Block Sparse Representations in Modified Fuzzy C-Regression Model Clustering Algorithm for TS Fuzzy Model Identification

Tanmoy Dam
Electrical Engineering Department
Indian Institute of Technology Kharagpur
Kharagpur-721302, India
Email: damtanmoy10@gmail.com

Alok Kanti Deb
Electrical Engineering Department
Indian Institute of Technology Kharagpur
Kharagpur-721302, India
Email: alokkanti@ee.iitkgp.ernet.in

***Abstract*—A novel objective function based clustering algorithm has been introduced by considering linear functional relation between input-output data and geometrical shape of input data. Noisy data points are counted as a separate class and remaining good data points in the data set are considered as good clusters. This noise clustering concept has been taken into the proposed objective function to obtain the fuzzy partition matrix of product space data. Block orthogonal matching pursuit algorithm is applied to determine the optimal number of rules from the over specified number of rules (clusters). The obtained fuzzy partition matrix is used to determine the premise variable parameters of Takagi-Sugeno (TS) fuzzy model. Once, the premise variable parameters and optimal number of rules (clusters) are identified then formulate the rule construction for identification of linear coefficients of consequence parameters. The effectiveness of the proposed algorithm has been validated on two benchmark models.**

I. Introduction

Fuzzy set theory has been extensively used in nonlinear system modeling and model based control. Among various fuzzy models, TS fuzzy model is quite popular in system identification domain because of its simple, interpretable structure to describe the complex nonlinear model in a simple way. Apart from system modeling, it is found in many areas such as fuzzy control, fault tolerant control [1]–[3] etc.

TS fuzzy model mainly consists of two aspects: determinations of structure and parameter identification i.e. determine premise variable parameters and identification of consequence variable coefficients. In structure identification, clustering algorithm has been found for partitioning the data space. Batch processing algorithms (such as Least Square (LS), weighted recursive least square (WRLS)) are used to estimate of consequence variable parameters [1], [8].

Partitioning the input-output data space into fuzzy subspaces is a primary task in TS fuzzy model. Fuzzy clustering algorithms are the most desirable technique for partitioning the data space to obtain fuzzy subspaces and translate them into fuzzy rules i.e. each rule in TS fuzzy model is associated with a fuzzy cluster [4]. Fuzzy C-Means (FCM), Gustafson-Kessel (G-K), Gath-Geva (G-G) , Fuzzy C-Regression Model (FCRM), Enhanced Fuzzy C-Regression Model (EFCRM) are popular clustering algorithms used in structure identification part and LS, WRLS, orthogonal least square (OLS) technique has been applied to estimate the consequence coefficients [8]. The (FCM) [5] clustering algorithm is partitioning a data space in spherical nature. If the input-output data is arbitrary in shape then FCM based TS fuzzy model could not give adequate model accuracy [6]. The modified version of FCM algorithm, known as, Gustafson-Kessel (G-K) algorithm, is considers data distribution using the euclidian distance metric [7] whereas FCM algorithm considers the simple euclidian distance metric. However, FCM, G-K, G-G algorithm is a prototype-based clustering algorithm since their data partitioning has been done by the distance point shape center to data points. FCRM algorithm partitions data space by hyperplane nature [6] which gives better performance in comparison to prototype-based clustering algorithm in TS fuzzy model [9]. Meanwhile, FCRM algorithm suffers two major problems: first, algorithm convergence takes much time due to bad initialization of parameters and second, the objective function is highly sensitive to noisy data.

The initialization problem has been addressed in [10] while taking partial derivative of objective function in terms of local linear model. But, input data distribution has not been considered that affects the estimation of fuzzy model parameters [8]. The noisy data points also affects estimation of model parameters [11].

Modification of FCRM objective function has been done by considering two issues. First, input data distribution has been incorporated into the objective function. Noise Clustering (NC) concept have been introduced into the proposed objective function to handle noisy data points that makes the algorithm robust. In NC, the noisy points are part of a noise cluster and remaining good data points are parts of good clusters.

The paper is organized as follows: Section II describes general structure of TS fuzzy modeling to represent complex nonlinear system. In section III, Modified Fuzzy C-Regression Model Noise Clustering (MFCRM-NC) algorithm has been proposed and in-depth explanations are given. Block Sparse representations for reduction of rules have been introduced. Identifications of premise and consequence variables parameter are also described. In section IV, the effectiveness of MFCRM-NC algorithm based TS fuzzy model has been validated on

978-1-4799-7560-0/15 $31.00 © 2015 IEEE
DOI 10.1109/SSCI.2015.237

IEEE computer society

two complex nonlinear models. Section V contains concluding remarks.

II. TS Fuzzy model descriptions

A complex nonlinear dynamical system can be approximated by TS fuzzy model. Implication of a TS fuzzy model is stated as a set of "IF-THEN"rules and each rule has the following form,

$$\begin{aligned} P^i:\ & IF\, x_{k1}\, is\, A_1{}^i and ... and\, x_{kn}\, is\, A_n{}^i \\ & THEN \quad y_k{}^i = p_0{}^i + x_{k1}.p_1{}^i + ... + x_{kn}.p_n{}^i \\ & \qquad = p_0{}^i + \sum_{j=1}^{n} x_{kj}.p_j{}^i,\ (i = 1, 2, ..., C\ rules) \\ & \qquad = [1\, \mathbf{x_k}{}^T]\theta^i = \bar{\mathbf{x}}_k \theta^i \end{aligned} \tag{1}$$

where, $\theta^i = [p_0{}^i, p_1{}^i, p_2{}^i, ..., p_n{}^i]^T \in \mathbb{R}^{(n+1)}$ is the coefficients of i-th consequence parameter, $\mathbf{x}_k = [x_{k1}, x_{k2}, ..., x_{kn}]^T \in \mathbb{R}^n$ is the k-th sample of premise variable and $y_k{}^i$ is k-th sample of output variable generated by i-th rule.

Here, the dumbbell shape membership function for premise variable is defined as,

$$\begin{aligned} A_j{}^i(x_{kj}) = \exp[-\tfrac{(x_{kj} - v_j{}^i)^2}{(\sigma_j{}^i)^2}] \in [0, 1], \\ (i = 1, 2, ..., C\ and\ j = 1, 2, ..., n) \end{aligned} \tag{2}$$

where, $\mathbf{v}^i = [v_1{}^i, v_2{}^i, ..., v_j{}^i]^T \in \mathbb{R}^n$ is center and $\sigma^{\mathbf{i}} = [\sigma_1{}^i, \sigma_2{}^i, ..., \sigma_j{}^i]^T \in \mathbb{R}^n$ is width of the membership function.

The estimated output can be represented by weighted average of all active rules ,

$$\hat{y}_k = \frac{\sum_{i=1}^{C} w_k{}^i . y_k{}^i}{\sum_{i=1}^{C} w_k{}^i} \text{ for } k - \text{th sample} \tag{3}$$

$$w_k{}^i = \min\{A_j{}^i(x_{kj})\},\ k = 1, 2, ..., N \tag{4}$$

where, the overall truth value $(w_k{}^i)$ of each rule can be obtained through logical "AND" operator [10].

III. Architecture of the proposed TS fuzzy model

In the section, a complete framework of TS fuzzy model has been described through a simple architecture. The complete architecture of top-down approach has been depicted in Fig. 1. The architecture consists of two parts: First part is related with structure identification where a novel objective function based clustering algorithm has been presented. Optimal number of partitions are obtained by block sparse representations of consequence vector. Last part is associated with parameter identification where the premise variables parameters (mean and width of the membership function) have been identified from the proposed objective function. Then, construct the rule for reducing the generalization error and determine the consequence parameters by an orthogonal least square method.

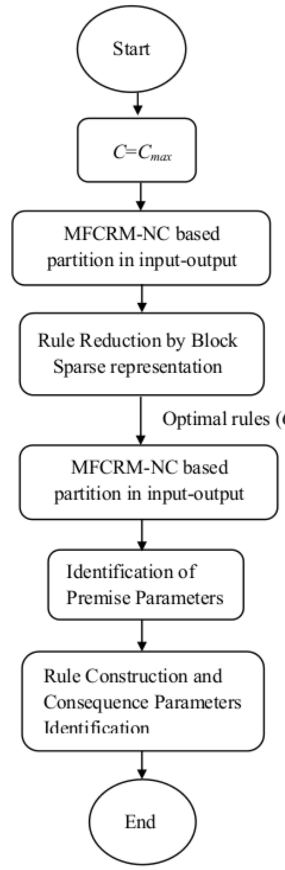

Fig. 1: Architecture of the MFCRM based TS Fuzzy model

A. *The Fuzzy C-Regression Model Clustering Algorithm*

Let, the data set,

$$\begin{aligned} D = \{(\mathbf{x_k}, y_k);\ \mathbf{x_k} \in \mathbb{R}^n, y_k \in \mathbb{R} \\ k = 1, 2, ..., N\} \end{aligned} \tag{5}$$

having N number of input-output patterns from which i-th fuzzy regression model can be drawn [12], [13]. The i-th hyperplane based regression model is expressed as follows:

$$\begin{aligned} y^i{}_k &= F^i(\mathbf{x}_k : \theta^i) \\ &= p_0{}^i + x_{k1}.p_1{}^i + x_{k2}.p_2{}^i + ... + x_{kn}.p_n{}^i \\ &= [1\ \mathbf{x}_k{}^T].\theta^i = \bar{\mathbf{x}}_k \theta^i \end{aligned} \tag{6}$$

where, $\mathbf{x}_k = [x_{k1}, x_{k2}, ..., x_{kn}]^T \in \mathbb{R}^n$ is the k-th regression vector and $y^i{}_k$ is the i-th regression model output. $\theta^i = [p_0{}^i, p_1{}^i, p_2{}^i, ..., p_n{}^i]^T \in \mathbb{R}^{(n+1)}$ is the coefficients of linear regressor. The modeling error [12] between actual output (y_k) and regression output $(F^i(\mathbf{x}_k : \theta^i))$ can be defined as

$$E_{ik}{}^2(\theta^i) = (y_k - y^i{}_k)^2 \tag{7}$$

Definition 1. *Let, D be a finite set and $2 \le C < N$ be number* *of clusters, then fuzzy partition matrix is defined for set D,*

$$U_{fc} = \{U \in R^{C\times N} | \mu_{ik} \in (0,1);\ \sum_{i=1}^{C} \mu_{ik} = 1; \\ 0 < \sum_{k=1}^{N} \mu_{ik} < N, \forall i = 1,2,...,C\} \tag{8}$$

where, μ_{ik} represents membership value of k-th data point belonging to i-th cluster.

The distance error with membership value (μ_{ik}) has been used in the FCRM objective function that has to be minimized over the clusters,

$$\min J_1(D:U,\theta) = \sum_{i=1}^{C}\sum_{k=1}^{N} {\mu_{ik}}^m {E_{ik}}^2(\theta^i) \tag{9}$$

subject to,

$$\sum_{i=1}^{C} \mu_{ik} = 1,\ \forall k = 1,2,...,N \tag{10}$$

where, $m \in (0,\infty)$ is the fuzzy exponent weight. If fuzzy exponent (m) value is set to 1, partition becomes hard partitions i.e. fuzzy membership weight (μ_{ik}) of a data point belonging to clusters is either 1 or 0. For avoiding the trivial solution of U, the m value is set at 2.

The modeling error based FCRM algorithm has been applied on input-output data to create the fuzzy subspaces as a fuzzy membership value. The coefficients of each linear model are obtained by Least Square (LS) solution [14].

Algorithm 1: FCRM algorithm for TS fuzzy model identification

Inputs:
Fuzziness ($m = 2$), Iteration number ($l = 0$).
Termination constant ($\xi = 0.001$)
$E_{ik}(\theta^i)$: Error distance between actual output and regressor model output for i-th cluster.
$U^0 \in \mathbb{R}^{C\times N}$: Fuzzy partition matrix for first iterat--ion.
Output:
U^l: Fuzzy partition matrix for l-th iteration.
Steps:
1. Update the consequence parameters by using LS solution.
2. Calculate the modeling error $E_{ik}(\theta^i)$ by using (7).
3. Update U^l with distance metric $E_{ik}(\theta^i)$.

$$U^l = \begin{cases} \dfrac{1}{\sum_{q=1}^{C}\left(\frac{{E_{ik}}^2(\theta^i)}{{E_{qk}}^2(\theta^q)}\right)^{\frac{1}{m-1}}} & \text{If } E_{ik}(\theta^i) > 0 \\ 0 & \text{otherwise} \end{cases}$$

until $\left\|U^l - U^{l-1}\right\|_2 \le \xi$ then stop;
otherwise, $l = l + 1$ and go to step 1.

B. The Modified Fuzzy C-Regression Model Clustering Algorithm

The solution of (9) is obtained by an iterative approach. FCRM objective function can get terminated at local minimum for poor initializations of cluster algorithm parameters [13]. Li have not considered premise data distribution into the FCRM objective function that affects estimation of model parameters [8]. The objective function of FCRM algorithm has been modified by considering input data distribution [8]. The premise variable partition has been done by G-K clustering algorithm.

$${G_{ik}}^2 = [\mathbf{x}_k - \mathbf{v}^i]^T M_i [\mathbf{x}_k - \mathbf{v}^i] \tag{11}$$

where, the cluster center is $\mathbf{v}^i$ and M_i is symmetric positive definite matrix. If $M_i = I$, then distance matric is a normal euclidian distance [15].

The objective function based clustering algorithm is sensitive to noisy data points [16]. Several methods are found in clustering algorithm for handling noisy data points [16]–[18]. One of them, in noise clustering (NC) algorithm [16], outlier data points are part of a noise cluster and other noise free data points are considered as parts of good cluster. The membership value of a data point belonging to noise cluster (†) is defined as,

$$\mu_{\dagger k} = 1 - \sum_{i=1}^{C} \mu_{ik} \tag{12}$$

where, $\mu_{\dagger k}$ is the membership value for k-th data points belonging to noise cluster. It has a fixed distance (γ) from all the data points i.e. it is represented as a fabricate prototype. The NC concept has been introduced in the modified FCRM clustering for handling the noisy data points as separate class. Thus, the constrained membership function is relaxed as

$$\sum_{i=1}^{C} \mu_{ik} < 1 \tag{13}$$

The Modified Fuzzy C-Regression-Noise Clustering (MFCRM-NC) algorithm objective function is defined as,

$$\min\ J_2(D;U,\theta,\mathrm{v},M_i) = \sum_{i=1}^{C}\sum_{k=1}^{N} {\mu_{ik}}^m[{E_{ik}}^2(\theta^i) + {G_{ik}}^2] \\ + \sum_{k=1}^{N} {\mu_{\dagger k}}^m \gamma^2 \tag{14}$$

subject to,

$$\sum_{i=1}^{C} \mu_{ik} + \mu_{\dagger k} = 1 \text{ and } |M_i| = \rho \\ i = 1,2,...,C \text{ and } k = 1,2,...,N \tag{15}$$

where, the noise cluster distance (γ) can be computed as [16],

$$\gamma^2 = \delta * \frac{\sum_{i=1}^{C}\sum_{k=1}^{N} [{E_{ik}}^2(\theta^i) + {G_{ik}}^2]}{CN} \tag{16}$$

where, δ is a user defined constant depending on system model data.

The analytical solution's of constrained optimization problem (14-15) is obtained by the Lagrangian approach converting it to unconstrained optimization problem by introducing N lagrangian multipliers, λ_k. The mathematical expressions for the Lagrangian form is followed as,

$$L(U,\lambda,\theta,\mathrm{v}) = J_2(D;U,\theta,\mathrm{v},M_i) - \sum_{i=1}^{C}\beta_i(|M_i|-\rho) \\ -\sum_{k=1}^{N}\lambda_k(\sum_{i=1}^{C}\mu_{ik}+\mu_{\dagger k}-1) \tag{17}$$

Differentiating (17) with respect to μ_{ik}, $\mu_{\dagger k}$, and λ_k and setting the derivative to zero.

$$\begin{cases} \frac{\partial L(...)}{\partial \mu_{1k}} = m(\mu_{1k})^{m-1}[{E_{1k}}^2(\theta^1)+{G_{1k}}^2]-\lambda_k = 0 \\ - \\ - \\ - \\ \frac{\partial L(...)}{\partial \mu_{Ck}} = m(\mu_{Ck})^{m-1}[{E_{Ck}}^2(\theta^C)+{G_{Ck}}^2]-\lambda_k = 0 \end{cases} \tag{18}$$

$$\frac{dL(...)}{d\mu_{\dagger k}} = m(\mu_{\dagger k})^{m-1}\gamma^2-\lambda_k = 0 \tag{19}$$

$$\frac{dL(...)}{d\lambda_k} = \sum_{i=1}^{C}\mu_{ik}+\mu_{\dagger k}-1 = 0 \tag{20}$$

From (18) and (19),

$$\mu_{ik} = \left[\frac{\lambda_k}{m}\right]^{1/(m-1)}\left[\frac{1}{{E_{ik}}^2(\theta^i)+{G_{ik}}^2}\right]^{1/(m-1)} \tag{21}$$

$$\mu_{\dagger k} = \left[\frac{\lambda_k}{m}\right]^{1/(m-1)}\left[\frac{1}{\gamma^2}\right]^{1/(m-1)} \tag{22}$$

From (18) and (19),

$$\left[\frac{\lambda}{m}\right]^{1/(m-1)} = \frac{\mu_{ik}}{\left[\frac{1}{{E_{ik}}^2(\theta^i)+{G_{ik}}^2}\right]^{1/(m-1)}} \tag{23}$$

$$\left[\frac{\lambda}{m}\right]^{1/(m-1)} = \frac{\mu_{\dagger k}}{\left[\frac{1}{\gamma^2}\right]^{1/(m-1)}} \tag{24}$$

Rearranging the (20),

$$\sum_{i=1}^{C}\mu_{ik}+\mu_{\dagger k} = 1 \tag{25}$$

Put the values of (23) and (24) into (25),

$$\left[\frac{\lambda}{m}\right]^{1/(m-1)}\left(\begin{array}{c}\sum_{q=1}^{C}\left[\frac{1}{{E_{qk}}^2(\theta^q)+{G_{qk}}^2}\right]^{1/(m-1)} \\ +\left[\frac{1}{\gamma^2}\right]^{1/(m-1)}\end{array}\right) = 1 \tag{26}$$

Put (23) into (26) and rearranging to obtain μ_{ik},

$$\mu_{ik} = \frac{1}{\sum_{q=1}^{C}\left(\frac{{E_{ik}}^2(\theta^i)+{G_{ik}}^2}{{E_{qk}}^2(\theta^q)+{G_{qk}}^2}\right)^{\frac{1}{m-1}}+\left(\frac{{E_{ik}}^2(\theta^i)+{G_{ik}}^2}{\gamma^2}\right)^{\frac{1}{m-1}}} \tag{27}$$

Taking partial derivative of (17) w.r.t $\mathbf{v^i}$,

$$\frac{\partial L(...)}{\partial \mathbf{v}^i} = \sum_{k=1}^{N}{\mu_{ik}}^m[\mathbf{x_k}-\mathbf{v}^i] = 0 \\ \Rightarrow \quad \mathbf{v}^i = \frac{\sum_{k=1}^{N}{\mu_{ik}}^m\mathbf{x_k}}{\sum_{k=1}^{N}{\mu_{ik}}^m} \tag{28}$$

Updating (17) by M_i ,

$$\frac{\partial L(...)}{\partial M_i} = \sum_{k=1}^{N}{\mu_{ik}}^m[\mathbf{x}_k-\mathbf{v}^i][\mathbf{x}_k-\mathbf{v}^i]^T \\ -\beta_i\,|M_i|\,{M_i}^{-1} = 0 \\ \Rightarrow {M_i}^{-1} = \frac{\sum_{k=1}^{N}{\mu_{ik}}^m[\mathbf{x}_k-\mathbf{v}^i][\mathbf{x}_k-\mathbf{v}^i]^T}{\beta_i|M_i|} \\ = \sqrt[n]{(\frac{1}{\rho|F_i|})}\ F_i. \tag{29}$$

where, the covariance matrix, F_i, is defined as,

$$F_i = \frac{\sum_{k=1}^{N}{\mu_{ik}}^m[\mathbf{x}_k-\mathbf{v}^i]\,[\mathbf{x}_k-\mathbf{v}^i]^T}{\sum_{k=1}^{N}{\mu_{ik}}^m} \tag{30}$$

The necessary conditions for minimization of (17) w.r.t to θ^i

$$\frac{\partial L(...)}{\partial \theta^i} = \sum_{k=1}^{N}{\mu_{ik}}^m\frac{\partial {E_{ik}}^2(\theta^i)}{\partial \theta^i} = 0 \tag{31}$$

where, $\frac{\partial {E_{ik}}^2(\theta^i)}{\partial \theta^i}$ is obtained from (7) by taking partial derivative ,

$$\frac{\partial {E_{ik}}^2(\theta^i)}{\partial \theta^i} = 2E_{ik}(\theta^i)\frac{\partial E_{ik}(\theta^i)}{\partial \theta^i} = [y_k-\bar{\mathbf{x}}_k.\theta^i].\bar{\mathbf{x}}_k \tag{32}$$

Rearranging (31- 32) and finally making it in matrix form,

$$\sum_{k=1}^{N}{\mu_{ik}}^m[y_k-\bar{\mathbf{x}}_k.\theta^i].\bar{\mathbf{x}}_k = 0 \\ \Rightarrow X^T\Lambda^iY-(X^T\Lambda^iX)\theta^i = 0,\ \forall i=1,...,C \\ \Rightarrow \theta^i = (X^T\Lambda^iX)^{-1}X^T\Lambda^iY \tag{33}$$

Here, X, Y, and Λ^i are expressed as follows,

$$X = \begin{bmatrix} 1 & {\mathbf{x_1}}^T \\ 1 & {\mathbf{x_2}}^T \\ \vdots & \vdots \\ 1 & {\mathbf{x_N}}^T \end{bmatrix}_{N\times(n+1)} \qquad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}_{N\times 1}$$

$$\Lambda^i = \begin{bmatrix} \mu_{i1} & 0 & ... & 0 \\ 0 & \mu_{i2} & ... & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & ... & 0 & \mu_{iN} \end{bmatrix}_{N\times N}$$

C. Sparse Block Structure Representations for TS Fuzzy Rule Reduction

Definition 2. *Let, $U \in U_{fc}$ be fuzzy partition matrix for set D. The TS fuzzy model with C rule creates a dictionary (Z^i) for i-th rule and it is defined as,*

$$Z^i = U^iX \in \mathbb{R}^{N\times(n+1)} \tag{34}$$

Algorithm 2: MFCRM-NC algorithm for TS fuzzy model identification

Inputs:
Fuzziness ($m = 2$) , Iteration number ($l = 0$), Initialize cluster number (C).
Cluster center= $\mathbf{v}^i$, Linear Coefficients= θ^i.
Termination constant ($\xi = 0.001$).
$U^0 \in \mathbb{R}^{C \times N}$: Fuzzy partition matrix for first iteration.
Outputs:
U^l: Fuzzy partition matrix while algorithm converges.
$\mathbf{v}^i$: Cluster center for premise variable data partition.
Steps
Repeat
Compute the modeling error ${E_{ik}}^2(\theta^i)$ by using (7) for C-th cluster.
Compute the premise variable partition ${G_{ik}}^2$ by using (11) for C-th cluster.
Update fuzzy covariance matrix F_i by using (30).
Update cluster center $\mathbf{v}^i$ by using (28).
Update noise cluster distance (γ^2) by using (16).
Update linear coefficients θ^i by using (33).
Update fuzzy partition matrix U^l by using (27).
$l = l + 1$.
Until: U or $\mathbf{v}$ stabilize.

where, $X \in \mathbb{R}^{N\times(n+1)}$ and $U^i \in \mathbb{R}^{N\times N}$ is a diagonal matrix for $i = 1, 2, ..., C$. The estimated output $\hat{y} = [\hat{y}_1, \hat{y}_2, ..., \hat{y}_N]$ is obtained as,

$$\hat{y} = \sum_{i=1}^{C} Z^i\theta^i = Z\theta \tag{35}$$

where, $\theta = [\theta^1, \theta^2, ..., \theta^C] \in \mathbb{R}^{(n+1)\times C}$ is the consequence parameters vector and $Z = [Z^1, Z^2, ..., Z^C] \in \mathbb{R}^{N(n+1)\times C}$ is the corresponding TS fuzzy model dictionary.

Minimizing square of modeling error between actual output and estimated model output has been utilized to determine the consequence parameters by using batch processing algorithm.

$$\min \frac{1}{2}e^2 = \min \frac{1}{2}\left\| y - \sum_{i=1}^{C} Z^i\theta^i \right\|^2 \tag{36}$$

Determination of optimal rules in an oversize dictionary based TS fuzzy model is an optimization problem [14]. Some of the rules in a dictionary are over specified that doesn't have substantial contribution on model accuracy i.e. some rules are assumed as redundant rules. Here, block sparse representations for consequence vector are adopted for rule reductions in proposed TS fuzzy modeling approach. Consequence parameters for each rule represent a block, and redundant blocks are simply assumed as zero.

Definition 3. *Let, $\boldsymbol{b} = (b_1, b_2, ..., b_m) \in \mathbb{R}^m$ be a vector, its sparsity defines non zero positions or entity.*

$$\|\mathrm{b}\|_0 = \lim_{a\to 0} \|\mathrm{b}\|_a^a = \lim_{a\to 0} \sum_{i=1}^{m} |b_i|^a = \hbar\{i : b_i \neq 0\} \tag{37}$$

where, $\hbar$ represents cardinality of the vector $\boldsymbol{b}$.

Definition 4. *The block structure representations for consequence vector is defined as,* $\theta = [\underbrace{{p_0}^1, {p_1}^1, ..., {p_n}^1}_{1-block}, ..., \underbrace{{p_0}^C, {p_1}^C, ., {p_n}^C}_{C-Block}]^T \in \mathbb{R}^{(n+1)\times C}$
$\|\theta\|_{a,0} = \left\| (\|\theta^1\|_a, \|\theta^2\|_a, ..., \|\theta^C\|_a) \right\|_0$

The block structure representations of consequence vector is a optimization problem and it is defined as,

$$\begin{aligned} &\min_{\theta} \|\theta\|_{2,0} \\ &\text{subject to, } \tfrac{1}{2}\left\| y - \sum_{i=1}^{C} Z^i\theta^i \right\|_2^2 < \partial \end{aligned} \tag{38}$$

where, less impact blocks (θ) are discarded from over specified number by the optimization algorithm in which model accuracy will retain high. The optimization problem (38) is NP hard [19]. There exists two approaches to obtain solutions. First approach to solving the problem by relaxation method that converts 0-norm constraints to 1-norm constraints. The convex optimization method [20] [21] can be applied directly to obtain the solution but TS fuzzy model objective should not be violated [22]. Second approach to solving the problem by greedy search algorithm, includes block orthogonal match pursuit (BOMP), block matching pursuit (BMP), group matching pursuit (GMP), group orthogonal matching pursuit (GOMP) and its several extensions [23]. Here, BOMP algorithm has been used for selecting more effective consequence coefficients block and less effective blocks are removed from over specified number. The BOMP has three steps: initialization, block selections and update. The block greedy optimization algorithm is described in Algorithm 3.

Algorithm 3: BOMP algorithm for TS fuzzy model identification

Input:
Z: Given TS model dictionary.
Outputs:
Update dictionary, consequence vector, rule selections.
Initializations:
index set $\Omega_0 = []$; dictionary matrix $\wp_0 = []$;
residual $r_0 = y$; and iterations number t.
Steps:Repeat
Block Selections: Find correlations l_t between dictionary and residual.
Update:
Update index set $\Omega_t = [\Omega_{t-1} \cup \mathrm{l_t}]$;
Update Dictionary matrix $\wp_t = [\wp_{t-1} \cup \mathrm{l_t}]$;
Determine the consequence parameters
$\theta = \arg\min_{\theta} \|y - \wp_t\theta\|_2$
Update residuals $r_t = y - \wp_t\theta$
until Termination criteria (residuals bounded ($\partial = 10^{-5}$) or fixed iterations (100)) true.
$t = t + 1$, go to Repeat.

In block selection steps, the highly correlated blocks (consequence vector) are selected, while the residual and consequence parameters are modified in updated steps. The block selection's data are applied in MFCRM-NC algorithm to obtain fuzzy partition matrix and premise variable mean.

D. Parameter Estimations

Gaussian membership function has been used for premise variable. The mean and width of gaussian function can be obtained by the following expressions.

$$\mathbf{v}^i = \frac{\sum_{k=1}^{N} \mu_{ik}{}^m \mathbf{x}_k}{\sum_{k=1}^{N} \mu_{ik}{}^m} \tag{39}$$

$$\sigma_j{}^i = \sqrt{\frac{2 * [\sum_{k=1}^{N} \mu_{ik}{}^m (x_{kj} - v_j{}^i)^2]}{\sum_{k=1}^{N} \mu_{ik}{}^m}} \tag{40}$$

Once, the premise parameters have been identified then construct rules from section (II) to identify the consequence parameters by the following matrix form.

$$y = \chi\pi^{opti} + \varepsilon \tag{41}$$

where, $\pi^{opti} = [p_0{}^1, p_1{}^1, ..., p_n{}^1, ..., p_0{}^{C_{opti}}, ..., p_n{}^{C_{opti}}]$, $y = [y_1, y_2, ..., y_N]$, $\varepsilon = [\varepsilon_1, \varepsilon_2, ..., \varepsilon_N]$ is the error vector with $\varepsilon = [y - \hat{y}]$ and C_{opti} is the optimal number of rules (clusters) obtained from BOMP algorithm. Regression vector is defined as, $\chi(\mathbf{x_k}) = [\psi_{1k}\ \psi_{1k}x_{k1}\ \psi_{1k}x_{k2}...\ \psi_{1k}x_{kn}\ \ ...\psi_{C_{opti}k}...\psi_{C_{opti}k}x_{kn}]$. where, x_{kn} is the n-th component of input variable for k-th sample and ψ is the normalized weight of membership function having the following expression,

$$\psi_{ik} = \frac{w_k{}^i}{\sum_{i=1}^{C_{opti}} w_k{}^i} \tag{42}$$

Weighted least square (WLS) technique has been applied to (41) to obtain the solution but it requires matrix inversion that may not be computationally possible all the time [13]. The orthogonal least square (OLS) approach [10] has been adopted to obtain solution of consequence parameters.

IV. Algorithm Validations Through Simulation Results

Two different performance indices have been used in this study i.e. Root Mean Square Error (RMSE) and Mean Square Error (MSE), defined as,

$$\text{RMSE} = \sqrt{\frac{\sum_{k=1}^{N} (y_k - \hat{y}_k)^2}{N}} \tag{43}$$

$$\text{MSE} = \frac{\sum_{k=1}^{N} (y_k - \hat{y}_k)^2}{N} \tag{44}$$

A. Validation on Nonlinear Dynamical System

The highly complex modified nonlinear model [24] [25] used for checking effectiveness of the proposed modeling approach is defined as,

$$y_k = \frac{y_{k-1}(y_{k-2} + 2)(y_k + 2.5)}{10 + y_{k-1}{}^2 + y_{k-2}{}^2} + u_k + \Delta_k \tag{45}$$

where, y_k is the model output and u_k is the model input which is bounded between $[-1 \quad +1]$. The model is also accounted with zero mean white noise ($\Delta_k,\ k \leq 1500$). The following input signal is expressed as,

$$u_k = \begin{cases} \sin(\frac{2k\pi}{250}) & \text{if, } k \leq 500 \\ 0.8 * \sin(\frac{2k\pi}{250}) + 0.2 * \sin(\frac{2k\pi}{25}) & \text{otherwise} \end{cases} \tag{46}$$

1500 samples were generated by simulation in which 500 samples were used to the train model. Fuzzy model parameters have been identified once, testing of model has been done by remaining 1000 samples. The product space variables of TS fuzzy model are selected as $[y_{k-1},\ y_{k-2},\ u_k,\ u_{k-1},\ y_k]$ where input and output variables are defined as $[y_{k-1},\ y_{k-2},\ u_k,\ u_{k-1}]^T$ and y_k respectively. In the actions of BOMP algorithm for selecting fuzzy rules from over specified rules, only 4 blocks of consequence vector remain nonzeros and others shrink to zero (Fig. 2). However, sparsity level of BOMP based MFCRM-NC was set to 4 [14]. Different SNR levels (SNR=1 dB, 5 dB, 10 dB, 20 dB)

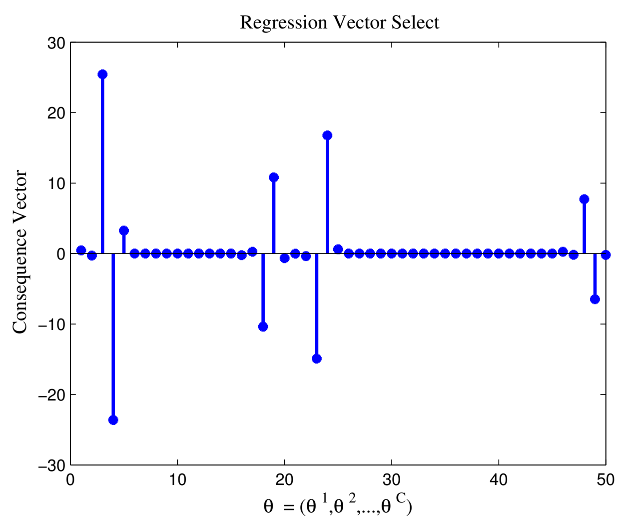

Fig. 2: Consequence block selects by BOMP algorithm

have been added to the plant model while collecting 1500 samples to classify the data space. In MFCRM-NC algorithm, the fabricated noise cluster distance is set as ($\delta = 0.1$). The evaluation performance index (RMSE-trn and RMSE-test) stands for training and testing data respectively.

Tables (I-IV) shows the comparative performance of MFCRM-NC based TS fuzzy model with different existing algorithms such as FCM [5], G-K [7], Fuzzy Model Identification (FMI) [26] and FCRM [12]. It is clearly seen from the results that the MFCRM-NC based TS fuzzy model gives better performance compared to results existing in literature.

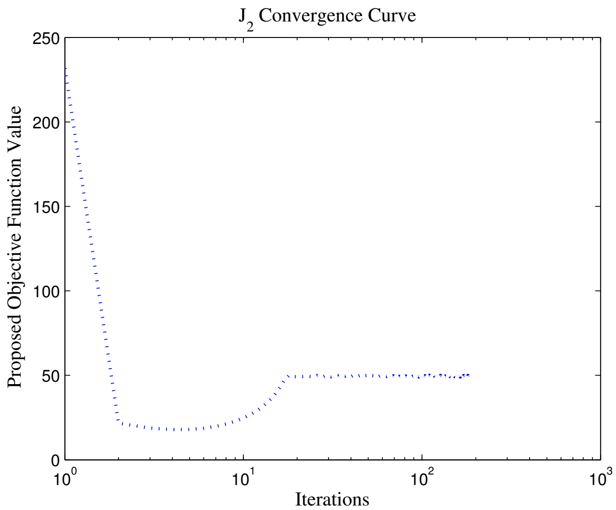

Fig. 3: Convergence curve of proposed objective function

TABLE I: 1 dB Noise

Model Name	RMSE-trn	RMSE-test	Rules
FCM	0.9330	1.3646	4
G-K	0.8932	1.0226	4
FMI	0.8953	1.0320	4
FCRM	0.8872	1.0095	4
MFCRM-NC	0.8699	0.9947	4

TABLE II: 5 dB Noise

Model Name	RMSE-trn	RMSE-test	Rules
FCM	0.6170	1.0364	4
G-K	0.5855	0.6505	4
FMI	0.6274	0.7021	4
FCRM	0.5805	0.6485	4
MFCRM-NC	0.5658	0.5998	4

TABLE III: 10 dB Noise

Model Name	RMSE-trn	RMSE-test	Rules
FCM	0.3453	0.7150	4
G-K	0.3330	0.3489	4
FMI	0.3956	0.4416	4
FCRM	0.3312	0.3694	4
MFCRM-NC	0.3233	0.3465	4

The noise clustering distance term was capable to handle highly noise points as a separate cluster. This noise cluster makes the algorithm robust in the presence of high noise. The convergence curve of proposed objective function is shown in Fig. (3) for 1 dB noisy data.

B. Box-Jenkins Model

The Box-Jenkins model [27] is a standard model in system identification domain. It consists of 296 input-output samples. Gas flow rate into furnace and CO_2 concentration in the outlet gases is considered as input (u) and output (y) respectively. The complete data set has been used for training the model where the selected input and output variables are $[y_{k-1}, u_{k-4}]^T$ and y_k respectively for comparing with other existing algorithms. Table V provides a comparative performance of proposed algorithm with other algorithms existing in literature. However, sparsity level for Box-Jenkins model was set to 2 [28]. Proposed algorithmic approach was also studied on two different noise levels (SNR=30 dB, 10 dB).

TABLE IV: 20 dB Noise

Model Name	RMSE-trn	RMSE-test	Rules
FCM	0.1393	0.3004	4
G-K	0.1326	0.1484	4
FMI	0.1457	0.2378	4
FCRM	0.1291	0.1495	4
MFCRM-NC	0.1183	0.1395	4

TABLE V: Box-Jenkins model identification comparison

Model	Inputs	Rules	MSE
Tong (80) [28]	2	19	0.469
Pedrycz (84) [28]	2	81	0.32
Xu (87) [28]	2	25	0.328
Joo (93) [26]	2	6	0.166
Chen (97) [26]	2	3	0.267
Liu (02) [28]	2	2	0.165
Zhang (06) [28]	2	2	0.160
MFCRM-NC, noiseless	2	2	0.152
MFCRM-NC, SNR=30 dB	2	2	0.153
MFCRM-NC, SNR=10 dB	2	2	0.277

It is clearly seen from the Table V that MFCRM-NC based TS fuzzy model gives reasonable better performance compared to existing algorithms in literature. For noiseless case, the noise clustering parameter (γ) has been used in MFCRM-NC objective function to reduce the sensitivity of data and provided a better MSE value.

V. Conclusion

In this paper, objective function in FCRM algorithm has been modified by considering input data distribution and noise clustering concept. The block sparse representations to consequence regressor vector for selecting the consequence blocks in such way that the system model accuracy is retained at high value. Block orthogonal matching pursuit algorithm (BOMP) has been carried out for selecting the most highly correlated block. Optimal blocks are selected, then identify the premise variable parameters and constructions rule to determine the consequence parameters by an orthogonal least square method. The simulation results show that MFCRM-NC based TS fuzzy model gives comparable accuracy compared to algorithms existing in literature with less number rules. The greedy search algorithm has been used to find out optimal number rules instead of relaxed convex optimization method. In further work, convex optimization method can be used to determine the optimal number rules.

References

[1] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Trans. on Systems, Man and Cybernetics*, no. 1, pp. 116–132, 1985.

[2] B. Jiang, Z. Gao, P. Shi, and Y. Xu, "Adaptive fault-tolerant tracking control of near-space vehicle using takagi–sugeno fuzzy models," *IEEE Trans. on Fuzzy Systems*, vol. 18, no. 5, pp. 1000–1007, 2010.

[3] S. Tong, T. Wang, and W. Zhang, "Fault tolerant control for uncertain fuzzy systems with actuator failures," *International Journal of Innovative Computing, Information and Control*, vol. 4, no. 10, pp. 2461–2474, 2008.

[4] S. L. Chiu, "A cluster extension method with extension to fuzzy model identification," *Third IEEE Conference on Fuzzy Systems*, pp. 1240–1245, 1994.

[5] J. C. Bezdek, R. Ehrlich, and W. Full, "Fcm: The fuzzy c-means clustering algorithm," *Computers & Geosciences*, vol. 10, no. 2, pp. 191–203, 1984.

[6] E. Kim, M. Park, S. Ji, and M. Park, "A new approach to fuzzy modeling," *IEEE Trans. on Fuzzy Systems*, vol. 5, no. 3, pp. 328–337, 1997.

[7] J. Abonyi, R. Babuška, and F. Szeifert, "Modified gath-geva fuzzy clustering for identification of takagi-sugeno fuzzy models," *IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 32, no. 5, pp. 612–621, 2002.

[8] T. Dam and A. K. Deb, "TS fuzzy model identification by a novel objective function based fuzzy clustering algorithm," *IEEE Symposium on Computational Intelligence in Ensemble Learning (CIEL)*, pp. 1–7, Dec 2014.

[9] J. Keller, R. Krisnapuram, and N. R. Pal, *Fuzzy models and algorithms for pattern recognition and image processing*. Springer Science & Business Media, 2005, vol. 4.

[10] C. Li, J. Zhou, B. Fu, P. Kou, and J. Xiao, "T-S fuzzy model identification with a gravitational search-based hyperplane clustering algorithm," *IEEE Trans. on Fuzzy Systems*, vol. 20, no. 2, pp. 305–317, 2012.

[11] C.-C. Chuang, S.-F. Su, and S.-S. Chen, "Robust TSK fuzzy modeling for function approximation with outliers," *IEEE Trans. on Fuzzy Systems*, vol. 9, no. 6, pp. 810–821, 2001.

[12] R. J. Hathaway and J. C. Bezdek, "Switching regression models and fuzzy clustering," *IEEE Trans. on Fuzzy Systems*, vol. 1, no. 3, pp. 195–204, 1993.

[13] C. Li, J. Zhou, X. Xiang, Q. Li, and X. An, "T-S fuzzy model identification based on a novel fuzzy c-regression model clustering algorithm," *Engineering Applications of Artificial Intelligence*, vol. 22, no. 4, pp. 646–653, 2009.

[14] M. Luo, F. Sun, H. Liu, and Z. Li, "A novel T-S fuzzy systems identification with block structured sparse representation," *Journal of the Franklin Institute*, vol. 351, no. 7, pp. 3508–3523, 2014.

[15] D. Gustafson and W. Kessel, "Fuzzy clustering with a fuzzy covariance matrix," *IEEE Conference on Decision and Control*, no. 17, pp. 761–766, 1978.

[16] R. N. Dave, "Characterization and detection of noise in clustering," *Pattern Recognition Letters*, vol. 12, no. 11, pp. 657–664, 1991.

[17] R. N. Dave, "Robust fuzzy clustering algorithms," *IEEE International Conference on Fuzzy Systems*, pp. 1281–1286, 1993.

[18] R. N. Dave and S. Sen, "Robust fuzzy clustering of relational data," *IEEE Trans. on Fuzzy Systems*, vol. 10, no. 6, pp. 713–727, 2002.

[19] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Review*, vol. 51, no. 1, pp. 34–81, 2009.

[20] M. Luo, F. Sun, and H. Liu, "Hierarchical structured sparse representation for T-S fuzzy systems identification," *IEEE Trans. on Fuzzy Systems*, vol. 21, no. 6, pp. 1032–1043, 2013.

[21] M. Luo, F. Sun, and H. Liu, "Sparse fuzzy c-regression models with application to ts fuzzy systems identification," *IEEE International Conference on Fuzzy Systems*, pp. 1571–1577, 2014.

[22] Y. C. Eldar, P. Kuppinger, and H. Bölcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. on Signal Processing*, vol. 58, no. 6, pp. 3042–3054, 2010.

[23] A. Majumdar and R. K. Ward, "Fast group sparse classification," *Canadian Journal of Electrical and Computer Engineering*, vol. 34, no. 4, pp. 136–144, 2009.

[24] B. Subudhi and D. Jena, "A differential evolution based neural network approach to nonlinear system identification," *Applied Soft Computing*, vol. 11, no. 1, pp. 861–871, 2011.

[25] M. Soltani, A. Chaari, and F. Ben Hmida, "A novel fuzzy c-regression model algorithm using a new error measure and particle swarm optimization," *International Journal of Applied Mathematics and Computer Science*, vol. 22, no. 3, pp. 617–628, 2012.

[26] J.-Q. Chen, Y.-G. Xi, and Z.-J. Zhang, "A clustering algorithm for fuzzy model identification," *Fuzzy Sets and Systems*, vol. 98, no. 3, pp. 319–329, 1998.

[27] G. E. Box, G. M. Jenkins, and G. C. Reinsel, *Time series analysis: forecasting and control*. John Wiley & Sons, 2011, vol. 734.

[28] M. Soltani, B. Aissaoui, A. Chaari, F. Ben Hmida, and M. Gossa, "A modified fuzzy c-regression model clustering algorithm for ts fuzzy model identification," *IEEE 8th International Multi-Conference on Systems, Signals and Devices (SSD)*, pp. 1–6, 2011.