Proceedings of the 7th International Conference on Applied Informatics
Eger, Hungary, January 28–31, 2007. Vol. 2. pp. 359–366.

Identification of dynamic systems by hinging hyperplane models

Tamas Kenesei, Balazs Feil, Janos Abonyi

University of Pannonia
e-mail: abonyij@fmt.uni-pannon.hu

Abstract

This article deals with the identification of nonlinear dynamic systems by hinging hyperplane models, which are represented by tree structured piecewise linear models.

This type of non-linear black-box models is relatively new, and its identification and application in the modeling of dynamic systems are not thoroughly examined and discussed so far. They can be an alternative to artificial neural nets but there is a clear need for an effective identification method, because the original identication algorithm given by Breimann suffers from convergency and range problems.

This paper presents a new identification technique based on a fuzzy clustering technique called Fuzzy c-Regression Clustering. This clustering technique applies linear models as prototypes and the model parameters and fuzzy membership degrees are identified simultaneously. To use this clustering procedure for the identification of hinging hyperplanes, there is a need to handle restrictions about the relative location of the hyperplanes: they should intersect each other in the operating regime covered by the data points.

After the theoretical survey the paper gives detailed technical analysis of the proposed technique with the help of the identification of nonlinear process systems.

Keywords: Hinging hyperplanes, fuzzy c-regression clustering, piecewise linear models, dynamic models, regression tree

1. Introduction

A lot of nonlinear regression techniques have been worked out so far (splines, artificial neural networks etc.). This article proposes a method for piecewise linear model identification applying hinging hyperplanes as linear submodels. Hinging hyperplane model is proposed by Breiman [3] and identification of this type of non-linear models is several times reported in the literature, because the original algorithm developed by Bremain suffers from convergency and range problems [8, 7]. Methods like the penalty of hinging angle were proposed to improve Breiman's algorithm [6], or Gauss-Newton algorithm can be used to obtain the final non-linear model [7]. The main goal of this paper is to present a new method for hinging hyperplane model identification with Fuzzy c-Regression Models. This approach yields simultaneous estimation of the parameters of c regression models, together with fuzzy partitioning of the data and as this paper presents with appropriate constrains FCRM can be used for hinge identification if $c = 2$.

In the application example the proposed hinge function based model will be used for dynamic system modeling. The Non-linear AutoRegressive with eXogenous input (NARX) model is frequently used with many non-linear identification methods, such as neural networks and fuzzy models (see in [1]). This will be used also in this paper to approximate the behavior of a water heater. This paper is organized as follows. Section 2 discusses hinge function approximation, and how the constrains can be incorporated into the FCRM identification approach. After that the resulted tree structured piecewise linear model is described. In Section 3.1 an application example is presented, and Section 4 concludes the paper.

2. Non-linear regression with Hinge functions and Fuzzy c-regression clustering

This section gives a brief description about what the hinging hyperplane approach means on the basis of [6], followed by the introduction of FCRM clustering, and describing how the constrains can be incorporated into the clustering process.

2.1. Function approximation with Hinge functions

Suppose two hyperplanes are given by: $y_k = \mathbf{x}_k^T \theta^+, y_k = \mathbf{x}_k^T \theta^-$, where $\mathbf{x}_k = [x_{k,0}, x_{k,1}, x_{k,2}, \dots, x_{k,n}]$ ($x_{k,0} \equiv 1$) is the kth regressor vector and y_k is the kth output variable ($k = 1, \dots, N$). These two hyperplanes are continuously joined together at $\{\mathbf{x} : \mathbf{x}^T (\theta^+ - \theta^-) = 0\}$ as can be seen in Figure 1. As a result they are called *hinging hyperplanes*. The joint $\triangle = \theta^+ - \theta^-$, multiples of $\triangle$ are defined *hinge* for the two hyperplanes, $y_k = \mathbf{x}_k^T \theta^+$ and $y_k = \mathbf{x}_k^T \theta^-$. The solid part of the two hyperplanes explicitly given by $y_k = \max(\mathbf{x}_k^T \theta^+, \mathbf{x}_k^T \theta^-)$ or $y_k = \min(\mathbf{x}_k^T \theta^+, \mathbf{x}_k^T \theta^-)$. For a sufficiently smooth function $f(\mathbf{x}_k)$, the approximation with hinge functions can get arbitrarily close if sufficiently large number of hinge functions are used. The sum of the hinge functions $\sum_{i=1}^{K} h_i(\mathbf{x}_k)$ constitutes a continuous piecewise linear function. The number of input variables n in each hinge function and the number in hinge functions K are two variables to be determined. The explicit form for representing a function $f(\mathbf{x}_k)$ with hinge functions becomes

$$f(\mathbf{x}_k) = \sum_{i=1}^{K} h_i(\mathbf{x}_k) = \sum_{i=1}^{K} \langle \max | \min \rangle \left(\mathbf{x}_k^T \theta_i^+, \mathbf{x}_k^T \theta_i^- \right) \tag{2.1}$$

where $\langle \max | \min \rangle$ means max or min.

2.2. Hinge search as an optimization problem

The essential hinge search problem can be viewed as an extension of the linear least-squares regression problem.

Given N data pairs as $\{\mathbf{x}_1, y_1\}$,$\{\mathbf{x}_2, y_2\}$,. . .,$\{\mathbf{x}_N, y_N\}$ from a function (linear or non-linear) $y_k = f(\mathbf{x}_k)$, the linear least-squares regression aims to find the best parameter vector $\widehat{\theta}$, by minimizing a quadratic cost function. The optimalization problem is deeply discoussed in [6], and in spite of using Gauss – Newton method to solve the optimalization problem, the proposed identification algorithm applies a much simpler optimization method, the so called alternating optimization which is a heuristic optimization technique and has been applied for several decades for many purposes, therefore it is an exhaustively tested method in non-linear parameter and structure identification as well. Within the hinge function approximation approach, the two linear submodels can be identified by the weighted linear least-squares approach, but their operating regimes (where they are valid) are still an open question.

For that purpose the FCRM method was used which is able to partition the data and determine the parameters of the linear submodels simultaneously. In this way, with the application of the alternating optimization technique and taking advantage of the linearity in $(y_k - \mathbf{x}_k^T\theta^+)$ and $(y_k - \mathbf{x}_k^T\theta^-)$, an effective approach is given for hinge function identification. The proposed procedure is attractive in the local minima point of view as well, because in this way although the problem is not avoided but transformed into a deeply discussed problem, namely the cluster validity problem. In the following section this method is discussed in general, and in Section 2.4 the hinge function identification and FCRM method are joined together.

2.3. Constrained prototype based FCRM

Fuzzy c-regression models, deeply discussed in the literature, yield simultaneous estimates of parameters of c regression models together with a fuzzy c-partitioning of the data. It is an important question how to incorporate constrains into the clustering procedure. These constrains can contain prior knowledge, or like in the hinge function identification approach, restrictions about the structure of the model (the relative location of the linear submodels).

This section deals with prototypes linear in the parameters. Therefore the parameters can be estimated by linear least-squares techniques. When linear equality and inequality constraints are defined on these prototypes, quadratic programming (QP) has to be used instead of the least-squares method. This optimization problem still can be solved effectively compared to other constrained nonlinear optimization algorithms.

The parameter constraints can be grouped into three categories:

- **Local constrains** are valid only for the parameters of a regression model, $\mathbf{\Lambda}_i \boldsymbol{\theta}_i \leqslant \boldsymbol{\omega}_i$.

- **Global constrains** are related to all of the regression models, $\mathbf{\Lambda}_{gl}\boldsymbol{\theta}_i \leqslant \boldsymbol{\omega}_{gl}$, $i = 1, \ldots, c$.
- **Relative constrains** define the relative magnitude of the parameters of two or more regression models,

$$\mathbf{\Lambda}_{rel,i,j} \begin{bmatrix} \boldsymbol{\theta}_i \\ \boldsymbol{\theta}_j \end{bmatrix} \leqslant \boldsymbol{\omega}_{rel,i,j} \tag{2.2}$$

For a throughout discussion how these constrains can be incorporated into the identification approach (see [1]).

2.4. Improvements of Hinge identification

For hinge function identification purposes, two prototypes have to be used by FCRM ($c = 2$), and these prototypes must be linear regression models. However, these linear submodels have to intersect each other within the operating regime covered by the known data points (within the hypercube expanded by the data). This is a crucial problem in the hinge identification area [6]. To take into account this point of view as well, constrains have to be taken into consideration as follows. Cluster centers $\mathbf{v}_i$ can also be computed from the result of FCRM as the weighted average of the known input data points

$$\mathbf{v}_i = \frac{\sum_{k=1}^{N} \mathbf{x}_k \mu_{i,k}}{\sum_{k=1}^{N} \mu_{i,k}} \tag{2.3}$$

where the membership degree $\mu_{i,k}$ is interpreted as a weight representing the extent to which the value predicted by the model matches y_k. These cluster centers are located in the "middle" of the operating regime of the two linear submodels. Because the two hyperplanes must cross each other (see also Figure 1), the following criteria can be specified:

$$\begin{aligned} &\mathbf{v}_1(\theta_1 - \theta_2) < 0 \quad \text{and} \quad \mathbf{v}_2(\theta_1 - \theta_2) > 0 \quad \text{or} \\ &\mathbf{v}_1(\theta_1 - \theta_2) > 0 \quad \text{and} \quad \mathbf{v}_2(\theta_1 - \theta_2) < 0. \end{aligned} \tag{2.4}$$

These relative constrains can be used to take into account the constrains above:

$$\mathbf{\Lambda}_{rel,1,2} \begin{bmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \end{bmatrix} \leqslant 0 \text{ where } \mathbf{\Lambda}_{rel,1,2} = \begin{bmatrix} \mathbf{v}_1 & -\mathbf{v}_1 \\ -\mathbf{v}_2 & \mathbf{v}_2 \end{bmatrix} \tag{2.5}$$

according to (2.2) and (2.4).

So far, the hinge function identification method is presented. The proposed technique can be used to determine the parameters of one hinge function. In general, there are two method to construct a piecewise linear model: additive and tree structured models [6]. In this paper the later will be used since the resulting binary tree structured hinge functions can have a simpler form to interpret and more convenient structure to implement.

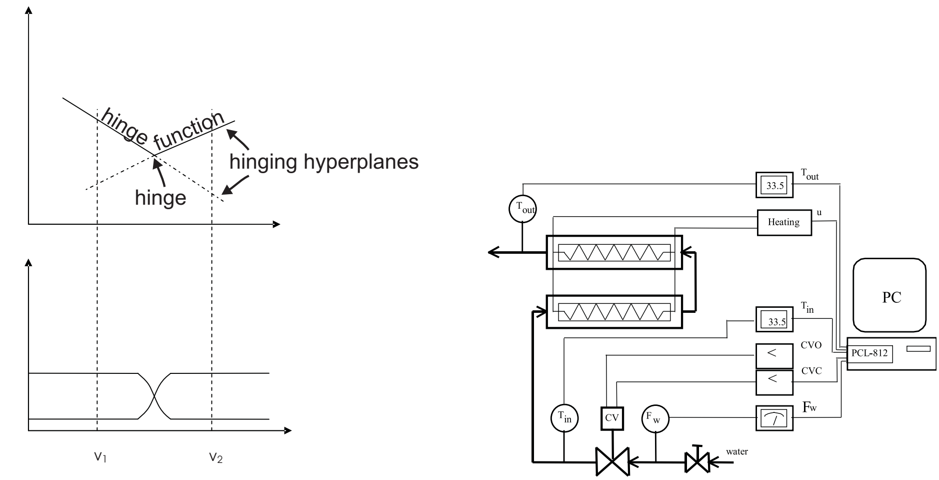

Figure 1: Basic definitions and hinge identification restrictions (left) and the simulated water heater (right)

3. Identification of dynamical systems by hinging hyperplane models

So far, *a general nonlinear modeling technique* was presented and a new identification approach was given for hinging hyperplane based nonlinear models: $\widehat{y} = f(\mathbf{x}(k), \theta)$ where $f(.)$ represents the hinging hyperplane based tree structured model and $\mathbf{x}(k)$ represents the input vector of the model. To identify a discrete-time *input-output model* for a dynamical system, the dynamic model structure has to be chosen or determined beforehand. A possible often applied structure is nonlinear autoregressive model with exogenous input (NARX) where the input vector of the model $\mathbf{x}(k)$ contains the delayed inputs and outputs of the system to be modeled [1]. In several practical cases a simpler and more specific model structure can be used to approximate the behavior of the system, which fits better the structure of the system. Therefore, it can be an advantage for the identification approach (models with simpler structure can be identified easier), and this model can be more accurate. One such special case of the NARX model is the Hammerstein model (see Figure 2), where the same static nonlinearity f is defined for all of the delayed control inputs (for the sake of simplicity, SISO models are considered):

$$\widehat{y} = \sum_{i=1}^{n_a} a_i y(k-i) + \sum_{j=1}^{n_b} b_j f(u(k-j)) \tag{3.1}$$

where $y()$ and $u()$ are the output and input of the system, respectively, and n_a and n_b are the output and input orders of the model. The parameters of the blocks of the Hammerstein model (static nonlinearity and linear dynamics) can be identified

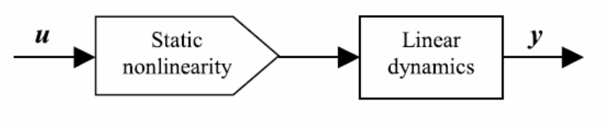

Figure 2: Hammerstein system

by the proposed method simultaneously if the same linear dynamic behavior can be guaranteed by all of the local hinging hyperplane based models. It can be done in an elegant way utilizing the flexibility of the proposed identification approach: global constrains can be formulated for the a_i and b_j parameters of the local models (for a detailed discussion what constrains have to be formulated, see [1]). In the following, the hinging hyperplane modeling technique is applied on a Hammerstein type system, and it will be shown that it is an effective tool for that purpose.

3.1. Application to the identification of Hammerstein systems

Modeling of a simulated water heater (Figure 1) is used to illustrate the advantages of the proposed hinging hyperplanes based models. The water flows through a pair of metal pipes containing a cartridge heater. The outlet temperature, T_{out}, of the water can be varied by adjusting the heating signal, u, of the cartridge heater [1]. The performance of the cartridge heater is given by:

$$Q(u) = Q_M \left[u - \frac{\sin(2\pi u)}{2\pi} \right] \tag{3.2}$$

where Q_M is the maximal power and u is the heating signal (voltage). As the equation above shows the heating performance is a static nonlinear function of the heating signal. Hence, the Hammerstein model is a good match to this process. The aim is to construct a dynamic prediction model from data for the output temperature (the dependent variable, $y = T_{out}$) as a function of the control input: the heating signal. The parameters of the Hammerstein model were chosen as $n_a = n_b = 2$. The performance of this modeling technique will be compared to linear and feedforward neural network models.

The modeling performances can be seen in Table 1. In this example a hinge function based tree with 4 leaves were generated. For the robust testing of the performance of the model building algorithm, 10 fold cross validation method is used. For comparison, a feedforward neural net and linear model was also trained and tested using the same data. The neural net contains one hidden layer with 4 neurons using tanh basis functions. As can be seen from the results, the training and test error are comparable with the errors of the proposed method. A very rigorous test of NARX models is free run simulation because the errors can be cumulated. It can be also seen on Figure 3 that the identified models (the proposed ones, linear

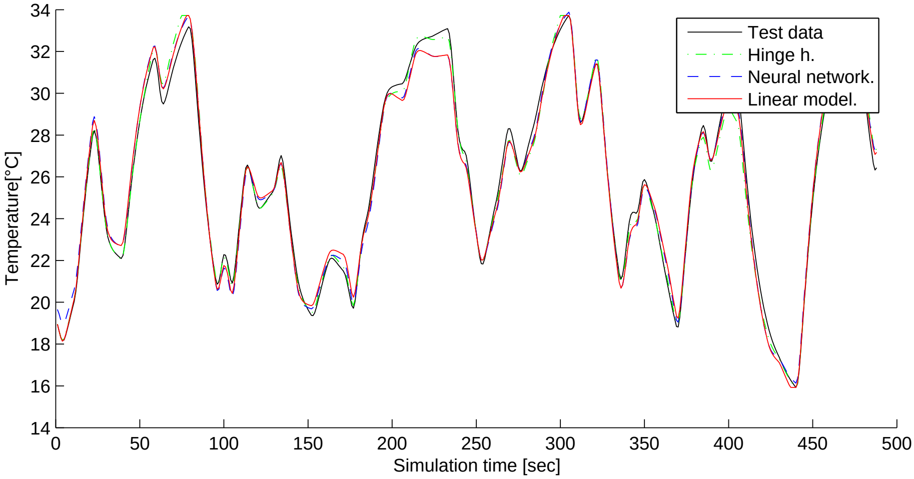

Figure 3: Free run simulation by the Hammerstein system (proposed hinge model, neural network, linear model)

	Training error	Test error	Free run
Linear model	0.0393	0.0449	0.387
Neural network	0.0338	0.0403	0.356
This paper	0.0367	0.0417	0.359

Table 1: Mean square errors of the generated models

models and the neural nets) perform very good also in free run simulation (the system output and the simulated ones can hardly be distinguished). Although the neural net seems to be more robust in this example, the proposed hinge model is much more interpretable than the neural net [5]. This confirms that both the proposed clustering based constrained optimization strategy and the hierarchial model structure has advantages over the classical gradient-based optimization of global hinging hyperplane models.

4. Conclusion

In this paper, a new approach was proposed to the identification of hinging hyperplane based models. The proposed Fuzzy c-Regression Clustering based technique can be used to determine the parameters of two hyperplanes that cross each other within the data space covered by samples. For that purpose, constrains have to be taken into account by clustering. This identification technique avoids the problems of the original hinging hyperplane identification method proposed by Breiman. To get a general nonlinear model, a binary tree structured model has been used where the internal nodes split the data space linearly based on the identified hinge. It has been shown that the flexibility of the identification method, i.e. the ability to take constrains into account, can also be utilized in the identification of Hammerstein systems. The simultaneous identification of the blocks of the Hammerstein model can be achieved in an elegant way. Modeling results for a laboratory water-heater have shown that this type of model can be an alternative to feedforward neural networks. An advantage of the proposed model structure is that by using hinge hyperplanes in a binary tree structure, the obtained model remains still interpretable.

Acknowledgments. The authors would like to acknowledge the sup-port of the Cooperative Research Centre (VIKKK) (project 2004-I) and Hungarian Research Found (OTKA T049534). Janos Abonyi is grateful for support of the Bolyai Fellowship of the Hungarian Academy of Sciences and the Öveges Fellowship of National Office of Re-search and Technology.

References

[1] Abonyi, J., Fuzzy model identification for control, *Birkhauser, Boston*, (2003).

[2] Bezdek, J., Pattern recognition with fuzzy objective function algorithms, *New York: Plenum*, (1981).

[3] Breiman, L., Hinging hyperplanes for regression, classification and function approximation, *IEEE Transactions on Information Theory*, (1993) Vol. 39, 3, 311–325.

[4] Hathaway, R., Bezdek, J., Switching regression models and fuzzy clustering, *IEEE Transactions on fuzzy systems*, 1 (3) (1993) 195–204.

[5] Nelles, O., Nonlinear system identification, *Springer-Verlag*, (2001).

[6] Niu, S., Pucar, P., Hinging hyperplanes for non-linear identification, `http://www.control.isy.liu.se`, (1995).

[7] Pucar, P., Sjoberg, J., Parameterization and conditioning of hinging hyperplane models, *Technical Report LiTH-ISY-R-1809*, Department of Electrical Engineering, Linkoping University, Linkoping, Sweden (ftp 130.236.24.1.), (1995).

[8] Ramirez, D. R., Camacho, E. F., Arahal, M. R., Implementation of min-max MPC using hinging hyperplanes. Application to a heat exchanger, *Control Engineering Practice*, (2004) Vol. 12, 1197–1205.

[9] Roll, J., Bemporad, A., Ljung, L., Identification of piecewise affine systems via mixed-integer programming, *Automatica*, Vol. 40, (2004), 37–50.

[10] Sjoberg, J., Nonlinear Toolbox, Version 1.0.