CONTINUOUS OPTIMIZATION APPROACHES TO THE IDENTIFICATION OF PIECEWISE AFFINE SYSTEMS

Eberhard Münz, Volker Krebs

Institut für Regelungs– und Steuerungssysteme,
Universität Karlsruhe (TH)
phone: +49 721 / 608 2462
fax: +49 721 / 608 2707
e–mail: {muenz,krebs}@irs.uni–karlsruhe.de

Abstract: Approaches for the identification of piecewise affine systems are presented. The identification problem for piecewise affine systems is formulated as a smooth constrained, and a nonsmooth unconstrained optimization problem. The main advantage of the new problem formulations is the fact that no mixed integer problems have to be solved within this method. The feasibility and performance of the procedures are demonstrated by identifying two piecewise affine characteristic maps.

Keywords: identification, modelling, optimization, hybrid systems

1. INTRODUCTION

The identification of piecewise affine systems is a field of growing interest and the increasing number of publications in this area emphasizes the relevance and importance of this topic.
In (Ferrari-Trecate *et al.*, 2001; Ferrari-Trecate *et al.*, 2003; Hoffmann, 1999; Münz and Krebs, 2002; Münz *et al.*, 2003; Münz *et al.*, 2004*a*; Münz *et al.*, 2004*b*) iterative identification procedures are proposed, where the identification procedure is split up into several tasks.
As an example, the method given in (Münz and Krebs, 2002) consists of two steps:

- First, the measured data is partitioned such that a local model is able to represent the data within one partition.
- In the second step, the shape and the location of these partitions are optimized.

In (Münz *et al.*, 2003; Münz *et al.*, 2004*b*), these two steps are extended by determination of gradients for the originally discontinuous problem and thus it becomes possible to apply numerical optimization methods which exploit gradient information for the partitioning and optimization.

Perhaps the most straightforward way of tackling the piecewise affine identification problem is to formulate a performance index and to minimize it with respect to its parameters. Such approaches, originally developed for linear systems, exist for piecewise affine systems with continuous transitions between the different submodels (e.g. (Roll, 2003)). For piecewise affine systems with jumps between different submodels other approaches are proposed in (Münz *et al.*, 2004*a*) and (Vidal *et al.*, 2003). As the optimization problem, resulting from the identification problem is mixed–integer in most cases, one would guess that it is a NP–equivalent problem. In this paper, it is shown that the problem can be reduced to a continuous nonlinear problem.

As a basis of this paper, the problem of piecewise affine identification is formalized in section 2. In

section 3 some methods are discussed for transforming the piecewise affine identification problem, which is originally mixed–integer, into a smooth nonlinear problem with linear constraints or into a nonsmooth but unconstrained problem. Since the resulting problem is not convex, some remarks for detecting local optima are made in section 4. The procedure is demonstrated identifying a static three–dimensional and a four–dimensional sample characteristic map in section 6. Finally, the results are discussed in section 7 and some conclusions are drawn.

2. PROBLEM STATEMENT

The usual approach for the identification of linear dynamic systems is to choose the parameters of a linear system such that they minimize the sum of a norm of the prediction error (Isermann, 1992); as a generalization of this approach the identification of piecewise affine models is interpreted as an optimization problem:
Consider a set of measured data at discrete time instants, $\underline{y}(1), \ldots, \underline{y}(N)$, $\underline{u}(1), \ldots \underline{u}(N-1)$ and assume that a subset of $\mathbb{R}^n$ forms the state space and a subset of $\mathbb{R}^m$ forms the space of the inputs. Then, it is the objective of the identification to find a model which maps the subset $\mathcal{I}$ of the space $\mathbb{R}^n \times \mathbb{R}^m$ to a subset of the space $\mathbb{R}^n$ according to the measured data.
In the following, the space $\mathcal{I} \subset \mathbb{R}^n \times \mathbb{R}^m$ is referred to as identification–space or $\mathcal{I}$–space.
The piecewise affine model $\mathcal{M}$ consists of M affine models

$$\underline{x}(k+1) = \underline{A}_j \underline{x}(k) + \underline{B}_j \underline{u}(k) + \underline{c}_j,\ j = 1 \ldots M \quad (1)$$

and the measurements $\underline{y}(k)$ are obtained from

$$\underline{y}(k) = \underline{x}(k) + \underline{\varepsilon}(k) \quad (2)$$

with $\underline{\varepsilon}$ as unavoidable noise. Each of these affine models is valid only on a subset $\mathcal{I}_j$ of $\mathcal{I}$ and the $\mathcal{I}_j$ form a complete and disjoint partition of $\mathcal{I}$; i.e.

$$\bigcup_{j=1}^{M} \mathcal{I}_j = \mathcal{I} \text{ and } \mathcal{I}_i \cap \mathcal{I}_j = \varnothing \quad \forall i \neq j.$$

With knowledge about the desired structure of the model, the identification problem can now be formulated as the minimization of a performance index

$$J = \sum_{i=1}^{N-1} \sum_{j=1}^{M} \left\| \left(\underline{y}(i+1) - \hat{\underline{y}}_j(i+1) \right) \right\| z_{ij}. \quad (3)$$

Herein, $\underline{y}(i+1)$ are the measured data corresponding to a certain state $\underline{x}(i+1)$ of the system according to equation 2. The quantities $\hat{\underline{y}}_j(i+1)$ are the predicted outputs of the identified submodel j and can be rewritten as

$$\hat{\underline{y}}_j(i+1) = \underline{A}_j \underline{y}(i) + \underline{B}_j \underline{u}(i) + \underline{c}_j.$$

The variables z_{ij} describe the affiliation of a data $y(i)$ to a submodel j. Thus, these variables are restricted to binary values. $\|\cdot\|$ denotes an arbitrary norm.
In general, this performance index has to be minimized with respect to the parameter matrices $\underline{A}_j, \underline{B}_j, \underline{c}_j$, and to the variables z_{ij}.
Since the norm is a nonlinear function, the objective of this problem is nonlinear. Furthermore, the problem is mixed integer, due to the binary variables z_{ij}
As general mixed–integer problems are hard to solve, the approach proposed in the sequel tries to transform this problem into a smooth optimization problem.

3. TRANSFORMATION OF THE PROBLEM

The problem stated in section 2 reads as follows:

$$\min \left\{ \sum_{i=1}^{N-1} \sum_{j=1}^{M} \left\| \left(\underline{y}(i+1) - \underline{A}_j \underline{y}(i) + \underline{B}_j \underline{u}(i) + \underline{c}_j \right) \right\| z_{ij} \right\}$$

s.t.

$$z_{ij} \in \{0, 1\} \quad (4)$$

$$\sum_{j=1}^{M} z_{ij} = 1 \quad \forall i$$

To solve this problem, a branch and bound algorithm could be applied, but the search tree for this problem increases exponentially with the number of data N and the number of submodels M.
Thus a relaxation of this problem is made, which replaces the integer constraint by $z_{ij} \in [0, 1]$. This leads to the relaxed problem

$$\min \left\{ \sum_{i=1}^{N-1} \sum_{j=1}^{M} \left\| \left(\underline{y}(i+1) - \underline{A}_j \underline{y}(i) - \underline{B}_j \underline{u}(i) - \underline{c}_j \right) \right\| z_{ij} \right\}$$

s.t.

$$z_{ij} \in [0, 1] \quad (5)$$

$$\sum_{j=1}^{M} z_{ij} = 1 \quad \forall i$$

Lemma 1. The global optimal value of problem (4) coincides with the global optimal value of the relaxed problem (5). The set of optimal solutions of problem (4) is a subset of the set of optimal solutions of problem (5).

PROOF Let $\underline{A}_j^*, \underline{B}_j^*, \underline{c}_j^*$ and $\underline{Z}^*$ be the parameters in a global optimum of problem 5. Suppose that the elements in $\underline{Z}^*$ are non–integer. Since the error of each data k to the different affine models is concerned, there is a set of submodels $\mathcal{M}_l$, $l \subset \{1, ..., M\}$ which show the least error to

that particular data k.
There are two cases: If the set of closest models has exactly one element, the corresponding parameter z_{kl} is increased by $\varepsilon > 0$; one of the other $z_{ki} > 0$ is decreased by $\varepsilon > 0$ to keep feasible. Hence the performance index decreases. This is a contradiction to the assumptions and thus a global optimal solution cannot violate the integer constraints in this case.
In the second case, the set of closest models contains more than one element. Then, the performance index does not change, if the percentage of the membership is changed between these closest models. In this case, the set of global optimal solutions increases by relaxation, but the global optimal value is not changed. Furthermore, the same contradiction as in case 1 holds, if one of the closest models and a non–closest model with membership $z_{ki} > 0$ are considered. Hence, an optimal solution in the second case yields memberships of the closest models which sum up to 1. An optimal integer solution is then obtained by setting the membership z_{ki} of one closest model to 1 and the others to 0. □

According to Lemma 1, by finding the global optimal solution of the nonlinear constrained problem (5), a solution for the mixed–integer problem (4) is obtained and thus, the original mixed integer piecewise affine identification problem is reduced to a nonlinear constrained optimization problem.
To solve this problem in practice, a particular norm has to be chosen. In this paper, the 1–norm and the 2–norm is considered.
If the 1–norm is chosen, the problem can be rewritten as

$$\min\left\{\sum_{i=1}^{N-1}\sum_{j=1}^{M}\left\|\left(\underline{y}(i{+}1)-\underline{A}_j\underline{y}(i)-\underline{B}_j\underline{u}(i)-\underline{c}_j\right)\right\|_1 z_{ij}\right\}$$

s.t.

$$\begin{aligned} &z_{ij}\in[0,1] \\ &\sum_{j=1}^{M} z_{ij}=1 \ \ \forall i, \end{aligned} \tag{6}$$

where the $\|\cdot\|_1$ denotes the 1–norm of a vector: $\|(x_1 \ x_2)^T\|_1 = |x_1| + |x_2|$.
This particular problem can be rewritten as a quadratic problem by eliminating the absolute values with help of slack variables. Since

$$\min|x|$$

is equivalent to

$$\begin{aligned} &\min s \\ &\text{s.t.} \\ &s \geq x \\ &s \geq -x \end{aligned}$$

problem (6) can be rewritten as

$$\begin{aligned} &\min\left\{\sum_{i=1}^{N-1}\sum_{j=1}^{M}\sum_{k=1}^{n} s_{ij}^k z_{ij}\right\} \\ &\text{s.t.} \\ &(s_{ij}^1,\ldots,s_{ij}^n)^T \geq \underline{y}(i{+}1)-\underline{A}_j\underline{y}(i)-\underline{B}_j\underline{u}(i)-\underline{c}_j \\ &(s_{ij}^1,\ldots,s_{ij}^n)^T \geq -\underline{y}(i{+}1)+\underline{A}_j\underline{y}(i)+\underline{B}_j\underline{u}(i)+\underline{c}_j \\ &z_{ij}\in[0,1] \\ &\sum_{j=1}^{M} z_{ij}=1 \ \ \forall i, \end{aligned} \tag{7}$$

In this representation, it is obvious, that the performance index of the piecewise affine identification problem is indefinite, since the Hesse–Matrix which can be seen directly from the performance index of this quadratic problem is symmetric and has only zeros on the prime diagonal. Hence the optimization problem does not have a unique minimum. This also can be seen from the fact, that for each feasible solution $\underline{x}_0 = (s_{11}^1,\ldots,s_{NM}^n,z_{11},\ldots,z_{NM})^T$ of (7), there are $M!$ equivalent points which are obtained by permuting the numbering of the submodels. Hence, there are at least $M!$ global minima. Further more, there are local optima in which an optimization algorithm may be trapped. A method to escape these local optima is discussed in the next section.

4. GLOBAL MINIMIZATION

The smooth optimization approach to piecewise affine identification problems discussed in the last section normally yields good results if the noise is low. But even for undisturbed problems, there may be local minima, as the following example shows.

Example 1. The measured data depicted in figure 1 is to be identified. The number of submodels $M = 2$ and the constants for each model $c_1 = 0$ and $c_2 = -6$ are known and thus, the parameters a_1 and a_2 are to estimate. This identification can be performed by minimizing the smooth problem (5). As another equivalent problem, it is possible to solve the continuous but nonsmooth problem

$$\min\left\{\sum_{i=1}^{N-1}\min_j\left\|\underline{y}(i{+}1)-\underline{A}_j\underline{y}(i)-\underline{B}_j\underline{u}(i)-\underline{c}_j\right\|\right\} \tag{8}$$

It becomes obvious that this problem is equivalent to (5) if a similar reasoning as for Lemma 1 is used. The performance index for the sample data set is plotted in figure 2; there, the quadratic norm is used. Apparently, there are two local minima and the right one coincides with the global minimum.

To be not trapped in such local minima, it is necessary to apply suitable optimization techniques,

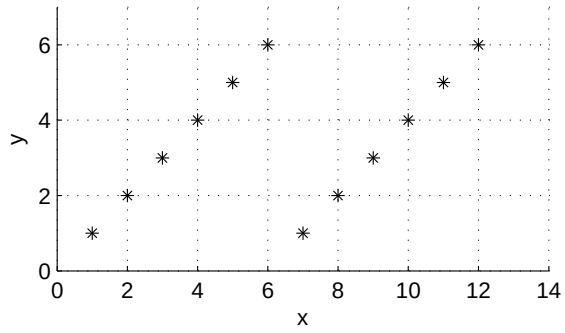

Fig. 1. measured data for example 1

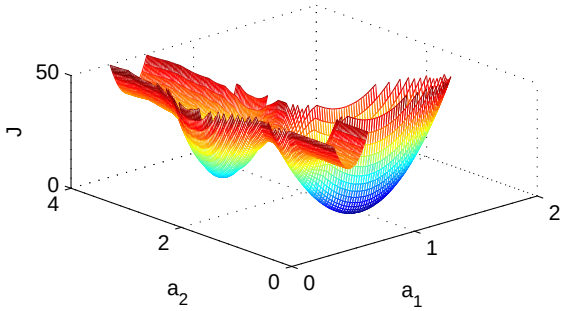

Fig. 2. performance index for example 1; function values larger than 50 are not plotted

as described in (Hansen, 1992; Ratz, 1992; Huyer and Neumaier, 1999). In particular the MCS–algorithm described in (Huyer and Neumaier, 1999) was applied in this work.
Even though, a global minimum for the performance index is found with these algorithms, it may happen in in the case of noisy data that some data are assigned to the "wrong" submodel. Thus, in the next section a method is described to assign the data to the submodels in a sensible way.

5. USING GRAPHS TO ASSIGN DATA TO SUBMODELS

If the data is assigned to the nearest submodel after minimizing performance index (5) or (8), it is possible in the case of noisy data that the assignment of the data is scattered and thus, there are no compact areas which belong to one submodel. To solve this problem, the following procedure can be applied.

In a first step a graph is needed, which indicates adjacency in the measured data. Two vertices of the graph, which are formed by the measured data, are linked by an edge, if these data are adjacent. To set up such a graph the Delaunay triangulation can be applied. Due to the properties of the Delaunay triangulation, it is asserted that no other data is contained in the circumsphere of a simplex and thus this triangulation produces a sensible adjacency graph.

In the second step, the vertices are assigned to the closest submodel according to the norm used for the global minimization. Afterwards, the graph is divided into M components; vertex i belongs to component j if the distance between submodel j and data i is minimum. All edges between vertices to different submodels are eliminated.In this way, M graphs are obtained.

In the third step, the largest connected component of each graph is searched and this largest component forms a center component to which all the remaining vertices, belonging not to a largest component, are assigned to in the fourth step. All edges between vertices belonging not to one of the largest connected components are removed, too

To do the assignment, all vertices adjacent to the center component j according to the eliminated edges become candidates for being assigned to submodel j. This way, M sets of candidates are obtained and the distances between the candidates and the corresponding affine model are calculated according to the norm used for the global minimization. Then, the candidate with the minimum distance between model and measured data is searched over all M sets of candidates and is assigned to the corresponding model. The sets of candidates have to be updated and step 4 is repeated until all vertices are assigned to a center component.

This procedure yields M connected components and hence compact subsets of the $\mathcal{I}$–space where one affine model is valid.

6. EXAMPLES

As Least Squares approaches for the identification of dynamical systems map the dynamic identification problem to a static identification problem, the approach proposed in this paper is demonstrated by using multidimensional static piecewise affine characteristic maps.

Example 2. The sample function used for the generation of the data is illustrated in figure 3. The function consists of three affine submodels $f_1(x_1, x_2) = 0.2x_1 + x_2 + 1$, $\{\underline{x} | x_2 \geq -5x_1 + 1.25 \wedge x_2 \geq x_1 - 0.1 \wedge x_2 \leq -0.5x_2 + 1\}$; $f_2(x_1, x_2) = 3$, $\{\underline{x} | x_2 \geq -5x_1 + 1.25 \wedge x_2 \geq x_1 - 0.1 \wedge x_2 \geq -0.5x_2 + 1\}$; $f_3(x_1, x_2) = 2$, otherwise. The noise added to the data is $N(\mu, \sigma^2) = N(0, 0.005)$ distributed.

This data was identified using the following procedure: The problem description (8) with a quadratic norm was used and minimized with MCS (Huyer and Neumaier, 1999). The intermediate result where the data is assigned to the nearest model can be seen in figure 4. It is visible, that

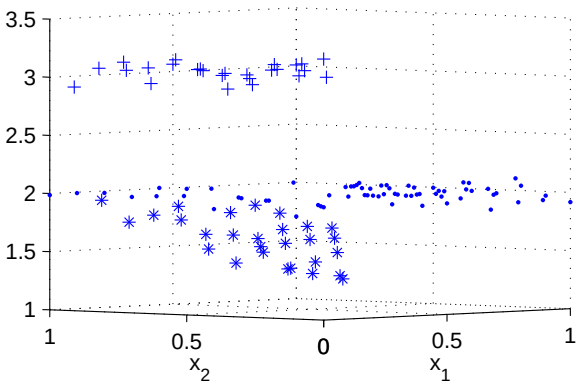

Fig. 3. data with noise considered in the identification process

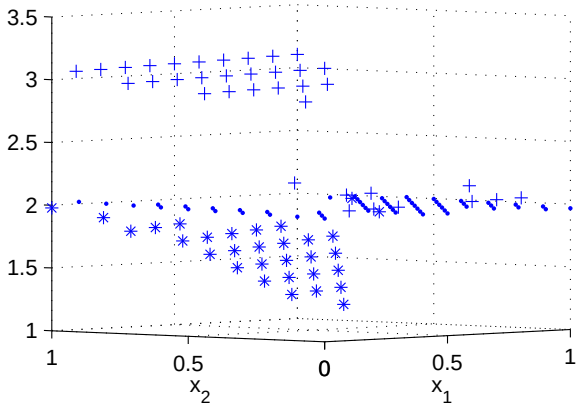

Fig. 4. after the first step of the identification: the data is assigned to the closest model

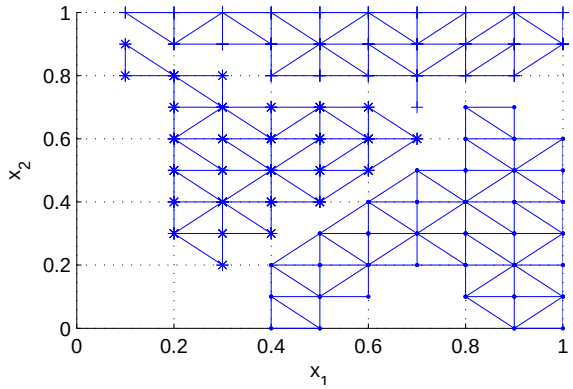

Fig. 5. largest components of the graph before assignment

some data, which should belong to model $(\cdot)$ is assigned to model $(+)$.

Carrying out the Delaunay triangulation and searching for the largest components of the resulting graph yields the largest components depicted in figure 5. The iterative assignment of the candidates yields the graph in figure 6. Using this fixed assignment of the data to the submodels, figure 7 is obtained after a least squares fit for each submodel. As it can be seen, the submodels and the assignment of the data is well identified.

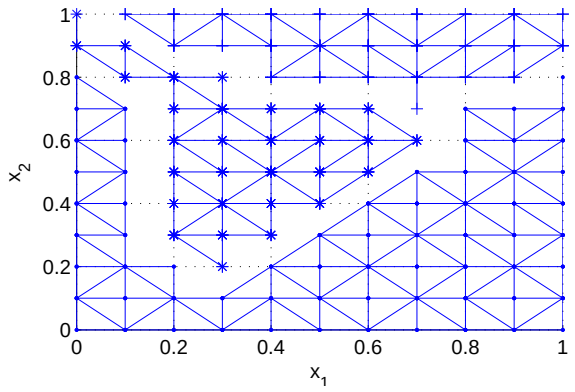

Fig. 6. largest components of the graph after assignment

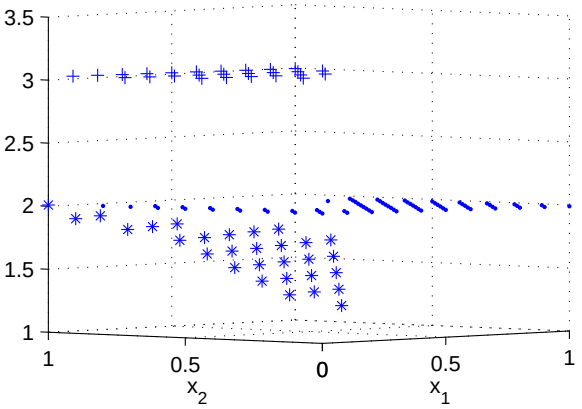

Fig. 7. identified model

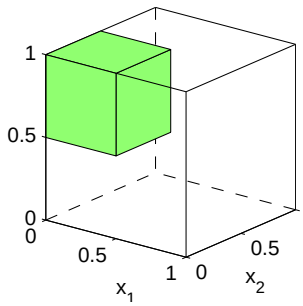

Fig. 8. $\mathcal{I}$–space for the 4D–example

Example 3. The next example deals with a piecewise affine function with two submodels depending on three variables $f_1(x_1, x_2, x_3) = 1$, $\{\underline{x} \,|\, x_1 < 0.5 \wedge x_2 < 0.5 \wedge x_3 > 0.5\}$; $f_2(x_1, x_2, x_3) = 0$, otherwise. This time, the noise is $N(0, 0.015)$ distributed. For the generation of sample data the $\mathcal{I}$–space was covered with 11^3 equidistant points in a 0.1 grid. In figure 8 the $\mathcal{I}$–space for the function with the two domains of the submodels used to generate the data is depicted. For the identification, again performance index (8) with a quadratic norm was used. After applying the identification to the data, the following affine parameters are obtained: $\hat{f}_1(x_1, x_2, x_3) = -0.0156x_1 + 0.1260x_2 - 0.0587x_3 + 1.02021$, $\hat{f}_2(x_1, x_2, x_3) = -0.0178x_1 + 0.0022x_2 - 0.0002x_3 + 0.01110$. All data are assigned correctly to the two affine submodels. To

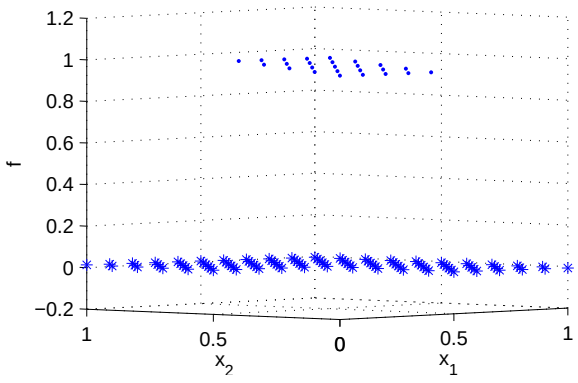

Fig. 9. identified model: cut at $x_3 = 0.8$

get an impression of the identified model, a cut at $x_3 = 0.8$ is shown in figure 9

7. CONCLUSIONS

The presented identification method proceeds in the steps global optimization of a performance index, where a suitable form can be chosen from 2 equivalent forms; the first is a smooth constrained formulation, the second is nonsmooth but unconstrained.

Since the number of parameters is quite small in the nonsmooth formulation, state of the art solvers as (Huyer and Neumaier, 1999) are able to solve the sample-problems within a few minutes.

The algorithms to create compact areas for the validity of one submodel work on graphs. Analysis of these algorithms shows that they are very efficient and fast.

To obtain good identification results, a number of submodels is necessary such that the real system can be approximated sufficiently good with the chosen number of models. I.e. the number of submodels has to be larger or equal to the number of submodels in the original system. Furthermore, the number of submodels has to be manageable with reasonable effort. Often, such a number is known from a priori knowledge. Thus, the importance of clustering techniques which guess the number of submodels from measured data exactly decreases in the field of piecewise affine identification.

The identification approaches introduced in this paper allow to determine the dynamics of the submodels and the validity of the submodels with sufficient accuracy and sufficient low computational effort. Available a priori knowledge may be used to support the different steps.

Since technical systems are often of higher order and similar characteristic maps are widespread in technical systems, e.g. in the models of turbo–chargers of diesel engines, it is of practical interest to have an approach that copes with such systems. As it was shown in the application, the presented algorithm is able to identify such systems.

Acknowledgement: We thank Ahn Son Phung who provided the code for the examples.

REFERENCES

Ferrari-Trecate, G., M. Muselli, D. Liberati and M. Morari (2001). Identification of piecewise affine and hybrid systems. In: *Proceedings of the ACC*. Arlington, USA. p. CD.

Ferrari-Trecate, G., M. Muselli, D. Liberati and M. Morari (2003). A clustering technique for the identification of piecewise affine and hybrid systems. *Automatica* **39**, 205–217.

Hansen, E. (1992). *Global Optimization Using Interval Analysis*. Marcel Dekker Inc.. NewYork.

Hoffmann, I. (1999). *Identifikation hybrider dynamischer Systeme*. Shaker Verlag.

Huyer, W. and A. Neumaier (1999). Global optimization by multilevel coordinate search. *Journal of Global Optimization* **14**, 331–355.

Isermann, R. (1992). *Identifikation dynamischer Systeme*. 2 ed.. Springer Verlag. Berlin.

Münz, E., G. Schullerus and V. Krebs (2004*a*). On the use of hybrid system identification methods for max-plus-linear system identification. In: *Proceedings of IFAC Workshop on Discrete Event Systems* (Janan Zaytoon, Ed.). Reims.

Münz, E., T. E. Hodrus and V. Krebs (2004*b*). Gradient based identification of hybrid systems. *Mathematical and Computer Modelling of Dynamical Systems*.

Münz, E., T. Hodrus and V. Krebs (2003). Top-down identification of hybrid characteristic maps. In: *Proceedings of ADHS*. Saint Malo.

Münz, Eberhard and Volker Krebs (2002). Identification of hybrid systems using a priori knowledge. In: *Proceedings of the IFAC World Congress*. Barcelona, Spain. p. CD.

Ratz, D. (1992). Automatische Ergebnisverifikation Bei Globalen Optimierungsproblemen. PhD thesis. Universität Karlsruhe.

Roll, J. (2003). Local and Piecewise Affine Approaches to System Identification. PhD thesis. Linköping University.

Vidal, R., S. Soatto and S. Sastry (2003). An algebraic geometric approach to the identification of linear hybrid systems. In: *Proceedings of IEEE Conference on Decision and Control*. Maui, Hawaii USA.