Signal segmentation and data classification

José RAGOT, Didier MAQUIN, Midzodzi PEKPE
Centre de Recherche en Automatique de Nancy. UMR CNRS 7039
Institut National Polytechnique de Lorraine
2, avenue de la forêt de Haye, F 54 516 Vandoeuvre les Nancy
{jragot, dmaquin, mpekpe}@ensem.inpl-nancy.fr

Abstract. The paper deals with a method for determining a switching combination of several local linear models using only the knowledge of the input-output data. The method is a direct optimisation of the sum of square errors between measured output and predicted model output ; in this procedure, the originality is based on the choice of an adapted criterion with particularly weights.

Keywords-component. Data, signal, classification, segmentation, switching.

1. INTRODUCTION

For the identification of nonlinear systems, there has been a large activity during the past years. In particular many interesting results have been reported in connection with multi-model [Gasso, 2001] and/or multiple models [Murray-Smith, 1997], [Mihaylova, 2001], hybrid systems [Bemporad, 2001], hinging hyperplanes [Pucar, 1998], [Breiman, 1993], hidden Markov models [Ding, 1997].

In the following we focuse the attention on PieceWise Auto Regressive Exogeneous models (PWARX). As it will be pointed out latter, if the partition of piecewise mapping is known, the problem of identification can easily be solved by using standard techniques of estimation. However, when the partition is unknown the problem becomes much more difficult (see for example the works of Gasso [Gasso, 2001] in the field of multi-models). Thus, there are two possibilities. Either a partitioning defining the local domains in which the system is constant, is a priori defined or the partitioning has to be estimated along with the local models. In the first case, the number of local domains has to be chosen very large. If the amount of input-output data is sufficient in each domain, the parameter estimation of local model is generally easy ; otherwize, problems of ill conditioning often occurs. In the second case, a few number of local domains are used, but the simultaneous estimation of their number and of the parameters of the local models generally leads to potentially many local minima which may make it difficult to apply local search routines.

This two difficulties have not received definitive solutions. Our contribution is to illustrate this problem in the case where the structure and the number of local models are known. Thus, we restrict the estimation problem to 1) the estimation of switching between the local models, 2) the estimation of the parameters of the local models.

To begin with, let us consider systems on the form :

$$y_k = \varphi_k^T \theta_j \qquad j = 1..s \tag{1a}$$

$$\text{if} \quad H_j^T \varphi_t \geq 0 \tag{1b}$$

where $\varphi_k \in \mathbb{R}^p$ is a regression vector and $\theta_j \in \mathbb{R}^r$ the parameter vector associated with the jth local model. The regression vector φ could, e.g., consists of old inputs and outputs :

$$\varphi_t = \left[y_{t-1} \ldots y_{t-n}, u_{t-1} \ldots u_{t-m},\ 1\right]^T . \tag{2}$$

The sets $S_j = \left\{H_j^T \varphi_t \geq 0\right\},\ j = 1..s$ are polyhedral partition of the $\varphi - space$.

Our problem, when we are given y_t and φ_t, $t = 1..N$, consists in finding the PWARX model that best matches the given data. The model (1) can be identified by solving the optimisation problem :

$$min \sum_{t=1}^{N} \left(\left\| y_t - \varphi_t^T \theta_j \right\|_2^2 \lambda_j(\varphi_t) \right) \tag{3}$$

$$\text{subject to : } \lambda_j(\varphi_t) = \begin{cases} 1 & if \quad H_j^T \varphi_t \geq 0 \\ 0 & otherwise \end{cases} \tag{4}$$

where θ_j and H_j, $j = 1..s$, are the unknown.

In our case, we limit the estimation problem to those of θ_j ; however, we need to simultaneously estimate the values of the function $\lambda_j(.)$ in order to know the time switching i.e. the data useful to estimate the parameters of the jth local model. In fact we are not involved with the explanation of the switching, i.e. the estimation of the H_j parameters.

This paper organised as follows : section 1 explains, through a simple example what is the problem to solve and the foregoing difficulties. Sections 3 constitutes the contribution of the paper. Some simulation examples provide an illustration of the proposed algorithms in section 4 and is followed by a conclusion.

2. INTRODUCTIVE EXAMPLE

Let us consider a first order system with input u and output y. Looking at the time evolution of the two signals (fig. 1) doesn't reveal an evident relationship between the two variables u and y. However it is reasonable to think that the parameters of the system have changed in respect to the time (see, for example, the evolution of u and y between times 6 and 10). To go further in this hypothesis, it would be interesting to test if a hidden evolution of the parameters exists.

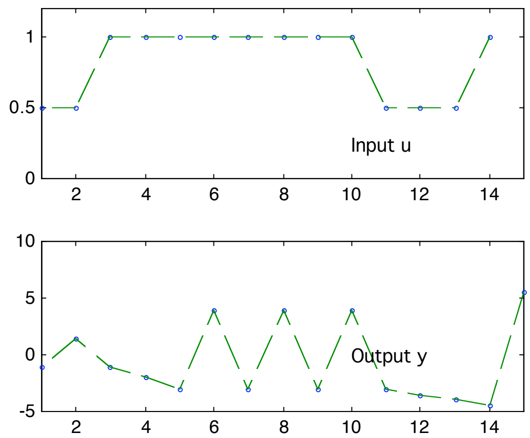

Figure 1. Input and output evolution

In fact considering the time varying model :

$$y_{k+1} = a_k y_k + b_k u_k \qquad k = 1..N-1 \tag{5}$$

we try to analyse the consistency of the data in regard to the structure of this model. For that purpose the data gathered in fig. 1 are used.

First we plot y_k / u_k versus y_{k+1} / u_k (fig. 2). It is thus clear that there are two subsets of data each being represented by a straight line. If we were able to identify this two subsets, i.e. to classify the input-output data, identifying the parameter models would be easy. Considering the presence of only two subsets, it is reasonable to hypothesis, that the system may be represented with two local models (with respective parameters a_1, b_1 and a_2, b_2) switching at particular instants :

$$y_{k+1} = \lambda_k(a_1 y_k + b_1 u_k) + (1 - \lambda_k)(a_2 y_k + b_2 u_k) \tag{6}$$
$$\lambda_k \in \{0,1\}.$$

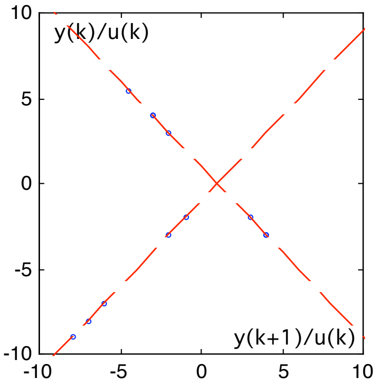

Figure 2. The two regimes of functioning

Now, we propose a systematic procedure to analyse the input-output data of an hybrid system in order to identify its parameters.

3. PROPOSED TECHNIQUE

Let y_k represents the output measurements of the underlying system and $y_{j,k}(\theta_j)$ the output of the jth local model parametrized by θ_j. To fit the local model to the data, we want to minimize the error function :

$$J = \sum_{k=1}^{N} \sum_{j=1}^{s} \left(\frac{y_{j,k}(\theta_j) - y_k}{p_{j,k}} \right)^2 \tag{7}$$
$$y_{j,k}(\theta_j) = \varphi_k^T \theta_j$$

where $y_{j,k}(\theta_j)$ is the jth local model output and where the weights $p_{j,k}$ have to be design such that the local model j is adapted only with the input-output data for which it is concerned. It can be seen that the cost function (19) represents a trade-off between local and global learning. Indeed, when the model output $y_{1,k}$ is closed to the measurement y_k then model j=1 matches the measurements and $\rho_{1,k}$ is greater than $\rho_{j,k}$. when $j \neq 1$

Obviously, the key point is the design of these weights. In the following a non parametric estimation is used because there is no need to parameterize the weighting functions, only their values being useful to separate the data according to the s local models. The ideal situation deals with the knowledge of the partitioning of the data into s groups, the first one gathering the data in accordance with the first model, and similarly for the othergroup. These two sets are noted S_j :

$$S_j = \left\{ (x_k, y_k), k = 1..N \;/\; (x_k, y_k) \text{ satisfy model } j \right\} \quad j = 1,2$$

Thus, the optimal weights are defined by :

$$p_{j,k} = \begin{cases} 1 & if \ (x_k, y_k) \in S_j \\ 0 & if \ (x_k, y_k) \notin S_j \end{cases} \qquad k = 1..N \qquad (8)$$

In fact our algorithm try to adapt the weights as closed as possible to the optimal ones.

The complete iterative algorithm is now described. Each iteration consists of two steps. The first one is to determine an estimation $\mu_{j,k}$ of the weighting functions $p_{j,k}$ given the local models. The second step is to identify the local models given the weights. Note that in [Verdult, 2001] a similar algorithm is used in the context of weighted combination of local linear state-space systems and using a extended Kalman smoother ; however, in that approach additional hypothesis is needed on the rate evolution of the weights.

1. Select a set of initial parameters $\theta_{0,j}$ for the s local models
   $\theta_j = \theta_{0,j}, \; j = 1..s$

2. Define the weights

$$\mu_{j,k} = \frac{1}{\left[\delta + \left(y_{j,k}(\theta_j) - y_k \right)^2 \right]^r} \qquad j = 1..s \quad k = 1..N$$

$$j = \arg \max_p (\mu_{p,k})$$

$$\mu_j = 1$$

$$\mu_p = 0, \quad p = 1..s, \quad p \neq j$$

$$W_j = diag\left(\mu_{j,1} \; \dots \; \mu_{j,N} \right), \quad j = 1,2$$

3. Compute the local model parameters

$$\theta_j = \left(X^T W_j X \right)^{-1} X^T W_j y$$

$$X = \left[\varphi_1 \; \cdots \; \varphi_N \right]^T$$

$$y = \left[y_1 \; \cdots \; y_N \right]$$

4. Go to step 2 until the convergence has been obtained

In step 1, the parameters of the s local models must have different initial values ; otherwise the identifying of the regimes would be unsuccessful. In step 2 the factor δ is used to avoid the nullity of the denominator of δ ; it must be chosen by the user as small as possible. The coefficient r enforced the weight ; it must be chosen by the user, a "good" value being $r = 2$.

4. RESULTS

For the proposed example, the model is identified on the data reported in figure 3, by solving the proposed estimation algorithm. Table 1 gathers the results : line 3 corresponds to the initialisation of the parameters, line 4 shows the estimated parameters at the first iteration, while line 5 gives the results at iteration 7 (where convergence has been reached), which may be compared to the true values at line 2.

Par.	a_1	b_1	a_2	b_2	a_3	b_3	a_4	b_4
true	1	2	-1	3	2	8	-2	6
ini.	2	1	1	1	5	5	-5	4
Iter. 1	0.38	4.27	-0.26	3.1	1.70	7.36	-1.62	5.93
iter. 7	0.99	2.46	-0.96	3.15	1.98	8.40	-1.99	6.08

Table 1. Parameters of the model

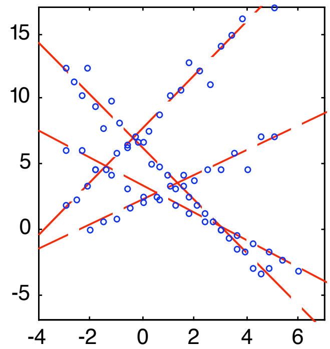

Figure 3. Measurement and estimation

Figure 4 shows the classification results. The weights for the 4 local models are represented for all the data at figure 4a ; figure 4b displays boolean classification derived for the comparison of the weigths.

Although the results of the estimation are satisfactorily, we have to keep in mind that this approach suffers of some drawbacks : in particular potentially local minima may be obtained. However, if some knowledge about the domain in which lie the parameters is available, then the algorithm reveals to be powerful. In practice, we can often make some educated guess of how to give initial values to the local models.

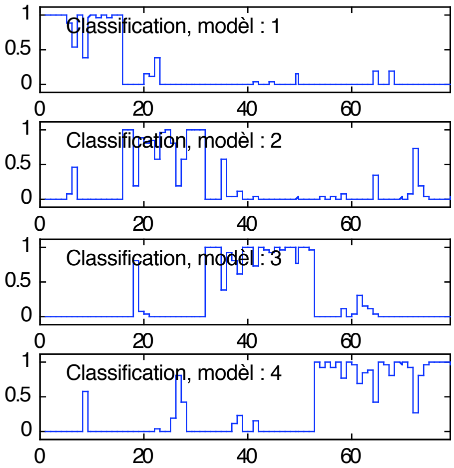

Figure 4a. Fuzzy classification

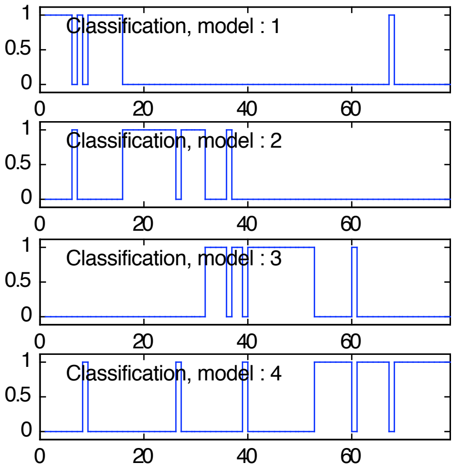

Figure 4. Boolean classification

5 CONCLUSION

In this paper, we have proposed a method for the identification of a linear combination of local linear systems from input and output data. In fact this combination is reduced to a switching between the two systems, the switching time being unknown.

The parameters of the local systems have been estimated by optimizing a cost function
which uses classical sum of squares of the deviation between the output measurements and the output of the model ; this has been performed with a judicious choice of the weighting function allowing the identification for each system with the appropriate set of data.
Because of the preliminary date of the research, the algorithm we proposed in this paper has only been evaluated on a number of experiments.

However, the experiments clearly show the potential of the two approaches and motivate for further developments. In the future, most of the developments will focuse on the determination of the structure of the models. Moreover, the degree of persistency of the exciting input signal remains an open problem. Comparisons of the approach with methods based on clustering will also be investigated as well as methods based on interacting multiple models (Mihaylova, 2001).

REFERENCES

[1] Ding Z., Hong L. An interactive multiple model algorithm with a switching markov chain. Math. Comput. Modelling, 25 (1), pp. 1-9, 1997.

[2] Bemporad A., Roll J., Ljung L. Identification of hybrid systems via mixed-integer programming. Proceedings of the 40th IEEE Conference on Decision and Control, pp. 786-792, 2001.

[3] Breiman L. Hinging hyperplanes for regression, classification and function approximation. IEEE Transations on Information Theory, 39 (3), pp. 999-1013, 1993.

[4] Gasso G., Mourot G. Ragot J. Structure identification in multiple model representation : elimination and merging of local models. IEEE Conference on Decision and Control, Orlando, Floride, 2001.

[5] Mihaylova L., Lampaert V., Bruyninckx H., Sweters J. Hysteresis functions identification by multiple model approach. International Conference MFI, Baden-Baden, 2001.

[6] Murray-Smith R., Johansen T.A. Multiple model approach in modelling and control. Taylor and Francis, 1997.

[7] Pucar P., Sjöberg J. On the hinge finding algorithm for higing hyperplanes. IEEE Transactions of INformation Theory, 44 (3), pp. 1310-1319, 1998.

[8] Verdult V., Verhaegen M. Identification of a weighted combination of multivariable local linear state-space systems from input and output data. Proceedings of the 40th IEEE Conference on Decision and Control, pp. 4760-4765, 2001.