10.2478/v10170-011-0029-8

Archives of Control Sciences
Volume 22(LVIII), 2012
No. 4*, pages 371–388*

Two steps piecewise affine identification of nonlinear systems

JURAJ ŠTEVEK, ALEXANDER SZŰCS, MICHAL KVASNICA, MIROSLAV FIKAR and ŠTEFAN KOZÁK

Given a set of input-output measurements, the paper proposes a method for approximation of a nonlinear system by a piecewise affine model (PWA). First step of the two-stage procedure is identification from input-output data, in order to obtain an appropriate nonlinear function in analytic form. The analytic expression of the model can be represented either by a static nonlinear function or by a dynamic system and can be obtained using a basis function expansion modeling approach. Subsequently we employ nonlinear programming to derive optimal PWA approximation of the identified model such that the approximation error is minimized. Moreover, we show that approximation of multivariate systems can be transformed into a series of one-dimensional approximations, which can be solved efficiently using standard optimization techniques.

Key words: piecewise affine systems, piecewise linear, approximation, nonlinear systems, orthogonal polynomials, basis function expansion

1. Introduction

Mathematical model of real plants plays a vital role in many control-oriented tasks, including, but not limited to, model validation, control design, and analysis of closed-loop systems. In order to perform these tasks accurately and efficiently, the model has to be capture the plant's behavior well enough in all operating regimes, while being simple enough to be able to processed by algorithms available in control theory. Nonlinear plant models are traditionally employed when high model accuracy is desired. However, the nonlinear nature of such models poses significant difficulty in control design and analysis, theory of which is usually based on linear models. Therefore it is a standard practice to linearize the nonlinear model around a selected operating point. The downside being that the linearized model is only accurate when the plant operates close to the selected

The authors are with Department of Process Control, Faculty of Chemical and Food Technology, Slovak University of Technology, Radlinskeho 9, 812 37 Bratislava, Slovak Republic. Corresponding author is J. Števek, e-mail: juraj.stevek@stuba.sk

The authors gratefully acknowledge the contribution of the Scientific Grant Agency of the Slovak Republic under the grants 1/1105/11, 1/0095/11, 1/0973/12, and the Slovak Research and Development Agency under the project APVV-0551-11.

Received 14.07.2012.

linearization point. More recently, the concept of hybrid systems [3] has been frequently adopted to model real plants. Hybrid systems allow to approximate nonlinear models by multiple linearizations around several operating points. Such an approximation, from a mathematical point of view, corresponds to a Piecewise Affine (PWA) system [15] composed of a finite number of distinct linear (or affine) dynamics with associated regions of validity. Compared to standard linear model, PWA models provide higher approximation accuracy, while the underlying piecewise linear nature of such models renders subsequent tasks (e.g. model validation, control synthesis, analysis, etc.) easier compared to full nonlinear models.

Parameters of PWA models, i.e. regions of validity and associated linear dynamics, can be obtained in two ways. If the analytical form of the nonlinear model to be approximated is fully available, one can manually select multiple linearization points and subsequently use the HYSDEL language [16] to describe the PWA system manually. Alternatively, one can use the method of [8] which employs nonlinear optimization to automatically determine regions of validity and corresponding linear dynamics. The resulting PWA approximation is optimal in the sense that it minimizes the error between the original nonlinear model and its PWA approximation. Disadvantage of these approaches is that they cannot be applied if the nonlinear plant characteristics are only available in the form of input-output data. Another class of approaches is based on hybrid identification techniques which apply clustering techniques to classify measured data into operating regions, see e.g. [13, 5]. The crucial downside of these approaches is twofold. First, they are computationally expensive, since they are based on solving mixed-integer optimization problems and are therefore limited to low-dimensional problems only. Second, there is no guarantee that the union of operating regions, obtained by clustering, cover the whole area of parameters of interest without creating "holes" where the model would be undefined. Several other methods have been proposed for fitting general piecewise linear functions to (multidimensional) data. A triangular nested partitioning was exploited in [7]. A non-parametric approach based on local piecewise (affine or polynomial) regression was presented in [6]. Both methods lead to regular division of the function domain. A heuristic method for fitting a convex piecewise linear function to a given set of data was presented in [10].

In this work we propose to use a two-stage optimization-based approach to derive PWA approximations of a nonlinear systems from input-output data. In the first stage we fit the data with an analytic function of a low complexity by employing the basis function expansion modeling approach (black box model). Output of this stage is an algebraic function f of the measured data which minimizes the fitting error while keeping the analytical form of f as simple as possible. In the second step we then apply our optimization-based procedure, introduced in [8], to derive optimal PWA approximation $\widetilde{f}$ of the fit f. We aim at obtaining the best possible PWA approximation $\widetilde{f}$ of a given complexity by minimizing the integrated squared error $\int (f(z) - \widetilde{f}(z))^2 \, \mathrm{d}z$. The crucial benefit of such a two-stage procedure stems from the fact that we can derive the PWA approximation with the rigorous property that the operating regions of the PWA

approximation cover the whole space of parameters of interests. Moreover, we show that approximation of multi-dimensional data boils down to a series of one-dimensional approximations, allowing for a fast and efficient construction of the PWA approximation. We also illustrate that the proposed procedure features fast training, allows to tune accuracy as a function of model complexity, and is applicable to a wide spectrum of static and dynamic systems.

The paper is divided into six sections. After formally stating the problem in Section 2, we give in Section 3 a direction how to use one-hidden layer neural network as a tool for obtaining analytical form of an investigated system from input-output measurements. The identification of the analytical form is characterized by convexity, fast convergence, and adjustable accuracy. In Section 4 we show that an optimal PWA approximation of the identified network can be formulated and solved as a nonlinear programming problem. We show that the problem boils down to solving a series of one-dimensional approximations. The last section serves for demonstrating the proposed method on two illustrative example, involving a static model of a Current-voltage characteristic as well as a dynamic model of a benchmark system.

2. Problem formulation

In this paper we aim at addressing the following problem. We are given T samples of input data $z_i \in \Omega \subset \mathbb{R}^{n_z}$ from some closed and bounded set Ω, and the corresponding measurements $y_i \in \mathbb{R}$, $i = 1,\dots,T$. We want to fit the data with a PWA function $\widetilde{f} : \mathbb{R}^{n_z} \to \mathbb{R}$ with N regions

$$\widetilde{f}(z) = \begin{cases} a_1^T z + c_1 & \text{if } z \in \mathcal{R}_1, \\ \vdots & \\ a_N^T z + c_N & \text{if } z \in \mathcal{R}_N, \end{cases} \tag{1}$$

which satisfies two design requirements:

R1: $\widetilde{f}$ is well defined [2] on Ω, i.e. it satisfies $\text{interior}(\mathcal{R}_i) \cap \text{interior}(\mathcal{R}_j) = \emptyset$, $\forall i \neq j$ and $\cup_i \mathcal{R}_i = \Omega$, $i = 1,\dots,N$.

R2: $\widetilde{f}$ is an optimal fit, i.e. it minimizes the fitting error $e_{\text{fit}} = \sum_{i=1}^{T} (y_i - \widetilde{f}(z_i))^2$.

Solving this problem (i.e. determining regions $\mathcal{R}_j \subseteq \mathbb{R}^{n_z}$ and parameters $a_j \in \mathbb{R}^{n_z}$, $c_j \in \mathbb{R}$, $j = 1,\dots,N$), however, is not trivial [8] if the input samples z_i are vectors, i.e. when $n_z > 1$. The difficulty being how to divide the domain Ω into non-overlapping regions $\mathcal{R}_j$ without creating "holes", i.e. guaranteeing that the union $\cup_j \mathcal{R}_j$ completely covers Ω if $\text{dimension}(\Omega) > 1$.

To overcome this difficulty, we propose to split the search for the PWA function $\widetilde{f}$ into two steps. In the first stage we fit the input data, represented by the (z_i, y_i) pairs, with a nonlinear function $y = f(z)$:

Problem 1 *Given are T samples of input data z_i and the corresponding measurements y_i, $i = 1,\dots,T$. Fit the data with a multivariable function $f : \mathbb{R}^{n_z} \to \mathbb{R}$ such that the fitting error $e_{fit} = \sum_{i=1}^{T}(y_i - f(z_i))^2$ is minimized.*

Once the analytical form of the fitting function f is available, in the second step we search for its optimal PWA approximation:

Problem 2 *Given is a nonlinear function $f : \mathbb{R}^{n_z} \to \mathbb{R}$ and its domain $\Omega \subset \mathbb{R}^{n_z}$. Find a well-defined PWA approximation $\widetilde{f}$ as in* (1) *such that the approximation error*

$$e_{aprx} = \int_{\Omega} (f(z) - \widetilde{f}(z))^2 dz \tag{2}$$

is minimized when the integral is evaluated over the whole domain Ω.

We will show in Section 4 how to solve Problem 2 if the following assumption holds:

Assumption 1 *The multivariable nonlinear function $f : \mathbb{R}^{n_z} \to \mathbb{R}$ can be represented as a sum of products of functions in single variables, i.e. $f(z_1,\dots,z_{n_z}) = \sum_{i=1}^{n_z} \alpha_i \left(\prod_{j=p_i}^{q_i} f_j(z_j)\right)$. Here, α_i are scalar coefficients and $f_j : \mathbb{R} \to \mathbb{R}$ are scalar-valued (possibly nonlinear) basis functions.*

To guarantee that f satisfies this assumption, in Section 3 we propose to solve Problem 1 by employing an approach based on neural networks with a predefined set of basis functions. The network then finds a simple combination of these basis functions which provide a good fit.

3. Neural network for input-output mapping of nonlinear systems

The first stage of the procedure requires identification of the non-linear system in the analytical form. This problem has been addressed in applied mathematics (multivariate function approximation), statistics (nonparametric multiple regression) and in computer science and engineering (artificial neural networks). It can be solved using a variety of methods. The process of the identification is based on objective of the non-linear regression. If the purpose of the regression analysis is to get a function for predicting future values of the system, then the accuracy is the only important aspect of the model. The measure of accuracy is often defined by the integral error or the expected error.

Depending on the application, other desirable properties of the approximation might include rapid computability and smoothness of the identified model, i.e. that its low order derivatives exists everywhere. In this work, moreover, it is required to obtain the model in the separable analytical form (Assumption. 1).

In the literature, two basic approaches for the non-linear regression exist:

- Parametric methods: the structure of the model is proposed a priori and unknown parameters are estimated optimizing the objective.
- Non-parametric methods: structure of the model is not formed explicitly, but is computed according to information derived from the data.

In the parametric modeling, the fundamental approach is to fit parametric function $f(x|\{a_j\}_1^P)$ to the training data most often by least-squares. The parameter estimates are given by

$$\{\widetilde{a}_j\}_1^P = \underset{\{a_j\}_1^P}{\operatorname{argmin}} \sum_{i=1}^{N} \left(y_i - f\left(x|\{a_j\}_1^P\right)^2 \right) \tag{3}$$

To the most extensively studied parametric models belongs additive function expansion

$$f(x) = \sum_{k=1}^{m} \alpha_k f_k(x) \tag{4}$$

with some basis functions f_k. The basis functions are mostly generated from one and the same "parent function" $\kappa(x)$ that is scaled and translated according to

$$f_k(x) = \kappa(\beta_k(x - \gamma_k)) \tag{5}$$

The basis functions are thus characterized by the scale (dilatation) parameters β_k and the location (translation) parameters γ_k [9]. Depending on the choice of basis functions we can deal with radial basis neural networks, one hidden layer sigmoidal neural networks, wavelets etc. In the literature, the basis function expansion belongs to black box model which interpretability does not play any important role. However, in the proposed procedure, it is possible to use any function expansion which satisfies user objectives and Assumption 1.

In the following, we will just point out some of the possible regression techniques applicable for further step.

Radial basis function networks: RBF neural network is an artificial neural network that uses RBF functions as activation function in three layers structure (Fig. 1). They are much used in function approximation, time series prediction, and control. The output, $y : \mathbb{R}^n \to \mathbb{R}$ of the network is given as

$$y(x) = \sum_{i=1}^{N} \alpha_i \rho(\|x - \gamma_i\|) \tag{6}$$

where N is the number of neurons in the hidden layer, γ_i is the center vector for neuron i, and α_i ate the weights of the linear output neuron. The norm is typically taken to be the Euclidean distance and the basis function is taken to be Gaussian

$$\rho(\|x - \gamma_i\|) = \exp\left(-\beta \|x - \gamma_i\|^2\right) \tag{7}$$

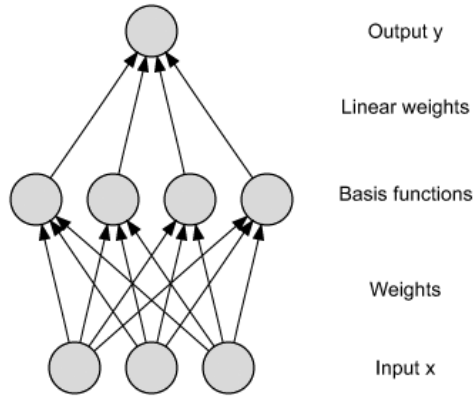


Figure 1. Architecture of a basis function network.

The analytical form (6) readily satisfies the Assumption 1.

Polynomial expansion: In system identification and control, an abundant attention has been given to polynomial model structures. Such a polynomial expansion can be modeled by three layer network depicted in Fig. 1. If the polynomial basis exhibits property of orthogonality we address an Orthogonal Activation Function based Neural Network (OAF-NN) [19, 12].

The hidden layer consists of neurons with orthogonal (preferably orthonormal) activation functions. The activation functions for these neurons belong to the same class of orthogonal functions and no two neurons have the same order of activation function.

The network output is defined by a linear combination of weights

$$\hat{y} = f(z,\hat{w}) = \sum_{n_1=0}^{N_1-1}\sum_{n_m=0}^{N_m-1} \hat{w}_{n_1\ldots n_m}\phi_{n_r\ldots n_m}(z) = \Phi^T(z)\hat{w}, \tag{8}$$

where $z = [z_1, z_2, \ldots, z_m]^T$ is an m-dimensional input vector, N_i is the number of neurons corresponding to the i-th input and $\hat{w}$ is the vector of weights between the hidden and output layers. Functions $\phi_{n_1\ldots n_m}(z)$ are orthogonal functions in m-dimensional space given by

$$\phi_{n_1\ldots n_m}(z) = \prod_{i-1}^{m}\phi_{n_i}(z_i), \tag{9}$$

where $\phi_i(z)$ are one-dimensional orthogonal functions implemented by each hidden layer neuron. For detailed topology of the network see [19]. Examples of orthonormal functions are normalized Fourier (harmonic) functions, Legendre polynomials, Chebyshev polynomials, Laguerre polynomials, Hermite polynomials, Walsh functions, and many others [17].

Due to the fact that the algebraic output of the OAF network satisfies Assumption 1, it is possible to apply the technique described in Section 4.

In this section we have reviewed some of the many possible approaches for identifying nonlinear system. Emphasis was on separability, accuracy and fast identification. In the next section we provide the optimization-based procedure for deriving PWA approximation from identified analytical form.

4. PWA approximation of neural network output

This section illustrates how to solve Problem 2 where the task is to approximate a given nonlinear multivariable function f by a PWA function $\widetilde{f}$ as in (1) such that the approximation error (2) is minimized. Through the section we will assume that the analytical form of f is known and that it satisfies Assumption 1. We remark that such an assumption is fulfilled if f is obtained by the neural network approach of Section 3. First, in Section 4.1 we illustrate how to derive optimal PWA approximation of scalar-valued nonlinear functions in a single variable, i.e. when $f : \mathbb{R} \to \mathbb{R}$. Subsequently, in Section 4.2 we extend the procedure to approximation of multivariable nonlinear functions $f : \mathbb{R}^{n_z} \to \mathbb{R}$. Worth noting is that we will solve Problem 2 by a sequence of one-dimensional approximations.

4.1. Functions in One Variable

First, we consider the one-dimensional case, i.e. approximating a nonlinear function $f : \mathbb{R} \to \mathbb{R}$ by a PWA function $\widetilde{f}(z) = a_i z + c_i$ if $z \in \mathcal{R}_i$, $i = 1, \ldots, N$. Here, the domain Ω of f is assumed to be a line segment $[\underline{z},\ \overline{z}]$. Regions $\mathcal{R}_i$ define the partition of such a line into N non-overlapping parts, i.e. $\mathcal{R}_1 = [\underline{z},\ r_1]$, $\mathcal{R}_2 = [r_1,\ r_2]$, …, $\mathcal{R}_{N-1} = [r_{N-2},\ r_{N-1}]$, $\mathcal{R}_N = [r_{N-1},\ \overline{z}]$ with $\cup_i \mathcal{R}_i = [\underline{z},\ \overline{z}]$. Solving Problem 2 then reduces to find the slopes $a_i \in \mathbb{R}$, offsets $c_i \in \mathbb{R}$ and breakpoints $r_i \in \mathbb{R}$ such that the approximation error is minimized, i.e.

$$\min_{a_i, c_i, r_i} \quad \int_{\underline{z}}^{\overline{z}} (f(z) - \widetilde{f}(z))^2 \, \mathrm{d}z \tag{10a}$$

$$\text{s.t.} \quad \widetilde{f}(z) = \begin{cases} a_1 z + c_1 & \text{if } z \in [\underline{z},\ r_1] \\ \vdots & \vdots \\ a_N z + c_N & \text{if } z \in [r_{N-1},\ \overline{z}] \end{cases} \tag{10b}$$

$$\underline{z} \leqslant r_1 \leqslant \cdots \leqslant r_{N-1} \leqslant \overline{z}, \tag{10c}$$

$$a_i r_i + c_i = a_{i+1} r_i + c_{i+1},\ i = 1, \ldots, N-1, \tag{10d}$$

where (10d) enforces continuity of $\widetilde{f}$ along the breakpoints r_i. The IF-THEN based nonlinear constraint (10b) can be eliminated by observing that, by definition, regions $\mathcal{R}_i$ are

non-overlapping and the integral in (10a) can hence be written as

$$\int_{\underline{z}}^{\overline{z}} \big(f(z)-\widetilde{f}(z)\big)^2\,\mathrm{d}z = \sum_{i=1}^{N}\Big(\int_{r_{i-1}}^{r_i}\big(f(z)-(a_iz+c_i)\big)^2\,\mathrm{d}z\Big), \tag{11}$$

with $r_0=\underline{z}$ and $r_N=\overline{z}$. The NLP (10) can therefore be written as

$$\min_{a_i,c_i,r_i} \quad \sum_{i=1}^{N}\Big(\int_{r_{i-1}}^{r_i}\big(f(z)-(a_iz+c_i)\big)^2\,\mathrm{d}z\Big) \tag{12a}$$

$$\text{s.t.} \quad \underline{z}\leqslant r_1\leqslant\cdots\leqslant r_{N-1}\leqslant\overline{z}, \tag{12b}$$

$$a_ir_i+c_i=a_{i+1}r_i+c_{i+1},\ i=1,\ldots,N-1. \tag{12c}$$

For simple nonlinear functions f, the integral in (12a) can be expressed in analytical form in unknowns a_i,c_i,r_i, along with the corresponding gradients. For more complex expressions, the integrals can be evaluated numerically, e.g. by using the trapezoidal rule. In either case, problem (12) can be solved to find a local optimum e.g. by using the `fmincon` solver of MATLAB. Alternatively, global optimization methods [4, 1] can be used that guarantees that an ε-neighborhood of the global optimum can be found.

Example 1 *Consider the function $f(z)=z^3$ for $-1.5\leqslant z\leqslant 1.5$. The analytic form of the integral* (12a) *is*

$$\sum_{i=1}^{N}\Big(c_i^2(r_i+r_{i-1})+a_ic_i(r_i^2-r_i^2)+\frac{a_i^2}{3}(r_i^3-r_{i-1}^3)-$$
$$-\frac{c_i}{2}(r_i^4-r_{i-1}^4)-\frac{2a_i}{5}(r_i^5-r_{i-1}^5)+\frac{1}{7}(r_i^7-r_{i-1}^7)\Big),$$

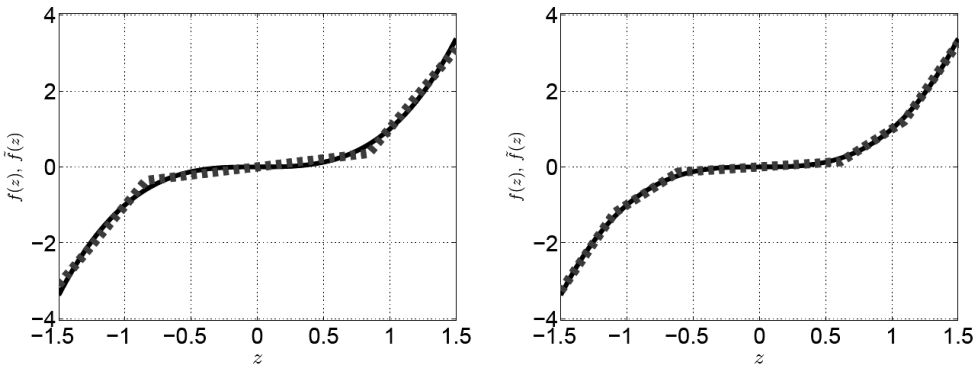
with $r_0=-1.5$ and $r_N=1.5$. The PWA approximation of $f(z)$ with $N=3$ regions was found by solving the NLP (12) *using* `fmincon`*.The obtained PWA approximation is then given by*

$$\widetilde{f}(z)=\begin{cases}4.1797z+3.1621 & \text{if } -1.5\leqslant z\leqslant -0.8423\\ 0.4257z & \text{if } -0.8423\leqslant z\leqslant 0.8423\\ 4.1797z-3.1621 & \text{if } \ 0.8423\leqslant z\leqslant 1.5\end{cases}$$

The approximation accuracy increases by a factor of 10 *when $N=5$ regions is used, as can be seen from Fig. 2.*

4.2. Multivariable Functions

The task is to approximate a given multivariable function $f(z_1,\ldots,z_{n_z}):\mathbb{R}^{n_z}\to\mathbb{R}$ with domain $\Omega\subset\mathbb{R}^{n_z}$ by a PWA function $\widetilde{f}(z_1,\ldots,z_{n_z})$, defined over the same domain, such that the approximation error (2) is minimized.



(a) PWA approximation with 3 regions. (b) PWA approximation with 5 regions.

Figure 2. Graph of $f(z) = z^3$ (solid line) and the PWA approximations $\widetilde{f}(z)$ (dashed lines).

Definition 1 [18] *A function* $f(z_1, \ldots, z_{n_z})$ *is called* separable *if it can be expressed as a sum of functions of a single variable, i.e.* $f(z_1, \ldots, z_{n_z}) = f_1(z_1) + \cdots + f_n(z_{n_z})$.

If $f(z_1, \ldots, z_{n_z})$ is readily separable (e.g. when $f(z_1, z_2) = e^{z_1} + \sin(z_2)$), its optimal PWA approximation can be obtained by applying the 1D scenario of Section 4.1 to the individual components of the function, i.e. $\widetilde{f}(z_1, \ldots, z_{n_z}) = \widetilde{f}_1(z_1) + \cdots + \widetilde{f}_n(z_{n_z})$. The total number of regions over which the PWA approximation $\widetilde{f}$ is defined is hence given by $\sum_{j=1}^{n_z} N_j$, where N_j is the pre-specified complexity of the j-th approximation $\widetilde{f}_j(z_j)$.

Any nonlinear non-separable function f satisfying Assumption 1 can be converted into the separable form by a simple change of variables [18]. To introduce the procedure, consider a non-separable function $f(z_1, z_2) = z_1 z_2$ with domain $\Omega := [\underline{z}_1,\ \bar{z}_1] \times [\underline{z}_2,\ \bar{z}_2]$. Define two new variables

$$y_1 = (z_1 + z_2), \quad y_2 = (z_1 - z_2). \tag{13}$$

Then it is easy to verify that $\frac{1}{4}(y_1^2 - y_2^2) = z_1 z_2$. The coordinate transformation therefore transforms the original function into a separable form, where both terms (y_1^2 and y_2^2) are now functions of a single variable. The procedure of Section 4.1 can thus be applied to compute PWA approximations of $f_{y_1}(y_1) := y_1^2$ and $f_{y_2}(y_2) := y_2^2$, where the function arguments relate to z_1 and z_2 via (13). Note that $f_{y_1}(\cdot)$ and $f_{y_2}(\cdot)$ have different domains, therefore their PWA approximations $\widetilde{f}_{y_1}(y_1) \approx y_1^2$ and $\widetilde{f}_{y_2}(y_2) \approx y_2^2$ will, in general, be different. Specifically, the domain of $f_{y_1}(\cdot)$ is $[\underline{y}_1,\ \bar{y}_1]$ with $\underline{y}_1 = \min\{z_1 + z_2 \mid \underline{z}_1 \leqslant z_1 \leqslant \bar{z}_1,\ \underline{z}_2 \leqslant z_2 \leqslant \bar{z}_2\}$ and $\bar{y}_1 = \max\{z_1 + z_2 \mid \underline{z}_1 \leqslant z_1 \leqslant \bar{z}_1,\ \underline{z}_2 \leqslant z_2 \leqslant \bar{z}_2\}$. Similarly, the domain of $f_{y_2}(\cdot)$ is $[\underline{y}_2,\ \bar{y}_2]$, which boundaries can be computed by respectively minimizing and maximizing $z_1 - z_2$ subject to the constraint $[z_1,\ z_2]^T \in \Omega$. The overall PWA approximation $\widetilde{f}(z_1, z_2) \approx z_1 z_2$ then becomes

$$\widetilde{f}(z_1, z_2) = \tfrac{1}{4}(\widetilde{f}_{y_1}(z_1 + z_2) - \widetilde{f}_{y_2}(z_1 - z_2)). \tag{14}$$

For any points z_1, z_2 the value $\widetilde{f}(z_1, z_2)$ is obtained by subtracting the PWA function $\widetilde{f}_{y_2}(\cdot)$ evaluated at the point $z_1 - z_2$ from $\widetilde{f}_{y_1}(\cdot)$ evaluated at $z_1 + z_2$, followed by a linear scaling.

The procedure naturally extends to multi-variable functions represented by the product of two nonlinear functions of a single variable, i.e. $f(z_1, z_2) = f_1(z_1) f_2(z_2)$. Here, the transformation (13) becomes

$$y_1 = f_1(z_1) + f_2(z_2), \quad y_2 = f_1(z_1) - f_2(z_2). \tag{15}$$

Therefore, $\frac{1}{4}(y_1^2 - y_2^2) = f(z_1, z_2)$ still holds. Let $f_{y_1}(y_1) := y_1^2$ and $f_{y_2}(y_2) := y_2^2$. The domain of $f_{y_1}(\cdot)$ is $[\underline{y}_1, \bar{y}_1]$ and $\mathrm{dom} f_{y_2}(\cdot) = [\underline{y}_2, \bar{y}_2]$ with

$$\underline{y}_1 = \min\{f_1(z_1) + f_2(z_2) \mid [z_1,\ z_2]^T \in \Omega\}, \tag{16a}$$
$$\bar{y}_1 = \max\{f_1(z_1) + f_2(z_2) \mid [z_1,\ z_2]^T \in \Omega\}, \tag{16b}$$
$$\underline{y}_2 = \min\{f_1(z_1) - f_2(z_2) \mid [z_1,\ z_2]^T \in \Omega\}, \tag{16c}$$
$$\bar{y}_2 = \max\{f_1(z_1) - f_2(z_2) \mid [z_1,\ z_2]^T \in \Omega\}, \tag{16d}$$

which can be computed by solving four NLP problems. Finally, since all expressions are now functions of a single variable, the PWA approximations $\widetilde{f}_1(z_1) \approx f_1(z_1)$, $\widetilde{f}_2(z_2) \approx f_2(z_2)$, $\widetilde{f}_{y_1}(y_1) \approx f_{y_1}(y_1)$, and $\widetilde{f}_{y_2}(y_2) \approx f_{y_2}(y_2)$ can be computed by solving the NLP (12). The overall optimal PWA approximation $\widetilde{f}(z_1, z_2) \approx f(z_1, z_2)$ then becomes

$$\widetilde{f}(z_1, z_2) = \tfrac{1}{4}\Big(\widetilde{f}_{y_1}\big(\widetilde{f}_1(z_1) + \widetilde{f}_2(z_2)\big) - \widetilde{f}_{y_2}\big(\widetilde{f}_1(z_1) - \widetilde{f}_2(z_2)\big)\Big). \tag{17}$$

The evaluation procedure is similar as above, i.e., given the arguments z_1 and z_2, one first evaluates $\widetilde{z}_1 = \widetilde{f}_1(z_1)$ and $\widetilde{z}_2 = \widetilde{f}_2(z_2)$. Subsequently, one evaluates $\widetilde{y}_1 = \widetilde{f}_{y_1}(\cdot)$ with the argument $\widetilde{z}_1 + \widetilde{z}_2$, then $\widetilde{y}_2 = \widetilde{f}_{y_2}(\cdot)$ at the point $\widetilde{z}_1 - \widetilde{z}_2$. Finally, $\widetilde{f}(z_1, z_2) = \frac{1}{4}(\widetilde{y}_1 - \widetilde{y}_2)$. Separation of multi-variable functions with more than two terms can be performed in an inductive manner. Consider $f(z_1, z_2, z_3) = f_1(z_1) f_2(z_2) f_3(z_3)$. First, approximate the product $f_1(z_1) f_2(z_2)$ by a PWA function of the form of (17), which requires four PWA approximations

$$\widetilde{f}_1(\cdot) \approx f_1(\cdot),\ \widetilde{f}_2(\cdot) \approx f_2(\cdot),\ \widetilde{f}_{y_1}(\cdot) \approx y_1^2,\ \widetilde{f}_{y_2}(\cdot) \approx y_2^2,$$

with y_1 and y_2 as in (15). Let $f_a(z_1, z_2) := f_1(z_1) f_2(z_2)$. Then $f(z_1, z_2, z_3) = f_a(z_1, z_2) f_3(z_3)$, which can again be approximated as a product of two functions. Specifically, define

$$y_3 = f_a(\cdot) + f_3(z_3), \quad y_4 = f_a(\cdot) - f_3(z_3), \tag{18}$$

and hence $f_a(z_1, z_2) f_3(z_3) = \frac{1}{4}(y_3^2 - y_4^2)$. The domains over which y_3^2 and y_4^2 are to be approximated are, respectively, $[\underline{y}_3, \bar{y}_3]$ and $[\underline{y}_4, \bar{y}_4]$ with

$$\underline{y}_3 = \min\{f_1(z_1) f_2(z_2) + f_3(z_3) \mid z \in \Omega\}, \tag{19a}$$

$$\bar{y}_3 = \max\{f_1(z_1)f_2(z_2) + f_3(z_3) \,|\, z \in \Omega\}, \tag{19b}$$

$$\underline{y}_4 = \min\{f_1(z_1)f_2(z_2) - f_3(z_3) \,|\, z \in \Omega\}, \tag{19c}$$

$$\bar{y}_4 = \max\{f_1(z_1)f_2(z_2) - f_3(z_3) \,|\, z \in \Omega\}, \tag{19d}$$

and $z = [z_1,\ z_2,\ z_3]^T$. Subsequently, three additional PWA approximations

$$\widetilde{f}_{y_3}(y_3) \approx y_3^2,\ \widetilde{f}_{y_4}(y_4) \approx y_4^2,\ \widetilde{f}_3(z_3) \approx f_3(z_3)$$

are to be computed over the corresponding domains. The aggregated optimal PWA approximation $\widetilde{f}(z_1, z_2, z_3) \approx f(z_1)f(z_2)f(z_3)$ consists of 7 individual approximations and is given by

$$\widetilde{f}(\cdot) = {}^1\!/_4\Big(\underbrace{\widetilde{f}_{y_3}\big(\hat{f}_a + \widetilde{f}_3(z_3)\big)}_{\hat{y}_3} - \underbrace{\widetilde{f}_{y_4}\big(\hat{f}_a - \widetilde{f}_4(z_3)\big)}_{\hat{y}_4}\Big). \tag{20}$$

Here, $\hat{f}_a$ is the function value $\widetilde{f}_a(z_1, z_2) \approx f_1(z_1)f_2(z_2)$ at z_1 and z_2, where $\widetilde{f}_a(\cdot)$ is obtained from (17), i.e.:

$$\hat{f}_a = {}^1\!/_4\Big(\underbrace{\widetilde{f}_{y_1}\big(\widetilde{f}_1(z_1) + \widetilde{f}_2(z_2)\big)}_{\hat{y}_1} - \underbrace{\widetilde{f}_{y_2}\big(\widetilde{f}_1(z_1) - \widetilde{f}_2(z_2)\big)}_{\hat{y}_2}\Big). \tag{21}$$

Then the overall PWA approximation $\widetilde{f}(z_1, z_2, z_3)$ can be evaluated, for any $z_1, z_2, z_3 \in \Omega$, by computing function values of respective approximations in the following order:

Step 1: $\hat{y}_1 = \widetilde{f}_{y_1}(\widetilde{f}_1(z_1) + \widetilde{f}_2(z_2))$,

Step 2: $\hat{y}_2 = \widetilde{f}_{y_2}(\widetilde{f}_1(z_1) - \widetilde{f}_2(z_2)$,

Step 3: $\hat{y}_3 = \widetilde{f}_{y_3}({}^1\!/_4(\hat{y}_1 - \hat{y}_2) + \widetilde{f}_3(z_3))$,

Step 4: $\hat{y}_4 = \widetilde{f}_{y_4}({}^1\!/_4(\hat{y}_1 - \hat{y}_2) - \widetilde{f}_3(z_3))$,

Step 5: $\widetilde{f}(z_1, z_2, z_3) = {}^1\!/_4(\hat{y}_3 - \hat{y}_4)$.

Such an inductive procedure can be repeated *ad-infinitum* to derive PWA approximations of any multi-variable function which satisfies Assumption 1. In general, the PWA approximation consists of $2p + n$ individual PWA functions, where n is the number of variables in $f(z_1, \ldots, z_{n_z})$ and p is the number of products between individual subfunctions $f_j(z_j)$. As an example, for $f(\cdot) := \alpha_1 f_1(z_1)f_2(z_2)f_4(z_4) + \alpha_2 f_3(z_3)f_5(z_5)$ we have $p = 3$. We remark that inclusion of scalar multipliers α_j into the PWA description (20)–(21) is straightforward and only requires linear scaling of the corresponding terms.

Remark 1 Since the approximation procedure is always transformed to a linearization of a sequence of one-dimensional functions and such transformation is driven by (17) the overall approximation error will be (directly) proportional to one-fourth of the sum of errors emerging during the approximation of the quadratic functions.

Remark 2 In general, we can guarantee continuity of the final PWA approximation regardless of the state space dimension, because in each case the approximation tools are linear or affine functions of the given independent variable, which are obviously continuous, since the domains for both functions are represented by the set of real numbers.

5. Examples

In this section, two identification experiments were conducted. The proposed algorithm has been implemented in MATLAB and utilized in experiments.

Example 2 (Static model of Current-Voltage characteristics). In this example we use OAF network in the first phase of identification and compare results with two existing approaches. The results of the identification study show the potential of the technique.

A simple modeling problem of a static system is considered. The two-dimensional nonlinear model of a GLASMONT *n*-channel transistor [14] was employed to prepare input-output data for the identification experiment. The Current-Voltage characteristics $I_{ds}(V_{gd}, V_{gs})$ of the system are captured in Fig. 3(a). The indices *ds*, *gd* and *gs* denote *drain-source*, *gate-drain*, and *gate-source*, respectively.

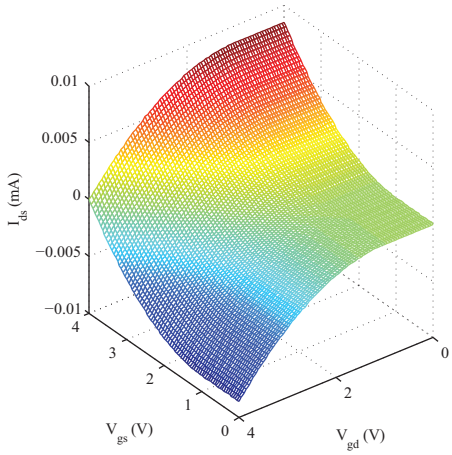
From a modeling point of view the system is quite unique. Prepared data does not exhibit any noise, typical for electrical measurements. This example is provided to illustrate and compare capabilities of the proposed modeling technique.

Chebyshev polynomials up to the fourth order were used in OAF network for this experiment. The first step of the identification gave the following formula of the OAF network

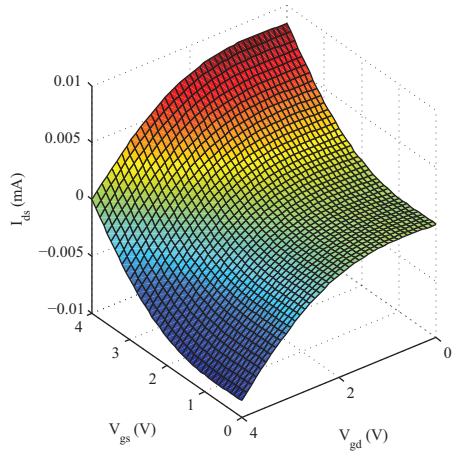
$$\begin{aligned} I_{\mathrm{OAF}}(V_{gd}, V_{gs}) = {} & 0.072\,{V_{gd}}^4 - 0.065\,{V_{gd}}^3 - 0.46\,{V_{gd}}^2 - 0.43\,V_{gd} \\ & - 0.072\,{V_{gs}}^4 + 0.065\,{V_{gs}}^3 + 0.46\,{V_{gs}}^2 + 0.43\,V_{gs} - 0.49 \cdot 10^{-15} \end{aligned} \tag{22}$$

The OAF model (22) has no mixed product terms, therefore it is in a Generalized Fourier Series form.

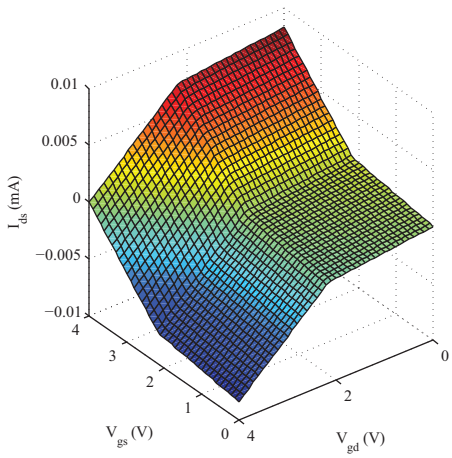
In the second step, four different approximations were performed, corresponding to four different subdivisions of the domain. The approximation results are compared with existing approaches [5, 7]. Tab. 5 compares three quantitative parameters of the identification. It summarizes final number of linear regions of the PWA model *Reg*, mean square error *MSE* and computation time *T*. The proposed technique is labeled as PWA-OAF in the table. Due to the fact that breaking points are defined for each dimension independently in case of PWA-OAF model, the results in each row of Tab. 5 are comparable. Results of HL-CPWL technique are taken from [7] without computational time. The results of K-mean cluster technique were obtained from simulation using HIT toolbox [5]. The clustering technique failed to complete in case of many linear regions. In Fig. 3, the nonlinear function I_{ds} and its OAF and PWA-OAF approximations are shown (each case corresponding to a given subdivision).



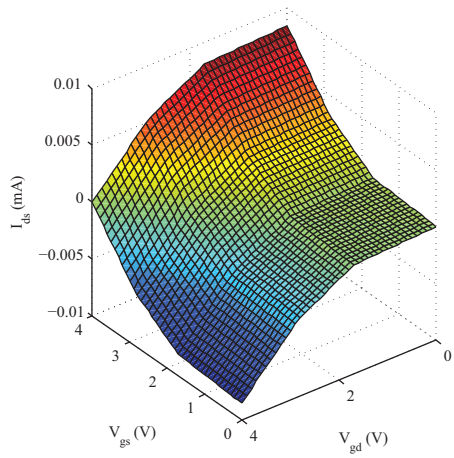
(a) The I-V characteristics of the GLASMONT nonlinear model [14].



(b) The I-V characteristics of the OAF approximation.



(c) The I-V characteristics of the PWA-OAF model (4 regions).



(d) The I-V characteristics of the PWA-OAF model (6 regions).

Figure 3. The Current-Voltage characteristics approximation results.

It is first observed that the proposed two-stage approximation does not raise MSE. The OAF technique provides tunable accuracy of the polynomial model (22). Its accuracy is influenced mainly by the number of neurons in hidden layer.

It is observed that independent linearization segments in each dimension are in general beneficial for convergence. The accuracy of approximation improves with increasing number of segments. Due to the convex training of the OAF network the overall identification process exhibits favorable computational times.

Table 1. Setup, results and comparison (Example 2).

PWA OAF model				HL CPWL model [7]			HIT-PWA model [5]		
Case	Reg.	MSE	T [s]	Reg.	MSE	T [s]	Reg.	MSE	T [s]
1	4	1.012e-7	2.159	4	9.31e-4	–	4	1.103e-7	19.96
2	6	1.659e-8	2.48	–	–	–	9	6.022e-8	39.93
3	8	5.4e-9	4.23	16	2.23e-4	–	16	5.22e-8	70.29
4	12	2.457e-9	4.72	36	5.6e-5	–	–	–	–

Example 3 (Narendra-Li Benchmark system). In this example we consider identification of a complex nonlinear discrete-time system with one input and one output. The system was originally proposed and discussed by Narendra and Li in [11] and has been considered in numerous discrete-time identification examples.

The discrete-time equations of the Narendra-Li system are:

$$\begin{aligned} x_1^+ &= \left(\frac{x_1}{1+x_1^2}+p_1\right)\sin(x_2) \\ x_2^+ &= x_2\cos(x_2)+x_1 e^{\frac{-(x_1^2+x_2^2)}{p_2}}+\frac{u^3}{1+u^2+p_3\cos(x_1+x_2)} \\ y &= \frac{x_1}{1+p_4\sin(x_2)+p_5\sin(x_1)} \end{aligned} \tag{23}$$

Two input-output data records with 300 samples each, one for estimation and one for validation purposes were prepared. The system was excited by a harmonic signal $u(t) = \sin(2\pi t/10)+\sin(2\pi t/25)$ which varies within the bounds $[-2,2]$ for $t = 0,1,\ldots,299$ seconds.

As in the previous example, the basis function expansion with OAF was use in the first step of the procedure. Chebyshev polynomials up to third order were used in OAF network for this experiment. The first step of the identification gave the following formula of the network

$$\begin{aligned} y(k) = {} & 0.41\,y^3(k-1)-y^2(k-1)\,(0.133\,u(k-1)+0.061) \\ & -y(k-1)\,\left(0.16\,u^2(k-1)+0.339\,u(k-1)-0.344\right) \\ & +1.16\,u^2(k-1)-1.55\,u^3(k-1)+1.196\,u(k-1)-0.13 \end{aligned} \tag{24}$$

The output consists of two mixed terms and two terms of single variable.

In second step of the approximation, each term was approximated individually. Three different approximations were performed, corresponding to different number of linearization regions for each term. Tab. 2 summarizes mean square error *MSE*, computational time T, and final number of regions for each term individually. Results in Tab. 2

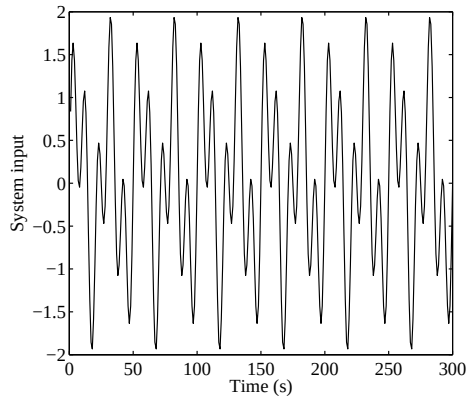
Table 2. Setup and results (Example 3).

PWA OAF model						
Case	MSE	Time	Regions			
			term 1	term 2	term 3	term4
1.	3.43e-1	9.68	15	15	7	7
2.	3.79e-1	8.67	13	13	5	5
3.	3.2e-0	4.23	8	8	2	2

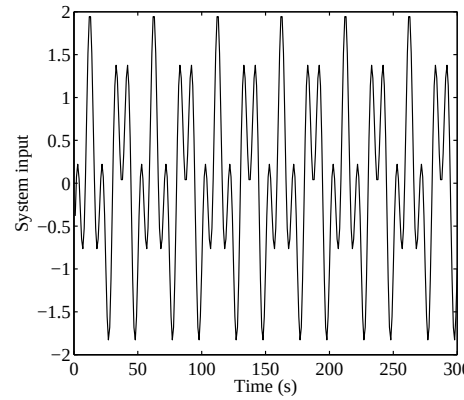
relate to model excited by validation signal $u(t) = \sin(2\pi t/10) - \sin(2\pi t/25)$. The experiment showed good generalization capabilities of the identified model by feeding in an unknown signal input not previously used for training (Fig. 4). The complexity of the final model depends on the segmentation of the algebraic formula (24). Yet, the mixed terms of the algebraic formula are linearized independently. It is possible to achieve a lower number of linear regions by approximation in the same points. It is also worth to mention that existing identification techniques which support identification of dynamic systems [5, 13] failed to converge.

6. Conclusion

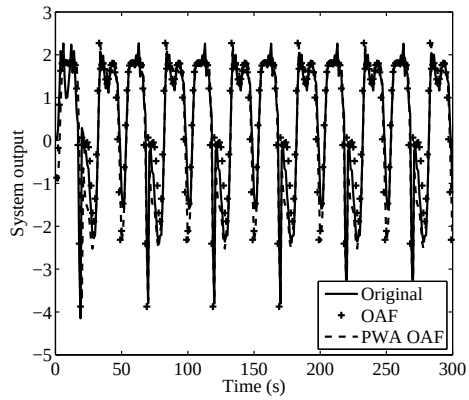
Piecewise affine (or linear) functions and approximations have many applications in global optimization, non-linear control, pattern detection or function compression. We have proposed a two step an approximation method for static and dynamic systems that transforms them into PWA systems of fixed complexity. The original system behavior is characterized by input-output measurements. The first step of the procedure identifies the system by a basis function expansion network. In the second step, we solve a series of one-dimensional problems to approximate the nonlinear system by PWA model. The procedure is optimal in regard to objectives in both stages. However, two steps of the method can introduce addition approximation error into final model. Also, no stability analysis of the proposed method was investigated. Numerical examples suggest, however, that the method works very well in practice. It also competes favorably with existing approximation methods in accuracy, number of linearization segments, and computational complexity.



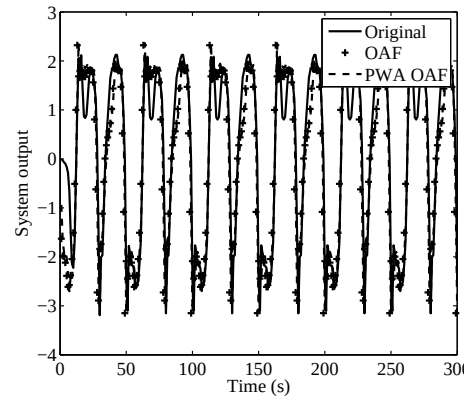
(a) Input training signal.



(b) Input validation signal.



(c) Time responses of the Narendra-Li system, OAF network and PWA-OAF model for training input signal.



(d) Time responses of the Narendra-Li system, OAF network and PWA-OAF model for validation input signal.

Figure 4. Narendra-Li benchmark system characteristics, example 3.

References

[1] C.S. ADJIMAN, I.P. ANDROULAKIS, C.D. MARANAS and C.A. FLOUDAS: A global optimization method, αBB for process design. *Computers & Chemical Engineering*, **20**(96), (1996), 419-424.

[2] A. BEMPORAD and M. MORARI: Control of systems integrating logic, dynamics, and constraints. *Automatica*, **35**(3), (1999), 407-427.

[3] M.S. BRANICKY: Studies in Hybrid Systems: Modeling, Analysis, and Control. PhD thesis, Massachusetts Institute of Technology, 1995.

[4] B. CHACHUAT, A.B. SINGER and P.I. BARTON: Global methods for dynamic optimization and mixed-integer dynamic optimization. *Industrial & Engineering Chemistry Research*, **45**(25), (2006), 8373-8392.

[5] G. FERRARI-TRECATE: Hybrid Identification Toolbox (HIT), 2005.

[6] J.H. FRIEDMAN: Multivariate adaptive regression splines. *Annals of Statistics*, **19**(1), (1991), 1-67.

[7] P. JULIAN, A. DESAGES and O. AGAMENNONI: High-level canonical piecewise linear representation using a simplicial partition. *IEEE Trans. on Circuits and Systems I: Fundamental Theory and Applications*, **46**(4), (1999), 463-480.

[8] M. KVASNICA, M. FIKAR and A. SZUCS: Automatic derivation of optimal piecewise affine approximations of nonlinear systems. In: *Proc. of the 18th IFAC World Congress*, (2011), 8675-8680.

[9] L. LJUNG: Approaches to identification of nonlinear systems. In: *Proc. of the 29th Chinese Control Conf. CCC 2010*, (2010), 1-5.

[10] A. MAGNANI and S.P. BOYD: Convex piecewise-linear fitting. *Optimization and Engineering*, **10**(1), (2008), 1-17.

[11] K.S. NARENDRA: Neural networks for control theory and practice. *Proc. of the IEEE*, **84**(10), (1996), 1385-1406.

[12] J.C.PATRA and A.C. KOT: Nonlinear dynamic system identification using Chebyshev functional link artificial neural networks. *IEEE Trans. on Systems Man and Cybernetics. Part B*, **32**(4), (2002), 505-511.

[13] J. ROLL, A. BEMPORAD and L. LJUNG: Identification of piecewise affine systems via mixed-integer programming. *Automatica*, **40**(1), (2004), 37-50.

[14] M.F. SEVAT: GLASMOST: A mosfet model of high numerical quality. In: *Int. Symp. on Circuits and Systems*, (1988), 2597-2600.

[15] E.D. SONTAG: Nonlinear regulation: The piecewise linear approach. *IEEE Trans. on Automatic Control*, **26**(2), (1981), 346-358.

[16] F.D. TORRISI and A. BEMPORAD: HYSDEL - A tool for generating computational hybrid models for analysis and synthesis problems. *IEEE Trans. on Control Systems Technology*, **12**(2), (2004), 235-249.

[17] V. TOTIK: Orthogonal polynomials. *J. of Computational and Applied Mathematics*, **0427** (2005), 84-86.

[18] H.P. WILLIAMS: Model Building in Mathematical Programming. Third edition. Volume 4 of *A Wiley - Interscience Publication*. John Wiley & Sons, 1999.

[19] C. ZHU, D. SHUKLA and F.W. PAUL: Orthogonal functions for systems identification and control. In C.T. Leondes (Ed.) Control and Dynamic Systems. Academic Press, London, 1997, 1-73.