representation using a simplicial partition." accepted in IEEE Trans. on Circ. and Syst., to appear, 1998.

[6] L. O. Chua and S. Kang, "Section-wise piecewise-linear functions: canonical representation, properties and applications," *Proc. IEEE*, vol. 65, pp. 915–929, 1977.

[7] J. N. Lin and R. Unbehahuen, "Canonical piecewise-linear approximations," *IEEE Trans. Circuits Syst.*, vol. 39, pp. 697–699, 1992.

[8] M. Chien and E. Kuh, "Solving nonlinear resistive networks using piecewise-linear analysis and simplicial subdivision," *IEEE Trans. Circuits Syst.*, vol. CAS-24, pp. 305–317, 1977.

[9] P. Julián, A. Desages, and O. Agamennoni, "On the high level representation of canonical piecewise linear functions," in *Proc. of ISCAS 98*, (Monterey, California), 1998.

[10] P. Julián, B. D. Amico, and A. Desages, "Orthonormal high level canonical pwl functions." submitted to ISCAS 99, 1999.

[11] P. Julián, M. Jordán, and A. Desages, "Canonical piecewise linear approximation of smooth functions," *IEEE Trans. on Circ. and Syst.*, vol. 45, pp. 567–571, 1998.

[12] J. N. Lin and R. Unbehahuen, "Adaptive nonlinear digital filter with canonical piecewise-linear function structure," *IEEE Trans. Circuits Syst.*, vol. 37, pp. 347–353, 1990.