New Optimal Approach for the Identification of Takagi-Sugeno Fuzzy Model

AGUSTÍN JIMÉNEZ, BASIL M AL-HADITHI AND FERNANDO MATÍA
Departamento de Automática, Ingeniería Electrónica e Informática Industrial
Intelligent Control Group, Universidad Politécnica de Madrid,
J.Gutierrez A bascal, 2. 28006-M adrid, Spain,

Abstract—A new and efficient approach is presented that lead to improve the local and global approximation (matching) and modelling capability of Takagi-Sugeno (T-S) fuzzy model. The main aim is obtaining high function approximation accuracy and fast convergence. The main problem encountered is that T-S model can not be applied when the membership functions are overlapped by pairs. This restricts the application of the T-S model because this type of membership function has been widely used during the last two decades in the stability, controller design of fuzzy systems and are popular in industrial control applications. The approach developed here can be considered as a generalized version of T-S fuzzy model with optimized performance in approximating nonlinear functions. Illustrative examples are chosen to evaluate the robustness and performance of the proposed method and the high accuracy obtained in approximating nonlinear systems locally and globally in comparison with the original T-S model. It is worth noting that proposed algorithm achieve a remarkable performance in comparison with the original T-S model.

Index Terms—Nonlinear systems, Fuzzy systems, Takagi-Sugeno fuzzy model, Universal approximators, optimization

1 Introduction
Nonlinear control systems based on the Takagi-Sugeno (T-S) fuzzy model [21], [22] have attracted lots of attention during the last twenty years (e.g., see [1], [2], [6], [9], [10], [12], [13], [14], [15], [16], [17], [19] and [24]. It provides a powerful solution for development of function approximation, systematic techniques to stability analysis and controller design of fuzzy control systems in view of fruitful conventional control theory and techniques.

This model is formed by using a set of fuzzy rules to represent a nonlinear system as a set of local affine models which are connected by fuzzy membership functions [3] and [4].

This fuzzy modelling method presents an alternative technique to represent complex nonlinear systems [7], [23], [27] and [29], and reduces the number of rules in modelling higher order nonlinear systems [9] and [22].

T-S fuzzy models are proved to be universal function approximators as they are able to approximate any smooth nonlinear functions to any degree of accuracy in any convex compact region [7], [11], [16], [23], [27] and [29]. This result provides a theoretical foundation for applying T-S fuzzy models to represent complex nonlinear systems approximately.

Great attention has been paid to the identification of T-S fuzzy models and several results have been obtained [5], [11], [18], [25] and [28]. They are based upon two kinds of approaches, one is to linearize the original nonlinear system in various operating points when the model of the system is known, and the other is based on the input-output data collected from the original nonlinear system when its model is unknown. The authors in [5] use a fuzzy clustering method to identify T-S fuzzy models, including identification of the number of fuzzy rules and parameters of fuzzy membership functions, and identification of parameters of local linear models by using a least squares method [20] and [26]. The goal is to minimize the error between T-S fuzzy models and the corresponding original nonlinear systems. The authors in [11] suggest a method to identify T-S fuzzy models. Their method aims at improving the local and global approximation of T-S model. However, this complicates the approximation in order to obtain both targets. It has been shown that constrained and regularized identification methods may improve interpretability of constituent local models as local linearizations, and locally weighted least squares method may explicitly address the trade-off between the local and global accuracy of T-S fuzzy models.

In [22], the authors develop an interesting method to identify nonlinear systems using input-output data. They divide the identification process into three steps; premise variables, membership functions and
consequent parameters. With respect to membership functions, they apply nonlinear programming technique using the complex method for the minimization of the performance index.

In [20] Sugeno and Tanaka developed a successive method for identifying T-S model. They combined the least square method, the complex method and an unbiased criterion; while parameter adjustment was adjustment rules and a weighted recursive least squares algorithm. In [26] the authors presented a method for generating fuzzy rules by learning from examples.

As we will be demonstrated in this article, the T-S model can not be applied when the membership functions are overlapped by pairs. This limits the usage of the model because as it was shown in the last two decades that the major part of the results obtained in the field of stability and controller synthesis are based on this type of membership functions. Moreover, the methods presented here are characterized by the high accuracy obtained in approximating nonlinear systems locally and globally in comparison with the original T-S model.

The rest of the paper is organized as follows. Section 2 presents T-S identification Method. Section 3 introduces restrictions of T-S identification Method. Section 4 demonstrates the proposed approach to improve and generalize the T-S model. Section 5 entails various examples to demonstrate the validity of the proposed approach. These examples show that the proposed approach are less conservative than those based on (standard) T-S model and illustrate the utility of the proposed approach in comparison with T-S model.

2 Identification of T-S Model

An interesting method of identification is presented in [22]. The idea is based on estimating the nonlinear system parameters minimizing a quadratic performance index. The method is based on the identification of functions of the following form:

\[ f: \mathbb{R}^n \rightarrow \mathbb{R} \]

\[ y = f(x_1, x_2, \ldots, x_n) \]

Each IF-THEN rule \( R_1^{(i_1 \ldots i_n)} \), for an \( n \)th order system can be rewritten as follows:

\[ S^{(i_1 \ldots i_n)}: \text{if } x_1 \text{ is } M_1^{i_1} \text{ and } x_2 \text{ is } M_2^{i_2} \text{ and } \ldots \text{ and } x_n \text{ is } M_n^{i_n} \]

\[ \text{then } \hat{y} = p_0^{(i_1 \ldots i_n)} + p_1^{(i_1 \ldots i_n)} x_1 + p_2^{(i_1 \ldots i_n)} x_2 + \ldots + p_n^{(i_1 \ldots i_n)} x_n \]

Where the fuzzy estimation of the output is:

\[
\hat{y} = \frac{\sum_{i=1}^n \sum_{i=1}^{\text{samples}} w^{(i_1 \ldots i_n)}(x_i) \left[ p_0^{(i_1 \ldots i_n)} + p_1^{(i_1 \ldots i_n)} x_1 + \ldots + p_n^{(i_1 \ldots i_n)} x_n \right]}{\sum_{i=1}^n \sum_{i=1}^{\text{samples}} w^{(i_1 \ldots i_n)}(x_i)}
\]

Let \( m \) be a set of input/output system samples \( \{x_{1k}, x_{2k}, \ldots, x_{nk}, y_k\} \). The parameters of the fuzzy system can be calculated by minimizing the following quadratic performance index:

\[
J = \sum_{k=1}^m (y_k - \hat{y}_k)^2 = \|Y - XP\|^2
\]

Where

\[ Y = \left[y_1, y_2, \ldots, y_m\right]^T \]

\[ P = \begin{bmatrix} p_0^{(1,1)} & p_1^{(1,1)} & \ldots & p_n^{(1,1)} \\
                    p_0^{(2,1)} & p_1^{(2,1)} & \ldots & p_n^{(2,1)} \\
                    \vdots & \vdots & \ddots & \vdots \\
                    p_0^{(m,1)} & p_1^{(m,1)} & \ldots & p_n^{(m,1)} \end{bmatrix}^T \]

\[ X = \begin{bmatrix} \beta_1^{(1,1)} \beta_1^{(2,1)} x_1 & \ldots & \beta_1^{(m,1)} x_n \\
                        \beta_2^{(1,1)} \beta_2^{(2,1)} x_1 & \ldots & \beta_2^{(m,1)} x_n \\
                        \vdots & \vdots & \ddots & \vdots \\
                        \beta_m^{(1,1)} \beta_m^{(2,1)} x_1 & \ldots & \beta_m^{(m,1)} x_n \end{bmatrix} \]

and

\[
\beta_k^{(i_1 \ldots i_n)} = \frac{w^{(i_1 \ldots i_n)}(x_k)}{\sum_{i=1}^n \sum_{i=1}^{\text{samples}} w^{(i_1 \ldots i_n)}(x_i)}
\]

If \( X \) is a matrix of complete rank, the parameters of the fuzzy system are obtained as follows:

\[
J = \|Y - XP\|^2 = (Y - XP)^T (Y - XP)
\]

\[ \nabla J = X^T (Y - XP) = X^T Y - X^T X P = 0 \]

\[ P = (X^T X)^{-1} X^T Y \]

3 Restrictions of T-S identification Method

The method proposed in [22] arises serious problems as it can not be applied in the most common case where the membership functions are those shown in fig. 1.
The membership functions
\( \mu_1(x_i) = \frac{x_{i2} - x_i}{x_{i2} - x_{i1}} \), \( \mu_2(x_i) = \frac{x_i - x_{i1}}{x_{i2} - x_{i1}} \) are defined in an interval \([x_{i1}, x_{i2}]\) which should verify:
\[
\begin{align*}
\mu_1(x_i) &= 1 & \mu_2(x_i) &= 0 \\
\mu_2(x_i) &= 0 & \mu_1(x_i) &= 1 \\
\mu_1(x_i) + \mu_2(x_i) &= 1
\end{align*}
\]

In this case, it can easily be demonstrated that the matrix \(X\) is not of complete rank and therefore \(X'X\) is not invertible, which makes the method of T-S invalid. This result can be easily proven as follows:

Supposing:
\[
f: \Re \to \Re
\]
applying T-S method, each row of the matrix \(X\) is of the form:
\[
X_k = \left[ \mu_k^1 \mu_k^2 x_k \right]
\]
verifying:
\[
\left[ \begin{array}{cccc}
-1 & 0 & 1 & 0 \\
0 & -1 & 1 & 0 \\
1 & 0 & -1 & 0
\end{array} \right] = 0
\]

The rank of \(X\) in this case is 3. In other words, the columns of \(X\) are linearly dependent which in turn makes impossible the use of the identification method proposed in [22].

Analyzing another example of two variables:
\[
f: \Re^2 \to \Re
\]
\[
y = f(x_1, x_2)
\]
Each row of the matrix \(X\) is of the form:
\[
X_k = \left[ \begin{array}{cccc}
\mu_k^{11} & \mu_k^{12} & \mu_k^{21} & \mu_k^{22} \\
\mu_k^{11} x_{k1} & \mu_k^{12} x_{k2} & \mu_k^{21} x_{k1} & \mu_k^{22} x_{k2} \\
\mu_k^{11} x_{k1}^2 & \mu_k^{12} x_{k2}^2 & \mu_k^{21} x_{k1}^2 & \mu_k^{22} x_{k2}^2 \\
\mu_k^{11} x_{k1}^3 & \mu_k^{12} x_{k2}^3 & \mu_k^{21} x_{k1}^3 & \mu_k^{22} x_{k2}^3
\end{array} \right]
\]

It can be noticed that the columns 1, 3, 4 and 6 have the same form as in the previous example multiplied by a constant \(\mu_k^{11}\) and therefore they are linearly dependent as well. The same thing happens with the columns 6, 9, 10 and 12, etc. In fact, the rank of the matrix in this case is 8.

The solution proposed in [22] avoids the occurrence of this situation. In order to identify a function in the interval \([x_{i1}, x_{i2}]\) using T-S method, certain intermediate points are chosen of the form:
\[
x_{i1} \in [x_{i1}, x_{i2}] \quad y \in [x_{i1}, x_{i2}]
\]

And they use membership functions which verify:
\[
\mu_k^1(x_i) = \left\{ \begin{array}{lr}
x_i - x_{i2} & x_{i1} \leq x_i \leq x_{i2} \\
x_{i1} - x_{i2} & x_{i2} \leq x_i \leq x_{i1}
\end{array} \right.
\]
\[
\mu_k^2(x_i) = \left\{ \begin{array}{lr}
x_i - x_{i1} & x_{i1} \leq x_i \leq x_{i2} \\
x_{i2} - x_{i1} & x_{i2} \leq x_i \leq x_{i1}
\end{array} \right.
\]

And thus:
\[
\mu_k^1(x_{i1}) = 1 \quad \mu_k^1(x_{i2}) = 0 \\
\mu_k^2(x_{i1}) = 0 \quad \mu_k^2(x_{i2}) = 1
\]

Which impedes that the domains of these functions being overlapped and therefore it can be observed that, except for some isolated points,
\[
\mu_k^1(x_i) + \mu_k^2(x_i) \neq 1
\]

And thus, in general, the matrix \(X\) will be of full rank and the method is applicable.

This solution can be clearly seen in [22] where the authors find the optimum membership functions minimizing the performance index and reducing the problem to a nonlinear programming one. For this reason, they use the well-known complex method for the minimization. This can obviously be observed in the illustrative examples selected by the authors in [22] where all the identified memberships are non overlapping ones.
4 Proposed Approach

The restriction of T-S identification method for the case presented in the previous section does not mean the non-existence of solutions. The problem comes from the fact that the solution should fulfil:

$$\nabla J = X^tY - X^tXP = 0$$  \hspace{1cm} (5)

But as it was shown above, the columns of the matrix $X$ are linearly dependent and consequently $X^tX$ is not an invertible matrix, therefore it is impossible to calculate $P$ through:

$$P = (X^tX)^{-1}X^tY$$  \hspace{1cm} (6)

Nevertheless, as the rows of $X^t$ are linearly dependent, the independent term in equation (5) will have the same dependence among its rows and thereupon the rank of the system matrix will be the same as the rank of the extended matrix by the independent term.

$$\text{rank}(X^tX) = \text{rank}(X^tX|X^tY)$$  \hspace{1cm} (7)

And so the system has solution. In other words, the system is a compatible indeterminate one, that is, if $P$ is a solution of (5) and $K$ is a Kernel of $X^tX$.

$$K = \text{Ker}(X^tX)$$  \hspace{1cm} (8)

Then $P^* = P + K\beta$ will also be a solution, where $\beta$ is any real vector whose dimension is same to that of the Kernel. Therefore, the problem is not the lack of a solution rather the existence of infinite solutions and the key idea is the ability to find one of them. Several proposals can be made to select a solution. In our case, we propose a solution with lower norm.

4.1 Reduced Matrix Approach-Optimum Solution

The approach implies the search for an exact and optimum solution at the expense of increasing the degree of complexity and computational cost. The problem is stated as follows:

$$\text{minimize} \quad \|P\|^2$$  \hspace{1cm} (9)

subject to

$$\nabla J = X^tY - X^tXP = 0$$

As already stated, the system of restriction equations is a compatible undetermined one, and therefore, there are linearly dependent restrictions upon others. What is proposed is to eliminate the linear restrictions until obtaining a system with all its restrictions being linearly independent. We obtain a reduced system of equations linearly independent

$$Y_r - X_r P = 0$$  \hspace{1cm} (10)

Then, Lagrange theorem can be applied, defining a lagrangian function

$$L = P^tP + \lambda^t(Y_r - X_r P)$$  \hspace{1cm} (11)

that fulfils:

$$\nabla \lambda L = 2P^t - \lambda^tX_r = 0 \quad \Rightarrow \quad 2P - \lambda X_r = 0$$

$$\nabla \lambda L = (Y_r - X_r P)^t = 0 \quad \Rightarrow \quad Y_r - X_r P = 0$$

This can also be represented in matrix form:

$$\begin{bmatrix} 2I & -X_r^t \\ X_r & 0 \end{bmatrix} \begin{bmatrix} P \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ Y_r \end{bmatrix}$$

And the solution will be:

$$\begin{bmatrix} P \\ \lambda \end{bmatrix} = \begin{bmatrix} 2I & -X_r^t \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ Y_r \end{bmatrix}$$

5 Examples

In the following examples, we will compare this method with the one proposed by T-S. Different non-linear functions will be proposed and the fuzzy models will be obtained assigning an interval $[x_{i1}, x_{i2}]$ for each variable $x_i$. In this interval, two fuzzy sets are defined whose membership functions are:

$$\mu_i^1 = \frac{x_i - x_{i1}}{x_{i1} - x_{i2}} \quad \mu_i^2 = \frac{x_{i2} - x_i}{x_{i2} - x_{i1}}$$

However, as demonstrated above, these membership functions can not be used directly in the method of T-S, as the resulting matrix $X$ would not be
complete rank. In order to compare these methods, we use a factor \(0 \leq \alpha \leq 1\), which determines two points within the interval, this means:

\[
\begin{align*}
    x_1^* &= x_1 - \alpha(x_2 - x_1) \\
    x_2^* &= x_1 + \alpha(x_2 - x_1)
\end{align*}
\]

And so, we define two fuzzy sets whose membership functions are:

\[
\mu_1(x) = \begin{cases}
    \frac{x_1 - x_1^*}{x_1 - x_2} & x_1 \leq x \leq x_2^* \\
    0 & x_2^* \leq x \leq x_2
\end{cases}
\]

\[
\mu_2(x) = \begin{cases}
    0 & x_1 \leq x \leq x_1^* \\
    \frac{x_2 - x_1}{x_2 - x_1^*} & x_1^* \leq x \leq x_2
\end{cases}
\]

which are those used with the direct method of T-S. As a measure of error for comparing these methods, the maximum of the absolute values of the errors is used, the same method applied in by T-S in [22].

**Example 1**

Consider the following simple nonlinear system:

\[
\dot{x} = x^2
\]

It is aimed to estimate this system:

\[
y = x^2 \quad x \in [0,1]
\]

Let us suppose that we define in this interval two fuzzy sets with their corresponding membership functions as follows:

\[
\mu_1(x) = 1 - x \\
\mu_2(x) = x
\]

The objective is to calculate the corresponding fuzzy model in an optimum form:

\[
S^1 : \text{if } x \text{ is } M_1^1 \text{ then } y = p_1^1 + p_1^1 x \\
S^2 : \text{if } x \text{ is } M_1^2 \text{ then } y = p_0^2 + p_1^2 x
\]

In order to identify the nonlinear function, we take 20 points uniformly distributed in the interval \([0, 1]\). Applying firstly the reduced matrix method, the product \(X'X\) is calculated as follows:

\[
\begin{align*}
    &6.8421 & 1.5789 & 3.1579 & 1.5789 \\
    &1.5789 & 0.6333 & 1.5789 & 0.9456 \\
    &3.1579 & 1.5789 & 6.8421 & 5.2632
\end{align*}
\]

Its rank is 3 and can be reduced to \(X_r\):

\[
\begin{align*}
    &6.8421 & 1.5789 & 3.1579 & 1.5789 \\
    &1.5789 & 0.6333 & 1.5789 & 0.9456 \\
    &3.1579 & 1.5789 & 6.8421 & 5.2632
\end{align*}
\]

The resultant \(P\) vector of the parameters in the consequent part is:

\[
\begin{align*}
    &0.0000 \quad -0.3333 \quad 0.3333 \quad 0.6667
\end{align*}
\]

This means that the resultant fuzzy rules are:

\[
S^1 : \text{if } x \text{ is } M_1^1 \text{ then } y = -0.3333x \\
S^2 : \text{if } x \text{ is } M_1^2 \text{ then } y = 0.3333 + 0.6667x
\]

The result in this case is obtained with an error of 7.0031e-016, which is practically zero. In the method of T-S, let us suppose that \(\alpha = 0.7\). The fuzzy model is:

\[
\begin{align*}
    &6.8421 & 1.5789 & 3.1579 & 1.5789 \\
    &1.5789 & 0.6333 & 1.5789 & 0.9456 \\
    &3.1579 & 1.5789 & 6.8421 & 5.2632
\end{align*}
\]

The identification error is 2.01e-002, which is higher than the one obtained in the proposed method. Increasing \(\alpha = 0.8\), the error is reduced to 1.60e-02, i.e., the identification is improved. When \(\alpha = 0.9\), the error becomes 8.1e-003, where the identification is again improved. The same occurs for \(\alpha = 0.95\) where the error is reduced to 4.4e-003.

In other words, as the factor \(\alpha\) is approaching unity the identification is improved, but can not reach the optimum which precisely occurs when \(\alpha = 1\). Since at this value, the matrix \(X\) is not of complete rank and therefore the matrix \(X'X\) is not invertible.

Even when \(\alpha\) is approaching unity, the condition number of the matrix \(X\) starts increasing which indicates that \(X'X\) is approaching the singularity and therefore its inverse is no longer numerically reliable. For instance, when \(\alpha = 0.999\), the condition number of \(X\) is 5.2523e+004 which shows clearly a non reliable result.

**Example 2**

Consider the following nonlinear system
Applying the reduced matrix method, the resultant fuzzy rules are:

\[ S^{11}: \text{if } x_1 \text{ is } M^1_1 \text{ and } x_2 \text{ is } M^1_2 \]
\[ \text{then } y = -0.0016 + 0.9510x_1 - 0.6985x_2 \]

\[ S^{12}: \text{if } x_1 \text{ is } M^1_1 \text{ and } x_2 \text{ is } M^2_2 \]
\[ \text{then } y = 0.0103 + 9510x_1 + 0.8015x_2 \]

\[ S^{21}: \text{if } x_1 \text{ is } M^1_1 \text{ and } x_2 \text{ is } M^1_2 \]
\[ \text{then } y = -1.2998 + 2.3090x_1 - 1.1312x_2 \]

\[ S^{22}: \text{if } x_1 \text{ is } M^1_1 \text{ and } x_2 \text{ is } M^2_2 \]
\[ \text{then } y = -0.7625 + 2.3090x_1 + 0.3688x_2 \]

The identification model is obtained with an error of 7.1497e-015. Applying the method presented by T-S, with 0.9 for \( \alpha \), the resultant fuzzy model is:

\[ S^{11}: \text{if } x_1 \text{ is } M^1_1 \text{ and } x_2 \text{ is } M^1_2 \]
\[ \text{then } y = -0.0085 + 0.5910x_1 - 1.0541x_2 \]

\[ S^{12}: \text{if } x_1 \text{ is } M^1_1 \text{ and } x_2 \text{ is } M^2_2 \]
\[ \text{then } y = 0.457 + 0.5910x_1 + 0.4273x_2 \]

\[ S^{21}: \text{if } x_1 \text{ is } M^1_1 \text{ and } x_2 \text{ is } M^1_2 \]
\[ \text{then } y = -0.948 + 1.9274x_1 - 1.3607x_2 \]

\[ S^{22}: \text{if } x_1 \text{ is } M^1_1 \text{ and } x_2 \text{ is } M^2_2 \]
\[ \text{then } y = -0.4880 + 1.9274x_1 + 0.1465x_2 \]

The identification error is 2.14e-02. But in the TS method, using \( \alpha = 0.8 \), which is the value that after several attempts provides better results, we get:

\[ S^{11}: \text{if } x_1 \text{ is } M^1_1 \text{ and } x_2 \text{ is } M^1_2 \]
\[ \text{then } y = -0.0016 + 0.9510x_1 - 0.6985x_2 \]

\[ S^{12}: \text{if } x_1 \text{ is } M^1_1 \text{ and } x_2 \text{ is } M^2_2 \]
\[ \text{then } y = 0.0103 + 9510x_1 + 0.8015x_2 \]

\[ S^{21}: \text{if } x_1 \text{ is } M^1_1 \text{ and } x_2 \text{ is } M^1_2 \]
\[ \text{then } y = -1.2998 + 2.3090x_1 - 1.1312x_2 \]

\[ S^{22}: \text{if } x_1 \text{ is } M^1_1 \text{ and } x_2 \text{ is } M^2_2 \]
\[ \text{then } y = -0.7625 + 2.3090x_1 + 0.3688x_2 \]

The identification error is 3.07e-02. In this case, with \( \alpha = 0.9 \) the results becomes worse because the condition number for the matrix \( X \) is 3.4381e+015 which is the same as that in the previous case, where the matrix \( X \) does not depend upon the function rather on the interval.

6 Conclusions

A new optimization method has been developed to improve the local and global approximation and modelling capability of Takagi-Sugeno (T-S) identification methodology model. The main problem encountered is that T-S model can not be applied when the membership functions are overlapped by pairs. This restricts the application of the T-S model because this type of membership function has been widely used during the last two decades in fuzzy control applications. An optimal solution has been proposed to reduce the error between the original system and the identified one. The results obtained have shown tangible improvement compared to the solution offered by T-S. Several illustrative examples have been presented to evaluate the validity and performance of the proposed method and the high accuracy obtained in approximating nonlinear systems locally and globally in comparison with the original
T-S model. The results obtained by applying the proposed method have demonstrated better results in comparison with the original T-S.

References


