Available online at www.sciencedirect.com
ScienceDirect
Fuzzy Sets and Systems 158 (2007) 2189 – 2202
FUZZY sets and systems
www.elsevier.com/locate/fss

Dependency between degree of fit and input noise in fuzzy linear regression using non-symmetric fuzzy triangular coefficients

Hong-Wei Ge*, Shi-Tong Wang

School of Information, Southern Yangtze University, Wuxi, Jiangshu, People's Republic of China

Received 2 May 2006; received in revised form 31 January 2007; accepted 14 March 2007
Available online 23 March 2007

Abstract

Fuzzy linear regression (FLR) model can be thought of as a fuzzy variation of classical linear regression model. It has been widely studied and applied in diverse fields. When noise exists in data, it is a very meaningful topic to reveal the dependency between the parameter h (i.e. the threshold value used to measure degree of fit) in FLR model and the input noise. In this paper, the FLR model is first extended to its regularized version, i.e. regularized fuzzy linear regression (RFLR) model, so as to enhance its generalization capability; then RFLR model is explained as the corresponding equivalent *maximum a posteriori* (MAP) problem; finally, the general dependency relationship that the parameter h with noisy input should follow is derived. Particular attention is paid to the regression model using non-symmetric fuzzy triangular coefficients. It turns out that with the existence of typical Gaussian noisy input, the parameter h is inversely proportional to the input noise. Our experimental results here also confirm this theoretical claim. Obviously, this theoretical result will be helpful to make a good choice for the parameter h, and to apply FLR techniques effectively in practical applications.

Keywords: Fuzzy linear regression model; Fuzzy regression analysis; Possibility theory; MAP; Optimal parameter choice

1. Introduction

Fuzzy linear regression (FLR) provides a means for tackling regression problems lacked of a significant amount of data to determine regression models and with vague relationships between the dependent variable and independent variables. Since the concept of FLR was first introduced by Tanaka et al. [15], the literature dealing with FLR has grown rapidly. For example, a modified version of Tanaka's fuzzy regression model was given in [8], where fuzzy regression for fuzzy input–output data was considered. In [5], fuzzy linear programming was introduced into the modified Tanaka's model. The important properties of fuzzy regression have been studied in [9,10,6]. More variants of Tanaka's model can be seen in [2,7,11–14]. Especially in [2], the support vector technique is introduced into fuzzy regression analysis to enhance its generalization capability. However, how to choose the parameter h (i.e. the threshold value used to measure degree of fit) for FLR model with noisy input still keeps an open problem. In practice, the input data often contain noise. With the existence of noisy input, one interesting and challenging issue is how to determine the parameter h in

* Corresponding author. Tel.: +86 510 88987789.
E-mail address: ghw8601@163.com (H.-W. Ge).

0165-0114/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.fss.2007.03.011

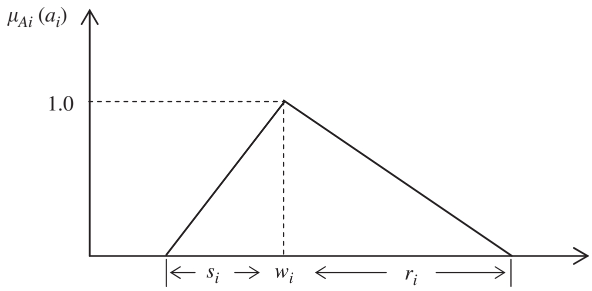

Fig. 1. Non-symmetric triangular fuzzy coefficients.

FLR model. A bad choice for h will heavily deteriorate the performance of FLR model. Therefore, in this paper, we will pay attention to deriving the dependency between the threshold h and the input noise.

In recent years, how to choose the loss function and the corresponding parameters for support vector regression (SVR) machine with noisy input has been studied well. Gao and Gunn pointed out that SVR problem could be transformed into the equivalent maximum a posteriori (MAP) problem [1]. Based on this idea, Kwok derived the linear dependency between ε and the input noise in ε-SVR [3], and Wang et al. investigated the theoretically optimal parameter choices for Huber-SVR with the Huber loss functions and r-SVR with the norm-r loss functions [16]. In this paper, based on the same idea, we first extend FLR model using symmetric and non-symmetric fuzzy triangular coefficients to its regularized version, and then explain this regularized model by using MAP framework [1,3,16], and finally study the dependency between h and the input noise. The rest of this paper is organized as follows. In Section 2, we will introduce the FLR model and the corresponding regularized fuzzy linear regression (RFLR) model and show the equivalent relationship between RFLR model and MAP. The analysis of the inverse linear dependency between h and the standard deviation of Gaussian noisy input is displayed in Section 3, while in Sections 4 and 5 the experimental results and some concluding remarks are provided, respectively.

2. FLR model and MAP

Fuzzy regression analysis using fuzzy linear models with symmetric triangular fuzzy number coefficients has been formulated earlier. Yen et al. extended the results of an FLR model that uses symmetric triangular coefficients to one with non-symmetric fuzzy triangular coefficients successfully [17]. The need for non-symmetric triangular coefficients arises due to the fact that during the regression using symmetric coefficients, the obtained regression line may not be the best-fitting line. This occurs because of the existence of large number of outliers and higher values of residuals. There are data sets that generate scatter plots in which the data do not fall symmetrically on both sides of the regression line.

2.1. FLR model

To introduce the nomenclature, the FLR technique [17] is summarized in the following.

Consider the function

$$Y = f(\mathbf{x}, \mathbf{A}) = A_0 + A_1x_1 + A_2x_2 + \cdots + A_nx_n, \tag{1}$$

where $\mathbf{x} = (1, x_1, x_2, \ldots, x_n)^{\mathrm{T}}$ is a vector of non-fuzzy inputs, and $\mathbf{A} = (A_0, A_1, \ldots, A_n)$ is a vector of fuzzy model parameters. Parameters A_i $(0 \leqslant i \leqslant n)$ are non-symmetric triangular fuzzy coefficients, and they can be described by the triplets $\{s_i, w_i, r_i\}$, where w_i is the point at which $\mu_{A_i}(a_i) = 1$, s_i is the left-side spread from the peak point w_i, and r_i represents the right-side spread as shown in Fig. 1.

Another representation is also possible, if we normalize the spreads. We can use either spread as the base to normalize the other one. Let us choose s_i as the base, then r_i can be expressed as $r_i = k_is_i$, where k_i are the skew factors and are positive real numbers. The selection of the values for k_i will be based on the knowledge of the problem and data characteristics. Then A_i can be described by the triplets $\{s_i, w_i, k_is_i\}$. If the values for k_i $(0 \leqslant i \leqslant n)$ are all selected to be 1, then A_i $(0 \leqslant i \leqslant n)$ become symmetric triangular fuzzy coefficients. In the rest of this paper, more attention will be paid to the FLR model using non-symmetric triangular fuzzy coefficients.

The membership functions for each A_i have the form

$$\mu_{A_i}(a_i) = \begin{cases} 1 - \dfrac{a_i - w_i}{r_i}, & w_i \leqslant a_i \leqslant w_i + r_i, \\ 1 - \dfrac{w_i - a_i}{s_i}, & w_i - s_i \leqslant a_i \leqslant w_i, \\ 0 & \text{otherwise.} \end{cases}$$

In vector notation, the fuzzy parameters $\mathbf{A}$ can be written as $\mathbf{A} = (\mathbf{s}, \mathbf{w}, \mathbf{r})$ where $\mathbf{s} = (s_0, s_1, \ldots, s_n)^{\mathrm{T}}$, $\mathbf{w} = (w_0, w_1, \ldots, w_n)^{\mathrm{T}}$, and $\mathbf{r} = (r_0, r_1, \ldots, r_n)^{\mathrm{T}}$. We assume $\mathbf{x}$ is non-negative throughout this paper. Following the principle of extension, the fuzzy membership function for the output can be obtained by

$$\mu_Y(y) = \begin{cases} 1 - \dfrac{y - \mathbf{w}^{\mathrm{T}}\mathbf{x}}{\mathbf{r}^{\mathrm{T}}|\mathbf{x}|}, & \mathbf{w}^{\mathrm{T}}\mathbf{x} \leqslant y \leqslant \mathbf{w}^{\mathrm{T}}\mathbf{x} + \mathbf{r}^{\mathrm{T}}|\mathbf{x}|, \\ 1 - \dfrac{\mathbf{w}^{\mathrm{T}}\mathbf{x} - y}{\mathbf{s}^{\mathrm{T}}|\mathbf{x}|}, & \mathbf{w}^{\mathrm{T}}\mathbf{x} - \mathbf{s}^{\mathrm{T}}|\mathbf{x}| \leqslant y \leqslant \mathbf{w}^{\mathrm{T}}\mathbf{x}, \\ 0 & \text{otherwise,} \end{cases}$$

where $|\mathbf{x}| = (1, |x_1|, |x_2|, \ldots, |x_n|)^{\mathrm{T}}$, $\mathbf{w}^{\mathrm{T}}\mathbf{x}$ is the point at which $\mu_Y(y) = 1$, the right-side spread of Y is $\mathbf{r}^{\mathrm{T}}|\mathbf{x}|(\mathbf{r}^{\mathrm{T}}|\mathbf{x}| \geqslant 0)$ and the left-side spread is $\mathbf{s}^{\mathrm{T}}|\mathbf{x}|(\mathbf{s}^{\mathrm{T}}|\mathbf{x}| \geqslant 0)$ (refer to [17]).

Assume we have a data set with n-dimensional non-fuzzy input $\mathbf{x}$ and one-dimensional non-fuzzy output variable y:

$$D = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_N, y_N)\}, \quad \mathbf{x}_i \in R^{n+1}, \quad y_i \in R, \quad i = 1, 2, \ldots, N. \tag{2}$$

The following FLR model can be established:

$$Y_i^* = \mathbf{A}^{*^{\mathrm{T}}}\mathbf{x}_i,$$

where $\mathbf{x}_i = (1, x_{i1}, x_{i2}, \ldots, x_{in})^{\mathrm{T}}$, $\mathbf{A}^* = (A_0^*, A_1^*, A_2^*, \ldots, A_n^*)^{\mathrm{T}}$, $i = 1, 2, \ldots, N$, and Y_i^* is the fuzzy estimate of Y_i. The objective is to minimize the fuzziness of the fuzzy linear model and the following linear programming problem can be formulated:

$$\Phi(\mathbf{s}, \mathbf{r}, \mathbf{w}, \Delta) = \min \quad \sum_i \mathbf{s}^{\mathrm{T}}|\mathbf{x}_i| + \sum_i \mathbf{r}^{\mathrm{T}}|\mathbf{x}_i| + C\sum_i (\Delta_i^- + \Delta_i^+),$$

$$\text{s.t.} \begin{cases} \dfrac{y_i - \mathbf{w}^{\mathrm{T}}\mathbf{x}_i}{\mathbf{r}^{\mathrm{T}}|\mathbf{x}_i|} \leqslant (1-h) + \Delta_i^-, \\ \dfrac{\mathbf{w}^{\mathrm{T}}\mathbf{x}_i - y_i}{\mathbf{s}^{\mathrm{T}}|\mathbf{x}_i|} \leqslant (1-h) + \Delta_i^+, \\ \Delta_i^-, \Delta_i^+ \geqslant 0, \\ i = 1, 2, \ldots, N, \end{cases} \tag{3}$$

where C is a predefined constant. Δ_i^- and Δ_i^+ denote the latent variables of the upper/lower bounds of output, respectively. $h(h \in [0, 1))$ is the threshold of fitness, $\mathbf{r}^{\mathrm{T}}|\mathbf{x}_i|$ is the right-side spread of Y_i^*, and $\mathbf{s}^{\mathrm{T}}|\mathbf{x}_i|$ is the left-side spread.

In order to enhance the generalization capability of FLR model and to be easy to analyze, just like SVR techniques in [16,4,2], a regularized term $\frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w}$ can be introduced into Eq. (3), i.e.

$$\Phi(\mathbf{s}, \mathbf{r}, \mathbf{w}, \Delta) = \min \sum_i \mathbf{s}^{\mathrm{T}}|\mathbf{x}_i| + \sum_i \mathbf{r}^{\mathrm{T}}|\mathbf{x}_i| + \frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} + C\sum_i (\Delta_i^- + \Delta_i^+). \tag{4}$$

Eq. (4) formulates the regularized version of FLR model, called RFLR model here. In the rest of this paper, how to choose the parameter h for RFLR model with noisy input will be studied. Obviously, the conclusion will also be very helpful for the practical applications of FLR model.

2.2. The regression model and MAP

In this subsection, in terms of the evidence theory [1,3], we will demonstrate that the RFLR model is equivalent to MAP estimation based on maximum likelihood estimation.

Now, the focus is on obtaining a weight vector $\mathbf{w}$ in the generalized linear regression model such as

$$y_i = \mathbf{w}^{\mathrm{T}}\mathbf{x}_i + n_i, \quad i = 1, 2, \ldots, N$$

for all the data in the data set (2), where all the data $\mathbf{x}_i$ follow distribution $p(\cdot)$ and all n_i are i.i.d noise following some distribution $\phi(\cdot)$. Thus, the corresponding density function on y can be denoted as $p(y|\mathbf{x}) = \phi(y - \tilde{\mathbf{w}}^{\mathrm{T}}\mathbf{x})$, where $\tilde{\mathbf{w}}$ is a weight vector and $\tilde{\mathbf{w}}^{\mathrm{T}}\mathbf{x}$ represents the mathematical expectation. The degree of such an approximation can be measured by a loss function $L(\mathbf{s}^{\mathrm{T}}\mathbf{x}, \mathbf{r}^{\mathrm{T}}\mathbf{x}, \mathbf{w}^{\mathrm{T}}\mathbf{x}, y)$, i.e.

$$L(\mathbf{s}^{\mathrm{T}}\mathbf{x}, \mathbf{r}^{\mathrm{T}}\mathbf{x}, \mathbf{w}^{\mathrm{T}}\mathbf{x}, y) = \begin{cases} 0, & \dfrac{y - \mathbf{w}^{\mathrm{T}}\mathbf{x}}{\mathbf{r}^{\mathrm{T}}|\mathbf{x}|} \leqslant 1 - h, \\ \dfrac{y - \mathbf{w}^{\mathrm{T}}\mathbf{x}}{\mathbf{r}^{\mathrm{T}}|\mathbf{x}|} - 1 + h, & \dfrac{y - \mathbf{w}^{\mathrm{T}}\mathbf{x}}{\mathbf{r}^{\mathrm{T}}|\mathbf{x}|} > 1 - h, \\ 0, & \dfrac{\mathbf{w}^{\mathrm{T}}\mathbf{x} - y}{\mathbf{s}^{\mathrm{T}}|\mathbf{x}|} \leqslant 1 - h, \\ \dfrac{\mathbf{w}^{\mathrm{T}}\mathbf{x} - y}{\mathbf{s}^{\mathrm{T}}|\mathbf{x}|} - 1 + h, & \dfrac{\mathbf{w}^{\mathrm{T}}\mathbf{x} - y}{\mathbf{s}^{\mathrm{T}}|\mathbf{x}|} > 1 - h. \end{cases}$$

Assume the loss function $L(\mathbf{s}^{\mathrm{T}}\mathbf{x}, \mathbf{r}^{\mathrm{T}}\mathbf{x}, \mathbf{w}^{\mathrm{T}}\mathbf{x}, y)$ leads to the following Gaussian probability density function on y:

$$p(y_i|\mathbf{x}_i, \mathbf{s}, \mathbf{r}, \mathbf{w}, \beta, h) = \frac{1}{c(\beta, h)} \exp[-\beta L(\mathbf{s}^{\mathrm{T}}\mathbf{x}_i, \mathbf{r}^{\mathrm{T}}\mathbf{x}_i, \mathbf{w}^{\mathrm{T}}\mathbf{x}_i, y_i)],$$

where $c(\beta, h) = \int\int_D \exp[-\beta L(\mathbf{s}^{\mathrm{T}}\mathbf{x}, \mathbf{r}^{\mathrm{T}}\mathbf{x}, \mathbf{w}^{\mathrm{T}}\mathbf{x}, y)]\,\mathrm{d}\mathbf{x}\,\mathrm{d}y$. Just as the mathematical analysis in [1,3] does, we adopt the Gaussian prior in the following analytical framework. With the Gaussian prior on $\mathbf{s}$, $\mathbf{r}$, and $\mathbf{w}$,

$$p(\mathbf{s}|\alpha, \mathbf{x}_i) = \frac{1}{M(\alpha)} \exp(-\alpha \mathbf{s}^{\mathrm{T}}|\mathbf{x}_i|),$$

where

$$M(\alpha) = \int \exp(-\alpha \mathbf{s}^{\mathrm{T}}|\mathbf{x}_i|)\,\mathrm{d}\mathbf{s},$$

$$p(\mathbf{r}|\eta, \mathbf{x}_i) = \frac{1}{N(\eta)} \exp(-\eta \mathbf{r}^{\mathrm{T}}|\mathbf{x}_i|),$$

where

$$N(\eta) = \int \exp(-\eta \mathbf{r}^{\mathrm{T}}|\mathbf{x}_i|)\,\mathrm{d}\mathbf{r},$$

$$p(\mathbf{w}|\gamma) = \frac{1}{F(\gamma)} \exp\left(-\frac{\gamma}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w}\right),$$

where

$$F(\gamma) = \int \exp\left(-\frac{\gamma}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w}\right)\mathrm{d}\mathbf{w},$$

and by applying the Bayes rule:

$$p(\mathbf{s}, \mathbf{r}, \mathbf{w}|D, \beta, h) \propto p(D|\mathbf{s}, \mathbf{r}, \mathbf{w}, \beta, h)\, p(\mathbf{s}|\alpha)\, p(\mathbf{r}|\eta)\, p(\mathbf{w}|\gamma),$$

i.e.

$$p(\mathbf{s},\mathbf{r},\mathbf{w}|D,\beta,h) \propto p(D|\mathbf{s},\mathbf{r},\mathbf{w},\beta,h)\left[\prod_{i=1}^{N} p(\mathbf{s}|\alpha,\mathbf{x}_i)p(\mathbf{x}_i)\right]\left[\prod_{i=1}^{N} p(\mathbf{r}|\eta,\mathbf{x}_i)p(\mathbf{x}_i)\right]p(\mathbf{w}|\gamma),$$

we have

$$\begin{aligned}&\ln p(\mathbf{s},\mathbf{r},\mathbf{w}|D,\beta,h)\\ &\quad= -\alpha\sum_{i=1}^{N}\mathbf{s}^{\mathrm{T}}|\mathbf{x}_i| - \eta\sum_{i=1}^{N}\mathbf{r}^{\mathrm{T}}|\mathbf{x}_i| - \frac{\gamma}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} - \beta\sum_{i=1}^{N}L(\mathbf{s}^{\mathrm{T}}\mathbf{x}_i,\mathbf{r}^{\mathrm{T}}\mathbf{x}_i,\mathbf{w}^{\mathrm{T}}\mathbf{x}_i,y_i) - N\ln C(\beta,h) + const.\end{aligned} \tag{5}$$

With setting $\gamma=\alpha$, $\eta=\alpha$, and $C=\beta/\alpha$, optimizing (4) can be interpreted as finding for the MAP estimate of $\mathbf{s}$, $\mathbf{r}$ and $\mathbf{w}$ at given values of β, h. That is to say, the RFLR model in (4) is equivalent to MAP estimation.

Generally speaking, it is not easy for us to get the MAP estimate $\hat{\mathbf{s}}$, $\hat{\mathbf{r}}$ and $\hat{\mathbf{w}}$ by directly solving (5), since it depends on the particular training set. In order to make the analysis easier and clearer, $(1/N)\sum_{i=1}^{N}L(\mathbf{s}^{\mathrm{T}}\mathbf{x}_i,\mathbf{r}^{\mathrm{T}}\mathbf{x}_i,\mathbf{w}^{\mathrm{T}}\mathbf{x}_i,y_i)$ in Eq. (5) is replaced by its expectation

$$\begin{aligned}E(L(\mathbf{s}^{\mathrm{T}}\mathbf{x},\mathbf{r}^{\mathrm{T}}\mathbf{x},\mathbf{w}^{\mathrm{T}}\mathbf{x},y)) &= \int\int_D L(\mathbf{s}^{\mathrm{T}}\mathbf{x},\mathbf{r}^{\mathrm{T}}\mathbf{x},\mathbf{w}^{\mathrm{T}}\mathbf{x},y)p(y|\mathbf{x})p(\mathbf{x})\,\mathrm{d}\mathbf{x}\,\mathrm{d}y\\ &= \int_D\left[\int_{-\infty}^{\mathbf{w}^{\mathrm{T}}\mathbf{x}-(1-h)\mathbf{s}^{\mathrm{T}}|\mathbf{x}|}\left(\frac{\mathbf{w}^{\mathrm{T}}\mathbf{x}-y}{\mathbf{s}^{\mathrm{T}}|\mathbf{x}|}-1+h\right)p(y|\mathbf{x})\,\mathrm{d}y\right.\\ &\quad\left.+\int_{\mathbf{w}^{\mathrm{T}}\mathbf{x}+(1-h)\mathbf{r}^{\mathrm{T}}|\mathbf{x}|}^{+\infty}\left(\frac{y-\mathbf{w}^{\mathrm{T}}\mathbf{x}}{\mathbf{r}^{\mathrm{T}}|\mathbf{x}|}-1+h\right)p(y|\mathbf{x})\,\mathrm{d}y\right]p(\mathbf{x})\,\mathrm{d}\mathbf{x},\end{aligned} \tag{6}$$

and $(1/N)\sum_{i=1}^{N}\mathbf{s}^{\mathrm{T}}|\mathbf{x}_i|$, $(1/N)\sum_{i=1}^{N}\mathbf{r}^{\mathrm{T}}|\mathbf{x}_i|$ by their expectations

$$E(\mathbf{s}^{\mathrm{T}}|\mathbf{x}|) = \int_D \mathbf{s}^{\mathrm{T}}|\mathbf{x}|p(\mathbf{x})\,\mathrm{d}\mathbf{x},$$

$$E(\mathbf{r}^{\mathrm{T}}|\mathbf{x}|) = \int_D \mathbf{r}^{\mathrm{T}}|\mathbf{x}|p(\mathbf{x})\,\mathrm{d}\mathbf{x}.$$

Thus, Eq. (5) becomes

$$\begin{aligned}M(\mathbf{s},\mathbf{r},\mathbf{w},\beta,h) &= \ln p(\mathbf{s},\mathbf{r},\mathbf{w}|D,\beta,h)\\ &= -\alpha N * E(\mathbf{s}^{\mathrm{T}}|\mathbf{x}|) - \eta N * E(\mathbf{r}^{\mathrm{T}}|\mathbf{x}|) - \frac{\gamma}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w}\\ &\quad -\beta N * E(L(\mathbf{s}^{\mathrm{T}}\mathbf{x},\mathbf{r}^{\mathrm{T}}\mathbf{x},\mathbf{w}^{\mathrm{T}}\mathbf{x},y)) - N\ln C(\beta,h) + const.\end{aligned} \tag{7}$$

In order to maximize (7), its derivatives with respect to $\mathbf{s}$, $\mathbf{r}$, $\mathbf{w}$, β, h must be zero. That is to say,

$$\left.\frac{\partial M(\mathbf{s},\mathbf{r},\mathbf{w},\beta,h)}{\partial\mathbf{s}}\right|_{\mathbf{s}=\hat{\mathbf{s}}} = -\alpha N * E(|\mathbf{x}|) - \beta N * \partial E(L(\hat{\mathbf{s}}^{\mathrm{T}}\mathbf{x},\mathbf{r}^{\mathrm{T}}\mathbf{x},\mathbf{w}^{\mathrm{T}}\mathbf{x},y))/\partial\mathbf{s} = 0, \tag{8}$$

$$\left.\frac{\partial M(\mathbf{s},\mathbf{r},\mathbf{w},\beta,h)}{\partial\mathbf{r}}\right|_{\mathbf{r}=\hat{\mathbf{r}}} = -\eta N * E(|\mathbf{x}|) - \beta N * \partial E(L(\mathbf{s}^{\mathrm{T}}\mathbf{x},\hat{\mathbf{r}}^{\mathrm{T}}\mathbf{x},\mathbf{w}^{\mathrm{T}}\mathbf{x},y))/\partial\mathbf{r} = 0, \tag{9}$$

$$\left.\frac{\partial M(\mathbf{s},\mathbf{r},\mathbf{w},\beta,h)}{\partial\mathbf{w}}\right|_{\mathbf{w}=\hat{\mathbf{w}}} = -\gamma\hat{\mathbf{w}} - \beta N * \partial E(L(\mathbf{s}^{\mathrm{T}}\mathbf{x},\mathbf{r}^{\mathrm{T}}\mathbf{x},\hat{\mathbf{w}}^{\mathrm{T}}\mathbf{x},y))/\partial\mathbf{w} = 0, \tag{10}$$

$$\frac{\partial M(\mathbf{s},\mathbf{r},\mathbf{w},\beta,h)}{\partial \beta} = \left[\left.\frac{\partial M(\mathbf{s},\mathbf{r},\mathbf{w},\beta,h)}{\partial \mathbf{s}}\right|_{\mathbf{s}=\hat{\mathbf{s}}}\right]\frac{\partial \hat{\mathbf{s}}}{\partial \beta} + \left[\left.\frac{\partial M(\mathbf{s},\mathbf{r},\mathbf{w},\beta,h)}{\partial \mathbf{r}}\right|_{\mathbf{r}=\hat{\mathbf{r}}}\right]\frac{\partial \hat{\mathbf{r}}}{\partial \beta}$$
$$+ \left[\left.\frac{\partial M(\mathbf{s},\mathbf{r},\mathbf{w},\beta,h)}{\partial \mathbf{w}}\right|_{\mathbf{w}=\hat{\mathbf{w}}}\right]\frac{\partial \hat{\mathbf{w}}}{\partial \beta} - N * E(L(\hat{\mathbf{s}}^{\mathrm{T}}\mathbf{x},\hat{\mathbf{r}}^{\mathrm{T}}\mathbf{x},\hat{\mathbf{w}}^{\mathrm{T}}\mathbf{x},y)) - N\frac{\partial C(\beta,h)/\partial\beta}{C(\beta,h)} = 0,$$

i.e.

$$E(L(\hat{\mathbf{s}}^{\mathrm{T}}\mathbf{x},\hat{\mathbf{r}}^{\mathrm{T}}\mathbf{x},\hat{\mathbf{w}}^{\mathrm{T}}\mathbf{x},y)) = -\frac{\partial C(\beta,h)/\partial\beta}{C(\beta,h)}, \tag{11}$$

$$\frac{\partial M(\mathbf{s},\mathbf{r},\mathbf{w},\beta,h)}{\partial h} = \left[\left.\frac{\partial M(\mathbf{s},\mathbf{r},\mathbf{w},\beta,h)}{\partial \mathbf{s}}\right|_{\mathbf{s}=\hat{\mathbf{s}}}\right]\frac{\partial \hat{\mathbf{s}}}{\partial h} + \left[\left.\frac{\partial M(\mathbf{s},\mathbf{r},\mathbf{w},\beta,h)}{\partial \mathbf{r}}\right|_{\mathbf{r}=\hat{\mathbf{r}}}\right]\frac{\partial \hat{\mathbf{r}}}{\partial h}$$
$$+ \left[\left.\frac{\partial M(\mathbf{s},\mathbf{r},\mathbf{w},\beta,h)}{\partial \mathbf{w}}\right|_{\mathbf{w}=\hat{\mathbf{w}}}\right]\frac{\partial \hat{\mathbf{w}}}{\partial h} - N\frac{\partial C(\beta,h)/\partial h}{C(\beta,h)}$$
$$-\beta N * \int_D \left[\int_{-\infty}^{\hat{\mathbf{w}}^{\mathrm{T}}\mathbf{x}-(1-h)\hat{\mathbf{s}}^{\mathrm{T}}|\mathbf{x}|} p(y|\mathbf{x})\,\mathrm{d}y + \int_{\hat{\mathbf{w}}^{\mathrm{T}}\mathbf{x}+(1-h)\hat{\mathbf{r}}^{\mathrm{T}}|\mathbf{x}|}^{+\infty} p(y|\mathbf{x})\,\mathrm{d}y\right] p(\mathbf{x})\,\mathrm{d}\mathbf{x} = 0,$$

i.e.

$$\int_D \left[\int_{-\infty}^{\hat{\mathbf{w}}^{\mathrm{T}}\mathbf{x}-(1-h)\hat{\mathbf{s}}^{\mathrm{T}}|\mathbf{x}|} p(y|\mathbf{x})\,\mathrm{d}y + \int_{\hat{\mathbf{w}}^{\mathrm{T}}\mathbf{x}+(1-h)\hat{\mathbf{r}}^{\mathrm{T}}|\mathbf{x}|}^{+\infty} p(y|\mathbf{x})\,\mathrm{d}y\right] p(\mathbf{x})\,\mathrm{d}\mathbf{x} = \frac{-\partial C(\beta,h)/\partial h}{\beta C(\beta,h)}. \tag{12}$$

When $\mathbf{s}=\hat{\mathbf{s}}$, $\mathbf{r}=\hat{\mathbf{r}}$, and $\mathbf{w}=\hat{\mathbf{w}}$, maximizing (7) actually becomes the following optimization problem:

$$\arg\min_{\alpha,\eta,\beta,h} \alpha E(\hat{\mathbf{s}}^{\mathrm{T}}|\mathbf{x}|) + \eta E(\hat{\mathbf{r}}^{\mathrm{T}}|\mathbf{x}|) + \beta E(L(\hat{\mathbf{s}}^{\mathrm{T}}\mathbf{x},\hat{\mathbf{r}}^{\mathrm{T}}\mathbf{x},\hat{\mathbf{w}}^{\mathrm{T}}\mathbf{x},y)) + \ln C(\beta,h). \tag{13}$$

3. The dependency relationship between h and the Gaussian noisy input

Please note $C(\beta,h)$ in Eq. (13). According to its corresponding MAP, we have

$$C(\beta,h) = \int\int_D \exp[-\beta L(\mathbf{s}^{\mathrm{T}}\mathbf{x},\mathbf{r}^{\mathrm{T}}\mathbf{x},\mathbf{w}^{\mathrm{T}}\mathbf{x},y)]\,\mathrm{d}\mathbf{x}\,\mathrm{d}y$$
$$= 2\left(\int_0^{1-h}\exp(0)\,\mathrm{d}t + \int_{1-h}^{+\infty}\exp(-\beta(t-1+h))\,\mathrm{d}t\right)$$
$$= \frac{2(1+(1-h)\beta)}{\beta}. \tag{14}$$

By substituting (14) into (11), we obtain

$$E(L(\hat{\mathbf{s}}^{\mathrm{T}}\mathbf{x},\hat{\mathbf{r}}^{\mathrm{T}}\mathbf{x},\hat{\mathbf{w}}^{\mathrm{T}}\mathbf{x},y)) = \frac{1}{\beta(1+(1-h)\beta)}, \tag{15}$$

and by substituting (14) into (12), we obtain

$$\int_D \left[\int_{-\infty}^{\hat{\mathbf{w}}^{\mathrm{T}}\mathbf{x}-(1-h)\hat{\mathbf{s}}^{\mathrm{T}}|\mathbf{x}|} p(y|\mathbf{x})\,\mathrm{d}y + \int_{\hat{\mathbf{w}}^{\mathrm{T}}\mathbf{x}+(1-h)\hat{\mathbf{r}}^{\mathrm{T}}|\mathbf{x}|}^{+\infty} p(y|\mathbf{x})\,\mathrm{d}y\right] p(\mathbf{x})\,\mathrm{d}\mathbf{x} = \frac{1}{1+(1-h)\beta}. \tag{16}$$

White noise often happens in real situations. Quite often, the i.i.d Gaussian distribution with zero mean and variance σ is utilized in most robustness analyses. This assumption is also taken here.

Let

$$p(y|\mathbf{x}) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(y-\tilde{\mathbf{w}}^{\mathrm{T}}\mathbf{x})^2}{2\sigma^2}\right],$$

and let $t = (y - \tilde{\mathbf{w}}^{\mathrm{T}}\mathbf{x})/\sqrt{2}\sigma$, $y - \tilde{\mathbf{w}}^{\mathrm{T}}\mathbf{x} \approx \hat{\mathbf{w}}^{\mathrm{T}}\mathbf{x} - \tilde{\mathbf{w}}^{\mathrm{T}}\mathbf{x} = \delta(\mathbf{x})$, $\varepsilon = 1 - h$, $A(\mathbf{x}) = \hat{\mathbf{s}}^{\mathrm{T}}|\mathbf{x}|$,

$$E = \int_D A(\mathbf{x}) p(\mathbf{x})\,\mathrm{d}\mathbf{x}, \quad B(\mathbf{x}) = \hat{\mathbf{r}}^{\mathrm{T}}|\mathbf{x}|, \quad G = \int_D B(\mathbf{x}) p(\mathbf{x})\,\mathrm{d}\mathbf{x}, \quad b_1(\mathbf{x}) = \frac{1}{\sqrt{2}}\left(\frac{\varepsilon \cdot A(\mathbf{x})}{\sigma} - \frac{\delta(\mathbf{x})}{\sigma}\right),$$

$$b_2(\mathbf{x}) = \frac{1}{\sqrt{2}}\left(\frac{\varepsilon \cdot B(\mathbf{x})}{\sigma} + \frac{\delta(\mathbf{x})}{\sigma}\right), \quad f(x) = \frac{2}{\sqrt{\pi}}\int_{\mathbf{x}}^{+\infty} \exp(-t^2)\,\mathrm{d}t.$$

Please note that

$$\int_D p(\mathbf{x})\,\mathrm{d}\mathbf{x} = 1,$$

$$\int_D \delta(\mathbf{x}) p(\mathbf{x})\,\mathrm{d}\mathbf{x} = 0,$$

$$\int_D \delta^2(\mathbf{x}) p(\mathbf{x})\,\mathrm{d}\mathbf{x} = \sigma^2,$$

$$\int_D \delta^3(\mathbf{x}) p(\mathbf{x})\,\mathrm{d}\mathbf{x} \approx 2\sigma^3,$$

$$\exp\left[-\frac{\delta^2(\mathbf{x})}{2\sigma^2}\right] \approx 1 - \frac{\delta^2(\mathbf{x})}{2\sigma^2}.$$

It can be shown that Eq. (16) reduces to

$$\begin{aligned}
\frac{1}{1+\varepsilon\beta} &= \int_D \left[\int_{-\infty}^{\hat{\mathbf{w}}^{\mathrm{T}}\mathbf{x}-\varepsilon\cdot A(\mathbf{x})} p(y|\mathbf{x})\,\mathrm{d}y + \int_{\hat{\mathbf{w}}^{\mathrm{T}}\mathbf{x}+\varepsilon\cdot B(\mathbf{x})}^{+\infty} p(y|\mathbf{x})\,\mathrm{d}y\right] p(\mathbf{x})\,\mathrm{d}\mathbf{x} \\
&= \int_D \frac{1}{\sqrt{\pi}}\left[\int_{-\infty}^{(\delta(\mathbf{x})-\varepsilon\cdot A(\mathbf{x}))/\sqrt{2}\sigma} \exp(-t^2)\,\mathrm{d}t + \int_{(\delta(\mathbf{x})+\varepsilon\cdot B(\mathbf{x}))/\sqrt{2}\sigma}^{+\infty} \exp(-t^2)\,\mathrm{d}t\right] p(\mathbf{x})\,\mathrm{d}\mathbf{x} \\
&= \int_D \frac{f(b_1)+f(b_2)}{2} p(\mathbf{x})\,\mathrm{d}\mathbf{x}.
\end{aligned} \tag{17}$$

From Eqs. (6) and (15), we have

$$\begin{aligned}
\frac{1}{\beta(1+\varepsilon\beta)} &= \int_D \left[\int_{-\infty}^{\hat{\mathbf{w}}^{\mathrm{T}}\mathbf{x}-\varepsilon\cdot A(\mathbf{x})} \left(\frac{\hat{\mathbf{w}}^{\mathrm{T}}\mathbf{x}-y}{A(\mathbf{x})} - \varepsilon\right) p(y|\mathbf{x})\,\mathrm{d}y + \int_{\hat{\mathbf{w}}^{\mathrm{T}}\mathbf{x}+\varepsilon\cdot B(\mathbf{x})}^{+\infty} \left(\frac{y-\hat{\mathbf{w}}^{\mathrm{T}}\mathbf{x}}{B(\mathbf{x})} - \varepsilon\right) p(y|\mathbf{x})\,\mathrm{d}y\right] p(\mathbf{x})\,\mathrm{d}\mathbf{x} \\
&= \int_D \left[\frac{1}{A(\mathbf{x})}\int_{-\infty}^{\hat{\mathbf{w}}^{\mathrm{T}}\mathbf{x}-\varepsilon\cdot A(\mathbf{x})} (\hat{\mathbf{w}}^{\mathrm{T}}\mathbf{x}-y) p(y|\mathbf{x})\,\mathrm{d}y + \frac{1}{B(\mathbf{x})}\int_{\hat{\mathbf{w}}^{\mathrm{T}}\mathbf{x}+\varepsilon\cdot B(\mathbf{x})}^{+\infty} (y-\hat{\mathbf{w}}^{\mathrm{T}}\mathbf{x}) p(y|\mathbf{x})\,\mathrm{d}y\right. \\
&\quad \left. -\varepsilon\int_{-\infty}^{\hat{\mathbf{w}}^{\mathrm{T}}\mathbf{x}-\varepsilon\cdot A(\mathbf{x})} p(y|\mathbf{x})\,\mathrm{d}y - \varepsilon\int_{\hat{\mathbf{w}}^{\mathrm{T}}\mathbf{x}+\varepsilon\cdot B(\mathbf{x})}^{+\infty} p(y|\mathbf{x})\,\mathrm{d}y\right] p(\mathbf{x})\,\mathrm{d}\mathbf{x} \\
&= \int_D \left[\frac{1}{A(\mathbf{x})}\int_{-\infty}^{\hat{\mathbf{w}}^{\mathrm{T}}\mathbf{x}-\varepsilon\cdot A(\mathbf{x})} (\hat{\mathbf{w}}^{\mathrm{T}}\mathbf{x}-y) p(y|\mathbf{x})\,\mathrm{d}y + \frac{1}{B(\mathbf{x})}\int_{\hat{\mathbf{w}}^{\mathrm{T}}\mathbf{x}+\varepsilon\cdot B(\mathbf{x})}^{+\infty} (y-\hat{\mathbf{w}}^{\mathrm{T}}\mathbf{x}) p(y|\mathbf{x})\,\mathrm{d}y\right. \\
&\quad \left. -\frac{\varepsilon}{2} f(b_1) - \frac{\varepsilon}{2} f(b_2)\right] p(\mathbf{x})\,\mathrm{d}\mathbf{x} \\
&= \int_D \left[\frac{1}{\sqrt{\pi}\cdot A(\mathbf{x})}\int_{(\varepsilon\cdot A(\mathbf{x})-\delta(\mathbf{x}))/\sqrt{2}\sigma}^{+\infty} (\sqrt{2}\sigma t + \delta(\mathbf{x}))\exp(-t^2)\,\mathrm{d}t\right.
\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{\sqrt{\pi}\cdot B(\mathbf{x})}\int_{(\varepsilon\cdot B(\mathbf{x})+\delta(\mathbf{x}))/\sqrt{2}\sigma}^{+\infty}(\sqrt{2}\sigma t-\delta(\mathbf{x}))\exp(-t^2)\,\mathrm{d}t-\frac{\varepsilon}{2}f(b_1)-\frac{\varepsilon}{2}f(b_2)\Bigg]\,p(\mathbf{x})\,\mathrm{d}\mathbf{x}\\
=&\int_D\Bigg[-\frac{\sqrt{2}\sigma}{2\sqrt{\pi}\cdot A(\mathbf{x})}\int_{(\varepsilon\cdot A(\mathbf{x})-\delta(\mathbf{x}))/\sqrt{2}\sigma}^{+\infty}\exp(-t^2)\mathrm{d}(-t^2)\\
&-\frac{\sqrt{2}\sigma}{2\sqrt{\pi}\cdot B(\mathbf{x})}\int_{(\varepsilon\cdot B(\mathbf{x})+\delta(\mathbf{x}))/\sqrt{2}\sigma}^{+\infty}\exp(-t^2)\mathrm{d}(-t^2)\\
&+\frac{\delta(\mathbf{x})}{\sqrt{\pi}\cdot A(\mathbf{x})}\int_{(\varepsilon\cdot A(\mathbf{x})-\delta(\mathbf{x}))/\sqrt{2}\sigma}^{+\infty}\exp(-t^2)\,\mathrm{d}t-\frac{\delta(\mathbf{x})}{\sqrt{\pi}\cdot B(\mathbf{x})}\int_{(\varepsilon\cdot B(\mathbf{x})+\delta(\mathbf{x}))/\sqrt{2}\sigma}^{+\infty}\exp(-t^2)\,\mathrm{d}t\\
&-\frac{\varepsilon}{2}f(b_1)-\frac{\varepsilon}{2}f(b_2)\Bigg]\,p(\mathbf{x})\,\mathrm{d}\mathbf{x}\\
=&\int_D\Bigg[\frac{1}{A(\mathbf{x})}\left(-\frac{\sigma b_1}{\sqrt{2}}f(b_1)+\frac{\sigma}{\sqrt{2\pi}}\exp(-b_1^2)\right)\\
&+\frac{1}{B(\mathbf{x})}\left(-\frac{\sigma b_2}{\sqrt{2}}f(b_2)+\frac{\sigma}{\sqrt{2\pi}}\exp(-b_2^2)\right)\Bigg]\,p(\mathbf{x})\,\mathrm{d}\mathbf{x}.
\end{aligned}
\tag{18}
$$

In the following, we will use the second-order Taylor series expansions for $f(x)$ and $\exp(-x^2)$, i.e.

$$f(x+u)=f(x)-\frac{2}{\sqrt{\pi}}\mathrm{e}^{-x^2}u+\frac{2}{\sqrt{\pi}}x\mathrm{e}^{-x^2}u^2+\mathrm{o}(u^3), \tag{19}$$

$$\exp(-(x+u)^2)=\exp(-x^2)(1-2xu+(2x^2-1)u^2)+\mathrm{o}(u^3), \tag{20}$$

i.e.

$$
\begin{aligned}
f\left(\frac{\varepsilon\cdot A(\mathbf{x})}{\sqrt{2}\sigma}-\frac{\delta(\mathbf{x})}{\sqrt{2}\sigma}\right)&\approx f\left(\frac{\varepsilon\cdot A(\mathbf{x})}{\sqrt{2}\sigma}\right)+\frac{2}{\sqrt{\pi}}\exp\left(-\frac{\varepsilon^2\cdot A^2(\mathbf{x})}{2\sigma^2}\right)\frac{\delta(\mathbf{x})}{\sqrt{2}\sigma}\\
&\quad+\frac{2}{\sqrt{\pi}}\frac{\varepsilon\cdot A(\mathbf{x})}{\sqrt{2}\sigma}\exp\left(-\frac{\varepsilon^2\cdot A^2(\mathbf{x})}{2\sigma^2}\right)\frac{\delta^2(\mathbf{x})}{2\sigma^2},\\
f\left(\frac{\varepsilon\cdot B(\mathbf{x})}{\sqrt{2}\sigma}+\frac{\delta(\mathbf{x})}{\sqrt{2}\sigma}\right)&\approx f\left(\frac{\varepsilon\cdot B(\mathbf{x})}{\sqrt{2}\sigma}\right)-\frac{2}{\sqrt{\pi}}\exp\left(-\frac{\varepsilon^2\cdot B^2(\mathbf{x})}{2\sigma^2}\right)\frac{\delta(\mathbf{x})}{\sqrt{2}\sigma}\\
&\quad+\frac{2}{\sqrt{\pi}}\frac{\varepsilon\cdot B(\mathbf{x})}{\sqrt{2}\sigma}\exp\left(-\frac{\varepsilon^2\cdot B^2(\mathbf{x})}{2\sigma^2}\right)\frac{\delta^2(\mathbf{x})}{2\sigma^2},
\end{aligned}
$$

$$\exp\left(-\left(\frac{\varepsilon\cdot A(\mathbf{x})}{\sqrt{2}\sigma}-\frac{\delta(\mathbf{x})}{\sqrt{2}\sigma}\right)^2\right)\approx\exp\left(-\frac{\varepsilon^2\cdot A^2(\mathbf{x})}{2\sigma^2}\right)\left(1+\frac{\varepsilon\cdot A(\mathbf{x})\cdot\delta(\mathbf{x})}{\sigma^2}+\left(\frac{\varepsilon^2\cdot A^2(\mathbf{x})}{\sigma^2}-1\right)\frac{\delta^2(\mathbf{x})}{2\sigma^2}\right),$$

$$\exp\left(-\left(\frac{\varepsilon\cdot B(\mathbf{x})}{\sqrt{2}\sigma}+\frac{\delta(\mathbf{x})}{\sqrt{2}\sigma}\right)^2\right)\approx\exp\left(-\frac{\varepsilon^2\cdot B^2(\mathbf{x})}{2\sigma^2}\right)\left(1-\frac{\varepsilon\cdot B(\mathbf{x})\cdot\delta(\mathbf{x})}{\sigma^2}+\left(\frac{\varepsilon^2\cdot B^2(\mathbf{x})}{\sigma^2}-1\right)\frac{\delta^2(\mathbf{x})}{2\sigma^2}\right).$$

Using Eqs. (19) and (20), after a little tedious computation, Eqs. (17) and (18) become

$$
\begin{aligned}
\frac{1}{1+\varepsilon\beta}&=\int_D\frac{f(b_1)+f(b_2)}{2}p(\mathbf{x})\,\mathrm{d}\mathbf{x}\\
&\approx\int_D\Bigg[\frac{1}{2}f\left(\frac{\varepsilon\cdot A(\mathbf{x})}{\sqrt{2}\sigma}\right)+\frac{1}{\sqrt{\pi}}\exp\left(-\frac{\varepsilon^2\cdot A^2(\mathbf{x})}{2\sigma^2}\right)\frac{\delta(\mathbf{x})}{\sqrt{2}\sigma}
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{1}{\sqrt{\pi}}\cdot\frac{\varepsilon\cdot A(\mathbf{x})}{\sqrt{2}\sigma}\exp\left(-\frac{\varepsilon^2\cdot A^2(\mathbf{x})}{2\sigma^2}\right)\frac{\delta^2(\mathbf{x})}{2\sigma^2}\\
&+\frac{1}{2}f\left(\frac{\varepsilon\cdot B(\mathbf{x})}{\sqrt{2}\sigma}\right)-\frac{1}{\sqrt{\pi}}\exp\left(-\frac{\varepsilon^2\cdot B^2(\mathbf{x})}{2\sigma^2}\right)\frac{\delta(\mathbf{x})}{\sqrt{2}\sigma}\\
&+\frac{1}{\sqrt{\pi}}\cdot\frac{\varepsilon\cdot B(\mathbf{x})}{\sqrt{2}\sigma}\exp\left(-\frac{\varepsilon^2\cdot B^2(\mathbf{x})}{2\sigma^2}\right)\frac{\delta^2(\mathbf{x})}{2\sigma^2}\Bigg]p(\mathbf{x})\,\mathrm{d}\mathbf{x}\\
=&\ \frac{1}{2}f\left(\frac{\varepsilon E}{\sqrt{2}\sigma}\right)+\frac{\varepsilon E}{2\sqrt{2\pi}\sigma}\exp\left(-\frac{\varepsilon^2E^2}{2\sigma^2}\right)+\frac{1}{2}f\left(\frac{\varepsilon G}{\sqrt{2}\sigma}\right)+\frac{\varepsilon G}{2\sqrt{2\pi}\sigma}\exp\left(-\frac{\varepsilon^2G^2}{2\sigma^2}\right),
\end{aligned}
\tag{21}
$$

$$
\begin{aligned}
&\frac{1}{\beta(1+\varepsilon\beta)}\\
&=\int_D\left[\frac{1}{A(\mathbf{x})}\left(-\frac{\sigma b_1}{\sqrt{2}}f(b_1)+\frac{\sigma}{\sqrt{2\pi}}\exp(-b_1^2)\right)+\frac{1}{B(\mathbf{x})}\left(-\frac{\sigma b_2}{\sqrt{2}}f(b_2)+\frac{\sigma}{\sqrt{2\pi}}\exp(-b_2^2)\right)\right]p(\mathbf{x})\,\mathrm{d}\mathbf{x}\\
&=\int_D\Bigg[\frac{\delta(\mathbf{x})}{2A(\mathbf{x})}f(b_1)-\frac{\delta(\mathbf{x})}{2B(\mathbf{x})}f(b_2)-\frac{\varepsilon}{2}f(b_1)-\frac{\varepsilon}{2}f(b_2)\\
&\quad+\frac{\sigma}{\sqrt{2\pi}A(\mathbf{x})}\exp(-b_1^2)+\frac{\sigma}{\sqrt{2\pi}B(\mathbf{x})}\exp(-b_2^2)\Bigg]p(\mathbf{x})\,\mathrm{d}\mathbf{x}\\
&\approx\int_D\Bigg[-\varepsilon\left(\frac{f(b_1)+f(b_2)}{2}\right)+\frac{\delta(\mathbf{x})}{2A(\mathbf{x})}f(b_1)-\frac{\delta(\mathbf{x})}{2B(\mathbf{x})}f(b_2)\\
&\quad+\frac{\sigma}{\sqrt{2\pi}A(\mathbf{x})}\exp\left(-\frac{\varepsilon^2\cdot A^2(\mathbf{x})}{2\sigma^2}\right)\left(1+\frac{\varepsilon\cdot A(\mathbf{x})\cdot\delta(\mathbf{x})}{\sigma^2}+\left(\frac{\varepsilon^2\cdot A^2(\mathbf{x})}{\sigma^2}-1\right)\frac{\delta^2(\mathbf{x})}{2\sigma^2}\right)\\
&\quad+\frac{\sigma}{\sqrt{2\pi}B(\mathbf{x})}\exp\left(-\frac{\varepsilon^2\cdot B^2(\mathbf{x})}{2\sigma^2}\right)\left(1-\frac{\varepsilon\cdot B(\mathbf{x})\cdot\delta(\mathbf{x})}{\sigma^2}+\left(\frac{\varepsilon^2\cdot B^2(\mathbf{x})}{\sigma^2}-1\right)\frac{\delta^2(\mathbf{x})}{2\sigma^2}\right)\Bigg]p(\mathbf{x})\,\mathrm{d}\mathbf{x}.
\end{aligned}
\tag{22}
$$

Using Eqs. (15), (17), and (19), this simplifies to

$$
\begin{aligned}
E(L(\hat{\mathbf{s}}^{\mathrm{T}}\mathbf{x},\hat{\mathbf{r}}^{\mathrm{T}}\mathbf{x},\hat{\mathbf{w}}^{\mathrm{T}}\mathbf{x},y))&=\frac{1}{\beta(1+\varepsilon\beta)}\\
&=-\frac{\varepsilon}{1+\varepsilon\beta}+\int_D\Bigg[\frac{\varepsilon}{\sqrt{2\pi}}\exp\left(-\frac{\varepsilon^2\cdot A^2(\mathbf{x})}{2\sigma^2}\right)\frac{\delta^3(\mathbf{x})}{2\sigma^3}\\
&\quad-\frac{\varepsilon}{\sqrt{2\pi}}\exp\left(-\frac{\varepsilon^2\cdot B^2(\mathbf{x})}{2\sigma^2}\right)\frac{\delta^3(\mathbf{x})}{2\sigma^3}\\
&\quad+\frac{1}{\sqrt{2\pi}\cdot A(\mathbf{x})}\exp\left(-\frac{\varepsilon^2\cdot A^2(\mathbf{x})}{2\sigma^2}\right)\left(\sigma+\left(\frac{\varepsilon^2\cdot A^2(\mathbf{x})}{\sigma^2}+1\right)\frac{\delta^2(\mathbf{x})}{2\sigma}\right)\\
&\quad+\frac{1}{\sqrt{2\pi}\cdot B(\mathbf{x})}\exp\left(-\frac{\varepsilon^2\cdot B^2(\mathbf{x})}{2\sigma^2}\right)\left(\sigma+\left(\frac{\varepsilon^2\cdot B^2(\mathbf{x})}{\sigma^2}+1\right)\frac{\delta^2(\mathbf{x})}{2\sigma}\right)\Bigg]p(\mathbf{x})\,\mathrm{d}\mathbf{x}
\end{aligned}
$$

$$= -\frac{\varepsilon}{1+\varepsilon\beta} + \frac{\varepsilon}{\sqrt{2\pi}}\exp\left(-\frac{\varepsilon^2E^2}{2\sigma^2}\right) + \frac{1}{\sqrt{2\pi}E}\exp\left(-\frac{\varepsilon^2E^2}{2\sigma^2}\right)\left(\sigma + \left(\frac{\varepsilon^2E^2}{\sigma^2}+1\right)\frac{\sigma}{2}\right)$$
$$-\frac{\varepsilon}{\sqrt{2\pi}}\exp\left(-\frac{\varepsilon^2G^2}{2\sigma^2}\right) + \frac{1}{\sqrt{2\pi}G}\exp\left(-\frac{\varepsilon^2G^2}{2\sigma^2}\right)\left(\sigma + \left(\frac{\varepsilon^2G^2}{\sigma^2}+1\right)\frac{\sigma}{2}\right). \tag{23}$$

From Eq. (23), we obtain

$$\frac{1}{\beta} = \frac{\sigma}{\sqrt{2\pi}E}\exp\left(-\frac{\varepsilon^2E^2}{2\sigma^2}\right)\left(\frac{3}{2}+\frac{\varepsilon E}{\sigma}+\frac{\varepsilon^2E^2}{2\sigma^2}\right) + \frac{\sigma}{\sqrt{2\pi}G}\exp\left(-\frac{\varepsilon^2G^2}{2\sigma^2}\right)\left(\frac{3}{2}-\frac{\varepsilon G}{\sigma}+\frac{\varepsilon^2G^2}{2\sigma^2}\right). \tag{24}$$

By substituting (21) into (23) and simplifying, we have

$$E(L(\hat{\mathbf{s}}^{\mathrm{T}}\mathbf{x}, \hat{\mathbf{r}}^{\mathrm{T}}\mathbf{x}, \hat{\mathbf{w}}^{\mathrm{T}}\mathbf{x}, y))$$
$$= -\frac{\varepsilon}{2}f\left(\frac{\varepsilon E}{\sqrt{2}\sigma}\right) + \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\varepsilon^2E^2}{2\sigma^2}\right)\left(\frac{3\sigma}{2E}+\varepsilon\right) - \frac{\varepsilon}{2}f\left(\frac{\varepsilon G}{\sqrt{2}\sigma}\right)$$
$$+\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\varepsilon^2G^2}{2\sigma^2}\right)\left(\frac{3\sigma}{2G}-\varepsilon\right). \tag{25}$$

We have defined $A(\mathbf{x}) = \hat{\mathbf{s}}^{\mathrm{T}}|\mathbf{x}|$, $E = \int_D A(\mathbf{x})p(\mathbf{x})\,\mathrm{d}\mathbf{x}$, $B(\mathbf{x}) = \hat{\mathbf{r}}^{\mathrm{T}}|\mathbf{x}|$, $G = \int_D B(\mathbf{x})p(\mathbf{x})\,\mathrm{d}\mathbf{x}$, $\eta = \alpha$, and $C = \beta/\alpha$, so we can obtain

$$\alpha E(\hat{\mathbf{s}}^{\mathrm{T}}|\mathbf{x}|) = \beta E C^{-1}) \tag{26}$$
$$\eta E(\hat{\mathbf{r}}^{\mathrm{T}}|\mathbf{x}|) = \beta G C^{-1}. \tag{27}$$

After substituting (14), (21), (24), (25)–(27) into (13), minimizing (13) becomes minimizing the following $g(\varepsilon/\sigma)$:

$$g\left(\frac{\varepsilon}{\sigma}\right) = \left(\frac{1}{\sqrt{2\pi}E}\exp\left(-\frac{\varepsilon^2E^2}{2\sigma^2}\right)\left(\frac{3}{2}+\frac{\varepsilon E}{\sigma}+\frac{\varepsilon^2E^2}{2\sigma^2}\right) + \frac{1}{\sqrt{2\pi}G}\exp\left(-\frac{\varepsilon^2G^2}{2\sigma^2}\right)\left(\frac{3}{2}-\frac{\varepsilon G}{\sigma}+\frac{\varepsilon^2G^2}{2\sigma^2}\right)\right)^{-1}$$
$$\times\left(-\frac{\varepsilon}{2\sigma}f\left(\frac{\varepsilon E}{\sqrt{2}\sigma}\right) + \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\varepsilon^2E^2}{2\sigma^2}\right)\left(\frac{3}{2E}+\frac{\varepsilon}{\sigma}\right) - \frac{\varepsilon}{2\sigma}f\left(\frac{\varepsilon G}{\sqrt{2}\sigma}\right)\right.$$
$$\left.+\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\varepsilon^2G^2}{2\sigma^2}\right)\left(\frac{3}{2G}-\frac{\varepsilon}{\sigma}\right) + \frac{E}{\sigma C}+\frac{G}{\sigma C}\right)$$
$$-\ln\left(\frac{1}{2}f\left(\frac{\varepsilon E}{\sqrt{2}\sigma}\right) + \frac{\varepsilon E}{2\sqrt{2\pi}\sigma}\exp\left(-\frac{\varepsilon^2E^2}{2\sigma^2}\right) + \frac{1}{2}f\left(\frac{\varepsilon G}{\sqrt{2}\sigma}\right) + \frac{\varepsilon G}{2\sqrt{2\pi}\sigma}\exp\left(-\frac{\varepsilon^2G^2}{2\sigma^2}\right)\right)$$
$$+\ln\left(\frac{\sigma}{\sqrt{2\pi}E}\exp\left(-\frac{\varepsilon^2E^2}{2\sigma^2}\right)\left(\frac{3}{2}+\frac{\varepsilon E}{\sigma}+\frac{\varepsilon^2E^2}{2\sigma^2}\right)\right.$$
$$\left.+\frac{\sigma}{\sqrt{2\pi}G}\exp\left(-\frac{\varepsilon^2G^2}{2\sigma^2}\right)\left(\frac{3}{2}-\frac{\varepsilon G}{\sigma}+\frac{\varepsilon^2G^2}{2\sigma^2}\right)\right) + const. \tag{28}$$

Obviously, when (ε/σ) takes some fixed value, (28) will achieve its minimum, which indicates that there is a linear dependency between ε and the standard variance σ of the Gaussian input noise. Thus, there is an inverse linear dependency between the parameter h in FLR model and σ because of $\varepsilon = 1 - h$.

4. Experimental studies

In this section, two experiments to validate the inverse linear dependency relationship between h and σ will be arranged as follows:

Experiment 1: The input–output data from Yen et al. [17] are used here and shown in Table 1.

Table 1
Input–output data [17]

i	1	2	3	4	5
x_{i1}	0.84	0.65	0.76	0.7	0.43
x_{i2}	0.86	0.52	0.57	0.3	0.6
y_i	3.54	4.05	4.51	2.63	1.9

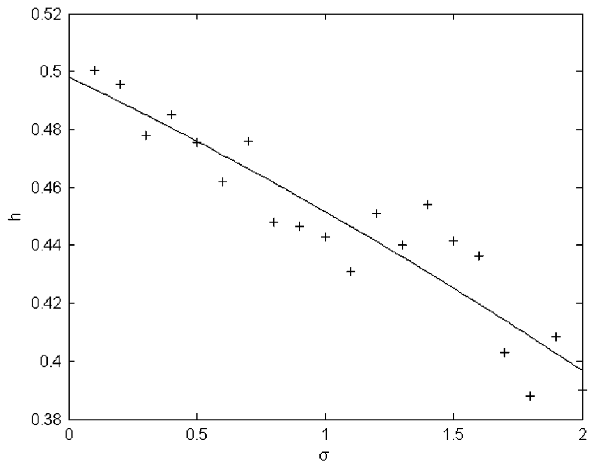

Fig. 2. The relationship between h and σ when $k = 5\%$.

First, for the data set, its RFLR model can be easily constructed, and by using the same method from [17], the following values for h and k_i are selected:

$$h = 0.5, \quad k_0 = 1.4, \quad k_1 = 1.6, \quad k_2 = 1.9.$$

Now, the corresponding regression values (l_i, c_i, r_i), $i = 1, 2, \ldots, 5$, can be obtained, where c_i denotes the center and l_i, r_i denote the left-side and right-side spreads, respectively. Next, in order to investigate the dependency relationship between h and the noisy input, let $y_i' = y_i + k \cdot n,\ i = 1, 2, \ldots, 5$, where k is a noise–signal ratio and $n \sim N(0, \sigma)$ represents the Gaussian noise. Then the corresponding sampling data set (x_i, y_i'), $i = 1, 2, \ldots, 5$, can be generated. Similarly, its corresponding regression values $(l_i', c_i', r_i'),\ i = 1, 2, \ldots, 5$, can be obtained by using the same RFLR model. In order to make the experimental results fair, σ is taken from $[0.1, 2.0]$ with the step length 0.1, and the Gaussian noise distribution is used to generate 20 groups of the corresponding sampling data sets for each given σ. For each given σ, h is taken as the average result of all 20 h values which can minimize $\sum_{i=1}^{10} \sqrt{(l_i - l_i')^2 + (c_i - c_i')^2 + (r_i - r_i')^2}$, respectively, for each group of the sampling data sets.

Figs. 2–4 depict the dependency relationships between h and σ for all 20 σ values with different k (see + in the figures), where the curves are used to roughly indicate the change tendencies between h and σ, respectively.

Experiment 2: The value for h is selected to be 0.5, the values for k_i $(i = 0, 1, 2)$ are selected to be 1 (corresponding to symmetric fuzzy triangular coefficients) and the same method in Experiment 1 is used. Figs. 5–7 demonstrate the corresponding results.

It can be easily seen from Figs. 2–7 that when noise is small, i.e. k and σ is comparatively small, there is an obvious inverse linear dependency relationship between h and σ. However, when k and/or σ are comparatively large, i.e. the data sets are seriously distorted, the inverse linear relationship between h and σ does not exist anymore. In other words, RFLR model may become ineffective for seriously distorted data sets. Of course, in Experiments 1 and 2, the value for h has been selected to be 0.5. If we select several different values for h and run the same program, the similar results

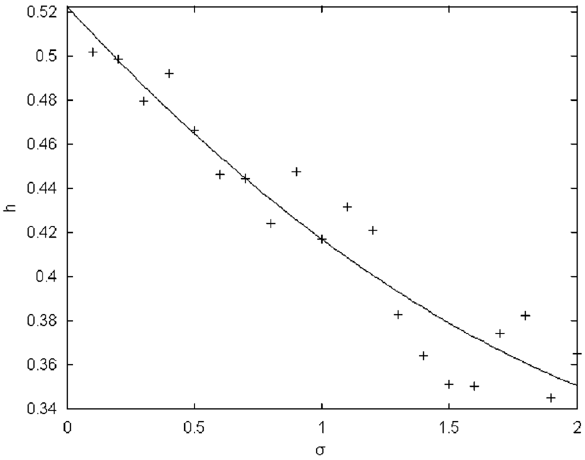

Fig. 3. The relationship between h and σ when $k = 10\%$.

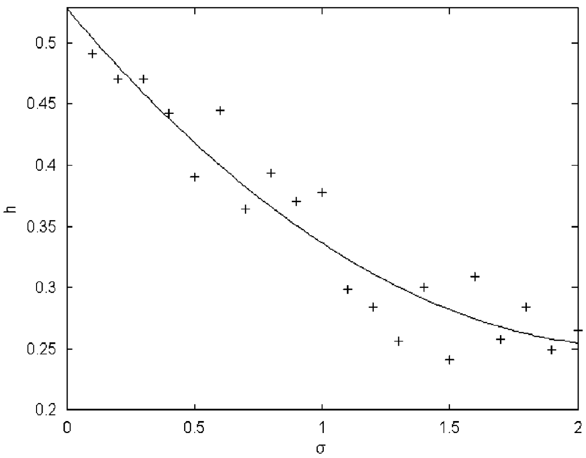

Fig. 4. The relationship between h and σ when $k = 15\%$.

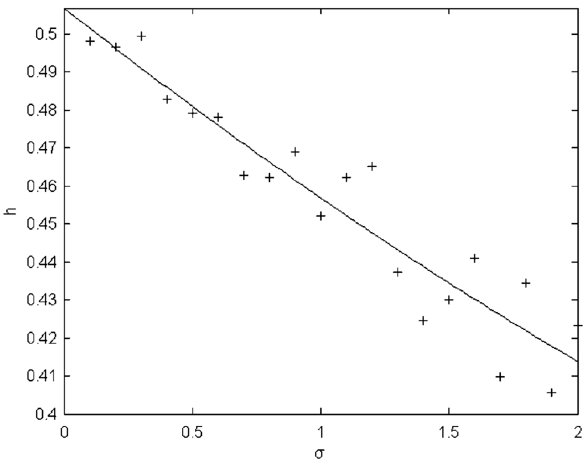

Fig. 5. The relationship between h and σ when $k = 5\%$.

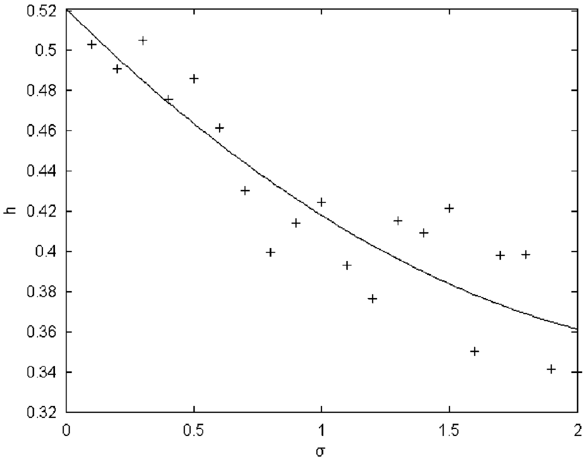

Fig. 6. The relationship between h and σ when $k = 10\%$.

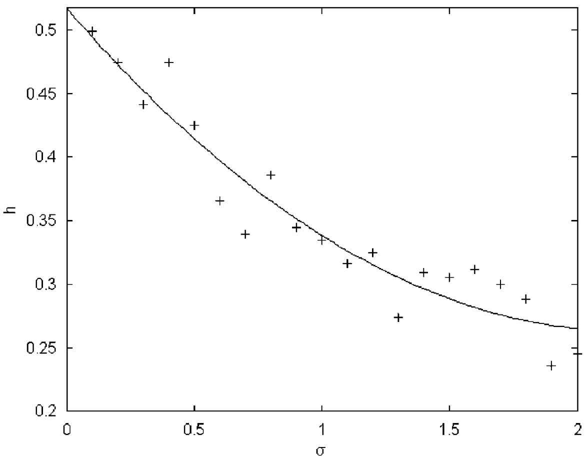

Fig. 7. The relationship between h and σ when $k = 15\%$.

can be obtained (in order to save the paper's space, these results are not shown here). In summary, the experimental results here validate the above-obtained conclusion on RFLR model.

5. Conclusions and future work

FLR model plays a pivotal role in fuzzy modeling. In this paper, attention is focused on the FLR model using non-symmetric fuzzy triangular coefficients. Based on the idea of SVR, the general dependency relationship between the parameter h and the input noise is studied. The FLR model is first extended to its regularized version and interpreted as the corresponding equivalent MAP problem. Accordingly, an approximately inverse proportional dependency between h and the standard deviation of Gaussian noisy input is analytically laid bare. The optimal choice of the parameter h is actually dependent on the variance of the input noise. Obviously, although our conclusion is based on RFLR model using non-symmetric triangular coefficients, the theoretical result here is very helpful for the practical applications of both FLR model and RFLR model using symmetric and non-symmetric fuzzy triangular coefficients.

Although Gaussian noise is typically adopted in most robustness analyses, there remain other types of noise such as Student-t-distribution noise and uniform noise in real data sets. The problem of choosing the optimal parameters in FLR model with such noisy input is interesting and is a topic for further study.

Acknowledgments

This work is supported by HongKong PolyU grant, National Science Foundation of China, JiangSu Natural Science Foundation, New Century Outstanding Young Scholar grant of Ministry of Education of China (NCET-04-0496), National KeySoft Laboratory at Nanjing University, The Key Laboratory of Computer Science at Institute of Software, CAS, China, The Key Laboratory of Computer Information Technologies at JiangSu Province, the 2004 Key Project of Ministry of Education of China, and National Key Laboratory of Pattern Recognition at Institute of Automation, CAS, China.

References

[1] J.B. Gao, S.R. Gunn, C.J. Ham's, A probabilistic framework for SVM regression and error bar estimation, Mach. Learning 46 (2002) 71–89.
[2] D.H. Hong, C. Hwang, Support vector fuzzy regression machines, Fuzzy Sets and Systems 138 (2003) 271–281.
[3] J.T. Kwok, I.W. Tsang, Linear dependency between ε and the input noise in ε-support vector regression, IEEE Trans. Neural Networks 14 (3) (2003) 544–553.
[4] M.H. Law, J.T. Kwok, Bayesian support vector regression, in: Proc. English Internat. Workshop on Artificial Intelligence and Statistics, Florida, 2001, pp. 239–244.
[5] G. Peters, Fuzzy linear regression with fuzzy intervals, Fuzzy Sets and Systems 63 (1994) 45–55.
[6] D.T. Redden, W.H. Woodall, Properties of certain fuzzy linear regression methods, Fuzzy Sets and Systems 64 (1994) 361–375.
[7] M. Sakawa, H. Yano, Multiobjective fuzzy linear regression analysis for fuzzy input–output data, Fuzzy Sets and Systems 47 (1992) 173–181.
[8] D. Savic, W. Pedrycz, Evaluation of fuzzy linear regression models, Fuzzy Sets and Systems 39 (1991) 51–63.
[9] H. Tanaka, Fuzzy data analysis by possibilistic linear model, Fuzzy Sets and Systems 24 (1987) 363–375.
[10] H. Tanaka, Possibilistic linear system and their application to the linear regression model, Fuzzy Sets and Systems 27 (1988) 275–289.
[11] H. Tanaka, H. Ishibuchi, Identification of possibilistic linear systems by quadratic membership functions of fuzzy parameters, Fuzzy Sets and Systems 41 (1991) 145–160.
[12] H. Tanaka, H. Ishibuchi, Possibilistic regression analysis based on linear programming, in: J. Kacprzyk, M. Fedrizzi (Eds.), Fuzzy Regression Analysis, Heidelberg, Germany, Omnitech, Warsaw, Poland, Springer, 1992, pp. 47–60.
[13] H. Tanaka, H. Ishibuchi, S. Yoshikawa, Exponential possibility regression analysis, Fuzzy Sets and Systems 69 (1995) 305–318.
[14] H. Tanaka, H. Lee, Interval regression analysis by quadratic programming approach, IEEE Trans. Fuzzy Systems 6 (4) (1998) 473–481.
[15] H. Tanaka, S. Uejima, K. Asai, Linear regression analysis with fuzzy model, IEEE Trans. Systems Man Cybernet. 12 (6) (1982) 903–907.
[16] S. Wang, Theoretically optimal parameter choices for support vector regression machines with noisy input, Internat. J. Soft Comput. 9 (10) (2005) 732–741.
[17] K.K. Yen, S. Ghoshray, G. Roig, A linear regression model using triangular fuzzy number coefficients, Fuzzy Sets and Systems 106 (1999) 167–177.