ROBUST NONLINEAR REGRESSION MODELING VIA L_<1->TYPE REGULARIZATION

Park, Heewon
Faculty of Global and Science Studies, Yamaguchi University

Konishi, Sadanori
Department of Mathematics, Faculty of Science and Engineering, Chuo University

https://doi.org/10.5109/2203027

出版情報：Bulletin of informatics and cybernetics. 48, pp.47-61, 2016-12. Research Association of Statistical Sciences
バージョン：
権利関係：

Bulletin of Informatics and Cybernetics, Vol. 48, 2016

ROBUST NONLINEAR REGRESSION MODELING VIA L_1-TYPE REGULARIZATION

By

Heewon PARK* and **Sadanori KONISHI†**

Abstract

The L_1-type regularization methods have drawn a large amount of attention for not only linear but also nonlinear regression modeling. By imposing lasso type penalties, the L_1-type regularization methods effectively select the number of basis functions, and thus we can perform well capturing the complex structure of data in nonlinear regression modeling. Although the L_1-type regularization approaches have been successfully used in various fields of research, their performances take a sudden turn for the worst in the presence of outliers, since the lasso type approaches are based on least squares or maximum likelihood estimator. To settle on the issue, we propose robust regularization methods for nonlinear regression modeling based on a novel least trimmed squares estimation. The proposed least trimmed squares regularization methods perform regression modeling based on s observations identified as non-outliers by outlier detection measures, and thus we can effectively perform robust nonlinear regression modeling without masking and swamping effects of outliers. We illustrate through Monte Carlo simulations and real world example that the proposed robust strategy effectively performs nonlinear regression modeling, even in the presence of outlier.

Key Words and Phrases: Basis expansion, L_1-type regularization, Least trimmed squares estimation, Nonlinear regression model, Robust regression

1. Introduction

Nonlinear regression modeling based on basis expansions has been widely used to analyze data with complex structure in various fields of research. A crucial issue of basis expansions is to express a regression model via a linear combination of known nonlinear functions (Konishi and Kitagawa, 2008; Tateishi et al., 2010). In recent year, L_1-type regularization methods (e.g., lasso (Tibshirani, 1996), elastic net (Zou and Hastie, 2005), and etc.) have drawn a large amount of attention for not only linear but also nonlinear regression modeling (Jiang et al., 2012; Arribas-Gil et al., 2014). By imposing L_1-type penalties to least squares loss function or maximum likelihood (ML) procedure, L_1-type regularization methods can perform feature selection and model estimation simultaneously. Furthermore, the lasso type approaches can effectively perform for high dimensional data analysis by overcoming the drawbacks of the existing methods (i.e., overfitting effect of maximum likelihood method and inapplicability of least squares (LS)

* Corresponding author, Faculty of Global and Science Studies, Yamaguchi University, 1677-1, Yoshida, Yamaguchi-shi, Yamaguchi Prefecture, 753-811, Japan. hwpark@yamaguchi-u.ac.jp

† Department of Mathematics, Faculty of Science and Engineering, Chuo University, 1–13–27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan. konishi@math.chuo-u.ac.jp

estimation). Although the ridge regression based on L_2-penalty has been also widely used for penalized regression modeling, the L_2-penalty cannot perform feature selection (Hoerl, Kennard, 1970). Thus, we consider not the ridge regression but the L_1-type penalty for nonlinear regression modeling.

It was, however, demonstrated that the performances of L_1-type regularization methods take a sudden for the worst in the presence of outliers, since the existing L_1-type approaches are based on least squares or likelihood functions. To settle on the issue, various robust L_1-type regularization methods have been proposed and successfully applied to linear regression modeling (Park et al., 2014; Lambert-Lacroix and Zwald, 2010; Zhang et al., 2009; Park and Konishi, 2016). However, the robust regularization approach for nonlinear regression modeling has been largely ignored, even though outliers may also significantly disturb the nonlinear regression modeling.

We propose robust L_1-type regularization methods for nonlinear regression modeling. In order to control outliers in modeling procedures, we consider a least trimmed squares (LTS) regularization method (Park et al., 2014). The existing LTS method is based only on s observations identified as non-outliers by using raw residuals (i.e., order statistics squared residuals) of least squares estimates, and thus we can reduce the influence of outliers. However, it was exposed that the residuals may fail to detect outliers at high leverage points, since high leverage points with outliers and residuals lead to masking and swamping effects (Riazoshams et al., 2009; She and Owen, 2011). To settle on the issue and improve the robustness of nonlinear regression modeling, we consider a novel LTS regularization method via not the raw residuals but various outlier detection measures (i.e., Studentized residual, COOK distance and DFFIT) based on hat matrix via linear approximation of basis functions (Riazoshams et al., 2009). We then propose robust L_1-type regularization approaches via a novel LTS loss function based on the outlier detection measures. In the proposed method, the nonlinear regression modeling is based on only s observations identified as non-outliers by outlier detection measures. It implies that the proposed robust strategy can effectively detect outliers without masking and swamping effect, and thus we can robustly perform nonlinear regression modeling, even in the presence of outliers. To the best of our knowledge, the proposed strategies are first attempt to incorporate the robust regularization approach in nonlinear regression modeling.

The rest of this paper is organized as follows. In Section 2, we describe the nonlinear regression model based on basis expansions with Gaussian basis functions. Section 3 demonstrates the sensitivity of the existing regularization method against outliers, and presents the proposed robust L_1-type regularized method for nonlinear regression modeling. In Section 4, we investigate the performance of the proposed robust strategy through numerical studies. Some concluding remarks are given in Section 5.

2. Nonlinear regression model

Suppose that we have n independent observations $\{(y_i, \boldsymbol{x}_i); i = 1, 2, ..., n\}$, where y_i are random response variables and $\boldsymbol{x}_i$ are p-dimensional vector of the explanatory variables. We consider the nonlinear regression model

$$y_i = u(\boldsymbol{x}_i) + \varepsilon_i \quad i = 1, 2, ..., n, \tag{1}$$

where $u(\cdot)$ is an unknown smooth function and the ε_i are independently distributed random error with mean zero and variance σ^2. It is assumed that the function $u(\cdot)$ can

be expressed as a linear combination of a prescribed set of m basis functions as follows,

$$u(\boldsymbol{x}_i; \boldsymbol{w}) = \sum_{j=1}^{m} w_j \phi_j(\boldsymbol{x}_i) = \boldsymbol{w}^T \boldsymbol{\phi}(\boldsymbol{x}_i), \tag{2}$$

where $\boldsymbol{\phi}(\boldsymbol{x}_i) = (\phi_1(\boldsymbol{x}_i), \phi_2(\boldsymbol{x}_i), ..., \phi_m(\boldsymbol{x}_i))^T$ is a vector of real-valued function (i.e., basis functions) and $\boldsymbol{w} = (w_1, ..., w_m)^T$ is an unknown coefficient parameter vector.

We consider the Gaussian function as a basis function, since it has various useful properties in analytical and practical viewpoints (Ando et al., 2008; Bishop, 1995). For the p-dimensional explanatory vector, the Gaussian basis functions are given by

$$\phi_j(\boldsymbol{x}; \boldsymbol{\mu}_j, h_j^2) = \exp\Big(- \frac{||\boldsymbol{x} - \boldsymbol{\mu}_j||^2}{2h_j^2}\Big), \quad j = 1, 2, ..., m, \tag{3}$$

where $\boldsymbol{\mu}_j$ is a p-dimensional vector determining the center of the basis function and h_j^2 is a bandwidth parameter that determines the dispersion. Although the parameter h_j^2 plays a key role in the basis function, the existing methods selected the parameter heuristically, and the heuristic methods for the parameter selection cannot always give effective results to control the amount of overlapping basis function (Ando et al., 2008). To settle on the drawback, we consider the following Gaussian function with hyperparameter proposed by Ando et al. (2008),

$$\phi_j(\boldsymbol{x}; \boldsymbol{\mu}_j, h_j^2, \nu) = \exp\Big(- \frac{||\boldsymbol{x} - \boldsymbol{\mu}_j||^2}{2\nu h_j^2}\Big), \quad j = 1, 2, ..., m, \tag{4}$$

where a hyperparameter ν adjusts the amount of overlapping among basis functions, and thus the basis function can effectively captures the structure in dataset.

2.1. Model estimation

The estimation of the nonlinear function $u(\boldsymbol{x})$ is based on two-step procedures to avoid local minimum and identification problem (Moody and Darken, 1989). In the first step, the center $\boldsymbol{\mu}_j$ and width parameter h_j^2 are determined by k-means clustering algorithm. The algorithm divides dataset into m clusters $\{C_1, C_2, ..., C_m\}$ corresponding to the number of the basis function and the center and width parameters are given by,

$$\hat{\boldsymbol{\mu}}_j = \frac{1}{n_j} \sum_{\boldsymbol{x}_i \in C_j} \boldsymbol{x}_i, \qquad \hat{h}_j^2 = \frac{1}{n_j} \sum_{\boldsymbol{x}_i \in C_j} ||\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}_j||^2, \tag{5}$$

where n_j is a number of observations in j^{th} cluster C_j. By replace $\boldsymbol{\mu}_j$ and h_j^2 with estimated $\hat{\boldsymbol{\mu}}_j$ and $\hat{h}_j^2$, we have the following m basis functions,

$$\phi_j(\boldsymbol{x}; \nu) = \exp\Big(- \frac{||\boldsymbol{x} - \hat{\boldsymbol{\mu}}_j||^2}{2\nu \hat{h}_j^2}\Big), \quad j = 1, 2, ..., m. \tag{6}$$

Then, the regression coefficients are estimated in the second step. The regression model based on the Gaussian basis function is represented by

$$\begin{aligned} y_i &= \sum_{j=1}^{m} w_j \phi_j(\boldsymbol{x}_i; \nu) + \varepsilon_i, \\ &= \boldsymbol{w}^T \boldsymbol{\phi}(\boldsymbol{x}_i; \nu) + \varepsilon_i, \qquad i = 1, 2, ..., n, \end{aligned} \tag{7}$$

where $\boldsymbol{\phi}(\boldsymbol{x}_i;\nu) = (\phi_1(\boldsymbol{x}_i;\nu), \phi_2(\boldsymbol{x}_i;\nu), ..., \phi_m(\boldsymbol{x}_i;\nu))^T$, $\boldsymbol{w} = (w_1, ..., w_m)^T$ and errors ε_i are independently, normally distributed with mean zero and variance σ^2.

The nonlinear regression model has been usually estimated by maximum likelihood or least squares method. However, it is well known that the existing methods yield unstable parameter estimation. Furthermore, the LS method cannot be used and the ML method suffers from overfitting results in high dimensional data analysis. To settle on the issues, various L_1-type regularization methods have been proposed.

In the L_1-type regularized regression modeling procedures, the unknown parameter vector is estimated by minimizing the following penalized least squares loss function,

$$\hat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}}\{\sum_{i=1}^{n}(y_i - \boldsymbol{w}^T\boldsymbol{\phi}(\boldsymbol{x}_i;\nu))^2 + P_\lambda(\boldsymbol{w})\}, \tag{8}$$

where $P_\lambda(\boldsymbol{w})$ is a L_1-type penalty, such as

- Lasso: $\lambda\sum_{j=1}^{m}|w_j|$,
- Elastic net: $\lambda\{\alpha\sum_{j=1}^{m}|w_j| + (1-\alpha)\sum_{j=1}^{m}w_j^2\}$,

where λ is a regularization parameter controlling model complexity and $\alpha \geq 0$. By imposing the L_1-type penalties, the regularization methods can overcome the drawbacks of the existing methods (i.e., instability and overfitting problems in high dimensional data analysis), and perform estimation and selection of basis functions simultaneously.

Although the regularization methods have been effectively performed for regression modeling, their performances take a sudden turn for the worst in the presence of outliers, since the methods based on least squares loss function or likelihood function, which are sensitive to outliers. We first show non-robustness of the ordinary regularization method in nonlinear regression modeling, and then propose a robust strategy for outlier-resistance nonlinear regression modeling via L_1-type regularization in next Section.

3. Robust nonlinear regularized regression modeling

The ordinary regularization methods for linear and nonlinear regression modeling are based on LS or ML method, and thus their procedures are significantly disturbed by outliers. In recent year, numerous studies demonstrated the drawback of the existing regularization methods, and attempted to overcome the drawbacks via robust statistical strategies for linear regression modeling (Park et al., 2013; Wang et al., 2007; Lambert-Lacroix and Zwald, 2010; Zhang et al., 2009). However, relatively little attention was paid to the robust regularization for nonlinear regression modeling.

We first demonstrate the non-robustness of the ordinary L_1-type regularization in nonlinear regression modeling. To show the adverse effect of outliers in nonlinear regression modeling, we generate random samples from a true model $f(x) = \exp(-2x)\cos\{3\pi\exp(x)\}$ and contaminate the dataset by replace 5% observations by outliers (i.e., $f(x) + N(1, 0.1\{\max(f(x)) - \min(f(x))\})$). Figure 1 shows the estimated curves in normal and contaminated datasets (i.e., triangle shape indicates the outliers) based on the ordinary lasso with Gaussian basis function. As shown in Figure 1, the fitted curve is significantly disturbed by outliers. It implies that the existing L_1-type regularization approach based on LS loss function cannot robustly perform for nonlinear regression modeling.

We propose a novel method for robust nonlinear regression modeling via L_1-type regularization method. To overcome the sensitivity against outliers of the existing methods, we consider various robust loss function instead of least squares loss function, such as M-function, least absolute loss function (LAD) and least trimmed squares (LTS) loss function. By replacing the least squares loss function with the robust methods, we can perform outlier-resistance nonlinear regression modeling even in the presence of outliers. However, the M-estimator (Susanti et al., 2014) and LAD estimator (Powell, 1984) do not have a high breakdown point, which is a main aim of the robust statistics. The breakdown point indicates the smallest fraction of contamination which can affect to the estimation procedure (Park et al., 2014; Rousseeuw and Leroy, 1987). It implies that the M-estimator and LAD methods cannot robustly perform well in a highly contaminated dataset.

Thus, we consider a least trimmed squares estimator having a maximum breakdown point $\{[(n-m)/2]+1\}/n$, which is asymptotically equal 50%, for $s=[(n+m+1)/2]$ (Rousseeuw and Leroy, 1987; Theorem 6), and refer to LTS-regularization strategy (Park et al., 2014) for nonlinear regression modeling,

$$\hat{\boldsymbol{w}}^{\mathrm{LTS}} = \arg\min_{\boldsymbol{w}} \{\sum_{i=1}^{s} r_{[i]}^2 + P_\lambda(\boldsymbol{w})\}, \tag{9}$$

where $r_{[i]}^2$ is the i^{th} order statistic of squared residuals and $r_i = y_i - \sum_{j=1}^{m} w_j \phi_j(\boldsymbol{x}_i;\nu)$. The LTS-regularization is based on only s observations without $n-s$ outliers, and thus

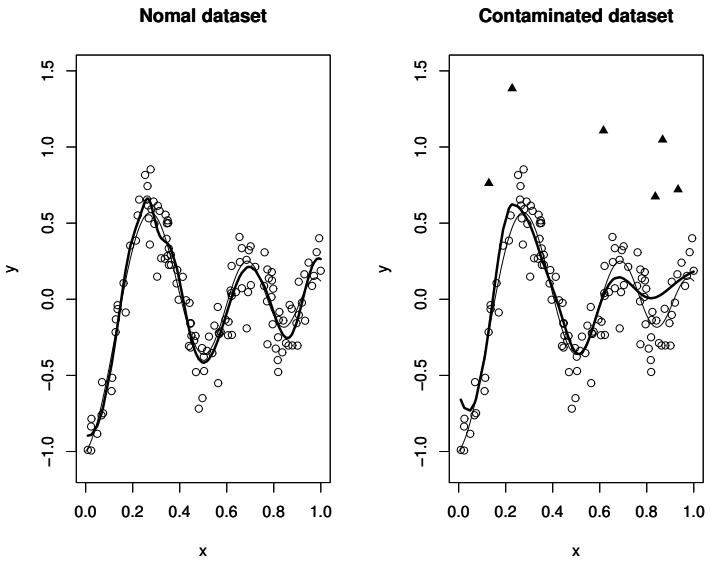

Figure 1: True function (normal line) and estimated curve (bold line) under the normal and contaminated datasets

we can robustly perform for nonlinear regression modeling.

In the LTS estimation procedure, determining $n-s$ outliers (i.e., outlier detection) is a crucial issue, since the regression modeling is based on s observations identified as non-outliers without interruption of outliers. However, it is well known that the squared residuals suffer from erroneous results of outlier detection at high leverage points. It implies that the LTS method based on the $r^2_{[i]}$ leads to masking and swamping effect of outliers in regression modeling. To settle on the issue, we propose a novel LTS regularization method based on effective outlier detection measures (Riazoshams et al., 2009).

3.1. Outlier detection in nonlinear regression model

We introduce methods for outlier detection in nonlinear regression modeling proposed by Riazoshams et al. (2009). In linear regression modeling, a hat matrix $W = X(X^TX)^{-1}X^T$ is widely used to detect outliers: the fitted $\hat{y}$ is dominated by $w_{ii}y_i$, and thus w_{ii} is interpreted as the amount of influence of $\hat{y}_i$ on y_i (i.e., amount of leverage of i^{th} observation) (Hoaglin and Welsh, 1978). Riazoshams et al. (2009) introduced a leverage matrix of nonlinear regression model via the following linear approximation form around the true value $\boldsymbol{w}^*$

$$u(\boldsymbol{x};\boldsymbol{w}) \cong u(\boldsymbol{x};\boldsymbol{w}^*) + \dot{\boldsymbol{\vartheta}}(\boldsymbol{w}-\boldsymbol{w}^*), \tag{10}$$

where the gradient matrix $\dot{\boldsymbol{\vartheta}}$ has elements $\frac{\partial u(\boldsymbol{x}_i;\boldsymbol{w}*)}{\partial w_j}$. By using the linear approximation, the leverage matrix is given as

$$H = \dot{\boldsymbol{\vartheta}}(\dot{\boldsymbol{\vartheta}}^T\dot{\boldsymbol{\vartheta}})^{-1}\dot{\boldsymbol{\vartheta}}^T. \tag{11}$$

The leverage matrix in nonlinear regression plays a similar role with the hat matrix in linear regression (Riazoshams et al., 2009; Cook and Weisberg, 1982; Kennedy and Gentle, 1980). However, the hat matrix cannot be used in high dimensional data situation (i.e., $m \gg n$), since the inverse matrix cannot be derived. Thus, we refer to a penalized hat matrix (Tharmaratnam et al., 2010),

$$H^* = \dot{\boldsymbol{\vartheta}}(\dot{\boldsymbol{\vartheta}}^T\dot{\boldsymbol{\vartheta}} + \gamma\boldsymbol{I}_m)^{-1}\dot{\boldsymbol{\vartheta}}^T, \tag{12}$$

where γ is a penalty parameter ($\gamma > 0$) and $\boldsymbol{I}_m$ = diag(1,1,...,1). In this study, we select the penalty parameter γ by the cross validation based on the ridge regression with response variables $\boldsymbol{y} = (y_1, y_2, ..., y_n)^T$ and predictors $\dot{\boldsymbol{\vartheta}}$. We then modified the outlier detection measures (Riazoharms et al., 2009) for nonlinear regression modeling based on the penalized hat matrix,

- Studentized residuals (cut-off value: $|t_i| > 2.5$):

$$t_i = \frac{r_i}{\hat{\sigma}\sqrt{1-h^*_{ii}}}, \tag{13}$$

- Cook's distance (cut-off value: $CD_i > 1$):

$$CD_i = \frac{t_i^2}{m}\frac{h^*_{ii}}{1-h^*_{ii}}, \tag{14}$$

- DFFIT (cut-off value: $DF_i > 2\sqrt{m/n}$):

$$DF_i = \Big(\sqrt{\frac{h^*_{ii}}{1-h^*_{ii}}}\Big)|d_i|, \tag{15}$$

where $d_i = \frac{r_i}{\hat{\sigma}_{(-i)}\sqrt{1-h^*_{ii}}}$, and $\hat{\sigma}_{(-i)}$ is standard deviation of the residual estimated without an i^{th} observation.

The residuals r_i are obtained from the LS, M or MM estimator (Susanti et al., 2014). For details on the M and MM estimators, see Susanti et al. (2014).

3.2. Robust nonlinear regression modeling via L_1-regularization

We propose novel least trimmed squares regularization methods for nonlinear regression modeling. The proposed methods perform nonlinear regression modeling based on only s observations identified as non-outliers. In order to overcome the drawback of the order statistics of squared residuals $r^2_{[i]}$ (i.e., masking and swamping effect) and improve robustness, we use the outlier detection measures (i.e., Studentized residuals, Cook's distance and DFFIT) for identifying outliers. In other words, we sort the values of the outlier detection measures $\boldsymbol{t} = (t_1, t_2, ..., t_n)^T$, $\boldsymbol{CD} = (CD_1, CD_2, ..., CD_n)^T$ and $\boldsymbol{DF} = (DF_1, DF_2, ..., DF_n)^T$ in ascending order, and then identify s non-outliers corresponding to the s smallest values of the measures,

- O^t: Observations i for $t_i \leq s^{th}$ smallest value of $\boldsymbol{t}$,
- O^{CD}: Observations i for $CD_i \leq s^{th}$ smallest value of $\boldsymbol{CD}$,
- O^{DF}: Observations i for $DF_i \leq s^{th}$ smallest value of $\boldsymbol{DF}$,

where s is a tuning constant (i.e., a number of observations used in regression modeling).

We then proposed novel least trimmed squares regularization methods for nonlinear regression modeling as follows,

- LTS.Ti Regularization:

$$\hat{\boldsymbol{w}}^{\text{LTS.Ti}} = \arg\min_{\boldsymbol{w}}\{\sum_{i\in O^t} r_i^2 + P_\lambda(\boldsymbol{w})\}, \tag{16}$$

- LTS.COOK Regularization:

$$\hat{\boldsymbol{w}}^{\text{LTS.COOK}} = \arg\min_{\boldsymbol{w}}\{\sum_{i\in O^{CD}} r_i^2 + P_\lambda(\boldsymbol{w})\}, \tag{17}$$

- LTS.DFFIT Regularization:

$$\hat{\boldsymbol{w}}^{\text{LTS.DFFIT}} = \arg\min_{\boldsymbol{w}}\{\sum_{i\in O^{DF}} r_i^2 + P_\lambda(\boldsymbol{w})\}, \tag{18}$$

By using the robust loss function based on outlier detection measures, we can effectively detect outliers, and thus robustly perform nonlinear regression modeling, even in the presence of outliers.

The nonlinear regression modeling based on the proposed LTS regularization methods are implemented by LARS (Least Angle Regression) algorithm (Efron et al., 2004). The equations (16), (17), (18) can be seen as ordinary L1-type regularization based on least squares loss function for dataset with only s observations. Thus, for LTS L_1-type regularized regression modeling, we apply the LARS algorithm for ordinary L_1-type regularized regression modeling (i.e., regularization method based on least squared loss function) with only s observations corresponding to the small values of the proposed outlier detection measures, and select the tuning parameter s by using a model selection criterion.

4. Numerical studies

We conduct simulation studies and real data analysis to investigate the performance of the proposed LTS regularization methods. In numerical studies, we consider the lasso as a L_1-type regularization method and tuning parameters are selected by the cross-validation. The raw residuals r_i in outlier detection measures are obtained by MM-estimator to improve robustness of regression modeling. We consider the tuning constant s as $n^s < s < n$, where n^s is $n-125\%$ of $\sum_{i=1}^{n} I(\text{measure}_i >$ cut-off value of each outlier detection measure), where $I(\cdot)$ is the indicator function.

4.1. Monte Carlo simulations

In this section, we show through Monte Carlo simulations the effectiveness of the proposed methods based on curve and surface fittings.

In the curve fitting, the random samples $\{(y_i, x_i); i = 1, 2, ..., 150\}$ are repeatedly generated based on a true regression model $y_i = u(x_i) + \varepsilon_i$. The design points are uniformly distributed in $[0, 1]$ and the errors ε_i are assumed to be independently distributed according to a normal distribution with mean 0 and standard deviance $\sigma = 0.1R_w$, where R_w is the range of $u(x)$ over $x \in [0, 1]$. We consider the following two functions for true nonlinear regression model,

(A) $u(x) = -4x(x-1) + \sin(4\pi x)/4$,

(B) $u(x) = -0.5\exp\{-50(x-0.3)^2\} + 0.5\exp\{-250(x-0.7)^2\}$.

In the surface fitting, the random samples $\{(y_i, x_{1i}, x_{2i}); i = 1, 2, ..., 150\}$ are generated from a true model $y_i = u(x_{1i}, x_{2i}) + \varepsilon_i$. The design points are uniformly distributed in $[0, 1] \times [0, 1]$. The errors ε_i are assumed to be independently distributed according to a normal distribution with mean 0 and standard deviance $\sigma = 0.1R_w$, where R_w is the range of $u(x_1, x_2)$ over $(x_1, x_2) \in [0, 1] \times [0, 1]$. The true models are given as

(C) $u(x_1, x_2) = \sin\{5x_1x_2\} + \cos\{3(x_1 + x_2)\}$,

(D) $u(x_1, x_2) = \sin(10\sqrt{x_1^2 + x_2^2})/(10\sqrt{x_1^2 + x_2^2})$.

We evaluate the proposed LTS regularization methods (i.e., LTS.Ti, LTS. COOK and LTS.DFFIT) compared with ordinary regularization based on least squares loss

function (i.e., LASSO) and the ordinary LTS regularization based on $r^2_{[i]}$ (i.e., LTS.R). We consider datasets with 5% and 10% outliers by replace y_i with $N(\mu^O, 1)$ where μ^O is $1.5 \times \max(y_i)$, and without outliers (i.e., 0% outlier). The number of basis function m was taken 10. In order to evaluate performances of each method, we compare the mean square errors MSE= $\sum_{i=1}^{150}\{u(x_i) - \hat{y}_i)\}^2/150$ and MSE= $\sum_{i=1}^{150}\{u(x_{1i}, x_{2i}) - \hat{y}_i\}^2/150$ in curve and surface fittings, respectively.

Tables 1 and 2 show median of MSE (i.e., column of "Med(·)"), selected parameters λ and ν, their median absolute deviation (MAD) = median($|A - \text{median}(A)|$), and averages of the numbers of selected basis functions (i.e., "Ave(No.B)") in 50 repeatedly generated datasets for curve and surface fittings, respectively. For the simulation settings, the nonlinear regression models by the existing and proposed methods are constructed with around 7∼9 basis functions as shown in column "Ave(No.B)". We can see through the results that the proposed LTS regularization methods also perform properly basis function selection, like ordinary Lasso. It can also be seen through Tables 1 and 2 that the robust regularization methods based on LTS loss function (i.e., LTS.R, LTS.Ti, LTS.COOK and LTS.DFFIT) robustly perform for nonlinear regression modeling compared with ordinary lasso in the viewpoint of prediction accuracy (i.e., as shown in column of "Med(MSE)"). We can also see that the proposed LTS regularization methods (i.e., LTS.Ti, LTS.COOK and LTS.DFFIT) based on the outlier detection measures

Table 1: Simulation results in Curve fittings for true models (A) and (B)

%	Model	Method	Med(λ)	MAD(λ)	Med(ν)	MAD(ν)	Med(MSE)	MAD(MSE)	Ave(No.B)
0%	(A)	LASSO	0.101	0.091	75.00	12.50	**0.674**	0.027	8.52
		LTS.R	0.066	0.045	75.00	12.50	0.676	0.028	8.86
		LTS.Ti	0.071	0.061	87.50	12.50	0.678	0.029	8.94
		LTS.COOK	0.051	0.040	75.00	25.00	0.677	0.032	8.72
		LTS.DFFIT	0.121	0.111	75.00	12.50	0.678	0.031	8.74
	(B)	LASSO	0.859	0.131	62.50	12.50	0.115	0.010	9.56
		LTS.R	0.884	0.106	75.00	25.00	**0.114**	0.011	9.74
		LTS.Ti	0.939	0.051	75.00	12.50	**0.114**	0.010	9.76
		LTS.COOK	0.990	0.000	75.00	12.50	0.115	0.010	9.88
		LTS.DFFIT	0.854	0.136	68.75	18.75	0.115	0.010	9.66
5%	(A)	LASSO	0.040	0.030	75.00	12.50	0.778	0.047	8.60
		LTS.R	0.061	0.051	75.00	12.50	0.710	0.036	8.68
		LTS.Ti	0.091	0.081	62.50	12.50	0.692	0.037	8.58
		LTS.COOK	0.051	0.040	62.50	12.50	0.691	0.036	8.28
		LTS.DFFIT	0.071	0.061	62.50	12.50	**0.686**	0.035	8.78
	(B)	LASSO	0.076	0.066	75.00	12.50	0.133	0.019	8.70
		LTS.R	0.328	0.303	75.00	25.00	0.116	0.016	9.20
		LTS.Ti	0.045	0.035	75.00	12.50	0.113	0.014	8.72
		LTS.COOK	0.126	0.116	62.500	12.50	0.113	0.014	8.86
		LTS.DFFIT	0.167	0.157	62.50	12.50	**0.111**	0.014	8.84
10%	(A)	LASSO	0.025	0.015	62.50	12.50	0.712	0.030	9.14
		LTS.R	0.126	0.106	75.00	12.50	0.700	0.027	9.06
		LTS.Ti	0.045	0.035	75.00	12.50	**0.698**	0.023	9.18
		LTS.COOK	0.101	0.091	68.75	6.25	**0.698**	0.024	8.50
		LTS.DFFIT	0.081	0.071	62.50	12.50	**0.698**	0.023	8.44
	(B)	LASSO	0.111	0.101	75.00	12.50	0.156	0.022	9.00
		LTS.R	0.667	0.323	62.50	12.50	0.118	0.014	9.48
		LTS.Ti	0.076	0.066	75.00	12.50	0.118	0.014	8.80
		LTS.COOK	0.116	0.106	62.50	12.50	0.120	0.015	8.76
		LTS.DFFIT	0.222	0.212	62.50	12.50	**0.115**	0.014	8.58

Table 2: Simulation results in Surface fittings for true models (C) and (D)

%	Model	Method	Med(λ)	MAD(λ)	Med(ν)	MAD(ν)	Med(MSE)	MAD(MSE)	Ave(No.B)
0%	(C)	LASSO	0.525	0.389	62.50	12.50	0.425	0.054	9.24
		LTS.R	0.707	0.237	75.00	12.50	0.432	0.059	9.82
		LTS.Ti	0.697	0.278	75.00	12.50	0.425	0.044	9.68
		LTS.COOK	0.747	0.242	75.00	12.50	**0.423**	0.047	9.40
		LTS.DFFIT	0.818	0.172	75.00	12.50	0.431	0.049	9.30
	(D)	LASSO	0.318	0.268	87.50	12.50	0.364	0.043	9.12
		LTS.R	0.697	0.268	75.00	18.75	0.364	0.040	9.36
		LTS.Ti	0.470	0.369	75.00	12.50	**0.358**	0.043	9.14
		LTS.COOK	0.510	0.414	75.00	25.00	**0.358**	0.045	9.52
		LTS.DFFIT	0.530	0.419	75.00	12.50	0.365	0.045	9.12
5%	(C)	LASSO	0.020	0.010	81.25	18.75	0.424	0.062	7.20
		LTS.R	0.374	0.343	75.00	12.50	0.431	0.049	9.12
		LTS.Ti	0.111	0.101	68.75	18.75	0.414	0.060	8.64
		LTS.COOK	0.354	0.343	68.75	18.75	0.409	0.050	8.66
		LTS.DFFIT	0.126	0.116	75.00	12.50	**0.408**	0.039	8.78
	(D)	LASSO	0.116	0.106	75.00	25.00	0.372	0.056	8.16
		LTS.R	0.525	0.338	75.00	12.50	0.349	0.047	9.26
		LTS.Ti	0.399	0.384	75.00	12.50	0.350	0.041	8.72
		LTS.COOK	0.192	0.182	75.00	12.50	**0.339**	0.032	8.24
		LTS.DFFIT	0.333	0.323	75.00	12.50	0.346	0.037	8.60
10%	(C)	LASSO	0.015	0.005	75.00	12.50	0.480	0.094	7.20
		LTS.R	0.404	0.374	81.25	18.75	0.426	0.065	8.76
		LTS.Ti	0.071	0.061	75.00	12.50	0.425	0.068	8.32
		LTS.COOK	0.091	0.081	62.50	12.50	**0.400**	0.049	7.92
		LTS.DFFIT	0.056	0.045	75.00	12.50	0.436	0.063	7.72
	(D)	LASSO	0.051	0.040	68.75	18.75	0.399	0.068	8.10
		LTS.R	0.369	0.328	87.50	12.50	0.359	0.056	9.12
		LTS.Ti	0.328	0.318	75.00	12.50	0.353	0.050	8.50
		LTS.COOK	0.020	0.010	75.00	18.75	**0.331**	0.040	7.68
		LTS.DFFIT	0.056	0.045	75.00	12.50	0.362	0.045	7.88

outperform nonlinear regression modeling compared with not only the ordinary lasso but also the LTS method based on $r^2_{[i]}$ (i.e., LTS.R) in overall. Furthermore, our methods show stable estimation results as shown in columns MAD(MSE) of Tables 1 and 2 in overall.

Figure 2 shows the true models and estimated curves by using each method in (A) and (B). The estimated curves are based on median of selected each tuning parameter (i.e., regularization parameter λ, tuning constant s and hyperparameter ν) in 50 datasets. We can see from Figure 2 that the ordinary lasso based on LS loss function is significantly disturbed by outliers. And, the ordinary LTS regularization method based on the order statistics of squared residuals is also disturbed by outliers even though the method can robustly performs compared with ordinary regularization. On the other hand, the proposed methods based on outlier detection measures robustly perform nonlinear regression modeling even in the presence of outliers.

4.2. Real world example: Birth data

We apply the proposed methods to real data analysis: Birth dataset which can be taken from package "*catdata*" in software *R*. The birth data provides various information about pregnancy and birth for 755 children who were born from 1990 to 2004. The birth data was collected from internet users recruited on french speaking pregnancy and birth

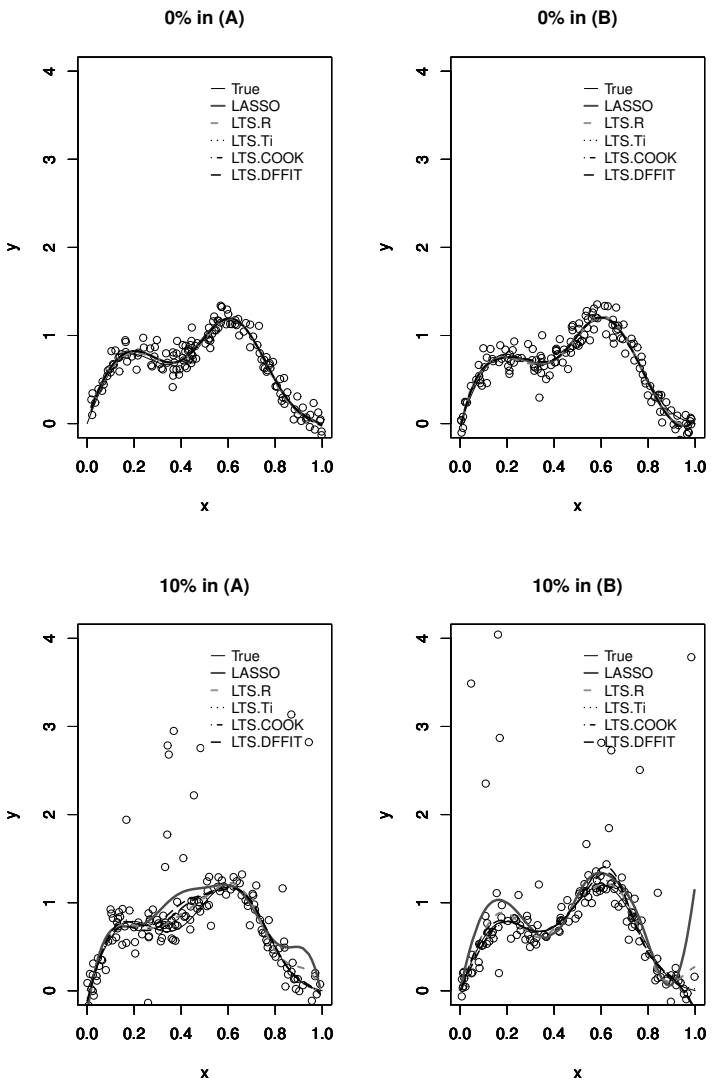

Figure 2: True and Estimated curves in (A) and (B) with 0% and 10% outliers

websites (Schauberger and Tutz, 2014). The dataset consists of 755 observations with 25 predictor variables about mother and children: sex, weight, height, head of circumference of child, month of birth, year of birth, country of birth, term of pregnancy, age of mother, number of pregnancies before, weight of mother before the pregnancy, height of mother, weight of mother after the pregnancy, days that child spent in intensive care unit and etc. (Schauberger and Tutz, 2014).

We consider a nonlinear regression model with the number of days spent in intensive care unit as a response variable and 7 predictor variables, which are expected to be strongly associated with the response variable: age, height and weight before the pregnancy of mother, weight, height and head circumference of children at the birth

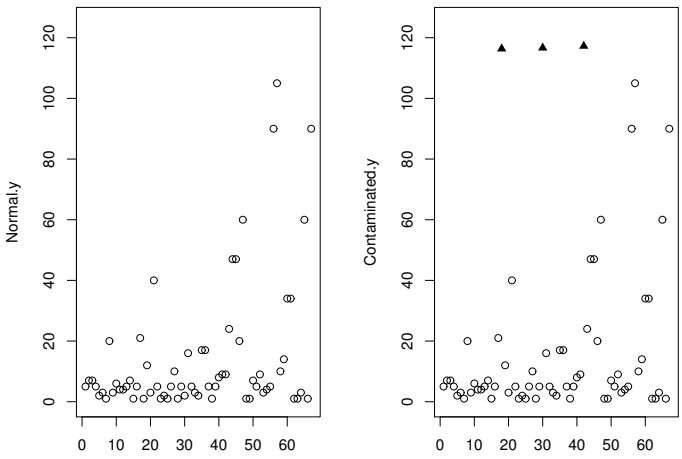

Figure 3: Ordinary Birth data and contaminated data with 5% outliers

and term of pregnancy. In our study, we use the data of children with the number of days spent in intensive care unit from 0 to 800, and thus dataset used in our study is based on 91 observations and 7 predictor variables. In order to evaluate robustness of the proposed methods, we contaminate the dataset by replace 5% observations of the response variable y with outliers ($N(1.1 \times \mathrm{y}, 1)$) as shown in Figure 3.

Table 3 shows Median Absolute Error (MAE) for measuring estimation accuracy of each method,

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \sum_{j=1}^{m} \hat{w}_j \phi_j(\boldsymbol{x}_i; \nu)|. \tag{19}$$

It can be seen through Table 3 that the robust regularization methods outperform fitting nonlinear regression model compared with ordinary regularization method based on LS loss function. Furthermore, the proposed methods based on the novel outlier detection measures, especially LTS.DFFIT, show the outstanding performance compared with not only ordinary regularization methods but also LTS regularization approaches based on the order statistics of squared residuals $r^2_{[i]}$. It implies that our methods are useful tools for nonlinear regression modeling, even in the presence of outliers.

Table 3: Birth data analysis results

	LASSO	LTS.R	LTS.Ti	LTS.COOK	LTS.DFFIT
MAE	9.736	4.866	4.436	4.593	**3.891**

5. Concluding remarks

We have introduced a novel method for robust nonlinear regression modeling via L_1-type regularization. Although the L_1-type regularization method has been widely applied and successfully performed nonlinear regression modeling, the existing method

is sensitive to outliers. To settle on the issue, we have considered a least trimmed squares regularization. In the proposed methodology, outliers are effectively detected by not the order statistics of squared residuals $r^2_{[i]}$ but the outlier detection measures, and thus the proposed LTS regularization methods can robustly perform nonlinear regression modeling based on s observations identified as non-outliers. Monte Carlo simulations and real world example have demonstrated that the proposed modeling strategies robustly perform nonlinear regression modeling in various situations. It implies that our method is an effective tool for nonlinear regression modeling in the presence of outliers.

We considered only a small number of basis functions in numerical studies. In order to incorporate the complex data structures in nonlinear regression modeling, the consideration of a large number of basis functions will be required. We consider the robust nonlinear regression modeling with a large number of basis functions as one of future works of this study.

In the robust sparse nonlinear regression modeling, tuning parameters (i.e., regularization parameters, tuning constant and hyperparameter) play a key role for not only estimation and selection of the number of basis functions but also outlier-resistance modeling. Although we selected the tuning parameters by using the cross validation, it is well known that the method is time consuming and suffers from overfitting in feature selection in the sparse regression modeling. In order to improve performance of regression modeling, the present study can be extended towards an information criterion for choosing the tuning parameters in robust nonlinear sparse regression modeling.

Acknowledgement

The authors would like to thank the associate editor and anonymous reviewers for the constructive and valuable comments that improved the quality of the paper.

References

Arribas-Gil, A., Bertin, K., Meza, C. and Rivoirard, V. (2014). Lasso type estimators for semiparametric nonlinear mixed effects models estimation. *Statistics and Computing*, **24**, 443-460.

Ando, T., Konishi, S. and Imoto S. (2008). Nonlinear regression modeling via regularized radial basis function networks. *Journal of Statistical Planning and Inference*, **138**, 3616-3633.

Bishop, C.M. (1995). *Neural Netowrks for Patterm Recognition*. Oxford University Press Oxford.

Cook, R.D. and Weisberg, S. (1982). *Residuals and Influence in Regression*. CHAPMAN and HALL.

Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least angle regresion *The Annals of Statistics*, **32**, 407-451.

Hoaglin, D.C. and Welsch, R. (1978). The hat matrix in regression and ANOVA. *Journal of the American Statistical Association*, **32**, 17-22.

Hoerl, A.E. and Kennard, R.W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Techonometrics*, **12**, 55-67.

Kennedy, W. and Gentle, J. (1980). *Statistical Computing*. Dekker, New York.

Konishi, S. and Kitagawa, G. (2008). *Information Criteria and Statistical Modeling*. Springer, New York.

Lambert-Lacroix, S. and Zwald, L. (2010). Robust regression through the Hubers criterion and adaptive lasso penalty. *Electronic Journal of Statistics*, **5**, 1015-1053.

Moody, J. and Darken, C.J. (1989). Fast learning in networks of locally-turned processing units. *Neural Computing and Applications*, **1**, 281-294.

Park, H., Sakaori, F. and Konishi, S. (2014). Robust sparse regression and tuning parameter selection via the efficient bootstrap information criteria. *Journal of Statistical Computation and Simulation*, **84**, 1596-1607.

Park, H. and Konishi, S. (2016). Robust solution path for high dimensional sparse regression modeling. *Communications in Statistics - Simulation and Computation*, **45**, 115-129.

Powell, J.L. (1984). Least abosulte deviations estiation for the censored regression model. *Journal of Econometrics*, **25**, 303-325.

Riazoshams, A.H., Habshah, B.M. and Adam, J.M.B. (2009). On the outlier detection in nonlinear regression. *World Academy of Science, Engineering and Technology*, **3**, 244-250.

Rousseeuw, P.J. and Leroy, A.M. (1987). *Robust Regression and Outlier Detection*. Wiley, New York.

Schauberger, G. and Tutz, G. (2014). *catdata* package manual of software *R*, Available at http://cran.r-project.org/web/packages/catdata/catdata.pdf.

She, Y. and Owen, A.B. (2011). Outlier detection using nonconvex penalized regressionm. *Journal of the American Statistical Association*, **106**, 626-39.

Susanti, Y., Pratiwi, H., Sulistijowati, S., and Liana, T. (2014). M estimation, S estimation, and MM estimation in robust regression. *International Journal of Pure and Applied Mathematics*, **91**, 349-360.

Tateishi, S., Matsui, H. and Konishi, S. (2010). Nonlinear regression modeling via the lasso-type regularization, *Journal of Statistical Planning and Inference*, **140**, 1125-1134.

Tharmaratnam, K., Claeskens, G., Croux, C. and Salibian-Barrera, M. (2010). S-estimation for penalized regression splines. *Journal of Computational and Graphical Statistics*, **19**, 609-625.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, **73**, 273-282.

Wang, H., Li, G. and Jiang, G. (2007). Robust regression shrinkage and consistent variable selection through the LAD-lasso. *Journal of Business & Economic Statistics*, **25**, 347-355.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B*, **67**, 301-320.

Jiang, X., Jiang, J. and Song, X. (2012). Oracle model selection for nonlinear models based on weighted composite quantile regression. *Statistica Sinica*, **22**, 1479-1506.

Zhang, Z.G., Chan, S.C., Zhou, Y. and Hu, Y. (2009). Robust linear estimation using M-estimation and weighted L1 regularization: model selection and recursive implementation. *Proceedings of the 2009 International Symposium on Circuits and Systems*, Taipei, Taiwan, 1193-1196.

Received January 7, 2016
Revised June 2, 2016