A MODIFICATION ON RIDGE ESTIMATION FOR FUZZY NONPARAMETRIC REGRESSION

R. FARNOOSH, J. GHASEMIAN AND O. SOLAYMANI FARD

Abstract. This paper deals with ridge estimation of fuzzy nonparametric regression models using triangular fuzzy numbers. This estimation method is obtained by implementing ridge regression learning algorithm in the Lagrangian dual space. The distance measure for fuzzy numbers that suggested by Diamond is used and the local linear smoothing technique with the cross-validation procedure for selecting the optimal value of the smoothing parameter is fuzzified to fit the presented model. Some simulation experiments are then presented which indicate the performance of the proposed method.

1. Introduction

In a great deal of literature on fuzzy regression analysis, most of research has focused on parametric forms of fuzzy regression, especially on the fuzzy linear regression models. In many practical situations, it may be unrealistic to predetermine a fuzzy parametric regression especially for a large dataset with a complicated underlying variation trend. In this respect, some other approaches have been developed to deal with the fuzzy regression problems without predefining a specific form of the underlying regression relationship. For example, Ishibuchi and Tanaka [13, 14] have suggested several fuzzy nonparametric regression methods by using the traditional back propagation networks. Cheng and Lee [1] have applied the radial basis function networks to fuzzy regression analysis.

In recent years, statistical nonparametric smoothing techniques have achieved significant development (see, for example, [6, 7, 8, 9]. These smoothing techniques are especially useful to deal with the nonparametric regression problems.

In multivariate analysis, the least-squares method is generally adopted in fitting a multiple linear regression model, but estimation of the least-squares is sometimes far from being perfect. One of the important causes leading to the result is the column vectors of the matrix of data which is considered independent is close to linear correlation. Approximate linear relationship among independent variables is called multicolinearity [16]. The multicolinearity among the independent variables leads to increase error in estimating of regression coefficients. The often used criterion to verify colinearity is simple correlation coefficient. When simple correlation

Received: July 2010; Revised: March 2011; Accepted: July 2011

Key words and phrases: Fuzzy regression, Ridge estimation, Fuzzy nonparametric regression, Local linear smoothing.

coefficient between two independent variables is large, the colinearity is considered [3, 16].

Ridge regression is one of the ways to overcome this problem, which was first used in the context of least square regression in [12]. Some papers such as Drucker et al. [4] and Saunders et al. [19] have used ridge regression in conjunction with a high dimensional feature space. The idea of ridge regression learning algorithm has also been used by Hong et al. [11] to fit some nonlinear fuzzy regression models. Moreover, Hong and Hwang [10] have developed the support vector fuzzy regression machines.

In this study, we focus on ridge regression. The local linear smoothing method, that is a special case of the local polynomial smoothing technique, is fuzzified to handle the fuzzy nonparametric regression with triangular fuzzy numbers based on the distance measure proposed by Diamond [2]. A distance based cross-validation procedure for selecting the optimal value of the smoothing parameter is also suggested.

The structure of paper is as follows: section 2 explains basic concepts of triangular fuzzy numbers and the local linear smoothing method. In section 3, fuzzy ridge nonparametric regression model will be presented. In section 4, we discuss the selection of kernel functions and the smoothing parameter. Finally, in the two last sections, some numerical examples and comments are given.

2. Multiple Fuzzy Nonparametric Regression Model

A fuzzy nonparametric regression model with multiple crisp input and triangular fuzzy output is considered in this section and, based on the local linear smoothing approach, a fitting procedure is proposed for this model.

2.1. **Basic Concepts.** Let $a = (m_a - \alpha_a, m_a, m_a + \beta_a)$ be a triangular fuzzy number with its center, left and right spread being, respectively, m_a, α_a and β_a. The membership function of a is

$$\mu_a(t) = \begin{cases} \frac{t-(m_a-\alpha_a)}{\alpha_a} & \text{if } m_a - \alpha_a \le t < m_x \\ \frac{m_a+\beta_a-t}{\beta_a} & \text{if } m_a \le t < m_a + \beta_a \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

In this paper, we denote the space of all fuzzy triangular numbers by $T(R)$, i.e. $T(R) = \{a : a = (m_a - \alpha_a, m_a, m_a + \beta_a)\}$. Now, consider the following multivariate fuzzy nonparametric regression model

$$Y = F(\mathbf{x})\{+\}\varepsilon = (m(\mathbf{x}) - \alpha(\mathbf{x}), m(\mathbf{x}), m(\mathbf{x}) + \beta(\mathbf{x}))\,\{+\}\varepsilon \tag{2}$$

In this model, $\mathbf{x} = (x_1, ..., x_p)$ is a $p-$dimensional crisp independent variable (input) whose domain is assumed to be $D \subseteq R^p$. $Y \in T(R)$ is a triangular fuzzy dependent variable (output). $F(\mathbf{x})$, a mapping from D to $T(R)$, is an unknown fuzzy regression function with its center, lower and upper limits being respectively $m(\mathbf{x}), l(\mathbf{x}) = m(\mathbf{x}) - \alpha(\mathbf{x})$ and $r(\mathbf{x}) = m(\mathbf{x}) + \beta(\mathbf{x})$. Moreover ε is an error term. Instead of being solely regarded as a random error with mean zero, ε may also be considered as a fuzzy error or a hybrid error containing both fuzzy and random components. $\{+\}$ is an operator whose definition depends on the fuzzy ranking method used.

2.2. **Local Linear Smoothing Method.** Let $a = (l_a, m_a, r_a)$ and $b = (l_b, m_b, r_b)$, $m_a, r_a, m_b, r_b \geq 0$ be any two triangular numbers in $T(R)$. Diamond [2] defined a distance between a and b as

$$d(a,b)^2 = (l_a - l_b)^2 + (m_a - m_b)^2 + (r_a - r_b)^2. \tag{3}$$

The distance (3) indeed measures the closeness between the membership functions of two triangular fuzzy numbers (for more details see [2] and [21]). The closeness between two fuzzy numbers can be measure as a new form that is named "possibilistic approach". Authors of [18, 17] have survey the linear regression based on this approach. We henceforth base the distance (3) to extend the local linear smoothing technique to fit the fuzzy nonparametric model (2). The main object in fuzzy nonparametric regression is to estimate $F(\mathbf{x})$ at any $\mathbf{x} \in D \subseteq R^p$ based on $(\mathbf{x}_i, Y_i), i = 1, 2, ..., n$. As pointed out by Kim and Bishu [15], the membership function of an estimated fuzzy output should be as close to that of the corresponding observed fuzzy number as possible. From this point of view, we shall estimate $m(\mathbf{x}), l(\mathbf{x})$ and $r(\mathbf{x})$ for each $\mathbf{x} \in D$ in the sense of best fit with respect to some distances that can measure the closeness between the membership functions of the estimated fuzzy output and the corresponding observed one. Suppose that $m(\mathbf{x}), l(\mathbf{x})$ and $r(\mathbf{x})$ have continuous partial derivatives with respect to each component x_i in the domain D of $\mathbf{x}$. Then, for a given $\mathbf{x_0} = (x_{01}, ..., x_{0p}) \in D$ and with Taylor's expansion, $m(\mathbf{x}), l(\mathbf{x})$ and $r(\mathbf{x})$ can be locally approximated in a neighborhood of $\mathbf{x}_0$, respectively by the following linear functions:

$$\begin{cases} l(\mathbf{x}) \approx \tilde{l}(\mathbf{x}) = l(\mathbf{x}_0) + l^{(x_1)}(\mathbf{x}_0)(x_1 - x_{01}) + ... + l^{(x_p)}(\mathbf{x}_0)(x_p - x_{0p}), \\ m(\mathbf{x}) \approx \tilde{m}(\mathbf{x}) = m(\mathbf{x}_0) + m^{(x_1)}(\mathbf{x}_0)(x_1 - x_{01}) + ... + m^{(x_p)}(\mathbf{x}_0)(x_p - x_{0p}), \\ r((\mathbf{x})) \approx \tilde{r}(\mathbf{x}) = r(\mathbf{x}_0) + r^{(x_1)}(\mathbf{x}_0)(x_1 - x_{01}) + ... + r^{(x_p)}(\mathbf{x}_0)(x_p - x_{0p}), \end{cases} \tag{4}$$

where $m^{(x_j)}(\mathbf{x}_0), l^{(x_j)}(\mathbf{x}_0)$ and $r^{(x_j)}(\mathbf{x}_0), j = 1, 2, ..., p$ are, respectively, the derivatives of $m(\mathbf{x}), l(\mathbf{x})$ and $r(\mathbf{x})$ with respect to x_j at $\mathbf{x}_0$.

Let $(\mathbf{x}_i, Y_i) = (x_{i1}, ..., x_{ip}, (l_{y_i}, m_{y_i}, r_{y_i})_{T(R)}), i = 1, 2, ..., n$ be a sample of the observed crisp inputs and triangular fuzzy outputs of the model (2) with the underlying fuzzy nonparametric regression function $F(\mathbf{x}) = (l(\mathbf{x}), m(\mathbf{x}), r(\mathbf{x}))$. Based on Diamond's distance (3) the following locally weighted least-squares is formulated. That is, minimize

$$\begin{aligned} &\textstyle\sum_{i=1}^n d^2 \left((l_{y_i}, m_{y_i}, r_{y_i})_{T(R)}, (\tilde{l}(\mathbf{x}_i), \tilde{m}(\mathbf{x}_i), \tilde{r}(\mathbf{x}_i))_{T(R)}\right) K_h(||\mathbf{x}_i - \mathbf{x}_0||) \\ &= \textstyle\sum_{i=1}^n (l_{y_i} - l(\mathbf{x}_0) - \sum_{j=1}^p l^{(x_j)}(\mathbf{x}_0)(x_{ij} - x_{0j}))^2 K_h(||\mathbf{x}_i - \mathbf{x}_0||) \\ &+ \textstyle\sum_{i=1}^n (m_{y_i} - m(\mathbf{x}_0) - \sum_{j=1}^p m^{(x_j)}(\mathbf{x}_0)(x_{ij} - x_{0j}))^2 K_h(||\mathbf{x}_i - \mathbf{x}_0||) \\ &+ \textstyle\sum_{i=1}^n (r_{y_i} - r(\mathbf{x}_0) - \sum_{j=1}^p r^{(x_j)}(\mathbf{x}_0)(x_{ij} - x_{0j}))^2 K_h(||\mathbf{x}_i - \mathbf{x}_0||) \end{aligned} \tag{5}$$

with respect to $m(\mathbf{x}_0), l(\mathbf{x}_0), r(\mathbf{x}_0)$ and $m^{(x_j)}(\mathbf{x}_0), l^{(x_j)}(\mathbf{x}_0), r^{(x_j)}(\mathbf{x}_0), j = 1, ..., p$ for the given kernel $K_h(.)$ and smoothing parameter h, where

$$K_h(||\mathbf{x}_i - \mathbf{x}_0||) = \frac{K(||\mathbf{x}_i - \mathbf{x}_0||/h)}{h}, i = 1, 2, ..., n,$$

are a sequence of weights at $\mathbf{x}_0$ whose role is to make the data that are close to $\mathbf{x}_0$ contribute more in estimating the parameters at $\mathbf{x}_0$ than those that are farther away with the adjustment of h. By solving this weighted least-squares problem, we can obtain not only the estimates of $m(\mathbf{x}_0), l(\mathbf{x}_0)$ and $r(\mathbf{x}_0)$ at $\mathbf{x}_0$, but also those of their respective derivatives $m^{(x_j)}(\mathbf{x}_0), l^{(x_j)}(\mathbf{x}_0), r^{(x_j)}(\mathbf{x}_0), j = 1, ..., p$. Since we mainly focus on estimating the underlying fuzzy nonparametric regression function $F(\mathbf{x}) = (l(\mathbf{x}), m(\mathbf{x}), r(\mathbf{x}))$ at $\mathbf{x}_0$, it is natural to take the solutions of $m(\mathbf{x}_0), l(\mathbf{x}_0)$ and $r(\mathbf{x}_0)$ in equation (5), denoted, respectively by $\hat{m}(\mathbf{x}_0), \hat{l}(\mathbf{x}_0)$ and $\hat{r}(\mathbf{x}_0)$ as the estimates of the center, the lower and the upper spread of $F(\mathbf{x})$ at $\mathbf{x}_0$. That is, the estimate of $F(\mathbf{x})$ at $\mathbf{x}_0$ is $\hat{F}(\mathbf{x}_0) = (\hat{l}(\mathbf{x}_0), \hat{m}(\mathbf{x}_0), \hat{r}(\mathbf{x}_0)) = (\hat{m}(\mathbf{x}_0) - \hat{\alpha}(\mathbf{x}_0), \hat{m}(\mathbf{x}_0), \hat{m}(\mathbf{x}_0) + \hat{\beta}(\mathbf{x}_0))$. It is observed that the equation (5) is a summation of the three parts and each part includes separately a different group of the unknown parameters, that is, $(m(\mathbf{x}_0), m^{(x_1)}(\mathbf{x}_0), ..., m^{(x_p)}(\mathbf{x_0})), (l(\mathbf{x}_0), l^{(x_1)}(\mathbf{x}_0), ..., l^{(x_p)}(\mathbf{x_0}))$ or $(r(\mathbf{x}_0), r^{(x_1)}(\mathbf{x}_0), ..., r^{(x_p)}(\mathbf{x_0}))$. Therefore, taking the partial derivatives of the objective function (5) with respect to these unknown parameters to be zero forms three groups of linear equations which include separately these parameters. Therefore, we can separately solve these three groups of linear equations to obtain the estimates of these parameters. In fact, according to the principle of the weighted least-squares and by utilizing matrix notations, we can immediately obtain

$$\left(\hat{l}(\mathbf{x}_0), \hat{l}^{(x_1)}(\mathbf{x}_0), ..., \hat{l}^{(x_p)}(\mathbf{x_0})\right)^{\mathbf{T}} = \left(\mathbf{X^T}(\mathbf{x_0})\mathbf{W}(\mathbf{x_0}; h)\mathbf{X}(\mathbf{x_0})\right)^{-1}\mathbf{X^T}(\mathbf{x_0})\mathbf{W}(\mathbf{x_0}; h)\mathbf{L}_y,$$
$$\left(\hat{m}(\mathbf{x}_0), \hat{m}^{(x_1)}(\mathbf{x}_0), ..., \hat{m}^{(x_p)}(\mathbf{x_0})\right)^{\mathbf{T}} = \left(\mathbf{X^T}(\mathbf{x_0})\mathbf{W}(\mathbf{x_0}; h)\mathbf{X}(\mathbf{x_0})\right)^{-1}\mathbf{X^T}(\mathbf{x_0})\mathbf{W}(\mathbf{x_0}; h)\mathbf{M}_y,$$
$$\left(\hat{r}(\mathbf{x}_0), \hat{r}^{(x_1)}(\mathbf{x}_0), ..., \hat{r}^{(x_p)}(\mathbf{x_0})\right)^{\mathbf{T}} = \left(\mathbf{X^T}(\mathbf{x_0})\mathbf{W}(\mathbf{x_0}; h)\mathbf{X}(\mathbf{x_0})\right)^{-1}\mathbf{X^T}(\mathbf{x_0})\mathbf{W}(\mathbf{x_0}; h)\mathbf{R}_y,$$

$$\mathbf{X}(\mathbf{x}_0) = \begin{pmatrix} 1 & x_{11} - x_{01} \ldots x_{1p} - x_{0p} \\ 1 & x_{21} - x_{01} \ldots x_{2p} - x_{0p} \\ \vdots & \vdots \\ 1 & x_{n1} - x_{01} \ldots x_{np} - x_{0p} \end{pmatrix} \mathbf{L}_y = \begin{pmatrix} l_{y_1} \\ l_{y_2} \\ \vdots \\ l_{y_n} \end{pmatrix} \mathbf{M}_y = \begin{pmatrix} m_{y_1} \\ m_{y_2} \\ \vdots \\ m_{y_n} \end{pmatrix} \mathbf{R}_y = \begin{pmatrix} r_{y_1} \\ r_{y_2} \\ \vdots \\ r_{y_n} \end{pmatrix} \tag{6}$$

and

$$\mathbf{W}(\mathbf{x}_0; h) = diag(K_h(||\mathbf{x}_1 - \mathbf{x}_0||)\, K_h(||\mathbf{x}_2 - \mathbf{x}_0||), ..., K_h(||\mathbf{x}_n - bfx_0||)).$$

Thus, the estimated fuzzy regression function $\hat{F}(\mathbf{x}_0)$ at $\mathbf{x}_0$ can be expressed by

$$\hat{F}(\mathbf{x}_0) = (\hat{l}(\mathbf{x}_0), \hat{m}(\mathbf{x}_0), \hat{r}(\mathbf{x}_0))_{T(R)} = (\mathbf{e}_1^T\mathbf{H}(\mathbf{x}_0; h)\mathbf{L}_y, \mathbf{e}_1^T\mathbf{H}(\mathbf{x}_0; h)\mathbf{M}_y, \mathbf{e}_1^T\mathbf{H}(\mathbf{x}_0; h)\mathbf{R}_y), \tag{7}$$

where

$$\mathbf{H}(\mathbf{x}_0; h) = (\mathbf{X}^T(\mathbf{x}_0)\mathbf{W}(\mathbf{x}_0; h)\mathbf{X}(\mathbf{x}_0))^{-1}\mathbf{X}^T(\mathbf{x}_0)\mathbf{W}(\mathbf{x}_0; h)$$

and $\mathbf{e}_1 = (1, 0, ..., 0)^T$, a $(p+1)-$dimensional vector with the first element being unity and the others being zero.

3. Fuzzy Ridge Nonparametric Regression Model

In most cases, due to multicolinearity among independent variables, either the matrix

$$\mathbf{X}^T(\mathbf{x}_0)\mathbf{W}(\mathbf{x}_0; h)\mathbf{X}(\mathbf{x}_0)$$

is a singular matrix or is very close to a singular matrix. In the least square method the name of this matrix is the weighted coefficients matrix of normal equations. In this paper, we use ridge regression to overcome this problem. Ridge regression gives computational efficiency in finding solutions of fuzzy regression models particularly for multivariable cases. In the following, we illustrate ridge regression procedures for model (4), which is based on the algorithms in dual variables. We need to slightly modify the formulation of ridge regression for crisp data.

Suppose that the observations consist of data pairs $(\mathbf{x}_i, Y_i), i = 1, 2, ..., n$, where $\mathbf{x}_i$ is $p-$vector of real numbers and each $Y_i \in T(R)$. Suppose $x_{ij} \geq 0$ be elements of $\mathbf{x}_i$. Let $\mathbf{A} = (a_1, ..., a_p)$ where $a_i = (m_{a_i}, \alpha_{a_i}, \beta_{a_i}), \alpha_{a_i}, \beta_{a_i} \geq 0, i = 1, ..., p$ and let $a_0 = (m_{a_0}, \alpha_{a_0}, \beta_{a_0}), \alpha_{a_0}, \beta_{a_0} \geq 0$. Without loss of generality, each equation of model (4) can be written in the following form:

$$G(\mathbf{x}) = a_0 + < \mathbf{A}, \mathbf{x} >= a_0 + a_1\mathbf{x}_1 + ... + a_p\mathbf{x}_p, a_0 \in T(R), \mathbf{A} \in T(R)^p, \mathbf{x} \in R^p \quad (8)$$

where $T(R)^p$ is the set of $p-$vectors of triangular fuzzy numbers. We consider

$$m_{\mathbf{A}} = (m_{a_0}, m_{a_1}, ..., m_{a_p}), \quad \alpha_{\mathbf{A}} = (\alpha_{a_0}, \alpha_{a_1}, ..., \alpha_{a_p}), \quad \beta_{\mathbf{A}} = (\beta_{a_0}, \beta_{a_1}, ..., \beta_{a_p}).$$

Then, by defining $||\mathbf{A}||^2 = ||m_{\mathbf{A}}||^2 + ||m_{\mathbf{A}} - \alpha_{\mathbf{A}}||^2 + ||m_{\mathbf{A}} + \beta_{\mathbf{A}}||^2$, we arrive at the following ridge regression learning procedure for this model (8) as follows:

$$\text{Minimize} \quad \lambda||\mathbf{A}||^2 + \sum_{k=1}^{3}\sum_{i=1}^{n} \xi_{ki}^2 \, K_h(||\mathbf{x}_i - \mathbf{x}_0||)$$

$$\text{Subject to} \begin{cases} m_{y_i} - < m_{\mathbf{A}}, \mathbf{x}_i >= \xi_{1i}, \\ (m_{y_i} - \alpha_{y_i}) - (< m_{\mathbf{A}}, \mathbf{x}_i > - < \alpha_{\mathbf{A}}, \mathbf{x}_i >) = \xi_{2i}, \\ (m_{y_i} + \beta_{y_i}) - (< m_{\mathbf{A}}, \mathbf{x}_i > + < \beta_{\mathbf{A}}, \mathbf{x}_i >) = \xi_{3i}. \end{cases} \quad (9)$$

Hence, we can construct a Lagrange function as follows:

$$\begin{aligned} L = \lambda||\mathbf{A}||^2 + \sum_{k=1}^{3}\sum_{i=1}^{n} \xi_{ki}^2 K_h(||\mathbf{x}_i - \mathbf{x}_0||) + \sum_{i=1}^{n} \theta_{1i}(m_{y_i} - < m_{\mathbf{A}}, \mathbf{x}_i > -\xi_{1i}) \\ + \sum_{i=1}^{n} \theta_{2i}(m_{y_i} - \alpha_{y_i} - (< m_{\mathbf{A}}, \mathbf{x}_i > - < \alpha_{\mathbf{A}}, \mathbf{x} >)\xi_{2i} \\ + \sum_{i=1}^{n} \theta_{3i}(m_{y_i} + \beta_{y_i} - (< m_{\mathbf{A}}, \mathbf{x}_i > - < \beta_{\mathbf{A}}, \mathbf{x} >)\xi_{3i}). \end{aligned} \quad (10)$$

It follows from the saddle point condition that the partial derivatives of L with respect to the primal variable $(\mathbf{A}, \xi_{ki}, k = 1, 2, 3)$ have to vanish for optimality,

$$\frac{\partial L}{\partial m_{\mathbf{A}}} = 2\lambda(3m_{\mathbf{A}} - \alpha_{\mathbf{A}} + \beta_{\mathbf{A}}) - \sum_{i=1}^{n} \theta_{1i}\mathbf{x}_i - \sum_{i=1}^{n} \theta_{2i}\mathbf{x}_i - \sum_{i=1}^{n} \theta_{3i}\mathbf{x}_i = 0, \quad (11)$$

$$\frac{\partial L}{\partial \alpha_{\mathbf{A}}} = 2\lambda(-m_{\mathbf{A}} + \alpha_{\mathbf{A}}) + \sum_{i=1}^{n} \theta_{2i}\mathbf{x}_i = 0, \quad (12)$$

$$\frac{\partial L}{\partial \beta_{\mathbf{A}}} = 2\lambda(m_{\mathbf{A}} + \beta_{\mathbf{A}}) - \sum_{i=1}^{n} \theta_{3i}\mathbf{x}_i = 0, \quad (13)$$

$$\frac{\partial L}{\partial \xi_{ki}} = 2\xi_{ki}K_h(||\mathbf{x}_i - \mathbf{x}_0||) - \theta_{ki} = 0\,, \quad k = 1, 2, 3, i = 1, ..., n. \quad (14)$$

Equations (3), (3) and (3) can be written as be follows:

$$m_{\mathbf{A}} = \frac{1}{2\lambda} \sum_{i=1}^{n} \theta_{1i}\mathbf{x}_i, \quad (15)$$

$$\alpha_{\mathbf{A}} = \frac{1}{2\lambda} \sum_{i=1}^{n} (\theta_{1i} - \theta_{2i})\mathbf{x}_i, \quad (16)$$

$$\beta_{\mathbf{A}} = \frac{1}{2\lambda} \sum_{i=1}^{n} (\theta_{3i} - \theta_{1i})\mathbf{x}_i. \quad (17)$$

We should take $\alpha_{\mathbf{A}} = \max\{0, \alpha_{\mathbf{A}}\}$ and $\beta_{\mathbf{A}} = \max\{0, \beta_{\mathbf{A}}\}$ in order to compute spreads. Here, 0 represents the corresponding zero vector. We use the same $\alpha_{\mathbf{A}}$ and $\beta_{\mathbf{A}}$ to avoid the abuse of notations. Substituting (11)-(14) into (10), we obtain the following simplified L:

$$L = -\frac{1}{4\lambda} \sum_{k=1}^{3} \sum_{i,j=1}^{n} \theta_{ki}\theta_{kj} < \mathbf{x}_i, \mathbf{x}_j > -\frac{1}{4} \sum_{k=1}^{3} \sum_{i=1}^{n} \frac{\theta_{ki}^2}{K_h(||\mathbf{x}_i - \mathbf{x}_0||)} + \sum_{i=1}^{n} m_{y_i}\theta_{1i} + \sum_{i=1}^{n} (m_{y_i} - \alpha_{y_i})\theta_{2i} + \sum_{i=1}^{n} (m_{y_i} + \beta_{y_i})\theta_{3i}. \quad (18)$$

Now we denote $\Theta, M, W, \tilde{W}$ and N respectively by

$\Theta = (\theta_{11}, ..., \theta_{1n}, \theta_{21}, ..., \theta_{2n}, \theta_{31}, ..., \theta_{3n})^T$,

$M = (m_{y_1}, ..., m_{y_n}, m_{y_1} - \alpha_{y_1}, ..., m_{y_n} - \alpha_{y_n}, m_{y_1} + \beta_{y_1}, ..., m_{y_n} + \beta_{y_n})^T$,

$W = W(\mathbf{x}_0; h) = diag(K_h(||\mathbf{x}_1 - \mathbf{x}_0||), K_h(||\mathbf{x}_2 - \mathbf{x}_0||), ..., K_h(||\mathbf{x}_n - \mathbf{x}_0||))$,

$$\tilde{W} = \begin{bmatrix} W & 0 & 0 \\ 0 & W & 0 \\ 0 & 0 & W \end{bmatrix}, \qquad N = \begin{bmatrix} WQ & 0 & 0 \\ 0 & WQ & 0 \\ 0 & 0 & WQ \end{bmatrix}$$

where Q is a $n \times n$ matrix of $Q_{ij} =< \mathbf{x}_i, \mathbf{x}_j >$ and 0 is the $n \times n$ zero matrix. Therefore, by differentiating L with respect to Θ, we have

$$\Theta = 2\lambda(N + \lambda I)^{-1}\tilde{W}M \tag{19}$$

Recalling the expression for model (8), (15), (16) and (17), we can obtain the prediction of $G(\mathbf{x})$ given the ridge regression procedure.

4. Selection of the Kernel Function and the Smoothing Parameter

When we use the above procedure to fit the fuzzy ridge nonparametric regression model (2), the regularization parameter λ, the kernel $K(.)$ and the smoothing parameter h in the $K_h(.)$ should be determined first. The role of the $K_h(||\mathbf{x}_i - \mathbf{x}_0||), i = 1, ..., n$ is to make the data that are close to the given point $\mathbf{x}_0$ contribute more to the estimate $\hat{F}(\mathbf{x})$ than those that are farther away. There are many types of kernel functions. In this work, we use Gaussian kernel

$$K(x) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{x^2}{2}) \tag{20}$$

and Epanechnikov's kernel

$$K(x) = \begin{cases} \frac{3}{4}(1 - x^2), & \text{if } |x| \leq 1 \\ 0, & \text{o.w} \end{cases} \tag{21}$$

To estimate $\hat{l}(\mathbf{x}), \hat{m}(\mathbf{x})$ and $\hat{r}(\mathbf{x})$, by using the role of the smoothing parameter h in the $K_h(.)$, we improve the degree of smoothness. Small h makes $\hat{m}(\mathbf{x}), \hat{l}(\mathbf{x})$ and $\hat{r}(\mathbf{x})$ too fluctuating and leads to over-fit and large h makes them too smooth and leads to lack-of-fit. Therefore, the proper selection of the smoothing parameter value is important in the local smoothing techniques. There have been a few approaches for selecting the optimal value of the smoothing parameter such as the cross-validation, generalized cross-validation, Bayesian and bootstrap methods (see [6, 8, 11], which are all possible candidates for the above local linear smoothing method. In this paper, a fuzzified cross-validation procedure is used to achieve this task because it is easy in implementing in our setting and shows a satisfactory performance by the simulations conducted in the following section. Based on Diamond's distance (3), the fuzzified cross-validation procedure can be described as follows. Let

$$\hat{F}_{(i)}(\mathbf{x}_i; h) = (\hat{l}_{(i)}(\mathbf{x}_i; h), \hat{m}_{(i)}(\mathbf{x}_i; h), \hat{r}_{(i)}(\mathbf{x}_i; h))_{T(R)}$$

be the predicted fuzzy ridge nonparametric regression function at input x_i computed by equation (7) with the smoothing parameter h in which the observation ith is eliminated in the process of implementing the fitting procedure. Here h is written in the expression of $\hat{F}_{(i)}(\mathbf{x}_i; h)$ in order to show that the predicted function at x_i is related to the smoothing parameter h. Compute $\hat{F}_{(i)}(\mathbf{x}_i; h)$for each $\mathbf{x}_i, i = 1, ..., n$ and let

$$\begin{aligned} CV(h) &= \frac{1}{n}\sum_{i=1}^{n} d^2(Y_i, \hat{F}_i(\mathbf{x}_i; h)) \\ &= \frac{1}{n}\sum_{i=1}^{n}((l_{y_i} - \hat{l}_{(i)}(\mathbf{x}_i; h))^2 + (m_{y_i} - \hat{m}_i(\mathbf{x}_i; h))^2 + (r_{y_i} - \hat{r}_i(\mathbf{x}_i; h))^2). \end{aligned} \tag{22}$$

However, because of the error term in model (2), $CV(h)$ cannot efficiently reflect the closeness between the underlying fuzzy nonparametric regression function $F(\mathbf{x})$ and its estimate. With this consideration, we further define a quantity for measuring the bias between the underlying fuzzy regression function and its estimate, which is

$$\begin{aligned} BIAS(h) &= \tfrac{1}{n}\sum_{i=1}^{n} d^2(F(\mathbf{x}),\hat{F}(\mathbf{x};h)) \\ &= \tfrac{1}{n}\sum_{i=1}^{n}((l(\mathbf{x})-\hat{l}(\mathbf{x};h))^2+(m(\mathbf{x})-\hat{m}(\mathbf{x};h))^2+(r(\mathbf{x})-\hat{r}(\mathbf{x};h))^2). \end{aligned} \quad (23)$$

Since $F(\mathbf{x})$ is certainly unknown then $BIAS(h)$ is not computable in practical applications. This quantity makes sense for examining the performance of the different methods by simulation. Both $CV(h)$ and $BIAS(h)$ will be reported in our simulations to numerically evaluate the performance of the proposed method. Choose h_0 as the optimal value such that

$$CV(h_0)=\min_{h>0} CV(h) \quad \text{and} \quad BIAS(h_0)=\min_{h>0} BIAS(h).$$

In practice, we may compute for a series of values of h to obtain h_0. The optimal value of h closely depends on the degree of smoothness of the regression function. A smoother regression function generally corresponds to a larger value of h while a more fluctuating regression function tends to select a smaller value of h. We note that the selected optimal value of h by the $CV(h)$ and $BIAS(h)$ closely depends on the degrees of smoothness of the center, the lower and the upper limit functions.

5. **Simulation Experiments**

In this section, the proposed method for three datasets is used to compare the results with the local linear smoothing (LLS) and the kernel smoothing (KS) methods. Gaussian (20) and Epanechnikov's kernels (21) are respectively performed to generate the weight sequence for all three methods and then these methods are compared by $CV(h)$ and $BIAS(h)$. Now the following examples from the literature [20] will be considered and the results of which are to be obtained.

Example 5.1. Consider the below function

$$g_1(x)=\frac{1}{5}x^2+2\exp(\frac{x}{10})$$

and equidistantly take $x_i=0.1i(i=1,2,...,100)$ on interval [0.1,10]. Let

$$\begin{cases} y_i=g_1(x_i)+\text{rand}[-0.5,0.5], \\ \sigma_i=\frac{1}{4}g_1(x_i)+\text{rand}[-0.25,0.25], \end{cases} \quad i=1,2,...,100,$$

where rand $[a_1,a_2]$ denotes a random number independently drawn from the uniform distribution on interval $[a_1,a_2]$ for each i. The observed fuzzy outputs are assumed to be symmetric triangular fuzzy numbers and they can be expressed with our notations as

$$y_i=(l_{y_i},m_{y_i},r_{y_i})_{T(R)}=(y_i-\sigma_i,y_i,y_i+\sigma_i)_{T(R)}, i=1,2,...,100.$$

$CV(h)$ and $BIAS(h)$ are used to numerically evaluate the performance of each method and the related results are summarized in Table 1. In order to make a graphical comparison, we showed in Figure 1 the center line, the lower and the upper limit lines of the real regression function and the corresponding lines estimated by the proposed method only with Gaussian kernel.

Method	Kernel	Smoothing Parameter	Ridge Parameter	CV	$BIAS$
The proposed method	Gauss	0.19	0.001	0.0039	0.0041
		0.51	0.001	0.6247	0.6261
		0.19	0.01	0.04	0.0435
		0.51	0.01	0.33	0.3351
		0.19	0.03	0.59	0.6271
		0.51	0.03	0.0071	0.0124
	(the best value)	0.28	0.01	2.4720e-5	5.7625e-6
	Epanechnikov	1.2	0.001	0.7936	0.7946
		0.4	0.001	0.0015	0.0016
		1.2	0.01	0.4372	0.4466
		0.4	0.01	0.066	0.0656
		1.2	0.03	0.023	0.0317
		0.4	0.03	0.74	0.7804
LLS	Gauss	0.51	-	0.2544	0.0317
	Epanechnikov	1.2	-	0.2607	0.0328
KS	Gauss	0.19	-	0.2212	0.0628
	Epanechnikov	0.4	-	0.2302	0.0606

TABLE 1. The Simulation Results Obtained by Different Methods

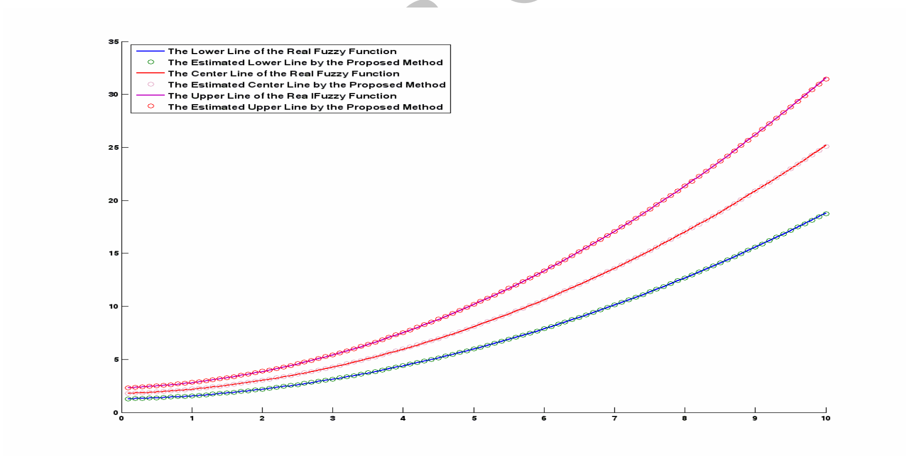


FIGURE 1. The Center Line, Lower and the Upper Limit Lines of the Real Function and Those Estimated by the Proposed Method with Gaussian kernel with Smoothing Parameter $h_0 = 0.28$ and Ridge Parameter$\lambda = 0.01$

Example 5.2. Consider the below function

$$g_2(x) = 10 + 5\sin(0.25\pi(1 - x^2))$$

and the same values of $x_i (i = 1, 2, ..., 100)$ as those in example 1. Let

$$\begin{cases} y_i = g_2(x_i) + \text{rand}[-0.5, 0.5], \\ \sigma_i = \frac{1}{3} g_2(x_i) + \text{rand}[-0.25, 0.25], \end{cases} \quad i = 1, 2, ..., 100,$$

The observed fuzzy outputs are

$$y_i = (l_{y_i}, m_{y_i}, r_{y_i})_{T(R)} = (y_i - \sigma_i, y_i, y_i + \sigma_i)_{T(R)}, i = 1, 2, ..., 100.$$

Table 2 shows the result for this dataset. Also the center line, the lower and the upper limit lines of the real regression function and the corresponding lines estimated by the proposed method only with Gaussian kernel are illustrated in Figure 2.

Method	Kernel	Smoothing Parameter	Ridge Parameter	CV	$BIAS$
The proposed method	Gauss	0.15	0.001	0.0031	0.0032
		0.21	0.001	0.0318	0.0321
		0.15	0.01	0.0162	0.0178
		0.21	0.01	2.8874e-4	3.2981e-4
		0.15	0.03	0.2692	0.2944
		0.21	0.03	0.1758	0.1982
	(the best value)	0.14	0.003	2.2701e-5	4.3991e-5
	Epanechnikov	0.34	0.001	0.0047	0.0048
		0.52	0.001	0.0597	0.0601
		0.34	0.01	0.013	0.0148
		0.52	0.01	0.0012	0.0016
		0.34	0.03	0.2556	0.2808
		0.52	0.03	0.1407	0.1598
LLS	Gauss	0.21	-	0.2252	0.0520
	Epanechnikov	0.52	-	0.2544	0.0552
KS	Gauss	0.15	-	0.2080	0.0821
	Epanechnikov	0.34	-	0.2337	0.0848

Table 2. The Simulation Results Obtained by Different Methods

Example 5.3. Consider the below function

$$\begin{cases} m(x_1, x_2) = 5, \\ l(x_1, x_2) = \frac{4}{625}(25 - (5 - x_1)^2)(25 - (5 - x_2)^2), \\ r(x_1, x_2) = 10 - l(x_1, x_2). \end{cases}$$

The domain of $x = (x_1, x_2)$ is $D = [0, 10] \times [0, 10]$ and the values of both x_1 and x_2 are equidistantly taken from 0 to 10 with increment 0.5. All possible combinations of the values of x_1 and x_2, which form the lattice points of size n = 441, are taken

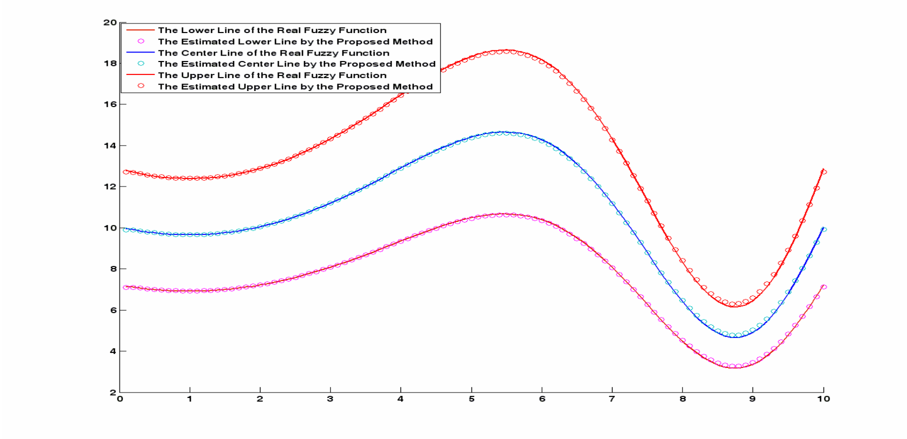


Figure 2. The Center Line, Lower and Upper Limit Lines of the Real Function and Those Estimated by the Proposed Method with Gaussian Kernel with Smoothing Parameter $h_0 = 0.14$ and Ridge Parameter $\lambda = 0.003$

as the crisp inputs of the independent variables x_1 and x_2. These lattice points are ordered in such a way that their Cartesian coordinates can be expressed as

$$(x_{i_1}, x_{i_2}) = (0.5mod(i-1, 21), 0.5int(i-1, 21)), \qquad i = 1, 2, ..., n,$$

where $mod(a, b)$ and $int(a, b)$ are respectively, the remainder and the integer part of a divided by b. Let

$$\begin{cases} y_i = m(x_{i_1}, x_{i_2}) + \varepsilon_i, \\ \sigma_i = m(x_{i_1}, x_{i_2}) - l(x_{i_1}, x_{i_2}) + \eta_i, \qquad i = 1, 2, ..., n, \end{cases}$$

where $\varepsilon_1, \varepsilon_2, ..., \varepsilon_n$ and $\eta_1, \eta_2, ..., \eta_n$, the observation errors of the centers and spreads of the observed fuzzy outputs, are independently drawn from the normal distributions $N(0, 0.5^2)$ and $N(0, 0.25^2)$, respectively.

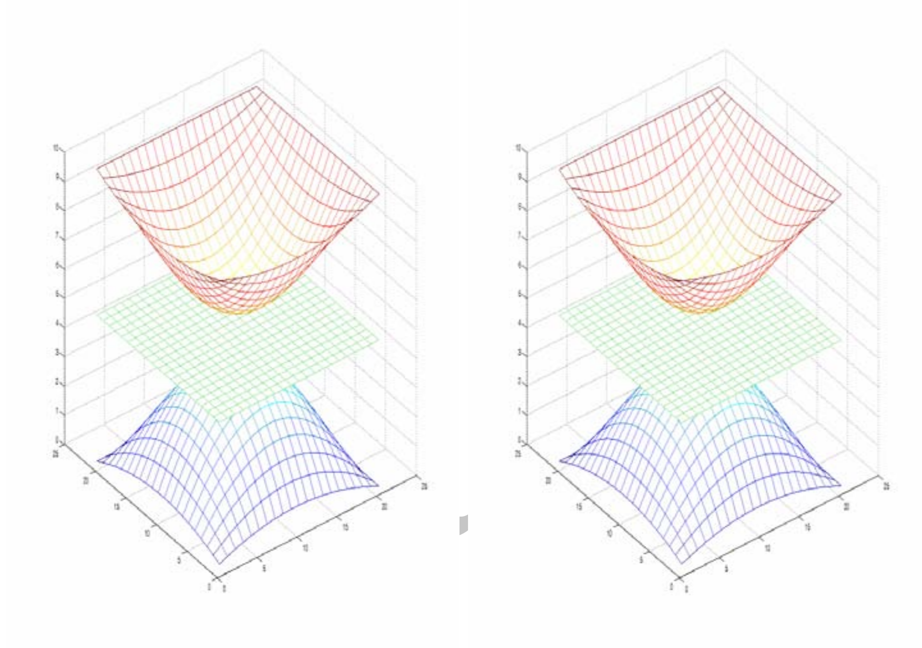
The observed fuzzy outputs are assumed to be symmetric triangular fuzzy numbers generated by

$$Y_i = (l_{y_i}, m_{y_i}, r_{y_i})_{T(R)} = (y_i - \sigma_i, y_i, y_i + \sigma_i)_{T(R)}, \qquad i = 1, 2, ..., n.$$

With the dataset $(x_{i_1}, x_{i_2}, Y_i), i = 1, 2, .., n$, the extended procedure was used to obtain the estimate of the fuzzy regression function $F(x_1, x_2)$. Table 3 shows the result for this dataset. The estimated fuzzy ridge nonparametric regression function was depicted in Figure 3(b). We observe from Figure 3 that the ridge method still produces a quite satisfactory estimate of the underlying fuzzy regression function in the case of two-dimensional input.

Method	Kernel	Smoothing Parameter	Ridge Parameter	CV	$BIAS$
The proposed method	Gauss	0.70	0.001	0.0292	0.0012
		0.55	0.001	0.0063	0.0013
LLS	Gauss	0.70	-	0.736	0.0654
KS	Gauss	0.55	-	0.7007	0.1037

TABLE 3. The Simulation Results Obtained by Different Methods



(a) (b)

FIGURE 3. (a): The Center Line, Lower and Upper Limit Lines of the Real Function; (b): The Estimation by the Proposed Method with Gaussian Kernel with Smoothing Parameter $h_0 = 0.55$ and Ridge Parameter $\lambda = 0.001$ (b)

6. Conclusions

In this study, we dealt with estimating ridge fuzzy nonparametric regression model with modeling the data with multivariate crisp input and triangular fuzzy output. The ridge estimation of fuzzy nonparametric regression models with the cross-validation procedure for selecting the optimal values of the smoothing and ridge parameter was proposed. Some simulation experiments were conducted to assess the performance of the proposed method. By comparing the results with those obtained by the LLS and KS methods, we found that the proposed method performs quite well in reducing the boundary effect and producing a satisfactory estimate of the underlying regression function. One of the advantages of this method is that we do not need to consider the underlying structure for the fuzzy nonparametric regression model used in this paper.

Moreover, we see that, for the three methods, the influence of the kernel is not significant because of the similar values of $CV(h)$ and $BIAS(h)$ for two kernels, which coincides with the empirical finding in the statistical nonparametric regression (see, for example [6]). The proposed method not only gives quite smooth estimates of the center line, the lower and the upper limit lines of the real regression function but also reduce the boundary effect significantly.

In future work, we intend to devise algorithms for estimating fuzzy nonparametric regression model with multiple fuzzy inputs, fuzzy output and the Gaussian fuzzy numbers. We observe from Tables 1, 2 and 3 that, in each case, the values of $CV(h)$ and $BIAS(h)$ in the proposed method are always smaller than that in the LLS and KS methods, which indicates that the proposed method tends to produce such estimates of the center, the lower and the upper limits of the fuzzy ridge nonparametric regression that are more close to their respective observations and gives less biased estimates of the center, the lower and the upper limits of the real function.

Acknowledgements. The authors would like to thank the editor and referees for constructive and helpful comments that improved the quality of the paper considerably.

References

[1] C. B. Cheng and E. S. Lee, *Fuzzy regression with radial basis function networks*, Fuzzy Sets and Systems, **119** (2001), 291-301.

[2] P. Diamond, *Fuzzy least squares*, Information Sciences, **46** (1988), 141-157.

[3] N. R. Draper and H. Smith, Applied Regression Analysis, Wiley, New York, 1980.

[4] H. Drucker, C. Burges, L. Kaufman, A. Smola and V. N. Vapnik, *Support vector regression machines, in: M. C. Mozer, M. I. Jordan, T. Petsche, Eds.,* Advances in Neural Information Processing Systems, MIT Press, Cambridge, MA, **9** (1996), 155-162.

[5] O. S. Fard and A. V. Kamyad, *Modified k-step method for solving fuzzy initial value problems*, Iranian Journal of Fuzzy Systems, **8(1)** (2011), 49-63.

[6] J. Fan and I. Gijbels, Local Polynomial Modeling and Its Applications, Chapman & Hall, London, 1996.

[7] W. Hardle, Applied Nonparametric Regression, Cambridge University Press, New York, 1990.

[8] J. D. Hart, Nonparametric Smoothing and Lack-of-fit Tests, Springer-Verlag, New York, 1997.

[9] T. J. Hastie and R. J. Tibshirani, Generalized Additive Models, Chapman & Hall, London, 1990.

[10] D. H. Hong and C. Hwang, *Support vector fuzzy regression machines,* Fuzzy Sets and Systems, **138** (2003), 271-281.

[11] D. H. Hong, C. Hwang and C. Ahn, *Ridge estimation for regression models with crisp inputs and Gaussian fuzzy output,* Fuzzy Sets and Systems, **142** (2004), 307-319.

[12] A. E. Hoerl and R. W. Kennard, *Ridge regression: biased estimates for nonorthogonal problems,* Technometrics, **12** (1970), 55-67.
[13] H. Ishibuchi and H. Tanaka, *Fuzzy regression analysis using neural networks*, Fuzzy Sets and Systems, **50** (1992), 257-265.
[14] H. Ishibuchi and H. Tanaka, *Fuzzy neural networks with interval weights and its application to fuzzy regression analysis,* Fuzzy Sets and Systems, **57** (1993), 27-39.
[15] B. Kim and R. R. Bishu, *Evaluation of fuzzy linear regression models by comparing membership functions,* Fuzzy Sets and Systems, **100** (1998), 343-352.
[16] R. X. Liu, J. Kuang, Q. Gong and X. L. Hou, *Principal component regression analysis with SPSS,* Computer Methods and Programs in Biomedicine, **71** (2003), 141-147.
[17] S. Pourahmad, S. M. T. Ayatollahi and S. M. Taheri, *Fuzzy logistic regression: a new possibilistic model and its application in clinical vague status*, Iranian Journal of Fuzzy Systems, **8** (2011), 1-17.
[18] H. Shakouri G and R. Nadimi, *A novel fuzzy linear regression model based on a non-equality possibility index and optimum uncertainty,* Applied Soft Computing, **9** (2009), 590-598.
[19] C. Saunders, A. Gammerman and V. Vork, *Ridge regression learning algorithm in dual variable,* Proceedings of the 15th International Conference on Machine Learning, (1998), 515-521.
[20] N. Wang, W. X. Zhang and C. L. Mei, *Fuzzy nonparametric regression based on local linear smoothing technique*, Information Sciences, **177** (2007), 3882-3900.
[21] M. S. Yang and C. H. Ko, *On a class of fuzzy c-numbers clustering procedures for fuzzy data*, Fuzzy Sets and Systems, **84** (1996), 49-60.

Rahman Farnoosh*, School of Mathematics, Iran University of Science and Technology, Narmak, Tehran-16846, Iran
E-mail address: `rfarnoosh@iust.ac.ir`

Javad Ghasemian, School of Mathematics, Iran University of Science and Technology, Narmak, Tehran-16846, Iran
E-mail address: `jghasemian@iust.ac.ir, jghasemian@gmail.com`

Omid Solaymani Fard, School of Mathematics and Computer Science, Damghan University, Damghan, Iran
E-mail address: `osfard@du.ac.ir, omidsfard@gmail.com`

*Corresponding author