Convergence of SOM Multiple Models Identifier

Alireza Fatehi* & Kenichi Abe**
Abe lab., Dept. of Electrical Eng.,
Tohoku University 5, 980-8579,
Sendai, Japan
* fatehi@abe.ecei.tohoku.ac.jp
** abe@abe.ecei.tohoku.ac.jp

ABSTRACT:

In [2], [3] an algorithm to derive a multiple models set for a plant by the use of the SOM were introduced. The statistic properties of the models are investigated in this paper. As a plant, we consider the linear time invariant one. The parameters of the plant at each step are selected randomly with a specified distribution. Based on this distribution, the point distribution of the parameters of the multiple models is derived for this plant and compared with the plant parameters distribution.

1. INTRODUCTION

The multiple modeling is a modeling method, which provides accurate models under parameter variation. Researchers have introduced various algorithms, mainly based on the adaptive estimation theory to derive suitable models for the plant [6], [7]. An algorithm to find the multiple linear models set for both linear time-variant [2] and nonlinear [3] plants was introduced. The algorithm is comprised of estimating a linear model at each instant of time and then giving these models to a self-organizing map (SOM) neural networks. The SOM gradually learns these instantaneous models and categorize them into a few numbers of models. Consequently, a set of models is evolved by the SOM which make the multiple models set.

The simulations show the ability of the algorithm on the plant identification. In this article, the analytic behaviour of it is going to be investigated. Among various mathematical aspects related to the algorithm, this research has concentrated on the convergence of the estimated parameters and the relation between the parameters of the plant and the estimated parameters. The algorithm uses the methods in the identification and SOM theories. Therefore it is reasonable to approach to its analysis by the known properties of these two theories.

The paper is organized as follows. The identification algorithm is reviewed in section 2. In section 3 the main problem of this research is stated which is dealt in section 4. To illustrate the results of section 4, the theoretical and simulation results for an example have been compared in section 5.

2. IDENTIFICATION ALGORITHM

The principle idea of the multiple modeling by SOM (MMS) algorithm, which is illustrated in fig.(1), is briefly explained in this section. Consider the linear time variant plant described by

\[ y(t) = \theta^T(t)\phi(t) + w(t) \]  \hspace{1cm} (2-1)

where \( \theta \in \mathbb{R}^p \) is the set of time varying parameters of the plant, \( \phi \in \mathbb{R}^p \) the state vector and \( w \) is the noise. It is desired to derive \( M \) linear local models such that each one can model the plant in a subspace of the plant parameters space.

The algorithm has two main steps:

**Step1** At each instant, a model is identified by one of the known and suitable identification methods. For that identification the regression vector consists of data in the last \( N \) steps, where \( N \) is the identification window. The parameters of these models are called instantaneous model parameters (IMP). In our study, the quadratic criteria prediction error (PE) estimation is used. So the instantaneous parameter estimations are described as:

Fig. 1 The structure of the identification algorithm
\[ \hat{\theta}_N(t) = \arg \min_{\theta} \left[ \frac{1}{N \tau_{\max}} \sum_{i=1}^{N} \frac{1}{2} (y(\tau) - \tilde{y}(\tau | \theta))^2 \right] \]

where

\[ \tilde{y}(\tau | \theta) = \theta^T(t) \phi(t) \]

It is evident that \( \hat{\theta}_N \) changes by time and as a consequence, after for example \( k \) steps there are \( k \) sets of estimated parameters.

**Step 2)** Derived IMPs are given to SOM neural networks, which have \( p \) dimensions inputs and \( M \) nodes or neurons. The SOM is an unsupervise-learning neural networks which defines a mapping from the input data space \( \mathbb{R}^p \) onto an array of \( M \) nodes. Every node is associated with a reference vector \( w_i = \{w_{i1}, w_{i2}, ..., w_{ip}\} \in \mathbb{R}^p \). The input to the SOM is compared to \( w_i \)'s by

\[ d^i = |w - \hat{\theta}| \]

The neuron with the least \( d^i \), that is the neuron which its reference vector is the nearest to the input vector, is selected as the winner neuron. The learning algorithm of SOM computes new reference vectors for the winner neuron and its so called neighbours so as to decrease the winner neuron and its neighbours reference vector distances with the input vector. Consequently, the input space is partitioned into \( M \) subspaces so that if the input is from a specific subspace the winner neuron will be its corresponding neuron.

In MMS, as the IMP is computed in the step (1) above, it is taught to the network. So gradually the SOM partition the space of the IMP, however, just the portion which has been spanned by IMPs up to time \( t \). SOM not only categorizes the inputs but also accumulates their information. Therefore, the reference vectors of the SOM are evolved into some local models parameters of the plant, which are called multiple models. To decrease the cumulative error between the plant and the models, it is desired that the point distributions of the multiple models approximate the plant parameters distributions.

### 3. STATEMENT OF THE PROBLEM

The multiple models set

\[ \hat{\Theta} = \{\hat{\Theta}^1, \hat{\Theta}^2, ..., \hat{\Theta}^M\}, \hat{\Theta}^i \in \mathbb{R}^p \]

is derived for the plant (2-1) by applying the algorithm of previous section. Although the algorithm has been applied to both time variant and nonlinear plants, the scope of the present analysis is restricted to the linear time variant case, where the sequence of the parameters \( \{\theta(t)\} \) of the plant are independent random variables with p.d.f.

\[ \Theta = f_\theta(\theta) \]

at every instance and the noise is also described by

\[ w \sim N(0, \sigma^2) \]

The rate of change of \( \theta \) requires supposing rather less than the sampling time of the plant observation. Considering these assumptions the main problem is stated as:

**MP)** “What is the relation between the plant parameters and the \( M \) estimation parameters sets?”

There are many aspects to consider for this problem. Among them the relation between the plant parameters p.d.f. and the point density of the \( M \) multiple models are noteworthy. Assume that the plant parameters are selected randomly with the p.d.f. (3-2) and assume also that the parameters of the \( M \) models (or the reference vectors) has the point density

\[ \hat{\Theta} \sim f_\theta(\hat{\theta}) \]

In this paper the relation between them are going to be determined.

### 4. MAIN RESULT

The SMM algorithm has two stages; 1) estimation of IM parameters, \( \hat{\theta}_N \), 2) learning these parameters by SOM to derive the multiple models, \( \hat{\theta} \). This implies that the mathematical theory of that is also based on the theories of both identification and SOM. The approach to find the relation between the \( f_\theta \) and \( f_\hat{\theta} \) is to combine the requisite properties of each of the two theories. From this standpoint the main problem of section 3 is divided to the following two sub-problems.

**SP1)** Assume that the plant parameters p.d.f. is \( f_\theta \). What will be the \( f_\hat{\theta} \), the p.d.f of the IMPs?

**SP2)** If the IMP with the p.d.f. \( f_\hat{\theta} \) are given to the SOM with \( M \) nodes, What will be \( f_\hat{\theta} \), the point distribution of the parameters of the \( M \) multiple models? In other words, what is the point distribution of the reference vectors of SOM?

Both above sub-problems have been already solved [4], [5]. In this section the main features of their solutions will be mentioned and then they will be combined to solve the main problem.

**Solution of SP1** As explained in [4], under some specific assumptions the p.d.f. of the IMPs of (2-2) is normal with specific mean and variance. Assume that

i) The plant parameters, \( \theta \), is constant,

ii) The plant is uniformly stable,

iii) The external input to the plant is bounded and deterministic,

iv) The elements of the sequence \( \{\theta(t)\} \) are jointly quasi-stationary,

v) The noise \( w(t) \) is a sequence of independent random variables with zero mean, variance \( \sigma^2 \) and bounded moments of order more than 4.

Under the above assumptions, it has been proved that for large \( N \)

\[ \sqrt{N} \hat{\theta}_N \text{ approaches to a normal distribution}, \]

\[ \lim_{N \to \infty} \sqrt{N} \hat{\theta}_N = N(\hat{\theta}^*, P) \]

in which
\[ \tilde{\theta}^* \rightarrow \arg \min_{\theta} V(\theta) \]  

\[ P = \sigma^2 E(\psi(t, \theta)\psi^T(t, \theta)) \]  

where
\[ V(\theta) = E - \frac{1}{2} \varepsilon^2(t, \theta) \]  
\[ \varepsilon(t, \tilde{\theta}) = y(t) - \tilde{y}(t \mid \tilde{\theta}) \]  
\[ \psi(t, \tilde{\theta}) = \frac{d}{d\theta} \tilde{y}(t \mid \theta) \bigg|_{\theta = \tilde{\theta}} \]

For large N the above relations may use as an approximation for the distribution of the estimated parameter, that is
\[ f_{\tilde{\theta}_N \mid \theta} (\tilde{\theta}_N \mid \theta) = \left( \frac{1}{\sqrt{2\pi P/N}} \right) \exp \left( - \frac{1}{2} (\tilde{\theta}_N - \theta)^2 / P/N \right) \]  

As is mentioned in assumption (i), (4-1) is valid if \( \theta \) is constant all through the N steps of identification. If \( \theta \) changes slowly it is possible to consider it constant for a period of time. Thus the class of the plants for which (4-1) is valid, is limited to those which the parameters change slowly compared to the identification window. Now, suppose that \( \theta(t) \) changes as described in (3-2) and remains constant through the identification window. Then for the p.d.f. of \( \tilde{\theta}_N \) we have
\[ f_{\tilde{\theta}_N \mid \theta} (\tilde{\theta}_N \mid \theta) = \int f_{\tilde{\theta}_N \mid \theta} (\tilde{\theta}_N \mid \theta) f_{\theta} (\theta) d\theta \]

**Solution of SP2** SOM is based on the vector quantization (VQ) method. VQ is a signal approximation method that generates a finite number of so-called codebook vectors (which in the SOM literature they are called reference vectors). This codebook vectors approximate a probability density function. Zador [7] (see also [5,pp49]) proved that the optimal selection of \( w' \) values is such that their point densities approximates to \( p(x)/[p(x)]^{p+2} \), where \( p(x) \) is the p.d.f. of the input data and \( p \) is the dimension of inputs, at least when the number of the output nodes is large. So if the p.d.f. of the IMP described as (4-8) then for large M
\[ f_{\tilde{\theta}} (\tilde{\theta}) = \frac{\left[ f_{\tilde{\theta}_N \mid \theta} (\tilde{\theta}_N \mid \theta) \right]^{p+2}}{\int \left[ f_{\tilde{\theta}_N \mid \theta} (\tilde{\theta}_N \mid \theta) \right]^{p+2} d\tilde{\theta}_N} \]  

where \( p \) is the number of the parameters. It implies that if the number of the parameters of the model is large, the output point density will be a better approximation of the IMP p.d.f.

**5. ILLUSTRATIVE EXAMPLE**

In this section, to illustrate the equation (4-9), the theoretical and the simulation p.d.f. of an example are derived and compared to each others. The example is adapted from [4]. Consider the system
\[ y(t) + a_y(t-1) = u(t-1) + e(t) \]  
where the input \( \{u(t)\} \) is a white noise with mean zero and variance \( \mu^2 \). \( \{e(t)\} \) is also a white noise with mean zero and variance \( \sigma^2 \) and \( u \) is independent of \( e \). We are going to identify the models with the structure
\[ \hat{y}(t) = -\hat{a}^t y(t-1) + u(t-1) \]

for the plant (5-1). In the first step, the instantaneous model
\[ \tilde{y}(t \mid \tilde{a}) = -\tilde{a}^t y(t-1) + u(t-1) \]

is identified, using a quadratic prediction error criteria. For this problem
\[ \varepsilon(t, \tilde{a}) = (\tilde{a} - a_y) y(t-1) + e(t) \]

where \( e(t) \) is the estimation of \( e(t) \) obtained by the algorithm. Also
\[ \psi(t, a) = \frac{d}{da} \tilde{y}(t \mid a) \]

For large N the mean and the variance of the \( \tilde{a}_N \) are derived as
\[ E\tilde{a}_N = a_y \]
\[ Cov\tilde{a}_N = \frac{1}{N} P_N \equiv \frac{\sigma^2}{\mu^2 + \sigma^2} \frac{1-a_y^2}{N} \]  

using the equations (4-1) to (4-6). The plant has one unknown parameter so the SOM reference vectors have also one element. Therefore, the p.d.f. of the \( \tilde{a} \) is proportion to 3rd root of the \( \tilde{a}_N \) p.d.f. In addition, as an effect of the \( Cov\tilde{a}_N \), the deviation of the models parameters from the plant ones changes by the variance of the noise and the external input and the identification window.

This example was simulated for 10,000 different parameters. The parameter of the plant was supposed to have a uniform p.d.f.
in [0.6,0.8]. Also the variance of the inputs \( \mu^2 = 1\), and of the noise \( \sigma^2 = 0.01\). A set of data in 10 steps were used for identification, i.e., the identification window \( N = 10\). The value of the parameter in the identification window was considered constant. The SOM has 150 neurons, which is quite large for this plant but are considered to justify the results. The learning rate factor for each neuron were 0.2 at the beginning and changes as 

\[
\alpha_i(t+1) = \frac{\alpha_i(t)}{1 + h_i \alpha_i(t)}
\]  

(5-8)

where \( \alpha_i(t) \) is the learning factor of the \( i \)th neuron, \( c \) is the winner neuron and \( h \) is the neighbourhood function.

The theoretical pdf of the plant, instantaneous models and the multiple models are shown in the figure (2). Figure (3) to (5) demonstrate the sample p.d.f. of each of the plant and models parameters together with their theoretic p.d.f.s.

6. CONCLUSION

This paper presents the relation between the plant parameters pdf and the point distribution of the multiple models which are obtained by MMS algorithm. This relation gives criteria of the goodness of the algorithm. It is reasonable to think on the pdf of the estimated parameters introduced by SOM as the best model for the plant. For this purpose, the relation between the pdf of the inputs to the SOM and the pdf of the weights of the winner neuron is required to be found.

REFERENCES