2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)
July 6-11, 2014, Beijing, China

Sparse Fuzzy c-regression Models with Application to T-S Fuzzy Systems Identification

Minnan Luo, Fuchun Sun and Huaping Liu
State Key Laboratory on Intelligent Technology and Systems
Department of Computer Science and Technology
Tsinghua University, Beijing, 100084, P.R. China
Email: luomn10@mails.tsinghua.edu.cn

***Abstract*—In this paper, the objective function of fuzzy c-regression models (FCRM) is modified to develop a novel fuzzy partition method on the basis of block structured sparse representation, namely as sparse fuzzy c-regression model. This method takes advantage of the block structured information in the objective function of FCRM and casts fuzzy partition into an optimization problem by making a tradeoff between traditional FCRM and the number of prototypes of hyper-plane with nonzero parameters. An alternating direction method of multipliers (ADMM) based algorithm is exploited to address the proposed optimization problem. Furthermore, based on sparse fuzzy c-regression models, a novel T-S fuzzy systems identification method is developed for reduction of fuzzy rules. Finally, examples on well-known benchmark data set are carried out to illustrate the effectiveness of the proposed methods.**

I. Introduction

Fuzzy partition is an exploratory data analysis technique that extends crisp clustering by assigning each object to all of the clusters with certain degree of memberships. It plays a significant role in feature analysis, systems identification and classifier design. For many real-world problems with impreciseness, uncertainty and vagueness, a fuzzy partitioning of the underlying space appears to be more realistic than "hard clustering". The idea of fuzzy partition is formulated by minimizing

$$f = \sum_{i=1}^{r} \sum_{j=1}^{N} u_{ij}^{m} d_{ij}$$

where fuzzifier $m > 1$ controls how many clusters may overlap; $d_{ij} = (x_j - p_i)^2$ denotes the distance of data object x_j to prototype p_i associated with cluster i; u_{ij} represents the membership degree of sample x_j to the i-th cluster and fuzzy partition matrix $U = (u_{ij}) \in \mathbb{R}^{r\times N}$ should satisfy $u_{ik} \in [0,1]$ and $\sum_{i=1}^{r} u_{ij} = 1$ for all $i = 1,2,\cdots,r$ and $j = 1,2,\cdots,N$. The distance measurement d_{ij} varies with respect to different cluster prototypes. For example, fuzzy c-means clustering (FCM) [1] represents each cluster with the center of cluster in the sense of geometrical structure, called hyper-spherical-shaped based fuzzy partition; and fuzzy c-elliptotypes, fuzzy c-varieties together with fuzzy c-regression models (FCRM) [3], [4] define the prototype as linear or nonlinear subspace of the data space, namely as hyperplane-shaped based fuzzy partition. Specifically, the FCRM method uses polynomials as cluster prototypes and divide the data set into a group of different regression models [5]. In addition, fuzzy partitions that simultaneously consider the mentioned two prototypes are also studied in [6]–[8].

Fuzzy partition techniques are extensively used to derive membership functions for fuzzy rule-based systems [4], [5], [9]. Specifically, each cluster is associated with one fuzzy rule involved in fuzzy models. With some prior knowledge, the number of clusters (fuzzy rules) is assumed to be fixed and assigned a priori in T-S fuzzy systems identification. However, for most unknown system, the appropriate and exact number of clusters (rules) may be unknown in practice [5], [10]. Usually, cluster validity criterion and its modifications with respect to different fuzzy partitions are developed to determine the number of clusters (fuzzy rules) [5], [11]–[13].

Motivated by the block structured information in T-S fuzzy model and the successful application of block structured sparse representation [14]–[17], we have done some effort to identify T-S fuzzy model on the basis of block structured sparse representation [18], [19]. These methods effectively take advantage of the block information and execute fuzzy rule reduction by selecting main important fuzzy rules and eliminating the redundant ones with block sparse representation. Nevertheless, the fuzzy system dictionary provided for fuzzy rules reduction is fixed once the parameters of fuzzy rules antecedent are determined by clustering. In this paper, we develop a novel fuzzy partition (clustering) method which simultaneously consider the number of prototypes of hyper-plane with nonzero parameters, namely as sparse fuzzy c-regression model. Moreover, a new T-S fuzzy systems identification method is exploited on the basis of sparse fuzzy c-regression model. Two phases are included in the proposed method. In the coarse learning phase, we partition the data set with sparse fuzzy c-regression model and obtain the fuzzy partition matrix and nonzero blocks of regression models. In the fine-tuning learning phase, we pick up the nonzero blocks and further adjust the corresponding consequent parameters of fuzzy rules to improve the modeling accuracy.

The contribution of this article is three-fold: Firstly, a new fuzzy partition method is proposed on the basis of sparse representation, namely as sparse fuzzy c-regression models This method takes advantage of the block structure information and casts fuzzy partition into an optimization problem by making a tradeoff between traditional objective function of

FCRM and the number of prototypes of hyper-plane with nonzero parameters. Secondly, an alternating direction method of multipliers (ADMM) based algorithm are exploited to address the optimization problem of sparse fuzzy c-regression model. This algorithm takes the form of a decomposition-coordination procedure, where the solutions to small local subproblems are coordinated to find a solution to a large global problem [2]. Thirdly, we apply the sparse fuzzy c-regression models for T-S fuzzy systems identification such that an appropriate number of fuzzy rules are associated with the fuzzy model.

The remainder of this paper is organized as follows. In Section II, notations and preliminaries about T-S fuzzy systems are recalled. Section III casts fuzzy partition into an optimization problem with block sparse representation, called sparse fuzzy c-regression model and an algorithm for sparse fuzzy c-regression model are exploited using alternating direction method of multipliers. In Section IV, a new method for T-S fuzzy systems identification is developed on the basis of sparse fuzzy c-regression model. Some examples on well-known benchmark data sets are carried out in Section V to illustrate the effectiveness of the proposed methods. Conclusions are given in Section VI.

II. NOTATIONS OF T-S FUZZY SYSTEMS IDENTIFICATION

T-S fuzzy system consists of "if then" fuzzy rules with fuzzy rule antecedent and consequent. Specifically, the consequent part is specified as an affine function rather than a fuzzy linguistic proposition. In such a way, T-S fuzzy model provides a reasonable framework for modeling by decomposition of a nonlinear system into a collection of local linear models [3]. More precisely, for n-dimension input variable $\mathbf{x} = (x_1 \ x_2 \ \cdots \ x_n)^\mathsf{T} \in \mathbb{R}^n$ of T-S fuzzy system with r fuzzy rules, the i-th fuzzy rule has the following form:

$$\begin{aligned} R^i: \ & \text{If } x_1 \text{ is } A_{i1}, \ x_2 \text{ is } A_{i2}, \cdots, x_n \text{ is } A_{in}, \text{ then} \\ & y = w_{i0} + w_{i1}x_1 + w_{i2}x_2 + \cdots + w_{in}x_n. \end{aligned}$$

$(i = 1,2,\cdots,r)$, where fuzzy subset $A_{ij}: \mathbb{R} \to [0,1]$ for $i = 1,2,\cdots,r$ and $j = 1,2,\cdots,n$.

Let $\mathbf{w}[i] = [w_{i0}, w_{i1}, w_{i2}, \cdots, w_{in}]^\mathsf{T}$ be the consequence parameter vector of the i-th fuzzy rule. Then, the value of corresponding fuzzy rule consequence is denoted by $g_i(\mathbf{x};\mathbf{w}_i) = \mathbf{w}[i]^\mathsf{T}(1 \ \mathbf{x}^\mathsf{T})^\mathsf{T}$ $(i = 1,2,\cdots,r)$. With weighted-average defuzzifier, T-S fuzzy system output is derived as

$$\hat{y} = \sum_{i=1}^{r} \phi_i(\mathbf{x}) g_i(\mathbf{x};\mathbf{w}_i) \tag{1}$$

where the firing strength of the i-th fuzzy rule with respect to input variable $\mathbf{x} = (x_1, x_2, \cdots, x_n)^\mathsf{T}$ is denoted by

$$\phi_i(\mathbf{x}) = \frac{A_i(\mathbf{x})}{\sum\limits_{k=1}^{r} A_k(\mathbf{x})}$$

with $A_i(\mathbf{x}) = \prod\limits_{j=1}^{n} A_{ij}(x_j)$ $(i = 1,2,,\cdots,r)$.

For multi-input-single-output (MISO) T-S fuzzy system identification, we denote the input-output data set by

$$\mathscr{D} = \{(\mathbf{x}_k^\mathsf{T}, y_k)^\mathsf{T}, \mathbf{x}_k = (x_{k1}, x_{k2}, \cdots, x_{kn})^\mathsf{T}, k = 1, \cdots, N\}$$

where $\mathbf{x}_k$ and y_k represent the k-th n-dimension input variable and output variable, respectively. Fuzzy partition is extensively studied to obtain a partitioning of input data, where the object can belong to all of the clusters with a certain degree of membership [3]. We assume the input-output data set $\mathscr{D}$ is divided by

$$\begin{aligned} \mathscr{M} = \{ & \mathbf{U} = (u_{ik}) \in \mathbb{R}^{r \times N} : u_{ik} \in [0,1] (\forall i, k), \\ & \sum_{i=1}^{r} u_{ik} = 1 (\forall k), 0 < \sum_{k=1}^{N} u_{ik} < N (\forall i)\} \end{aligned}$$

where u_{ik} denotes the membership degree of data object $\mathbf{x}_k$ to the i-th cluster $(i = 1,2,\cdots,r; k = 1,2,\cdots,N)$. Particularly, if $u_{ik} \in \{0,1\}$ $(\forall i,k)$, the fuzzy partition turns to a crisp one. In this paper, bell-shaped membership function is utilized to represent the membership function of fuzzy set $A_{ij}(x_j)$ in fuzzy rules' antecedent, i.e.,

$$\mu_{A_{ij}}(x_j) = \exp[-\frac{(x_j - c_{ij})^2}{\sigma_{ij}^2}] \tag{2}$$

for $i = 1,2,\cdots,r$ and $j = 1,2,\cdots,n$, where

$$c_{ij} = \frac{\sum_{k=1}^{N} u_{ik} x_{kj}}{\sum_{k=1}^{N} u_{ik}}, \text{ and } \sigma_{ij} = \sqrt{\frac{\sum_{k=1}^{N} u_{ik}(x_{kj} - c_{ij})^2}{\sum_{k=1}^{N} u_{ik}}}. \tag{3}$$

Given the determined parameters of fuzzy rule antecedent, the corresponding consequents are traditionally estimated as parameters of local models by weighted least-squares approach, i.e.,

$$\mathbf{w}[i] = \arg\min_{\mathbf{w}[i]} \sum_{k=1}^{N} u_{ik}^m [y_k - g_i(\mathbf{x}_k;\mathbf{w}_i)]^2 \tag{4}$$

$(i = 1,2,\cdots,r)$, where fuzzifier $m \in [1,\infty)$ controls how many clusters may overlap [5].

III. SPARSE FUZZY C-REGRESSION MODEL

In this section, we consider the block structured in fuzzy c-regression method (FCRM) and develop a new fuzzy partition model on the basis of block sparse representation, namely as sparse fuzzy c-regression model. Furthermore, an algorithm are exploited on the basis of alternating direction method of multipliers (ADMM) to address the optimization of sparse fuzzy c-regression model.

For better presentation hereafter, we first introduce some notations about mix-norm of vector with block structure.

Definition 1: For a vector $\mathbf{x} = (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n$, its L_0-norm is defined to count the sparsity number of the vector by

$$\|\mathbf{x}\|_0 = \lim_{p \to 0} \|\mathbf{x}\|_p^p = \lim_{p \to 0} \sum_{k=1}^{m} |x_k|^p = \sharp\{i : x_i \neq 0\}$$

where $\sharp$ denotes set cardinality.

Definition 2: Let $\theta = (\theta[1]^\mathsf{T}, \theta[2]^\mathsf{T}, \cdots, \theta[r]^\mathsf{T})^\mathsf{T} \in \mathbb{R}^{rd}$ be a vector formed from r blocks $\theta[i] \in \mathbb{R}^d$ $(i = 1, 2, \cdots, r)$. We define the general mixed $L_{2,p}$-norm as

$$\|\theta\|_{2,p} = \|(\|\theta[1]\|_2, \|\theta[2]\|_2, \cdots, \|\theta[r]\|_2)^\mathsf{T}\|_p = (\sum_{i=1}^{r} \|\theta[i]\|_2^p)^{\frac{1}{p}}$$

where $p = 1, 2, \infty$ here and in the following.

Specifically, we have

$$\|\theta\|_{2,1} = \sum_{i=1}^{r} \|\theta[i]\|_2.$$

$\|\theta\|_{2,0}$, indeed, represents the number of non-zero blocks for

$$\|\theta\|_{2,0} = \lim_{p \to 0} (\sum_{i=1}^{r} \|\theta[i]\|_2^p)^{\frac{1}{p}} = \sharp\{i : \|\theta[i]\|_2 \neq 0\}.$$

Different from the tradition FCRM clustering methods, the proposed sparse fuzzy c-regression model takes advantage of the block structure information and control the number of prototypes of hyper-plane with nonzero parameters simultaneously. We collect all of the parameters of hyper-plane $\theta[i]$ $(i = 1, 2, \cdots, r)$ (the local regression models parameters in FCRM) into a vector θ such that $\theta = (\theta[1]^\mathsf{T}, \theta[2]^\mathsf{T}, \cdots, \theta[r]^\mathsf{T})^\mathsf{T} \in \mathbb{R}^{r(n+1)}$. The idea of sparse fuzzy c-regression model can be formulated by the optimization problem

$$\begin{aligned} \min \quad & \sum_{i=1}^{r}\sum_{j=1}^{N} u_{ij}^m [y_j - g_i(\mathbf{x}_j; \theta[i])]^2 + \lambda \|\theta\|_{2,0} \\ \text{s.t} \quad & \sum_{i=1}^{r} u_{ij} = 1 \;\; (j = 1, 2, \cdots, N) \\ & u_{ij} \in [0, 1] \;\; (i = 1, 2, \cdots, r; j = 1, 2, \cdots, N) \end{aligned} \tag{5}$$

where the constant $\lambda > 0$ is used to control the trade-off between model fit (the first term) and the number of clusters (the second term). This optimization problem aims to make the consequent parameter vector θ as block sparse as possible for a reasonable trade off between tradition objective function of FCRM and the number of prototypes of hyper-plane with nonzero parameters.

A. ADMM algorithm for sparse fuzzy c-regression model

It has been proved that optimization problems with mixed $L_{2,0}$-norm is NP-Hard. The $L_{2,1}$-norm convex optimization is a popular strategy to approximate the optimization problems [14], [15]. In this sense, we approximate solution of the optimization problem (5) by minimizing

$$\begin{aligned} \min \quad & \sum_{i=1}^{r}\sum_{j=1}^{N} u_{ij}^m [y_j - g_i(\mathbf{x}_j; \theta[i])]^2 + \lambda \sum_{i=1}^{r} \|\theta[i]\|_2 \\ \text{s.t} \quad & \sum_{i=1}^{r} u_{ij} = 1 \;\; (j = 1, 2, \cdots, N) \\ & u_{ij} \in [0, 1] \;\; (i = 1, 2, \cdots, r; j = 1, 2, \cdots, N) \end{aligned} \tag{6}$$

With an appropriate Lagrange multiplier $\beta = (\beta_1, \beta_2, \cdots, \beta_N)^\mathsf{T} \in \mathbb{R}^N$, the optimization problem (6) is rewritten as

$$\min f(\mathbf{U}, \theta, \beta) + \lambda \sum_{i=1}^{r} \|\theta[i]\|_2 \tag{7}$$

where

$$f(\mathbf{U}, \theta, \beta) = \sum_{i=1}^{r}\sum_{j=1}^{N} u_{ij}^m [y_j - g_i(\mathbf{x}_j; \theta[i])]^2 + \sum_{j=1}^{N} \beta_j (\sum_{i=1}^{r} u_{ij} - 1).$$

Writing the problem in the format above enables us to apply alternating direction method of multipliers (ADMM) for general L_1 regularized loss minimization [2], where ADMM is an algorithm that is intended to blend the decomposability of dual ascent with the superior convergence properties of the method of multipliers [2].

Let $h(\hat{\theta}) = \lambda \sum_{i=1}^{r} \|\hat{\theta}[i]\|_2$, the optimization problem (7) can be rewritten in ADMM form as

$$\begin{aligned} & \min f(\mathbf{U}, \theta, \beta) + h(\hat{\theta}) \\ & \text{s.t} \quad \theta - \hat{\theta} = \mathbf{0} \end{aligned} \tag{8}$$

Furthermore, with the method of multipliers $\alpha = (\alpha[1]^\mathsf{T}, \alpha[2]^\mathsf{T}, \cdots, \alpha[r]^\mathsf{T})^\mathsf{T} \in \mathbb{R}^{r(n+1)}$, $\alpha[i]^\mathsf{T} \in \mathbb{R}^{n+1}$ $(i = 1, 2, \cdots, r)$, we form the augmented Lagrangian as

$$L_\rho(\mathbf{U}, \theta, \hat{\theta}, \alpha) = f(\mathbf{U}, \theta, \beta) + h(\hat{\theta}) + \alpha^\mathsf{T}(\theta - \hat{\theta}) + \frac{\rho}{2}\|\theta - \hat{\theta}\|_2^2.$$

Therefore, the following iterations for optimization (8) are consisted in the framework of ADMM,

$$\begin{cases} \mathbf{U}^{k+1} := \arg\min_{\mathbf{U}} L_\rho(\mathbf{U}, \theta^k, \hat{\theta}^k, \alpha^k) \\ \theta^{k+1} := \arg\min_{\theta} L_\rho(\mathbf{U}^{k+1}, \theta, \hat{\theta}^k, \alpha^k) \\ \hat{\theta}^{k+1} := \arg\min_{\hat{\theta}} L_\rho(\mathbf{U}^{k+1}, \theta^{k+1}, \hat{\theta}, \alpha^k) \\ \alpha^{k+1} := \alpha^k + \rho(\theta^{k+1} - \hat{\theta}^{k+1}) \end{cases}$$

where $k = 0, 1, 2, \cdots$ denotes the number of iterations. Let $\mathbf{u} = \frac{1}{\rho}\alpha$ be a scaled dual variable, then we have the scaled form of ADMM for the optimization problem as

$$\mathbf{U}^{k+1} := \arg\min_{\mathbf{U}} f(\mathbf{U}, \theta^k, \beta) \tag{9}$$

$$\theta^{k+1} := \arg\min_{\theta} f(\mathbf{U}^{k+1}, \theta, \beta) + \frac{\rho}{2}\|\theta - \hat{\theta}^k + \mathbf{u}^k\|_2^2 \tag{10}$$

$$\hat{\theta}^{k+1} := \arg\min_{\hat{\theta}} h(\hat{\theta}) + \frac{\rho}{2}\|\theta^{k+1} - \hat{\theta} + \mathbf{u}^k\|_2^2 \tag{11}$$

$$\mathbf{u}^{k+1} := \mathbf{u}^k + \theta^{k+1} - \hat{\theta}^{k+1} \tag{12}$$

Subsequently, we introduce the following theorems to calculate the optimal values of variables $\mathbf{U}, \theta$ and $\hat{\theta}$ at each iteration.

Subsequently, we analysis the optimization problems (9)-(11) and derive the corresponding solutions by the following theorems.

Theorem 1: The necessary condition for a minimum of

$$f(\mathbf{U}, \theta, \beta) = \sum_{i=1}^{r}\sum_{j=1}^{N} u_{ij}^m [y_j - g_i(\mathbf{x}_j; \theta[i])]^2 + \sum_{j=1}^{N} \beta_j (\sum_{i=1}^{r} u_{ij} - 1)$$

yields the membership update equation:

$$u_{ij} = \frac{1}{\sum_{l=1}^{r} (D_{ij}/D_{lj})^{\frac{1}{m-1}}}$$

where $D_{ij} = (y_j - g_i(\mathbf{x}_j; \theta[i]))^2$ for all $i = 1, 2, \cdots, r$ and $j = 1, 2, \cdots, N$.

Proof: It is evident that this optimization problem is, indeed, the traditional FCRM.

Theorem 2: For the optimization problem

$$\min_{\theta} f(\mathbf{U},\theta,\beta)+\frac{\rho}{2}\|\theta-\hat{\theta}+\mathbf{u}\|_2^2,$$

we have

$$\theta[i]=(\rho\mathbf{I}+\mathbf{X}^{\mathsf{T}}\mathbf{U}_i\mathbf{X})^{-1}(\mathbf{X}^{\mathsf{T}}\mathbf{U}_i\mathbf{y}+\hat{\theta}[i]-\mathbf{u}[i])\ (i=1,2,\cdots,r).$$

Proof: Because

$$\begin{aligned}&f(\mathbf{U},\theta,\beta)+\frac{\rho}{2}\|\theta-\hat{\theta}+\mathbf{u}\|_2^2\\=&\frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{N}u_{ij}^m[y_j-g_i(\mathbf{x}_j;\theta[i])]^2+\frac{\rho}{2}\|\theta-\hat{\theta}+\mathbf{u}\|_2^2\end{aligned}$$

the optimization problem above is separable with respect to the partition $\theta[1],\theta[2],\cdots,\theta[r]$. Therefore, it is equivalent to address the following optimization problems

$$\min_{\theta[i]} l_i(\mathbf{U},\theta,\hat{\theta},\mathbf{u}) \tag{13}$$

for $i=1,2,\cdots,r$, where

$$\begin{aligned}&l_i(\mathbf{U},\theta,\hat{\theta},\mathbf{u})\\=&\frac{1}{2}(\mathbf{y}+\mathbf{X}\theta[i])^{\mathsf{T}}\mathbf{U}_i(\mathbf{y}-\mathbf{X}\theta[i])+\frac{\rho}{2}\|\theta[i]-\hat{\theta}[i]+\mathbf{u}[i]\|_2^2\end{aligned}$$

with $\mathbf{U}_i=\mathrm{diag}(u_{i1}^m,u_{i2}^m,\cdots,u_{iN}^m)$, $\mathbf{y}=(y_1,y_2,\cdots,y_N)^{\mathsf{T}}$ and $\mathbf{X}=(\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_N)^{\mathsf{T}}$, $\mathbf{x}_i=(1,x_{i1},x_{i2},\cdots,x_{in})^{\mathsf{T}}$. Furthermore, compute the derivative of l_i with respect to $\theta[i]$ and set it to zero, i.e.,

$$\frac{\partial l_i}{\partial\theta[i]}=-\mathbf{X}^{\mathsf{T}}\mathbf{U}_i\mathbf{y}+\mathbf{X}^{\mathsf{T}}\mathbf{U}_i\mathbf{X}\theta[i]+\rho(\theta[i]-\hat{\theta}[i]+\mathbf{u}[i])=0.$$

It leads that

$$\theta[i]=(\rho\mathbf{I}+\mathbf{X}^{\mathsf{T}}\mathbf{U}_i\mathbf{X})^{-1}(\mathbf{X}^{\mathsf{T}}\mathbf{U}_i\mathbf{y}+\rho(\hat{\theta}[i]-\mathbf{u}[i])) \tag{14}$$

for $i=1,2,\cdots,r$. The proof is completed.

In order to address the optimization problem with respect to $\hat{\theta}$, we first introduce a lemma on sum-of-norms regularization [17], [20].

Lemma 1: [17] Let $\mathbf{x}=(\mathbf{x}[1]^{\mathsf{T}},\mathbf{x}[2]^{\mathsf{T}},\cdots,\mathbf{x}[r]^{\mathsf{T}})\in\mathbb{R}^n$, where $\mathbf{x}[i]^{\mathsf{T}}\in\mathbb{R}^{n_i}$ with $0\le n_i\le n\ (\forall i)$ and $\sum_{i=1}^r n_i=n$. The minimizer of optimization problem

$$\min_{\mathbf{x}\in\mathbb{R}^n}\ \frac{1}{2}\|\mathbf{c}-\mathbf{x}\|_2^2+\lambda\|\mathbf{x}\|_{2,1}$$

is obtained as $\mathbf{x}[i]=\mathsf{S}_{\lambda}(\mathbf{c}[i])$ for $i=1,2,\cdots,r$, where the vector soft thresholding operator $\mathsf{S}_{\kappa}:\mathbb{R}^m\to\mathbb{R}^m$ is defined as

$$\mathsf{S}_{\kappa}(\mathbf{a})=(1-\frac{\kappa}{\|\mathbf{a}\|})_{+}\mathbf{a}\text{ with }\mathsf{S}_{\kappa}(\mathbf{0})=\mathbf{0}.$$

Theorem 3: For the optimization problem

$$\min_{\hat{\theta}}\ h(\hat{\theta})+\frac{\rho}{2}\|\theta-\hat{\theta}+\mathbf{u}\|_2^2$$

we have $\hat{\theta}[i]=\mathsf{S}_{\lambda/\rho}(\theta[i]+\mathbf{u}[i])$ for $i=1,2,\cdots,r$.

Based on the analysis above, we update the variables $\mathbf{U}^{k+1},\theta^{k+1},\hat{\theta}^k$ and $\mathbf{u}^{k+1}$ at each iteration k in the framework of ADMM by

$$u_{ij}^{k+1}=\frac{1}{\sum_{l=1}^r(D_{ij}^k/D_{lj}^k)^{\frac{1}{m-1}}},\ D_{ij}^k=(y_j-g_i(\mathbf{x}_j;\theta[i]^k))^2 \tag{15}$$

$$\theta[i]^{k+1}=(\rho\mathbf{I}+\mathbf{X}^{\mathsf{T}}\mathbf{U}_i^{k+1}\mathbf{X})^{-1}(\mathbf{X}^{\mathsf{T}}\mathbf{U}_i^{k+1}\mathbf{y}+\rho(\hat{\theta}[i]^k-\mathbf{u}[i]^k)) \tag{16}$$

$$\hat{\theta}[i]^{k+1}=\mathsf{S}_{\lambda/\rho}(\theta[i]^{k+1}+\mathbf{u}[i]^k) \tag{17}$$

$$\mathbf{u}^{k+1}=\mathbf{u}^k+\theta^{k+1}-\hat{\theta}^{k+1} \tag{18}$$

for $i=1,2,\cdots,r$ and $j=1,2,\cdots,N$. Furthermore, we summarize the ADMM algorithm for sparse fuzzy c-regression model by Algorithm 1.

Algorithm 1 ADMM for sparse fuzzy c-regression model

Input: Data set $\mathscr{D}$, Initial number of fuzzy rules r, parameters $m\in(1,\infty)$ and λ.
Initialization: Initialize $\mathbf{U}^0$ and $\theta[1]^0,\cdots,\theta[r]^0$.

```
for t = 1, 2, ··· do
    if halting criterion is not true then
        Update U^{k+1} using (18).
        Update θ[i]^{k+1} (i = 1, 2, ··· , r) using (19).
        Update θ̂[i]^{k+1} (i = 1, 2, ··· , r) using (20).
        Update u^{k+1} using (21).
    end if
end for
```

IV. SPARSE FUZZY c-REGRESSION MODEL FOR T-S FUZZY SYSTEMS IDENTIFICATION

In this section, we develop a novel T-S fuzzy systems identification method on the basis of sparse fuzzy c-regression model. The method consists of two phases: (1) Coarse learning phase of fuzzy partition and determination the number of clusters; and (2) Fine-tuning learning phase of consequent parameters of fuzzy rules.

Specifically, in the coarse learning phase, given the initial number of clusters r that is assigned a priori, we partition the input-output data set with the sparse fuzzy c-regression model. In such a way, fuzzy partition matrix $\mathbf{U}$ and the block sparse parameter vector θ are obtained. In the fine-tuning learning phase, we first pick up the m $(m\le r)$ nonzero blocks of θ, denoted by $\theta[i_j]$ and the corresponding fuzzy partition vectors $(u_{i_j1},u_{i_j2},\cdots,u_{i_jN})$, where $i_j\in\{1,2,\cdots,r\}$ for all $j=1,2,\cdots,m$. Then we extract the parameters of bell-shaped membership function of the fuzzy rules antecedent using (3) and further adjust the corresponding consequent parameters with global least square or local least square [3] to improve the modeling accuracy. We have summarized the flow chart of the sparse fuzzy c-regression model based method for T-S fuzzy systems identification in Fig. 1.

V. APPLICATION EXAMPLES

In this section, a benchmark example of Mackey-Glass time series is carried out to illustrate the effectiveness of

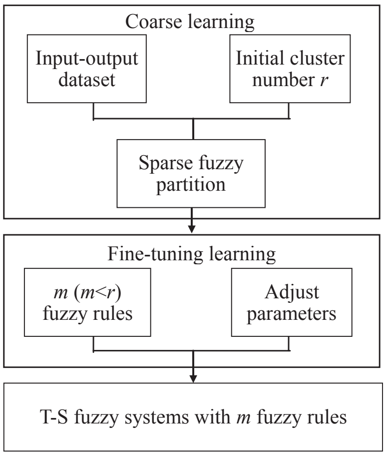

Fig. 1. The flow chart of sparse fuzzy c-regression model based method for T-S fuzzy systems identification.

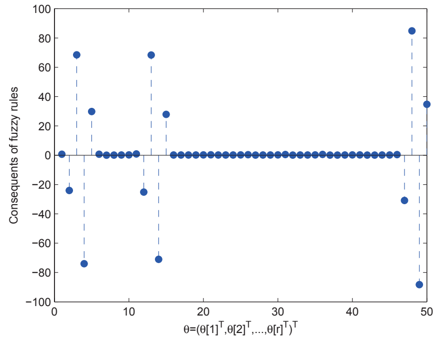

Fig. 2. Only three blocks of fuzzy rules consequent parameters remain nonzero by sparse fuzzy c-regression model.

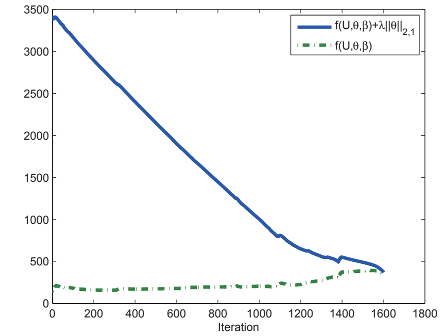

Fig. 3. The objective functions of sparse fuzzy c-regression model and traditional FCRM with iterations.

the proposed method for sparse fuzzy c-regression model and T-S fuzzy systems identification. The chaotic time series is generated by the Mackey-Glass delayed differential equation:

$$\frac{dx(t)}{dt} = \frac{0.2(t-\tau)}{1+x^{10}(t-\tau)} - 0.1x(t).$$

and has been used as a benchmark example in the areas of neural networks, fuzzy systems and hybrid systems [21], [22]. It is assumed that the time step $\Delta t = 0.1$, $x(0) = 1.2, \tau = 17$ and $x(t) = 0$ for $t \leq 0$. Based on Runge-Kutta (RK4) method, 85-steps ahead Mackey-Glass time series prediction is utilized in this example: predict the values $x(t+85)$ from input vectors $(x(t-18), x(t-12), x(t-6), x(t))$ for any value of the time t. The training data set is collected into

$$\mathscr{D}_{tr} = \{(x(t-18), x(t-12), x(t-6), x(t), x(t+85)) : \\ t = 201, 202, \cdots, 3200\}$$

and the testing data set is

$$\mathscr{D}_{te} = \{(x(t-18), x(t-12), x(t-6), x(t), x(t+85)) : \\ t = 5001, 5002, \cdots, 5500\}.$$

where $x(t-18), x(t-12), x(t-6), x(t)$ and $x(t+85)$ are the input variable and output variable of the time t, respectively.

In the ADMM algorithm for sparse fuzzy c-regression model, we set the parameters $\rho = 1.1$, $\lambda = 0.5$ and the initial number of clusters $r = 10$. The initial fuzzy partition of training data is obtained by FCM. In the process of ADMM for fuzzy partition and fuzzy rules selection, many blocks of fuzzy rule consequent parameters $\theta[i]$ $(i = 1, 2, \cdots, 10)$ shrink to zero and only three blocks remain nonzero (see Fig. 2). As a result, a T-S fuzzy model with three fuzzy rules are established after fine-tuning learning. The fuzzy rules are presented as following:

$$\begin{aligned} R^1 :\ & \text{If } x_1 \text{ is } A_{11}, x_2 \text{ is } A_{12}, \cdots, x_4 \text{ is } A_{14}, \text{ then} \\ & 0.0007 - 0.1676x_1 + 0.5052x_2 - 0.5212x_3 + 0.1839x_4 \\ R^2 :\ & \text{If } x_1 \text{ is } A_{21}, x_2 \text{ is } A_{22}, \cdots, x_6 \text{ is } A_{24}, \text{ then} \\ & y = 0.0004 - 0.1027x_1 + 0.3389x_2 - 0.3782x_3 + 0.1426x_4 \\ R^3 :\ & \text{If } x_1 \text{ is } A_{31}, x_2 \text{ is } A_{32}, \cdots, x_4 \text{ is } A_{34}, \text{ then} \\ & y = 0.0003 - 0.3140x_1 + 0.9934x_2 - 1.0590x_3 + 0.3803x_4 \end{aligned}$$

where the parameters c_{ij} and σ_{ij} of bell-shaped membership function for fuzzy rules' antecedent are illustrated by Table I. We also depict the objective function of traditional FCRM

TABLE I
ANTECEDENT PARAMETERS OF FUZZY RULES

Fuzzy rule	A_{i1}		A_{i2}	
	v_{i1}	σ_{i1}	v_{i2}	σ_{i2}
R^1	0.9921	0.0675	0.9897	0.0674
R^2	0.9312	0.0674	0.9252	0.0660
R^3	0.8925	0.0345	0.8831	0.0351

Fuzzy rule	A_{i3}		A_{i4}	
	v_{i3}	σ_{i3}	v_{i4}	σ_{i4}
R^1	0.9869	0.0670	0.9833	0.0661
R^2	0.9194	0.0643	0.9140	0.0623
R^3	0.8742	0.0362	0.8657	0.0379

TABLE II
COMPARISON RESULTS OF DIFFERENT FUZZY MODELS

Fuzzy models	Number of fuzzy rules(nodes, clusters)	NDEI
DENFIS	27	0.404
exTs	9	0.361
eTs	9	0.372
eTs+ [21]	8	0.438
rGK [22]	3	0.4667
rGK [22]	10	0.3787
Our model	3	0.3582

TABLE III
COMPARISON RESULTS OF MODELS FOR BOSTON HOUSING DATA

Fuzzy Model	Number of fuzzy rules	RMSE (training)	RMSE (testing)
SELM [24]	8192	24.12	24.38
RBFNN [23]	36	6.36	6.94
RBFNN + CFC [23]	36	5.52	6.91
Linguistic modeling [23]	36	4.12	5.32
SSEM [25]	4	4.23	5.36
Our model	3	3.6651	4.4293

and the proposed sparse fuzzy c-regression model in Fig. 3. It is guaranteed that the objective function of sparse fuzzy c-regression model keeps decreasing with iterations. Furthermore, we compare our fuzzy model with other excellent models on the nondimensional error index (NDEI) which is defined as the ratio of the root mean square error (RMSE) to the standard deviation of the target data (see Table II). It indicates that our fuzzy model performance better over other models with only three fuzzy rules.

A. *Example with Boston Housing Data*

The Boston Housing Data is obtained from the UCI Machine Learning Repository, where 506 samples are collected to concerns the housing values in suburbs of Boston. The main purpose of this experiment is to predict the median value of owner-occupied homes in \$1000's (the last column) from 1 binary-valued attribute and 12 continuous attributes (the first 13 columns). We split the data set into the training (60%) and testing set (40%). This pattern was also studied in [23]–[25], With cross validation method and the proposed algorithm in this paper, T-S fuzzy model with 13 input variables and one output variable is learned from the data set.

In the ADMM algorithm for sparse fuzzy c-regression model, we set the parameters $\rho = 1.1$, $\lambda = 0.5$ and the initial number of clusters $r = 10$. We compare our fuzzy model with other excellent models over the number of fuzzy rules, root mean square error (RMSE) for training data and testing date in Table 3. It indicates that our model performs better with just three T-S fuzzy rules; better accuracy on for both training data and testing data are also obtained. We also execute the simulation for 50 times by randomly dividing the data set into training data (60%) and testing data (40%). Over 50 times experiments, the averaged RMSE of training data and the corresponding standard deviation are obtained as 3.6724 and 0.3230, respectively; the values of testing data are derived as 4.5868 and 0.5763. It is evident that our model shows more robust results in terms of accuracy for both training data and testing data.

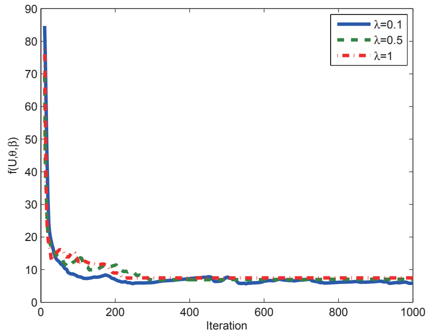

Fig. 4. With different values of λ, the variance of traditional objective function of FCRM within 1000 iterations.

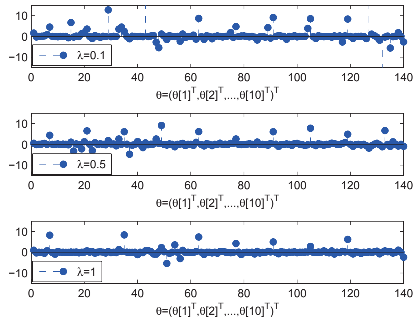

Fig. 5. With different values of λ, the sparsity of consequent parameters within 1000 iterations.

In the framework of sparse fuzzy c-regression model, parameter λ plays an important role by balancing a tradeoff between traditional objective function of FCRM and the number of clusters. There is no doubt that the first term of sparse fuzzy c-regression model decreases with the increase of the number of clusters. We study the influence of different λ on sparse fuzzy c-regression model and show the results by Fig. 4. It indicates that the traditional objective function of FCRM is convergent with different values of λ. Specifically, the consequent parameters becomes more sparser with bigger value of λ in the 1000-th iteration.

VI. CONCLUSIONS

In this paper, a new fuzzy partition method is developed on the basis of block structured sparse representation, and is applied to T-S fuzzy systems identification. Taking advantage of the block structured information, this method cast fuzzy partition into an optimization problem by making a tradeoff between traditional objective function of FCRM and the number of prototypes of hyper-plane with nonzero parameters. An ADMM based algorithm is exploited to address the sparse fuzzy c-regression model. In such a way, an appropriate number of fuzzy rules for T-S fuzzy systems identification is technically learned in the process of sparse fuzzy c-regression model. In future work, we will further investigate the issue of fuzzy system identification on the basis of sparse representation.

REFERENCES

[1] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Functional Algorithm*. New York: Plenum, 1981.

[2] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers”, *Foundations and Trends in Machine Learning* vol. 3, no. 1, pp. 1-122, 2010.

[3] R. Babuška, *Fuzzy modeling for control*, Kluwer Academic Publishers, Boston, 2001.

[4] E. Kim, M. Park, S. Ji, M. Park, “A new approach to fuzzy modeling”, *IEEE Transactions on Fuzzy Systems*, vol. 5, pp. 327-337, 1997.

[5] C. Kung, J. Su, “Affine Takagi-Sugeno fuzzy modelling algorithm by fuzzy c-regression models clustering with a novel cluster validity criterion”, *IET Control Theory* & *Applications*, vol. 1, no. 15, pp. 1255-1265, 2007.

[6] F. Höppner, F. KlawonnX, “Improved fuzzy partitions for fuzzy regression models”, *International Journal of Approximate Reasoning*, vol. 32, pp. 85-102, 2003.

[7] N. Zhao, F. Sun, “Modified clustering algorithm for fuzzy modeling using fuzzy *c*-regression model”, in *Proc. the 2005 International Conference on Computational Intelligence for Modelling, Control and Automation, and International Conference on Intelligent Agents, Web Technologies and Internet Commerce*, pp. 140-145, 2005.

[8] C. Chuang, C. Hsiao, J. T. Jeng, “Adaptive fuzzy regression clustering algorithm for TSK fuzzy modeling”, in *Proc. the 2003 IEEE International Symposium on Computational Intelligence in Rnbotirs and Automation*, (2003)200-206.

[9] C. S. Li, J. Z. Zhou, B. Fu, P. Kou and J. Xiao, “T-S fuzzy model identification with gravitational search based hyper-plane clustering algorithm,” *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 2, pp. 305-317, 2012.

[10] C. C. Chuang, S. F. Su and S. S. Chen, “Robust TSK fuzzy modeling for function approximation with outliers”, *IEEE Trans. Fuzzy Syst.*, vol. 9, no. 6, pp. 810-821, 2001.

[11] X. Xie, G. Beni, “A validity measure for fuzzy clustering”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 8, pp. 841-846, 1999.

[12] A. Flores-Sintas, J. Cadenas, F. Martin, “Partition validity and defuzzification”, *Fuzzy Sets Systems*, vol. 112, no. 3, pp. 433-447, 2000.

[13] A. Devillez, P. Billaudel, G. Lecolier, “A fuzzy hybrid hierarchical clustering method with a new criterion able to find the optimal partition”, *Fuzzy Sets Systems*, vol. 128, no. 3, pp. 323-338, 2002.

[14] M. Yuan, Y. Lin, “Model selection and estimation in regression with grouped variables”, *J. Roy. Stat. Soc. B*, vol. 68, no. 1,49-67, 2006.

[15] P. Zhao, G. Rocha and B. Yu, “Grouped and hierarchical model selection through composite absolute penalties”, *Annals of Statistics*, vol. 37, no. 6, pp. 3468-3497, 2009.

[16] A. C. Lozano, G. Świrszcz, N. Abe, “Grouped orthogonal matching pursuit for variable selection and prediction”, in *NIPS*, 2009.

[17] E. van Den Berg, M. Schmidt, M. P. Friedlander and K. Murphy, “Group sparsity via linear-time projection”, Technical Report TR-2008-09, Department of Computer Science, University of British Columbia.

[18] M. N. Luo, F. C. Sun and H. P. Liu, “A dynamic T-S fuzzy systems identification algorithm based on sparsity regularization”, in *Proc. 2012 IEEE International Symposium on Digital Object Identifier*, pp. 721-726, 2012.

[19] M. N. Luo, F. C. Sun and H. P. Liu, “Hierarchical Structured Sparse Representation for TS Fuzzy Systems Identification”, *IEEE Transactions on Fuzzy Systems*, vol. 21, no. 6, pp. 1032-1043, 2013.

[20] H. Ohlsson, L. Ljung, and S. Boyd, “Segmentation of ARX-models using sum-of-norms regularization”, *Automatica*, vol. 46, pp. 1107-1111, 2010.

[21] P. Angelov, “Evolving Takagi-Sugeno fuzzy systems from streaming data (eTs+)”, in Evolving Intelligent Systems: Methodology and Applications (eds P. Angelov, D. P. Filev and N. Kasabov), John Wiley & Sons, Inc., Hoboken, NJ, USA. doi: 10.1002/9780470569962.ch2, 2010.

[22] D. Dovžan, I. Škrjanc, “Recursive clustering based on a Gustafson-Kessel algorithm”, *Evolving Systems*, vol. 2, no. 1,pp. 15-24, 2011.

[23] W. Pedrycz, “Linguistic models as a framework of user-centric system modeling”, *IEEE Transactions on System, Man, and Cybernetics: Part A*, vol. 36, pp. 727-745, 2006.

[24] D. Wang, X. J. Zeng, J. A. Keane, “A structure evolving learning method for fuzzy systems”, *Evolving Systems*, vol. 1, pp. 83-95, 2010.

[25] D. Wang, X. J. Zeng, J. A. Keane, “A simplified structure evolving method for Mamdani fuzzy system identification and its application to high-dimensional problems”, *Information Sciences*, vol. 220, no. 20, pp. 110-123, 2013.

[26] J. C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms. New York: Plenum, 1981.