Identification of Hinging Hyperplane Models by Fuzzy c-Regression Clustering

Tamas Kenesei, Balazs Feil, Janos Abonyi

University of Pannonia, Department of Process Engineering,
P.O. Box 158, H-8201 Veszprem, Hungary, abonyij@fmt.uni-pannon.hu
www.fmt.vein.hu/softcomp

Abstract. *This article deals with the identification of the so called hinging hyperplane models. This type of non-linear black-box models is relatively new, and its identification is 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. This paper presents a new identification technique for that purpose that is 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. The proposed method identifies a hinging hyperplane model first that contains two linear submodels, and after that the two halves of the model (the two linear hyperplanes) are treated separately: two other hinging hyperplane models are identified on the basis of the operating regions of the first two linear submodels. Following these steps, a tree structured piecewise linear model is identified, where the branches correspond to linear division of the operating regime, and the leaves correspond to linear models. In this way a piecewise linear model is constructed.*

1 Introduction

The problem of nonlinear function approximation has attracted much attention during the past years [1], because in real life data sets the relationship between the input and output variables is often nonlinear, which can be obtained via nonlinear regression.

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. A piecewise linear model [2,3] contains many linear submodels, each of them operates a specific range of the whole operating region. The proposed approach applies hinging hyperplanes as linear submodels. Hinging hyperplane model is proposed by Breiman [4], and several application examples have been published in the literature, e.g. it can be used in model predictive control [5], or identification of piecewise affine systems via mixed-integer programming [6].

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 [5,7]. Methods like the penalty of hinging angle were proposed to improve Breiman's algorithm [1], or Gauss-Newton algorithm can be used to obtain the final non-linear model [8]. Clustering technique has been applied for identification purposes as well in [9] and [3].

The main goal of this paper is to present a new method for hinging hyperplane model identification. The proposed technique uses the Fuzzy c-Regression Clustering. The Fuzzy c-Regression Model (FCRM) approach yields simultaneous estimation of the parameters of c regression models, together with fuzzy partitioning the data. In this clustering, the cluster prototypes are functions instead of geometrical objects (like points as in Fuzzy c-Means clustering, or ellipsoids as in Gath-Geva clustering). Therefore, if the number of prototypes c is equal to two, FCRM can be used to identify hinging hyperplanes, if the relative location of the two linear regression models correspond to a hinge function, in other words: they should intersect each other in the current operating region filled by the data point available. For that purpose, constraints must be taken into account within the clustering procedure. Taking constrains into account is important because in this way even prior knowledge can be incorporated and several problems can be effectively solved in the clustering procedure (e.g. good initialization, avoiding local minima and determining the number of clusters). In this paper a method is proposed with which the constrains can be incorporated in the clustering procedure, and this is used within the hinge function identification approach. The proposed clustering based hinge function identification approach uses the alternating optimization technique to determine the parameters of the model because this is used by the applied fuzzy clustering technique like several other partitioning clustering methods [10]. It 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.

This paper is organized as follows. Section 2 discusses hinge function approximation, the applied Fuzzy c-Regression Clustering technique, and how the constrains can be incorporated into the identification approach. After that the resulted tree structured piecewise linear model is described. In Section 3 some application examples are 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 [1], followed by the Fuzzy c-Regression Model (FCRM) clustering definitions.

2.1 Function Approximation with Hinge Functions

Suppose two hyperplanes are given by:

$$y_k = \mathbf{x}_k^T \theta^+, y_k = \mathbf{x}_k^T \theta^- \tag{1}$$

where $\mathbf{x}_k = [x_{k,0}, x_{k,1}, x_{k,2}, \ldots, x_{k,n}]$, $x_{k,0} \equiv 1$ is the kth regressor vector and y_k is the kth output variable ($k = 1, \ldots, N$). 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/shaded part of the two hyperplanes explicitly given by

$$y_k = \max(\mathbf{x}_k^T \theta^+, \mathbf{x}_k^T \theta^-) \text{ or } y_k = \min(\mathbf{x}_k^T \theta^+, \mathbf{x}_k^T \theta^-) \tag{2}$$

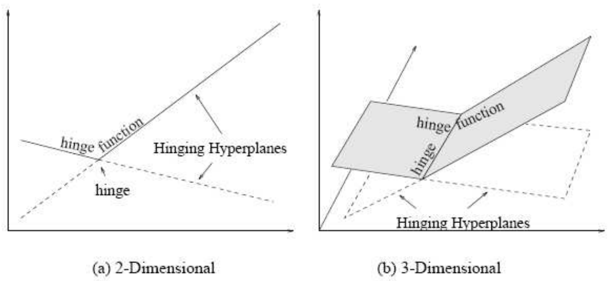

Fig. 1. Basic hinge definitions

For a sufficiently smooth function $f(\mathbf{x}_k)$, which can be linear or non-linear, assuming that the regression data $\{\mathbf{x}_k, y_k\}$ are available for $k = 1, \ldots, N$ and assuming that $\hat{f}(\mathbf{w})$ is the Fourier transform of $f(\mathbf{x})$, then the function $f(\mathbf{x}_k)$ can be represented as the sum of a series of hinge functions $h_i(\mathbf{x}_k), i = 1, 2, \ldots, K$ are defined as the hinge function. 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

$$f(\mathbf{x}_k) = \sum_{i=1}^{K} h_i(\mathbf{x}_k) = \sum_{i=1}^{K} \langle \max | \min \rangle \left(\mathbf{x}_k^T \theta_i^+, \mathbf{x}_k^T \theta_i^- \right) \tag{3}$$

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 $\{\mathbf{x}_1, y_1\}, \{\mathbf{x}_2, y_2\}, \ldots, \{\mathbf{x}_N, y_N\}$ from a function (linear or non-linear)

$$y_k = f(\mathbf{x}_k) \tag{4}$$

the linear least-squares regression aims to find the best parameter vector $\widehat{\theta}$, by minimizing a quadratic cost function

$$\widehat{\theta} = \arg\min_{\theta} \sum_{k=1}^{N} \left(y_k - \mathbf{x}_k^T \theta\right)^2 \tag{5}$$

with which, the regression model gives the best linear approximation to y. For nonsingular data matrix

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_N^T \end{bmatrix} \tag{6}$$

the linear least squares estimate $y = \mathbf{x}^T \theta$ is always uniquely available.

The hinge search problem, on the other hand, aims to find the two parameter vectors θ^+ and θ^-, defined by

$$\left[\theta^+,\ \theta^-\right] = \arg\min_{\theta^+,\ \theta^-} \sum_{k=1}^{N} \left[y_k - \langle \max | \min \rangle \left(\mathbf{x}_k^T \theta^+, \mathbf{x}_k^T \theta^-\right)\right]^2 \tag{7}$$

or in an equivalent form

$$\left[\theta^+,\ \theta^-\right] = \arg\min_{\theta^+,\ \theta^-} \sum_{k=1}^{N} \left[\langle \max | \min \rangle \left(y_k - \mathbf{x}_k^T \theta^+, y_k - \mathbf{x}_k^T \theta^-\right)\right]^2 \tag{8}$$

A brute force application of Gauss-Newton method can solve the above optimization problem. However, two problems exist [1]:

1. High computational requirement. The Gauss-Newton method is computationally intensive. In addition, since the cost function is not continuously differentiable, the gradients required by Gauss-Newton method can not be given analytically. Numerical evaluation is thus needed which has high computational demand.
2. Local minima. There is no guarantee that the global minimum can be obtained. Therefore appropriate initial condition is crucial.

The proposed identification algorithm applies a much simpler optimization method, the so called alternating optimization which is 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 (Problem 1). The proposed procedure is attractive in the local minima point of view (Problem 2) 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 two sections this method is discussed briefly in general, and in Section 2.5 the hinge function identification and FCRM method are joined together.

2.3 Fuzzy c-Regression Models

Fuzzy c-regression models yield simultaneous estimates of parameters of c regression models together with a fuzzy c-partitioning of the data. The regression models take the following general form

$$y_k = f_i(\mathbf{x}_k, \theta_i) \tag{9}$$

where the local functions f_i are parameterized by θ_i. The membership degree $\mu_{i,k} \in U$ is interpreted as a weight representing the extent to which the value predicted by the model $f_i(\mathbf{x}_k, \theta_i)$ matches y_k. This prediction error is defined by:

$$E_{i,k} = \left(y_k - f_i(\mathbf{x}_k; \theta_i)\right)^2, \tag{10}$$

but other measures can be applied as well, provided they fulfill the minimizer property stated by Hathaway and Bezdek [11]. The family of objective functions for fuzzy c-regression models is defined by

$$E_m(\mathbf{U}, \{\theta_i\}) = \sum_{i=1}^{c} \sum_{k=1}^{N} (\mu_{i,k})^m E_{i,k}(\theta_i) \tag{11}$$

where $m \in \langle 1, \infty)$ denotes a weighting exponent which determines the fuzziness of the resulting clusters. One possible approach to the minimization of the objective function (11) is the group coordinate minimization method that results in the following algorithm:

- **Initialization** Given a set of data $Z = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)\}$ specify c, the structure of the regression models (10) and the error measure (11). Choose a weighting exponent $m > 1$ and a termination tolerance $\varepsilon > 0$. Initialize the partition matrix randomly.

- **Repeat** for $l = 1, 2, \ldots$

  Step 1 Calculate values for the model parameters θ_i that minimize the cost function $E_m(\mathbf{U}, \{\theta_i\})$.

  Step 2 Update the partition matrix

$$\mu_{i,k}^{(l)} = \frac{1}{\sum_{j=1}^{c} (E_{i,k}/E_{j,k})^{2/(m-1)}}, \; 1 \leq i \leq c, \; 1 \leq k \leq N. \tag{12}$$

  until $||\mathbf{U}^{(l)} - \mathbf{U}^{(l-1)}|| < \varepsilon$.

A specific situation arises when the regression functions f_i are linear in the parameters θ_i, $f_i(\mathbf{x}_k; \theta_i) = \mathbf{x}_{i,k}^T \theta_i$, where $\mathbf{x}_{ik}$ is a known arbitrary function of $\mathbf{x}_k$. In this case, the parameters can be obtained as a solution of a set of weighted least-squares problem where the membership degrees of the fuzzy partition matrix $\mathbf{U}$ serve as the weights.

The N data pairs and the membership degrees are arranged in the following matrices.

$$\mathbf{X}_i = \begin{bmatrix} \mathbf{x}_{i,1}^T \\ \mathbf{x}_{i,2}^T \\ \vdots \\ \mathbf{x}_{i,N}^T \end{bmatrix}, \; \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \; \mathbf{W}_i = \begin{bmatrix} \mu_{i,1} & 0 & \cdots & 0 \\ 0 & \mu_{i,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_{i,N} \end{bmatrix}. \tag{13}$$

The optimal parameters θ_i are then computed by:

$$\theta_i = [\mathbf{X}^T \mathbf{W}_i \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{W}_i \mathbf{y}. \tag{14}$$

The next section gives solutions how to incorporate constrains into the clustering procedure presented in this section. 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).

2.4 Constrained Prototype based FCRM

This section deals with prototypes linear in the parameters. Therefore, as it was shown, 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, $\Lambda_i \theta_i \leq \omega_i$.
- **Global constrains** are related to all of the regression models, $\Lambda_{gl} \theta_i \leq \omega_{gl}$, $i = 1, \ldots, c$.

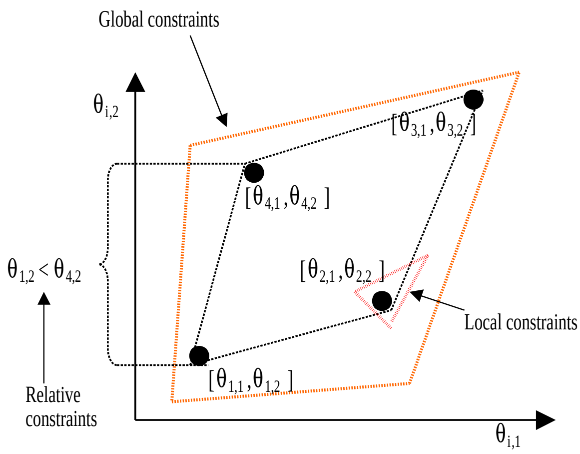

Fig. 2. Example for local, global and relative constraints.

- **Relative constrains** define the relative magnitude of the parameters of two or more regression models,

$$\Lambda_{rel,i,j} \begin{bmatrix} \theta_i \\ \theta_j \end{bmatrix} \leq \omega_{rel,i,j} \tag{15}$$

An example for these types of constrains are illustrated in Figure 2.

In order to handle relative constraints, the set of weighted optimization problems has to be solved simultaneously. Hence, the constrained optimization problem is formulated as follows:

$$\min_{\theta} \left\{ \frac{1}{2} \theta^T \mathbf{H} \theta + \mathbf{c}^T \theta \right\} \tag{16}$$

with $\mathbf{H} = 2\mathbf{X}^T \mathbf{W} \mathbf{X}$, $\mathbf{c} = -2\mathbf{X}^T \mathbf{W} \mathbf{y}'$, where

$$\mathbf{y}' = \begin{bmatrix} y \\ y \\ \vdots \\ y \end{bmatrix}, \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_c \end{bmatrix}, \mathbf{X} = \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_c \end{bmatrix}, \mathbf{W} = \begin{bmatrix} W_1 & 0 & \cdots & 0 \\ 0 & W_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & W_c \end{bmatrix}. \tag{17}$$

and the constraints on θ:

$$\Lambda \theta \leq \omega \tag{18}$$

with

$$\Lambda = \begin{bmatrix} \Lambda_1 & 0 & \cdots & 0 \\ 0 & \Lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Lambda_c \\ \Lambda_{gl} & 0 & \cdots & 0 \\ 0 & \Lambda_{gl} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Lambda_{gl} \\ \{\Lambda_{rel}\} & & & \end{bmatrix}, \ \omega = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_c \\ \omega_{gl} \\ \omega_{gl} \\ \vdots \\ \omega_{gl} \\ \{\omega_{rel}\} \end{bmatrix}. \tag{19}$$

2.5 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 [1]. 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

$$\mathbf{v}_i = \frac{\sum_{k=1}^{N} \mathbf{x}_k \mu_{i,k}}{\sum_{k=1}^{N} \mu_{i,k}}. \tag{20}$$

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 3), the following criteria can be specified:

$$\mathbf{v}_1(\theta_1 - \theta_2) < 0 \text{ and } \ \mathbf{v}_2(\theta_1 - \theta_2) > 0 \tag{21}$$

or

$$\mathbf{v}_1(\theta_1 - \theta_2) > 0 \text{ and } \ \mathbf{v}_2(\theta_1 - \theta_2) < 0 \tag{22}$$

These relative constrains (15) can be used to take into account the constrains above:

$$\Lambda_{rel,1,2} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \leq 0 \text{ where } \ \Lambda_{rel,1,2} = \begin{bmatrix} \mathbf{v}_1 & -\mathbf{v}_1 \\ -\mathbf{v}_2 & \mathbf{v}_2 \end{bmatrix} \tag{23}$$

according to (21).

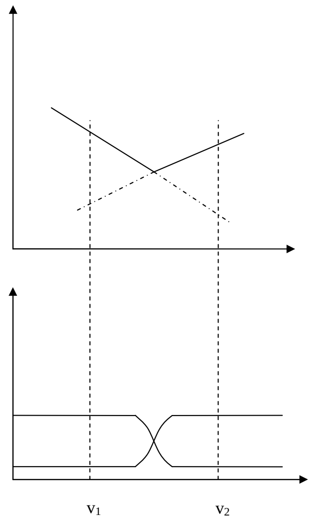

Fig. 3. Hinge identification restrictions.

2.6 Tree Structured Piecewise Linear Models

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 [1]. 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.

The basic idea of the binary tree structure is as follows. Divide the data set into an estimation set and validation set. With the estimation data set, based on certain splitting rules, grow and sufficiently large binary tree. Then use the validation data set to prune the tree into a right size. The estimation data is recursively partitioned into subsets, while each subset leads to a model. As a result, this type of model is also called the recursive partitioning model. For example given a simple symmetrical binary tree structure model, the first level contains one hinge function, the second level contains 2 hinge functions, the third level contains 4 hinges, and in general the kth level contains $2^{(k-1)}$ hinge functions. Any of the following criteria can be used to determine whether the tree growing process:

1. The loss function becomes zero. This corresponds to the situation where the size of the data set is less or equal to the dimension of the hinge. Since the hinging hyperplanes are located by linear least–squares. From least–squares theory, when the number of data is equal to the number of parameters to be determined, the result would be exact, given the matrix is not singular.
2. $J = J^{+} + J^{-}$ During the growth of the binary tree, the loss function is always non-increasing, i.e. $J \geq J^{+} + J^{-}$. When there is no decrease in loss function is observed, when the tree growing should be stopped.

3 Application Examples

In this example, the proposed method is used to approximate two simple univariate nonlinear functions, $y = sin(3x)$ and $y = x^3$.

The proposed method identifies a hinging hyperplane model first that contains two linear submodels and after that for each operating regions of these submodels two other hinging hyperplane models are identified. Following these steps, a tree structured piecewise linear model is identified, where the branches correspond to linear division of the operating regime, $\mathbf{x}^T\,(\theta^+ - \theta^-) \leq 0$, and the leaves correspond to linear models. In this way a piecewise linear model is constructed (see Figure 4 and Figure 5).

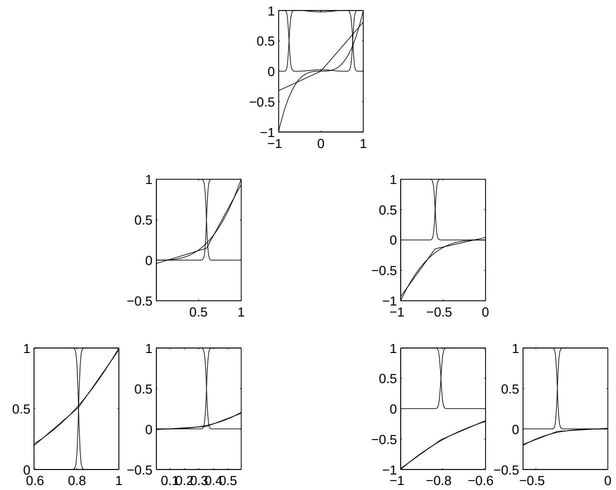

Fig. 4. Approximation of the $y = x^3$ function by a binary tree of hinging hyperplane models.

As can be seen in these figures, the proposed clustering algorithm is able to effectively partition of the operating region of the local models and the resulted partitioning is easily interpretable thanks to the tree-like structure of the model.

For comparison the nonlinear identification toolbox of Jonas Sjoberg [12] has been used and global hinging hyperplane models with eight hinges have been identified based on the same training data. Surprisingly, the resulted classical models gave extremely bad modeling performance (see Figure 6 and Table 3). 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.

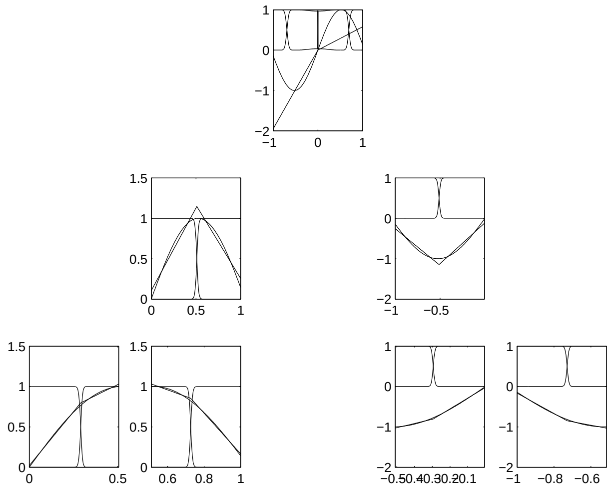

Fig. 5. Approximation of the $y = sin(3x)$ function by a binary tree of hinging hyperplane models.

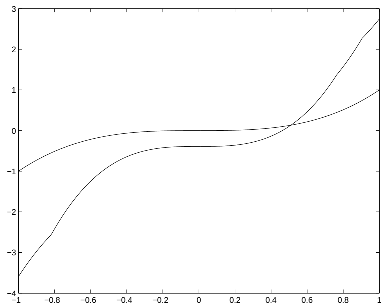

Fig. 6. Approximation of the $y = x^3$ function by a classical global hinging hyperplane model.

	$sin(3x)$	x^3
Sjoberg	0.1052	1.1198
This paper	$7.4771\ 10^{-4}$	$1.7032\ 10^{-4}$

Table 1. Mean square errors of the hinging hyperplane models

4 Conclusion

Hinging hyperplane models can be an alternative to artificial neural nets. This paper presents a new identification technique for hierarchial hinging hyperplane models based Fuzzy c-Regression Clustering. This clustering technique applies linear models as prototypes and the model parameters and fuzzy membership degrees are identified simultaneously. Since hyperplanes should intersect each other in the operating regime covered by the data points a constrained version of the fuzzy c-regression clustering has been developed. The proposed method identifies a hinging hyperplane model first that contains two linear submodels, and after that the two halves of the model (the two linear hyperplanes) are treated separately: two other hinging hyperplane models are identified on the basis of the operating regions of the first two linear submodels. Following these steps, a tree structured piecewise linear model is identified, where the branches correspond to linear division of the operating regime, and the leaves correspond to linear models. In this way a piecewise linear model is constructed.

Acknowledgement

The authors would like to acknowledge the support of the Cooperative Research Centre (VIKKK) (project 2004-I) and Hungarian Research Found (OTKA T049534). Janos Abonyi is grateful for the support of the Bolyai Research Fellowship of the Hungarian Academy of Sciences.

References

1. S. Niu, P. Pucar, Hinging hyperplanes for non-linear identification 1995, chttp://www.control.isy.liu.se
2. X. Sun, S. Wang, A special kind of neural networks: continuous piecewise linear functions, Lecture Notes in Computer Science 3496 (2005) 375–379.
3. S. Jakubek, N. Keuth, A local neuro-fuzzy network for high-dimensional models and optimization,Engineering Applications of Artificial Intelligence (2006) 705–717
4. L. Breiman, Hinging hyperplanes for regression, classification and function approximation, IEEE Transactions on Information Theory 39 (3) (1993) 311–325.
5. D. Ramirez, E. Camacho, M. Arahal, Implementation of min-max mpc using hinging hyperplanes. application to a heat exchanger., Control Engineering Practice 12 (2004) 1197–1205.
6. J. Roll, A. Bemporad, L. Ljung, Identification of piecewise affine systems via mixed-integer programming, Automatica 40 (2004) 37–50.
7. P. Pucar, J. Sjoberg, Parameterization and conditioning of hinging hyperplane models. Technical Report LiTH-ISY-R-1809, Department of Electrical Engineering, Linkoping University, Linkoping, Sweden, 1995. Available by anonymous ftp 130.236.24.1.
8. P. Pucar, M. Millnert, Smooth hinging hyperplanes - an alternative to neural nets. In Proc. 3rd ECC, Italy, 1995.
9. G. Ferrari-Trecate, M. Muselli, D. Liberati, M. Morari, A clustering technique for the identification of piecewise affine systems, Lecture Notes in Computer Science" 2034 (2001)
10. J. Bezdek, Pattern recognition with fuzzy objective function algorithms, New York: Plenum, 1981.
11. R. Hathaway, J. Bezdek, Switching regression models and fuzzy clustering, IEEE Transactions on fuzzy systems 1 (3) (1993) 195–204.
12. J. Sjoberg, Nonlinear Toolbox, Version 1.0