Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 19, No. 2, pp. 315–334, 2023

NUMERICAL OPTIMIZATION AND COMPUTATION FOR SECOND-ORDER LEAST SQUARES ESTIMATION*

Xin Wang, Lingchen Kong and Liqun Wang

Abstract: The second-order least squares (SLS) estimation is a parameter estimation method for nonlinear regression model based on second-order moment information. Its optimization is a non-convex problem, even for the linear regression. Existing research does not propose a systematic and complete calculation method for the optimization corresponding to this estimation. Although this is a smooth optimization, the objective function is non-convex, which causes traditional methods to easily fall into local solutions or fail to obtain the desired accuracy. In this paper, we propose a systematic calculation method for SLS estimation, which is called alternate updating (AU) method. First, we give the assumptions needed for this estimation in linear regression and analyze some potential properties. Second, we design an alternate updating method based on a strong first-order optimality condition and establish its convergence. In the end, the effectiveness of the alternating updating method is demonstrated by numerical simulations.

Key words: *second-order least squares estimation, strong first-order optimality condition, alternate updating method*

Mathematics Subject Classification: *49M05, 65K05, 90C26, 90C30*

1 Introduction

Regression problem is one of the important problems in statistical machine learning, and it has a wide range of applications in the fields of management, medicine, economics, agriculture and so on. There are many estimations for unknown regression parameter, such as ordinary least squares (OLS) estimation, least absolute deviation (LAD), Huber regression, ridge regression. The OLS estimation is the most common and widely used for regression problem. The LAD and Huber regression are robust methods based on the absolute value measurement. The estimator produced by ridge regression is biased, but it can overcome the collinearity of data. However, these methods only use the first-order information of data. Wang and Leblanc [13] extended OLS estimation by including in the criterion function the distance of the squared response variable to its second conditional moment, and proposed the SLS estimation.

The SLS estimation was first proposed to deal with the measurement error problems in nonlinear regression models in [11], [12]. However, these results can not be applied to the tradition nonlinear regression problem under standard conditions. Wang and Leblanc improved the theories of SLS estimation and compared it with the OLS estimation in general

*This work was supported by the National Natural Science Foundation of China (12071022).

nonlinear models in [13]. Their conclusions show that the SLS estimator is asymptotically more efficient than the OLS estimator if the third moment of the random error is nonzero or the distribution of error is asymmetric, and both estimators have the same asymptotic covariance matrix if the error distribution is symmetric. In practice, it is difficult to determine what the distribution of random error is. Therefore, the SLS estimator is more efficient than or as efficient as the OLS estimator. Furthermore, Abarin and Wang made a comparison between generalized method of moments (GMM) and SLS estimation in nonlinear models in [1], and they established the SLS estimation in censored regression models in [2]. Moreover, there are some studies based on SLS estimation (See [6], [7]). However, there is little research on calculation methods. The traditional Newton method cannot guarantee its convergence for solving such non-convex optimization problem, and the first-order line search method cannot achieve a desired accuracy. So, it is still an interesting and challenging task to solve the non-convex optimization problem corresponding to SLS estimation. In this paper, we attempt to design a systematic numerical method to solve the optimization of SLS estimation.

Optimality conditions can be used to characterize the information of local or global minimizer. The standard optimality condition can only describe local information on a small scale. It follows from the definition of the SLS estimation that the global minimization is necessary. Thus, the standard optimality condition may be weak. For some specific composite optimization problem (COP),

$$\min_{x} f_1(x) + f_2(x),$$

there are some better results. Xu et al. [14] designed a fast solver for the $\ell_{1/2}$ regularized minimization problem. Inspired by this research, Peng et al. [10] proposed the global necessary optimality condition of a class of matrix optimization. Zhou et al. [15] established a global necessary optimality condition for ℓ_0-regularized optimization. Their conclusions show that this optimality condition is stronger than the standard optimality. Moreover, Peng et al. [10] used the proximal gradient algorithm (PGA) (See [14], [10], [4], [8]) to solve this optimization, and established convergence of their algorithm. The premise of these theories and such algorithms is that L-smooth, i.e. the gradient function $\nabla f_1(x)$ is Lipschitz continuous with the Lipschitz constant $L > 0$. However, the optimization of the SLS estimation does not satisfy this condition. Fortunately, under mild conditions, this problem can be overcome. We will give a specific analysis later. In addition, notice that there are two variables we need to calculate. The frameworks of AU method in [16], [17], [18] can be referred. These algorithms provide a reference for us to solve the optimization problem of the SLS estimation.

In this paper, we attempt to design a systematic and complete calculation method for the optimization problem corresponding to SLS estimation in linear regression. First, we give the assumptions of SLS estimation and analyze the complex nature of these assumptions. Second, under these implicit properties, the L-smooth condition can be weakened to the local Lipschitz continuity, and a necessary optimality condition stronger than the standard optimality condition of this problem is proved. Finally, we discuss the numerical computation of this problem, and propose an AU method to solve it. The convergence of this AU method is establied. Finally, some numerical simulations verify the effectiveness of the AU method and the superiority of SLS estimation in linear regression. Our work provides not only an effective calculation method for SLS estimation, but also theoretical support for regularized SLS estimation.

The rest of the paper is organized as follows. In Section 2, we introduce the SLS estimation in linear regression problems and its optimization. In Section 3, we define a class of stationary points and analyze its relationship with the minimizer. In Section 4, we discuss the computation of the optimization problem of SLS estimation, and propose an AU method to solve it. In addition, the convergence of the AU method is established. In Section 5, the effectiveness of AU method is demonstrated by numerical simulation. The final conclusions and discussion are given in Section 6.

2 The SLS Estimation and Optimization Model

Consider the linear regression model

$$y = \mathbf{x}^T\beta + \varepsilon,$$

where $y \in \mathbb{R}$ is the response variable, $\mathbf{x} \in \mathbb{R}^p$ is the predictor variable, $\beta \in \mathbb{R}^p$ is the unknown regression parameter, and ε is the random error satisfying $E(\varepsilon|\mathbf{x}) = 0$ and $E(\varepsilon^2|\mathbf{x}) = \sigma^2$.

Suppose $\{\mathbf{x}_i, y_i\}$, $i = 1, \ldots, n$ is an *i.i.d.* random sample, the optimization problem of the second-order least squares (SLS) estimation is described as

$$\min_{\beta\in\Theta,\sigma^2\in\Sigma} Q_n(\beta,\sigma^2) := \sum_{i=1}^{n} \rho_i^T(\beta,\sigma^2)W_i(\mathbf{x}_i)\rho_i(\beta,\sigma^2), \tag{2.1}$$

where $\rho_i(\beta,\sigma^2) = (y_i - \mathbf{x}_i^T\beta, y_i^2 - (\mathbf{x}_i^T\beta)^2 - \sigma^2)^T$ and $W_i(\mathbf{x}_i)$ is a 2×2 nonnegative definite matrix which may depend on $\mathbf{x}_i$. Here, we assume that $W_i(\mathbf{x}_i)$ is positive definite (see Lemma 9 in Section 5). Further, we denote $\gamma = (\beta^T, \sigma^2)^T$ and assume that the true parameter value $\gamma_{\text{true}} = (\beta_{\text{true}}^T, \sigma_{\text{true}}^2)^T$ lies in the parameter space $\Gamma = \Theta\times\Sigma \subseteq \mathbb{R}^{p+1}$. Note that if the weights are taken as W_i=[1,0;0,0], then the SLS estimator degenerates to the ordinary least squares (OLS) estimator $\beta_{\text{OLS}} = (X^TX)^{-1}X^T\mathbf{y}$, where $X = [\mathbf{x}_1, \ldots, \mathbf{x}_n]^T$ and $\mathbf{y} = (y_1, \ldots, y_n)^T$. It is also worthwhile to note that the problem (2.1) is non-convex optimization even for linear regression.

The SLS estimation was developed for general nonlinear regression models by [13], who established the asymptotic theories under the following regularity assumptions.

Assumption 1. The regression function $m(\mathbf{x};\beta)$ is a measurable function of $\mathbf{x}$ for every $\beta \in \Theta$, and is continuous in $\beta \in \Theta$ for $\mu-$almost all $\mathbf{x}$.

Assumption 2. $E\|W(\mathbf{x})\|_2(\sup_\Theta)(m^4(\mathbf{x};\beta) + 1) < \infty$.

Assumption 3. The parameter space $\Gamma = \Theta \times \Sigma$ is compact.

Assumption 4. For any $\gamma \in \Gamma$, $E[\rho(\gamma) - \rho(\gamma_{\text{true}})]^TW(\mathbf{x})[\rho(\gamma) - \rho(\gamma_{\text{true}})] = 0$ if and only if $\gamma = \gamma_0$.

Assumption 5. β_{true} is an interior point of Θ and $m(x;\beta)$ is twice continuously differentiable in Θ for $\mu-$almost all $\mathbf{x}$. Furthermore, the first and second derivatives of $m(\mathbf{x};\beta)$ satisfy

$$E\|W(\mathbf{x})\|_2 \sup_\Theta \|\nabla_\beta m(\mathbf{x};\beta)\|_2^4 < \infty,\ E\|W(\mathbf{x})\|_2 \sup_\Theta \|\nabla_\beta^2 m(\mathbf{x};\beta)\|_2^4 < \infty.$$

Assumption 6. The matrix $A = E[\nabla_\gamma\rho^T(\gamma_{\text{true}})W(\mathbf{x})\nabla_\gamma\rho(\gamma_{\text{true}})]$ is nonsingular, where

$$\nabla_\gamma\rho^T(\gamma_{\text{true}}) = -\begin{pmatrix} \nabla_\beta m(x;\beta_{\text{true}}) & 2m(x;\beta_{\text{true}})\nabla_\beta m(x;\beta_{\text{true}}) \\ 0 & 1 \end{pmatrix}.$$

However, for linear regression model $m(\mathbf{x};\beta) = \mathbf{x}^T\beta$, some of these assumptions are naturally satisfied, and some other conditions can be simplified or relaxed. In the following we provide some detailed discussion. First, Assumption 1 is obviously satisfied, and Assumption 2 and 5 hold under Assumption 3 and the following assumption.

Assumption 7. $E[\|W(\mathbf{x})\|_2(\|\mathbf{x}\|_2^4 + 1)] < \infty$.

Then, the second derivative of Q_n at true parameter γ_{true} is given by

$$\nabla_\gamma^2 Q_n(\gamma_{\text{true}}) = 2\sum_{i=1}^n [\nabla\rho_i^T(\gamma_{\text{true}})W_i\nabla\rho_i(\gamma_{\text{true}}) + (\rho_i^T(\gamma_{\text{true}})W_i \otimes I_{p+1})\nabla\text{vec}(\nabla\rho_i^T(\gamma_{\text{true}}))],$$

where $\otimes$ is Kronecker product, I_{p+1} is identity matrix, and

$$\nabla_\gamma\text{vec}(\nabla_\gamma\rho_i^T(\gamma_{\text{true}})) = -\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 0 \\ 2\mathbf{x}_i\mathbf{x}_i^T & \mathbf{0} \\ \mathbf{0} & 0 \end{pmatrix},$$

and 0 is the zero in $\mathbb{R}$ and $\mathbf{0}$ is the zero vector or the zero matrix. By Assumption 6,

$$\frac{1}{n}\nabla_\gamma^2 Q_n(\gamma_{\text{true}}) = \frac{2}{n}\sum_{i=1}^n [\nabla\rho_i^T(\gamma_{\text{true}})W_i\nabla\rho_i(\gamma_{\text{true}})] + o_p(1)$$

is positive definite when n is sufficiently large. By some straightforward calculation, it follows from the positive definite of the sub-matrix $\nabla_\beta^2 Q_n(\gamma_{\text{true}})$ of $\nabla_\gamma^2 Q_n(\gamma_{\text{true}})$ that X^TX is a positive definite matrix, which implies that X has full column rank. Further, there is at least one $i \in \{1, \ldots, n\}$ such that $|\mathbf{x}_i^T\beta| \to \infty$ for any β satisfying $\|\beta\|_2 \to \infty$. Indeed, this can be seen by the definition of OLS estimator and

$$\begin{aligned} & L_n(\beta) \\ = \; & L_n(\beta_{\text{OLS}}) + \langle\nabla_\beta L_n(\beta_{\text{OLS}}), \beta - \beta_{\text{OLS}}\rangle + \frac{1}{2}(\beta - \beta_{\text{OLS}})^T(\nabla_\beta^2 L_n(\beta_{\text{OLS}}))(\beta - \beta_{\text{OLS}}) \\ = \; & L_n(\beta_{\text{OLS}}) + \frac{1}{2}(\beta - \beta_{\text{OLS}})^T(X^TX)(\beta - \beta_{\text{OLS}}) \\ \geq \; & L_n(\beta_{\text{OLS}}) + \frac{1}{2}\lambda_{\min}(X^TX)\|\beta - \beta_{\text{OLS}}\|_2^2 \text{ for any } \beta \in \mathbb{R}^p, \end{aligned}$$

where $L_n(\beta) = \sum_{i=1}^n (y_i - \mathbf{x}_i^T\beta)^2$ is the OLS loss, $\lambda_{\min}(X^TX)$ is the minimum eigenvalue of X^TX. It follows that, if $\|\beta\|_2 \to \infty$, then $L_n(\beta) \to \infty$ and therefore $|\mathbf{x}_i^T\beta| \to \infty$ for at least one i. Further, we can show the following result.

Theorem 2.1. *Suppose that $\{W_i\}$ is given. The function Q_n in problem (2.1) is proper, closed, coercive.*

Proof. For any given $\gamma \in \mathbb{R}^p \times \mathbb{R}_+$, we have that $Q_n(\gamma) \geq 0$ and $Q_n(\gamma) < \infty$. Thus, Q_n is proper. In addition, from Corollary 2.9 in [3], Q_n is closed.

Next, we will verify the coerciveness of Q_n. The analysis process is divided into two cases.

Case 1: $\|\beta\|_2$ is infinite, σ^2 is finite or infinite. By the previous analysis, there is at least one $i \in \{1, \ldots, n\}$ such that $|\mathbf{x}_i^T\beta| \to \infty$ if $\|\beta\|_2 \to \infty$. For this i, we have that

$$\begin{aligned}
&\lim_{\|\beta\|_2\to\infty} \frac{(W_i)_{1,1}(y_i - \mathbf{x}_i^T\beta)^2 + (W_i)_{2,2}(y_i^2 - (\mathbf{x}_i^T\beta)^2 - \sigma^2)^2}{2\sqrt{(W_i)_{1,1}(W_i)_{(2,2)}}|y_i - \mathbf{x}_i^T\beta||y_i^2 - (\mathbf{x}_i^T\beta)^2 - \sigma^2|} \\
= &\lim_{\|\beta\|_2\to\infty} \left(\frac{(W_i)_{1,1}|y_i - \mathbf{x}_i^T\beta|}{2\sqrt{(W_i)_{1,1}(W_i)_{2,2}}|y_i^2 - (\mathbf{x}_i^T\beta)^2 - \sigma^2|} + \frac{(W_i)_{2,2}|y_i^2 - (\mathbf{x}_i^T\beta)^2 - \sigma^2|}{2\sqrt{(W_i)_{1,1}(W_i)_{2,2}}|y_i - \mathbf{x}_i^T\beta|}\right) \\
= &\lim_{\|\beta\|_2\to\infty} \left(\frac{(W_i)_{1,1}\frac{|y_i - \mathbf{x}_i^T\beta|}{|\mathbf{x}_i^T\beta|}}{2\sqrt{(W_i)_{1,1}(W_i)_{2,2}}\frac{|y_i^2 - (\mathbf{x}_i^T\beta)^2 - \sigma^2|}{|\mathbf{x}_i^T\beta|}} + \frac{(W_i)_{2,2}\frac{|y_i^2 - (\mathbf{x}_i^T\beta)^2 - \sigma^2|}{|\mathbf{x}_i^T\beta|}}{2\sqrt{(W_i)_{1,1}(W_i)_{2,2}}\frac{|y_i - \mathbf{x}_i^T\beta|}{|\mathbf{x}_i^T\beta|}}\right) \\
= &\ \infty.
\end{aligned}$$

It follows that

$$\begin{aligned}
&\lim_{\|\beta\|_2\to\infty} (W_i)_{1,1}(y_i - \mathbf{x}_i^T\beta)^2 + (W_i)_{2,2}(y_i^2 - (\mathbf{x}_i^T\beta)^2 - \sigma^2)^2 + \\
&\qquad\qquad 2(W_i)_{1,2}(y_i - \mathbf{x}_i^T\beta)(y_i^2 - (\mathbf{x}_i^T\beta)^2 - \sigma^2) \\
\geq &\lim_{\|\beta\|_2\to\infty} (W_i)_{(1,1)}(y_i - \mathbf{x}_i^T\beta)^2 + (W_i)_{2,2}(y_i^2 - (\mathbf{x}_i^T\beta)^2 - \sigma^2)^2 - \\
&\qquad\qquad 2\sqrt{(W_i)_{1,1}(W_i)_{2,2}}|y_i - \mathbf{x}_i^T\beta||y_i^2 - (\mathbf{x}_i^T\beta)^2 - \sigma^2| = \infty,
\end{aligned}$$

Thus,

$$\lim_{\|\gamma\|_2\to\infty} Q_n(\gamma) \geq \lim_{\|\gamma\|_2\to\infty} \rho_i^T(\beta, \sigma^2)W_i\rho_i(\beta, \sigma^2) = \infty.$$

Case 2: $\|\beta\|_2$ is finite, σ^2 is infinite. Notice that $Q_n(\gamma)$ is a quadratic function about σ^2. It is trivial that

$$\lim_{\|\gamma\|_2\to\infty} Q_n(\gamma) = \infty.$$

Thus, Q_n is coercive. This conclusion holds. □

It follows from Theorem 2.14 in [3] that Q_n attains its minimal value over $\mathbb{R}^p \times \mathbb{R}_+$, and there exists an $M > 0$ such that the unique minimizer of Q_n is in $(\mathbb{R}^p \times \mathbb{R}_+) \cap B_{\|\cdot\|_2}[\mathbf{0}, M]$, where $B_{\|\cdot\|_2}[\mathbf{0}, M] := \{\gamma \in \mathbb{R}^{p+1} : \|\gamma\|_2 \leq M\}$ denotes a closed ball in $\mathbb{R}^{p+1}$ with a center of $\mathbf{0}$ and a radius M. The parameter set can be set to $(\mathbb{R}^p \times \mathbb{R}_+) \cap B_{\|\cdot\|_2}[0, M]$, where M is sufficiently large and we can give a suitable M through OLS estimator. Then, Assumption 2 can be omitted. Moreover, by the coerciveness of Q_n, the minimizer of Q_n over $(\mathbb{R}^p \times \mathbb{R}_+) \cap B_{\|\cdot\|_2}[0, M]$ is the minimizer of Q_n over $\mathbb{R}^p \times \mathbb{R}_+$. By Assumption 4, Q_n has a unique minimizer in parameter set when n is sufficiently large. Thus, we can derive the SLS estimator by solving the following problem,

$$\min_{\beta\in\mathbb{R}^p, \sigma^2\in\mathbb{R}_+} Q_n(\beta, \sigma^2). \tag{2.2}$$

Finally, based on the above analysis, the asymptotic properties of SLS estimator in linear regression can be established only under Assumptions 4, 6 and 7.

Theorem 2.2. *(Theorems 1 and 2 in [13]) Let γ_{true} be the true parameter. Then*

(i)(Consistency) under Assumptions 4 and 7, the SLS estimator $\gamma_{\text{SLS}} \xrightarrow{a.s.} \gamma_{\text{true}}$, as $n \to \infty$.

(ii)(Asymptotic normality) under Assumptions 4, 6 and 7, as $n \to \infty$, $\sqrt{n}(\gamma_{\text{SLS}} - \gamma_{\text{true}}) \xrightarrow{L} N(0, A^{-1}BA^{-1})$, where

$$B = E\left(\ \nabla_\gamma\rho^T(\gamma_{\text{true}})W(\mathbf{x})\rho(\gamma_{\text{true}})\rho^T(\gamma_{\text{true}})W(\mathbf{x})\nabla_\gamma\rho(\gamma_{\text{true}})\ \right).$$

$\boxed{3}$ Optimality

In this section, we propose a necessary optimality condition for problem (2.2), which is stronger than the standard optimality condition of problem (2.2). The standard first-order stationary point can be defined as follows.

Definition 3.1. We say that $\widehat{\gamma}$ is a standard first-order stationary point of problem (2.2) if

$$0 \in \nabla_\gamma Q_n(\widehat{\gamma}) + N_{\mathbb{R}^p\times\mathbb{R}_+}(\widehat{\gamma}), \tag{3.1}$$

where $N_{\mathbb{R}^p\times\mathbb{R}_+}(\gamma)$ denotes the normal cone to $\mathbb{R}^p \times \mathbb{R}_+$ at γ.

By some calculation, the specific expression of $N_{\mathbb{R}^p\times\mathbb{R}_+}(\widehat{\gamma})$ can be obtained. Thus, (3.1) is equivalent to

$$\nabla_\beta Q_n(\widehat{\beta}, \widehat{\sigma}^2) = 0 \text{ and } \langle \nabla_{\sigma^2} Q_n(\widehat{\beta}, \widehat{\sigma}^2), \sigma^2 - \widehat{\sigma}^2\rangle \geq 0 \text{ for any } \sigma^2 \geq 0.$$

It is obvious that each local minimizer of the problem (2.2) must be a standard first-order stationary point. However, the SLS estimator is the global minimizer of problem (2.2). Based on definitions of the global necessity condition in [10], [15], we can give the global necessity condition of problem (2.2). Before that, we review the definition of proximal mapping.

Definition 3.2. Given a function $f : \mathbb{R}^p \to (-\infty, \infty]$, the proximal mapping of f is the operator given by

$$\operatorname{prox}_f(\beta) = \arg\min_{u\in\mathbb{R}^p} f(u) + \frac{1}{2}\|u-\beta\|_2^2, \quad \textit{for any } \beta \in \mathbb{R}^p. \tag{3.2}$$

For a fixed sequence $\{W_i\}$, the $Q_n(\gamma)$ can be rewritten as

$$\sum_{i=1}^n [(W_i)_{1,1}(y_i - \mathbf{x}_i^T\beta)^2 + (W_i)_{2,2}(y_i^2 - (\mathbf{x}_i^T\beta)^2 - \sigma^2)^2 + 2(W_i)_{1,2}(y_i - \mathbf{x}_i^T\beta)(y_i^2 - (\mathbf{x}_i^T\beta)^2 - \sigma^2)].$$

For convenience, let

$$h(\beta) := \sum_{i=1}^n (W_i)_{1,1}(y_i - \mathbf{x}_i^T\beta)^2, \text{ and}$$
$$g(\beta,\sigma^2) := \sum_{i=1}^n [(W_i)_{2,2}(y_i^2 - (\mathbf{x}_i^T\beta)^2 - \sigma^2)^2 + 2(W_i)_{1,2}(y_i - \mathbf{x}_i^T\beta)(y_i^2 - (\mathbf{x}_i^T\beta)^2 - \sigma^2)].$$

Thus, the first-order stationary point we consider is defined as follows.

Definition 3.3. We say that $\widehat{\gamma} \in \mathbb{R}^p \times \mathbb{R}_+$ is a first-order stationary point of problem (2.2) if there exists a constant $\widehat{L} > 0$, satisfying the following conditions:

$$\begin{cases} \widehat{\beta} = \operatorname{prox}_{\frac{1}{L}h}(\widehat{\beta} - \frac{1}{L}\nabla_\beta g(\widehat{\beta}, \widehat{\sigma}^2))), & \textit{for any } L \geq \widehat{L}, \\ \widehat{\sigma}^2 = \arg\min_{\sigma^2\geq 0} Q_n(\widehat{\beta}, \sigma^2). \end{cases} \tag{3.3}$$

Note that if $\widehat{\gamma}$ is the minimizer of problem (2.2), then $\widehat{\sigma}^2$ is the minimizer of

$$\min_{\sigma^2 \geq 0} Q_n(\widehat{\beta}, \sigma^2).$$

Hence, we can derive the second item in (3.3). By Definition 3.2, we can see that

$$\begin{cases} \widehat{\beta} = \arg\min_{\beta \in \mathbb{R}^p} \frac{1}{L} h(\beta) + \frac{1}{2} \|\beta - (\widehat{\beta} - \frac{1}{L} \nabla_\beta g(\widehat{\beta}, \widehat{\sigma}^2))\|_2^2, \\ \widehat{\sigma}^2 = \arg\min_{\sigma^2 \geq 0} Q_n(\widehat{\beta}, \sigma^2), \end{cases}$$

which implies (3.1) holds. Thus, the first-order stationary point of problem (2.2) is a standard first-order stationary point. In contrast, the operator in the first term in (3.3) corresponds to a strong convex optimization. If $\widehat{\gamma}$ is a standard first-order stationary point and $\nabla_\beta Q_n(\widehat{\beta}, \widehat{\sigma}^2) = 0$, then $\widehat{\beta}$ is the minimizer of

$$\min_{\beta \in \mathbb{R}^p} \frac{1}{L} h(\beta) + \frac{1}{2} \|\beta - (\widehat{\beta} - \frac{1}{L} \nabla_\beta g(\widehat{\beta}, \widehat{\sigma}^2))\|_2^2$$

for any given $0 < L < \infty$. Thus, by Definition 3.3, the standard first-order stationary point is weaker than first-order stationary point.

Next, we analyze the relationship between the global minimizer and the first-order stationary point of problem (2.2).

Theorem 3.4. *Let $\widehat{\gamma}$ be the minimizer of problem (2.2). Then there exists a constant $\widehat{L} > 0$ such that $\widehat{\gamma}$ is a first-order stationary point of problem (2.2) for any $L \geq \widehat{L}$.*

Proof. Suppose $\widehat{\gamma}$ is a minimizer of problem (2.2), then $\widehat{\sigma}^2$ is the minimizer of problem

$$\min_{\sigma^2 \geq 0} \ Q_n(\widehat{\beta}, \sigma^2),$$

which yields the second item of the definition of the first stationary point.

Further, let γ_M be a point in $\mathbb{R}^n \times \mathbb{R}_+$ satisfying $Q_n(\gamma_M) > Q_n(\widehat{\gamma})$. By the coerciveness of Q_n, there exists $M > 0$ such that

$$Q_n(\gamma) > Q_n(\gamma_M) \text{ for any } \gamma \text{ satisfying } \|\gamma\|_2 > M.$$

This implies that $\|\gamma\|_2 \leq M$ for any γ satisfying $Q_n(\gamma) \leq Q_n(\gamma_M)$. Since $\widehat{\gamma}$ is a minimizer of problem (2.2), then $Q_n(\widehat{\gamma}) \leq Q_n(\gamma_M)$. Thus, $\widehat{\gamma} \in (\mathbb{R}^p \times \mathbb{R}_+) \cap B_{\|\cdot\|_2}[\mathbf{0}, M]$ and $\|\widehat{\beta}\|_2 < \sqrt{M^2 - \widehat{\sigma}^4}$. Define a auxiliary function

$$F(\beta, \widehat{\beta}, \widehat{\sigma}^2, L) := g(\widehat{\beta}, \widehat{\sigma}^2) + \langle \nabla_\beta g(\widehat{\beta}, \widehat{\sigma}^2), \beta - \widehat{\beta} \rangle + \frac{L}{2} \|\beta - \widehat{\beta}\|_2^2 + h(\beta).$$

The function F is strong convex with respect to β. Let $\widetilde{\beta}$ be the minimizer of F. Then we have

$$\langle \nabla_\beta g(\widehat{\beta}, \widehat{\sigma}^2), \widetilde{\beta} - \widehat{\beta} \rangle + \frac{L}{2} \|\widetilde{\beta} - \widehat{\beta}\|_2^2 + h(\widetilde{\beta}) < h(\widehat{\beta}), \text{ for any} L > 0,$$

which along with $h \geq 0$ yields that

$$\frac{L}{2} \|\widetilde{\beta} - \widehat{\beta}\|_2^2 - \|\nabla_\beta g(\widehat{\beta}, \widehat{\sigma}^2)\|_2 \|\widetilde{\beta} - \widehat{\beta}\|_2 - h(\widehat{\beta}) < 0.$$

By simple calculation, we have

$$\|\widetilde{\beta}-\widehat{\beta}\|_2 < \frac{\|\nabla_\beta g(\widehat{\beta},\widehat{\sigma}^2)\|_2 + \sqrt{\|\nabla_\beta g(\widehat{\beta},\widehat{\sigma}^2)\|_2^2 + 2Lh(\widehat{\beta})}}{L}.$$

Furthermore,

$$\|\widetilde{\beta}\|_2 < \|\widehat{\beta}\|_2 + \frac{\|\nabla_\beta g(\widehat{\beta},\widehat{\sigma}^2)\|_2 + \sqrt{\|\nabla_\beta g(\widehat{\beta},\widehat{\sigma}^2)\|_2^2 + 2Lh(\widehat{\beta})}}{L} \le \|\widehat{\beta}\|_2 + I.$$

Note that $\|\widehat{\beta}\|_2 \le \sqrt{M^2-\widehat{\sigma}^4}$ and $I \to 0$ as $L \to \infty$. Thus, there exists a constant $\underline{L}$ such that $I \le M - \sqrt{M^2-\widehat{\sigma}^4}$ for any $L \ge \underline{L}$ and $\|\widetilde{\beta}\|_2 \le M$.

In addition, let $\Psi_\beta := \{\beta \in \mathbb{R}^n : \|(\beta^T,\widehat{\sigma}^2)^T\|_2 \le M\}$ and $L_M := \sup_{\beta\in\Psi_\beta} \|\nabla_\beta^2 g(\beta,\widehat{\sigma}^2)\|_2$. For any $\beta \in \Psi_\beta$, if $L \ge L_M$, we have

$$\begin{aligned}
Q_n(\beta,\widehat{\sigma}^2) &= h(\beta) + g(\beta,\widehat{\sigma}^2)\\
&= h(\beta) + g(\widehat{\beta},\widehat{\sigma}^2) + \langle \nabla_\beta g(\widehat{\beta},\widehat{\sigma}^2), \beta-\widehat{\beta}\rangle\\
&\quad + \frac{1}{2}(\beta-\widehat{\beta})^T \nabla_\beta^2 g(\xi,\widehat{\sigma}^2)(\beta-\widehat{\beta})\\
&= F(\beta,\widehat{\beta},\widehat{\sigma}^2,L) + \frac{1}{2}(\beta-\widehat{\beta})^T \nabla_\beta^2 g(\xi,\widehat{\sigma}^2)(\beta-\widehat{\beta}) - \frac{L}{2}\|\beta-\widehat{\beta}\|_2^2 \qquad (3.4)\\
&\le F(\beta,\widehat{\beta},\widehat{\sigma}^2,L) + \frac{1}{2}\|\nabla_\beta^2 g(\xi,\widehat{\sigma}^2)\|_2 \|\beta-\widehat{\beta}\|_2^2 - \frac{L}{2}\|\beta-\widehat{\beta}\|_2^2\\
&\le F(\beta,\widehat{\beta},\widehat{\sigma}^2,L) + \frac{L_M}{2}\|\beta-\widehat{\beta}\|_2^2 - \frac{L}{2}\|\beta-\widehat{\beta}\|_2^2\\
&\le F(\beta,\widehat{\beta},\widehat{\sigma}^2,L),
\end{aligned}$$

where $\xi = u + \alpha(\beta-u)$ for some $\alpha \in (0,1)$. In other words, as long as L is sufficiently large, (3.4) can be established on Ψ_β.

For any $L \ge \widehat{L} := \max(\underline{L}, L_M)$, we have that

$$Q_n(\widetilde{\beta},\widehat{\sigma}^2) \le F(\widetilde{\beta},\widehat{\beta},\widehat{\sigma}^2,L) \le F(\widehat{\beta},\widehat{\beta},\widehat{\sigma}^2,L) = Q_n(\widehat{\beta},\widehat{\sigma}^2) \le Q_n(\widetilde{\beta},\widehat{\sigma}^2). \qquad (3.5)$$

Therefore, by the strong convexity of F, (3.5) implies that $\widetilde{\beta} = \widehat{\beta}$. From Definition 3.2, we can get that

$$\widehat{\beta} = \mathrm{prox}_{\frac{1}{L}h}(\widehat{\beta} - \frac{1}{L}\nabla_\beta g(\widehat{\beta},\widehat{\sigma}^2))),$$

which is the first term of the definition of the first-order stationary point. Thus, $\widehat{\gamma}$ is a first-order stationary point of problem (2.2) for any $L \ge \widehat{L}$. □

4 Alternate Updating Method

In this section, we consider the numerical optimization for problem (2.2) and propose an AU method, which is a combination of the proximal gradient method and classical alternate updating method. The framework of the AU method is shown below.

Algorithm 1 The alternate updating method

Initialization: Let $0 < L_{\min} < L_{\max}$, $\alpha > 1$, $c > 0$. Choose an initial point (β_0, σ_0^2). Set

$k = 0$.
Step1: Choose $L_k^0 \in [L_{\min}, L_{\max}]$ and set $L_k = L_k^0$.
(1a): Solve subproblems

$$\begin{cases} \beta_k(L_k) = \arg\min_{\beta\in\mathbb{R}^p} h(\beta) + \langle \nabla_\beta g(\beta_k, \sigma_k^2), \beta - \beta_k\rangle + \frac{L_k}{2}\|\beta - \beta_k\|_2^2, \\ \sigma_k^2(L_k) = \arg\min_{\sigma^2>0} Q_n(\beta_{k+1}(L_k), \sigma^2). \end{cases} \tag{4.1}$$

(1b): Go to **Step2**, if

$$Q_n(\beta_k(L_k), \sigma_k^2(L_k)) \le Q_n(\beta_k, \sigma_k^2) - [\frac{c}{2}\|\beta_k(L_k) - \beta_k\|_2^2 + \sum_{i=1}^n (W_i)_{2,2}(\sigma_k^2(L_k) - \sigma_k^2)^2]. \tag{4.2}$$

(1c): $L_k = \alpha L_k$ and go to (**1a**).
Step2: Set $\beta_{k+1} \leftarrow \beta_k(L_k)$, $\sigma_{k+1}^2 \leftarrow \sigma_k^2(L_k)$, $k \leftarrow k+1$ and go to **Step1**.

By simple calculation, subproblems (4.1) have closed-form solutions:

$$\begin{cases} \beta_k(L_k) = (2\widehat{X}^T\widehat{X} + L_k I_n)^{-1}(L_k\beta_k + 2\widehat{X}^T\widehat{y} - \nabla_\beta g(\beta_k, \sigma_k^2)), \\ \sigma_k^2(L_k) = \max\{0, \frac{\sum_{i=1}^n 2(W_i)_{1,2}(y_i - \mathbf{x}_i^T\beta_k(L_k)) + (2(W_i)_{2,2})(y_i^2 - (\mathbf{x}_i^T\beta_k(L_k))^2)}{\sum_{i=1}^n 2(W_i)_{2,2}}\}. \end{cases}$$

where $\widehat{X} = [\widehat{\mathbf{x}}_1, \dots, \widehat{\mathbf{x}}_n]^T$, $\widehat{\mathbf{x}}_i = \sqrt{(W_i)_{1,1}}\mathbf{x}_i$ and $\widehat{y}_i = \sqrt{(W_i)_{1,1}}y_i$, which makes iteration easy.

As we know that the convergence of proximal gradient methods relies on the assumption that g is L-smooth, i.e. the gradient function ∇g is globally lipschitz continuous with lipschitz constant L. Unfortunately, ∇g does not satisfy this assumption in problem (2.2). However, this problem can be overcome under the coerciveness of Q_n. Next, we establish the convergence of the AU method. The convergence analysis refers to Proposition A.1. in [5].

We first define the following quantities:

$$A := \sup_{\|\gamma\|_2 \le M_0} \|\nabla g(\gamma)\|_2,\ B := \sup_{\|\beta\|_2 \le M_0} h(\beta),\ L_g := \sup_{\|\gamma\|_2 \le M_0 + \Delta} \|\nabla^2 g(\gamma)\|_2,$$

where M_0 and Δ are given constants, which can be seen from the following theorem.

Theorem 4.1. *Let the sequence $\{\beta_k, \sigma_k^2\}$ be generated by the AU method. The following statements hold:*
(i) $Q(\beta_{k+1}, \sigma_{k+1}^2) \le Q(\beta_0, \sigma_0^2)$ for all $k \ge 0$.
(ii) $\{L_k\}$ is bounded.
(iii) For each $k > 0$, the descent criterion (4.2) holds after at most

$$\left\lfloor \frac{\log(\overline{L} + c) - \log(L_{\min})}{\log(\alpha)} + 1 \right\rfloor$$

inner iterations.

Proof. (i) When $k = 0$, by the coerciveness of Q_n, there exist $M_0 > 0$ such that $\|\gamma_0\|_2 < M_0$ and $Q_n(\gamma) > Q_n(\gamma_0)$ for any γ satisfying $\|\gamma\|_2 > M_0$, which implies that $\|\gamma\|_2 \le M_0$ for any γ satisfying $Q_n(\gamma) \le Q_n(\gamma_0)$. For any $L > 0$, set

$$\beta_0(L) = \arg\min_{\beta\in\mathbb{R}^p} h(\beta) + \langle \nabla_\beta g(\beta_0, \sigma_0^2), \beta - \beta_0\rangle + \frac{L}{2}\|\beta - \beta_0\|_2^2,$$

which along with $h \geq 0$ yields

$$\frac{L}{2}\|\beta_0(L) - \beta_0\|_2^2 - \|\nabla_\beta g(\beta_0, \sigma_0^2)\|_2 \|\beta_0(L) - \beta_0\|_2 - h(\beta_0) \leq 0.$$

Thus, we have

$$\|\beta_0(L) - \beta_0\|_2 \leq \frac{\|\nabla_\beta g(\beta_0, \sigma_0^2)\|_2 + \sqrt{\|\nabla_\beta g(\beta_0, \sigma_0^2)\|_2^2 + 2Lh(\beta_0)}}{L}.$$

Furthermore,

$$\|\beta_0(L)\|_2 \leq \|\beta_0\|_2 + \frac{A + \sqrt{A^2 + 2LB}}{L},$$

which implies that there exists constant $\underline{L}$ such that $\|\beta_0(L)\|_2 \leq M_0 + \Delta$ for any $L \geq \underline{L}$. Indeed, $\|\beta_0\|_2 \leq M_0$ and

$$\frac{A + \sqrt{A^2 + 2LB}}{L} \to 0 \text{ as } L \to \infty.$$

On the other hand, let $\overline{L} = \max\{\underline{L}, L_g\}$. Then

$$\begin{aligned} g(\beta_0(L), \sigma_0^2) &\leq g(\beta_0, \sigma_0^2) + \langle \nabla_\beta g(\beta_0, \sigma_0^2), \beta_0(L) - \beta_0 \rangle + \frac{\overline{L}}{2}\|\beta_0(L) - \beta_0\|_2^2 \\ &=: f(\beta_0(L), \beta_0, \sigma_0^2, \overline{L}). \end{aligned}$$

It follows that

$$\begin{aligned} Q_n(\beta_0(L), \sigma_0^2) &= g(\beta_0(L), \sigma_0^2) + h(\beta_0(L)) \\ &\leq f(\beta_0(L), \beta_0, \sigma_0^2, \overline{L}) + h(\beta_0(L)) \\ &\leq f(\beta_0(L), \beta_0, \sigma_0^2, L) + h(\beta_0(L)) - \frac{c}{2}\|\beta - \beta_0\|_2^2 \\ &\leq f(\beta_0, \beta_0, \sigma_0^2, L) + h(\beta_0(L)) - \frac{c}{2}\|\beta - \beta_0\|_2^2 \\ &= Q_n(\beta_0, \sigma_0^2) - \frac{c}{2}\|\beta - \beta_0\|_2^2 \qquad (4.3) \end{aligned}$$

for any $L \geq \overline{L} + c$, which implies that $\|(\beta_0(L)^T, \sigma_0^2)^T\|_2 \leq M_0$.

In addition, by simple calculation, if $\sigma_1^2 > 0$, then

$$Q_n(\beta_0(L), \sigma_0^2(L)) - Q_n(\beta_0(L), \sigma_0^2) = -\sum_{i=1}^n (W_i)_{2,2}(\sigma_0^2(L) - \sigma_0^2)^2. \qquad (4.4)$$

Otherwise, if $\sigma_1^2 = 0$, then

$$Q_n(\beta_0(L), \sigma_0^2(L)) - Q_n(\beta_0(L), \sigma_0^2) \leq -\sum_{i=1}^n (W_i)_{2,2}(\sigma_0^2(L) - \sigma_0^2)^2. \qquad (4.5)$$

Combining (4.3), (4.4), (4.5), the descent criterion (9) holds and $Q_n(\beta_1, \sigma_1^2) \leq Q_n(\beta_0, \sigma_0^2)$ for any $L_1 \geq \overline{L} + c$ when $k = 0$.

We now suppose that statements (i) hold for all $k \leq K$ for some $K > 0$. Thus, $\|\gamma_k\|_2 \leq M_0$. Repeat the above analysis, we have $Q_n(\beta_{k+1}, \sigma_{k+1}^2) \leq Q_n(\beta_k, \sigma_k^2)$ for any $L_{k+1} \geq \overline{L} + c$

when $k = K + 1$. Further, by induction hypothesis, we have $Q(\beta_{k+1}, \sigma^2_{k+1}) \le Q(\beta_0, \sigma^2_0)$ and the statement (i) holds for any $L_{k+1} \ge \overline{L} + c$.

(ii) Note that $\underline{L}$, L_g and $\overline{L}$ are bounded if M_0 and Δ are given, and their values are fixed. When $L_k \ge \overline{L} + c$, the inequality (4.2) must be established. Thus, we have $L_k \le \alpha(\overline{L} + c)$ for any $k \ge 0$. Hence, the statement (ii) holds.

(iii) Let J_k denote the number of inner iterations at the kth iteration. Then,

$$L_{\min}\alpha^{J_k-1} \le L^0_k \alpha^{J_k-1} \le \overline{L} + c.$$

It follows that

$$J_k \le \log_\alpha(\frac{\overline{L} + c}{L_{\min}}) = \frac{\log(\overline{L} + c) - \log(L_{\min})}{\log(\alpha)}.$$

Thus, the statement (iii) holds. □

Finally, we establish the convergence of AU method.

Theorem 4.2. *Let the sequence $\{\beta_k, \sigma^2_k\}$ be generated by the AU method. The following statements hold:*

(i) $\{\beta_k, \sigma^2_k\}$ *is bounded;*

(ii) $\lim_{k\to\infty} \|\beta_{k+1} - \beta_k\|^2_2 = 0$ *and* $\lim_{k\to\infty} |\sigma^2_{k+1} - \sigma^2_k| = 0$.

(iii) $\lim_{k\to\infty} L_k\|\beta_{k+1} - \beta_k\|^2_2 = 0$.

(iv) *Any accumulation point of $\{\beta_k, \sigma^2_k\}$ is a first-order stationary point of problem (2.2).*

Proof. (i) By the (i) in Theorem 4.1,

$$Q(\beta_k, \sigma^2_k) \le Q(\beta_0, \sigma^2_0) \text{ for all } k \ge 0.$$

Since Q_n is coercive, we have $\|(\beta^T_k, \sigma^2_k)^T\|_2 \le M_0$ and $\{\beta_k, \sigma^2_k\}$ is bounded.

(ii) It follows from descent criterion (4.2) that

$$\frac{c}{2}\|\beta_{k+1} - \beta_k\|^2_2 + \sum_{i=1}^{n}(W_i)_{2,2}(\sigma^2_k(L) - \sigma^2_k)^2 \le Q_n(\beta_k, \sigma^2_k) - Q_n(\beta_{k+1}, \sigma^2_{k+1}),$$

and $\lim_{k\to\infty} Q_n(\gamma_k) = \zeta$. Thus,

$$\begin{aligned} \sum_{k=1}^{\infty}\frac{c}{2}\|\beta_{k+1} - \beta_k\|^2_2 + \sum_{i=1}^{n}(W_i)_{2,2}(\sigma^2_k(L) - \sigma^2_k)^2 \quad &\le \quad \sum_{k=1}^{\infty}(Q_n(\beta_k, \sigma^2_k) - Q_n(\beta_{k+1}, \sigma^2_{k+1})) \\ &= \quad Q_n(\beta_0, \sigma^2_0) - \zeta \\ &\le \quad Q_n(\beta_0, \sigma^2_0) < \infty, \end{aligned}$$

which implies that

$$\lim_{k\to\infty}\frac{c}{2}\|\beta_{k+1} - \beta_k\|^2_2 + \sum_{i=1}^{n}(W_i)_{2,2}(\sigma^2_k(L) - \sigma^2_k)^2 = 0.$$

Thus, the statement (ii) holds.

(iii) From (ii) in Theorem 4.1, we can see that $\{L_k\}$ is bounded. Thus,

$$\lim_{k\to\infty} L_k\|\beta_{k+1}-\beta_k\|_2 = 0.$$

(iv) By the boundedness of $\{(\beta_k,\sigma_k^2)\}$, for any accumulation point $(\widehat{\beta},\widehat{\sigma}^2)$, there exists a subsequence $\{\beta_{k_j},\sigma_{k_j}^2\}$ such that $\lim_{k_j\to\infty}\beta_{k_j}=\widehat{\beta}$ and $\lim_{k_j\to\infty}\sigma_{k_j}^2=\widehat{\sigma}$, where $k_j\to\infty$ as $j\to\infty$. From statement(ii), we have that

$$\begin{aligned}
&\|\beta_{k_j+1}-\widehat{\beta}\|_2 \le \|\beta_{k_j+1}-\beta_{k_j}\|_2+\|\beta_{k_j}-\widehat{\beta}\|_2\to 0,\ \text{as } k_j\to\infty,\\
&|\sigma_{k_j+1}^2-\widehat{\sigma}^2| \le |\sigma_{k_j+1}^2-\sigma_{k_j}^2|+|\sigma_{k_j}^2-\widehat{\sigma}^2|\to 0,\ \text{as } k_j\to\infty.
\end{aligned}$$

Thus, $\beta_{k_j+1}\to\widehat{\beta}$ and $\sigma_{k_j+1}^2\to\widehat{\sigma}^2$. From the AU method, we have

$$\begin{aligned}
\beta_{k_j+1} &= \arg\min_{\beta\in\mathbb{R}^p} h(\beta)+\langle\nabla_\beta g(\beta_{k_j},\sigma_{k_j}^2)),\beta-\beta_{k_j}\rangle+\frac{L_{k_j}}{2}\|\beta-\beta_{k_j}\|_2^2,\\
\sigma_{k_j+1}^2 &= \arg\min_{\sigma^2\ge 0} Q_n(\beta_{k_j+1},\sigma^2).
\end{aligned}$$

Furthermore,

$$\begin{aligned}
h(\beta_{k_j+1}) \quad + \quad & \langle\nabla_\beta g(\beta_{k_j},\sigma_{k_j}^2)),\beta_{k_j+1}-\beta_{k_j}\rangle+\frac{L_{k_j}}{2}\|\beta_{k_j+1}-\beta_{k_j}\|_2^2 \le\\
& h(\beta)+\langle\nabla_\beta g(\beta_{k_j},\sigma_{k_j}^2)),\beta-\beta_{k_j}\rangle+\frac{L_{k_j}}{2}\|\beta-\beta_{k_j}\|_2^2,
\end{aligned}$$

Taking limit $k_j\to\infty$, using continuity of the h and $\nabla_\beta g$, and by the Theorem 3.4, we immediately obtain that there exist an $\widetilde{L}$ such that

$$\begin{aligned}
\widehat{\beta} &= \arg\min_{\beta\in\mathbb{R}^p} h(\beta)+\langle\nabla_\beta g(\widehat{\beta},\widehat{\sigma}^2)),\beta-\widehat{\beta}\rangle+\frac{L}{2}\|\beta-\widehat{\beta}\|_2^2,\quad \forall L\ge\widetilde{L},\\
\widehat{\sigma}^2 &= \min_{\sigma^2\ge 0} Q_n(\widehat{\beta},\widehat{\sigma}^2).
\end{aligned}$$

By simplifying above formulas, the accumulation point of $\{\beta_k,\sigma_k^2\}$ is a first-order stationary point of problem (2.2). □

5 Numerical Simulations

In this section, we study the performance of the AU method for solving problem (2.2) by numerical simulations. Before that, we give the calculation method of the optimal weighting matrix in Q_n.

Lemma 5.1. *(Corollary in [13]) If $\sigma_{\text{true}}^2(\mu_4-\sigma_{\text{true}}^4)-\mu_3^2\neq 0$, then the optimal weighting matrix $\{\widehat{W}_i\}$, $i=1,\dots,n$ is given by*

$$\begin{aligned}
\widehat{W}_i &= \frac{1}{\sigma_{\text{true}}^2(\mu_4-\sigma_{\text{true}}^4)-\mu_3^2}\\
&\times \begin{pmatrix} \mu_4+4\mu_3\mathbf{x}_i^T\beta_{\text{true}}+4\sigma_{\text{true}}^2(\mathbf{x}_i^T\beta_{\text{true}})^2-\sigma_{\text{true}}^4 & -\mu_3-2\sigma_{\text{true}}^2\mathbf{x}_i^T\beta_{\text{true}}\\ -\mu_3-2\sigma_{\text{true}}^2\mathbf{x}_i^T\beta_{\text{true}} & \sigma_{\text{true}}^2 \end{pmatrix},
\end{aligned}$$

where $\mu_3 = E(\varepsilon^3|X)$ and $\mu_4 = E(\varepsilon^4|X)$. Moreover, the asymptotic covariance matrix of the most efficient SLS estimator is given by

$$C = \begin{pmatrix} V(\beta_{\text{SLS}}) & \frac{\mu_3}{\mu_4-\sigma_{\text{true}}^4} V(\sigma_{\text{SLS}}^2) G_2^{-1} G_1 \\ \frac{\mu_3}{\mu_4-\sigma_{\text{true}}^4} V(\sigma_{\text{SLS}}^2) G_1^T G_2^{-1} & V(\sigma_{\text{SLS}}^2) \end{pmatrix},$$

where

$$V(\beta_{\text{SLS}}) = (\sigma_{\text{true}}^2 - \frac{\mu_3^2}{\mu_4 - \sigma_{\text{true}}^4})(G_2 - \frac{\mu_3^2}{\sigma_{\text{true}}^2(\mu_4 - \sigma_{\text{true}}^4)} G_1 G_1^T)^{-1},$$
$$V(\sigma_{\text{SLS}}^2) = \frac{(\mu_4 - \sigma_{\text{true}}^4)(\sigma_{\text{true}}^2(\mu_4 - \sigma_{\text{true}}^4) - \mu_3^2)}{\sigma_0^2(\mu_4 - \sigma_{\text{true}}^4 - \mu_3^2 G_1^T G_2^{-1} G_1)},$$

and

$$G_1 = E(\mathbf{x}), \quad G_2 = E(\mathbf{x}\mathbf{x}^T).$$

The Lemma 5.1 provides the optimal weight matrix sequence and the asymptotic covariance matrix of the most efficient SLS estimator. However, some parameters in the above lemma are unknown. We use the two-stage procedure in [13] to calculate the optimal weight matrix sequence, and derive the SLS estimator.

We consider numerical simulation with three error distributions: normal distribution $N(0,1)$, student t distribution, chi square distribution $(\chi^2(3) - 3)/\sqrt{6}$. We also apply the SLS estimation on a real dataset. The mean squared error (MSE) and Variance (Var) are used to compare the quality of the SLS estimator and the OLS estimator. They are defined as follows:

Replicate $N_s = 100$ times simulations. For each $j \in \{1, \ldots, p\}$, calculate the mean estimator

$$\overline{\beta}_j = \frac{1}{N_s} \sum_{i=1}^{N_s} \beta_{i,j}.$$

The MSE for each coefficient is calculated by

$$\text{MSE}(\beta_j) = \frac{1}{N} \sum_{i=1}^{N} (\beta_{i,j} - (\beta_{\text{true}})_j)^2, \; j = 1, \ldots, p.$$

where $\beta_{i,j}$ denote the $j-$th element of the estimator in the $i-$th simulation. The Var for each coefficient is calculated by

$$\text{Var}(\beta_j) = \frac{1}{N} \sum_{i=1}^{N} (\beta_{i,j} - (\overline{\beta})_j)^2, \; j = 1, \ldots, p.$$

For the AU method, the OLS estimator is taken as the initial point. The stopping criterion as follows:

$$\frac{\|\beta_k - \beta_{k+1}\|_2}{\max(1, \|\beta_{k+1}\|_2)} \leq \nu,$$

or the maximum iterative time of 5000s is reached.

All the numerical experiments were performed in MATLAB (R2019b) on a laptop with an Intel(R) Core(TM)i5-6200 CPU(2.40GHz) and 8GB of RAM.

5.1 Linear model without intercept

We consider linear regression without an intercept term

$$y_i = \mathbf{x}_i^T \beta + \varepsilon_i,\ \ i = 1, \ldots, m.$$

where $\mathbf{x}_i^T$ is normal with mean 0 and its correlation between $\mathbf{x}_i$ and $\mathbf{x}_j$ is $0.5^{|i-j|}$. Let $p = 4$, $m = 50,\ 100,\ 200,\ 500$, and $\beta_{\text{true}} = (2.5, 0.6, -0.5, -2.3)^T$.

The simulation results for this linear regression are presented in Tables 1, 2, 3. All results show that the SLS estimator and the OLS estimator are close to the true parameter as the sample size increases, and both Var and MSE are decreasing. From the first two tables, the two estimators and their two evaluation indicators are very close, and the tiny gap between them decreases as the sample size increases. These gaps are caused by the finiteness of samples and calculation errors. However, this will not affect the theoretical equivalence of the two estimators. This conclusion also implies the effectiveness of our proposed algorithm. In the case of $\varepsilon_i \sim (\chi^2(3) - 3)/\sqrt{6}$, the MSE and Var of the SLS estimator are smaller than the OLS estimator, and this gap will not decrease significantly as the sample size increases, which means that when the random error distribution is asymmetric, the SLS eatimator is asymptotically more efficient than the OLS eatimator.

These conclusions not only show the superiority of the SLS estimation for linear regression but also verify the effectiveness of our calculation method.

Table 1: Simulation results with $\varepsilon(i) \sim N(0, 1)$.

	SLSE	Var	MSE	OLSE	Var	MSE
m=50						
β_1	2.498	3.693e-03	3.696e-03	2.498	2.892e-03	2.896e-03
β_2	0.601	4.876e-03	4.878e-03	0.601	3.630e-03	3.631e-03
β_3	-0.498	4.885e-03	4.888e-03	-0.499	3.779e-03	3.779e-03
β_4	-2.307	3.565e-03	3.617e-03	-2.305	2.820e-03	2.844e-03
m=100						
β_1	2.501	1.488e-03	1.489e-03	2.501	1.345e-03	1.346e-03
β_2	0.599	1.933e-03	1.933e-03	0.600	1.678e-03	1.678e-03
β_3	-0.499	1.860e-03	1.862e-03	-0.498	1.625e-03	1.627e-03
β_4	-2.301	1.612e-03	1.614e-03	-2.301	1.427e-03	1.428e-03
m=200						
β_1	2.501	6.491e-04	6.497e-04	2.501	6.184e-04	6.192e-04
β_2	0.599	9.239e-04	9.246e-04	0.599	8.959e-04	8.967e-04
β_3	-0.500	9.431e-04	9.431e-04	-0.500	9.026e-04	9.026e-04
β_4	-2.299	7.590e-04	7.593e-04	-2.300	7.323e-04	7.324e-04
m=500						
β_1	2.500	2.734e-04	2.734e-04	2.500	2.680e-04	2.681e-04
β_2	0.600	3.264e-04	3.264e-04	0.600	3.228e-04	3.229e-04
β_3	-0.500	3.523e-04	3.523e-04	0.500	3.454e-04	3.454e-04
β_4	-2.300	2.575e-04	2.577e-04	-2.300	2.546e-04	2.548e-04

5.2 Linear model with intercept

We consider linear regression with intercept term

$$y_i = \mathbf{x}_i^T \beta + \beta_{in} + \varepsilon_i,\ \ i = 1, \ldots, m.$$

Table 2: Simulation results with $\varepsilon(i) \sim t(5)$.

	SLSE	Var	MSE	OLSE	Var	MSE
m=50						
β_1	2.502	5.924e-03	5.926e-02	2.505	5.088e-03	5.109e-03
β_2	0.599	6.506e-03	6.507e-02	0.596	5.771e-03	5.784e-03
β_3	-0.502	6.962e-03	6.966e-02	-0.501	6.309e-03	6.309e-03
β_4	-2.298	5.435e-03	5.439e-02	-2.299	5.118e-03	5.119e-03
m=100						
β_1	2.502	2.322e-03	2.327e-03	2.502	2.275e-03	2.278e-03
β_2	0.599	3.198e-03	3.199e-03	0.599	3.196e-03	3.197e-03
β_3	-0.503	2.807e-03	2.816e-03	-0.502	2.812e-03	2.818e-03
β_4	-2.298	2.157e-03	2.162e-03	-2.298	2.123e-03	2.127e-03
m=200						
β_1	2.501	1.046e-03	1.048e-03	2.501	1.076e-03	1.077e-03
β_2	0.599	1.336e-03	1.337e-03	0.599	1.414e-03	1.415e-03
β_3	-0.499	1.421e-03	1.422e-03	0.500	1.511e-03	1.511e-03
β_4	-2.300	1.106e-03	1.106e-03	-2.300	1.168e-03	1.168e-03
m=500						
β_1	2.500	4.379e-04	4.380e-04	2.500	4.491e-04	4.491e-04
β_2	0.600	5.741e-04	5.743e-04	0.595	5.949e-04	5.952e-04
β_3	-0.499	5.697e-04	5.704e-04	-0.499	5.750e-04	5.753e-04
β_4	-2.301	4.616e-04	4.626e-04	-2.301	4.737e-04	4.748e-04

Table 3: Simulation results with $\varepsilon(i) = (\chi^2(3) - 3)/\sqrt{6}$.

	SLSE	Var	MSE	OLSE	Var	MSE
m=50						
β_1	2.500	2.182e-03	2.182e-03	2.500	3.272e-03	3.272e-03
β_2	0.602	2.317e-03	2.321e-03	0.602	3.787e-03	3.790e-03
β_3	-0.500	2.515e-03	2.515e-03	-0.500	3.741e-03	3.741e-03
β_4	-2.302	2.104e-03	2.108e-03	-2.300	2.859e-03	2.859e-03
m=100						
β_1	2.501	8.688e-04	8.704e-04	2.502	1.434e-03	1.438e-03
β_2	0.601	1.007e-03	1.009e-03	0.600	1.700e-03	1.700e-03
β_3	-0.501	1.072e-03	1.074e-03	-0.502	1.834e-03	1.837e-03
β_4	-2.300	8.866e-04	8.866e-04	-2.299	1.434e-03	1.434e-03
m=200						
β_1	2.501	3.878e-04	3.882e-04	2.500	6.606e-04	6.607e-04
β_2	0.600	4.917e-04	4.918e-04	0.600	8.261e-04	8.261e-04
β_3	-0.500	4.860e-04	4.861e-04	-0.500	8.486e-04	8.488e-04
β_4	-2.301	3.663e-04	3.670e-04	-2.300	6.788e-04	6.790e-04
m=500						
β_1	2.500	1.479e-04	1.480e-04	2.500	2.824e-04	2.825e-04
β_2	0.599	1.865e-04	1.868e-04	0.600	3.356e-04	3.357e-04
β_3	-0.500	1.874e-04	1.874e-04	-0.500	3.241e-04	3.242e-04
β_4	-2.300	1.525e-04	1.525e-04	-2.300	2.826e-04	2.827e-04

where $\beta_{in} \in \mathbb{R}$ is the intercept term. Set $\widehat{X} = [X\vdots\mathbf{1}]$, where $\mathbf{1}$ is the vector in $\mathbb{R}^m$ with all components are 1, the OLS estimator is calculated by $(\widehat{X}^T\widehat{X})^{-1}\widehat{X}^T y$. We set $\beta_0 = [2.5, 0.6, -0.5]^T$. The rest of the settings are the same as Section 5.1.

Simulation results are given in displayed in Tables 4, 5, 6. For β_i, $i = 1, 2, 3$, we have the same conclusion as in the previous section. However, the OLS estimator and SLS estimator for β_{in} have similar performance. Indeed, it follows from Lemma 1 in [7], the existence of the intercept term yields to $G_1^T G_2^{-1} G_1 = 1$. Further, Theorem 4 in [13] explains this phenomenon.

Table 4: Simulation results with $\varepsilon(i) \sim N(0,1)$.

	SLSE	Var	MSE	OLSE	Var	MSE
m=50						
β_{in}	-2.298	2.231e-02	2.232e-02	-2.296	2.107e-02	2.109e-02
β_1	2.508	3.508e-02	3.515e-02	2.509	3.015e-02	3.023e-02
β_2	0.601	4.155e-02	4.155e-02	0.600	3.493e-02	3.493e-02
β_3	-0.503	3.613e-02	3.614e-02	-0.501	2.983e-02	2.983e-02
m=100						
β_{in}	-2.300	1.000e-02	1.000e-02	-2.299	9.916e-03	9.916e-03
β_1	2.499	1.564e-02	1.564e-02	2.500	1.455e-02	1.455e-02
β_2	0.603	1.839e-02	1.834e-02	0.602	1.742e-02	1.743e-02
β_3	-0.502	1.547e-02	1.548e-02	-0.503	1.470e-02	1.471e-02
m=200						
β_{in}	-2.299	5.083e-03	5.084e-03	-2.299	5.111e-03	5.112e-03
β_1	2.504	6.739e-03	6.756e-03	2.504	6.586e-03	6.603e-03
β_2	0.601	8.427e-03	8.428e-03	0.601	8.172e-03	8.173e-03
β_3	-0.504	6.828e-03	6.842e-03	-0.504	6.737e-03	6.752e-03
m=500						
β_{in}	-2.298	2.051e-03	2.054e-03	-2.298	2.051e-03	2.055e-03
β_1	2.501	2.434e-03	2.435e-03	2.500	2.446e-03	2.446e-03
β_2	0.598	3.476e-03	3.479e-03	0.598	3.468e-03	3.471e-03
β_3	-0.499	2.715e-03	2.715e-03	-0.499	2.716e-03	2.717e-03

5.3 Real data example

The first two sets of simulations not only verified the effectiveness of the AU method for problem (2.2), but also demonstrated that the SLS estimator is asymptotically more efficient than the OLS estimator if the third moment of the random error is nonzero. We applied the proposed method on a real data set in this subsection.

The paper [9] provids a data set, which included the house price information and the 13 predictor variables. This data set was taken from the StatLib library which is maintained at Carnegie Mellon University. We can download it from https://archive.ics.uci.edu/ml/machine-learning-databases/housing/. In this data set, MEDV (Median value of owner-occupied homes in $1000's$) is the response variable, and CRIM (per capita crime rate by town), ZN (proportion of residential land zoned for lots over 25,000 sq.ft), INDUS (proportion of non-retail business acres per town), CHAS (Charles River dummy variable (= 1 if tract bounds river; 0 otherwise)), NOX (nitric oxides concentration (parts per 10 million)), RM (average number of rooms per dwelling), AGE (proportion of owner-occupied units built prior

Table 5: Simulation results with $\varepsilon(i) \sim t(5)$.

	SLSE	Var	MSE	OLSE	Var	MSE
m=50						
β_{in}	-2.303	3.139e-02	3.140e-02	-2.300	3.360e-02	3.360e-02
β_1	2.503	5.255e-02	5.256e-02	2.511	4.937e-02	4.950e-02
β_2	0.611	6.211e-02	6.222e-02	0.609	6.216e-02	6.223e-02
β_3	-0.504	5.440e-02	5.442e-02	-0.504	5.558e-02	5.560e-02
m=100						
β_{in}	-2.299	1.679e-02	1.679e-02	-2.299	1.786e-02	1.786e-02
β_1	2.502	2.360e-02	2.361e-02	2.501	2.472e-02	2.472e-02
β_2	0.602	2.858e-02	2.859e-02	0.603	2.942e-02	2.943e-02
β_3	-0.506	2.145e-02	2.149e-02	-0.508	2.185e-02	2.192e-02
m=200						
β_{in}	-2.300	8.926e-03	8.926e-03	-2.300	9.256e-03	9.261e-03
β_1	2.499	9.726e-03	9.727e-02	2.501	1.030e-02	1.03e-02
β_2	0.597	1.320e-02	1.321e-02	0.597	1.398e-02	1.399e-02
β_3	-0.500	1.072e-02	1.072e-02	-0.500	1.117e-02	1.117e-02
m=500						
β_{in}	-2.298	3.478e-03	3.483e-03	-2.298	3.530e-03	3.535e-03
β_1	2.502	4.095e-03	4.101e-03	2.502	4.306e-03	4.311e-03
β_2	0.596	5.084e-03	5.101e-03	0.596	5.185e-03	5.200e-03
β_3	-0.500	4.383e-03	4.383e-03	-0.500	4.523e-03	4.523e-03

Table 6: Simulation results with $\varepsilon(i) = (\chi^2(3) - 3)/\sqrt{6}$.

	SLSE	Var	MSE	OLSE	Var	MSE
m=50						
β_{in}	-2.341	1.991e-02	2.158e-02	-2.302	2.021e-02	2.021e-02
β_1	2.497	1.686e-02	1.687e-02	2.499	2.808e-02	2.808e-02
β_2	0.600	2.173e-02	2.173e-02	0.599	3.576e-02	3.576e-02
β_3	-0.501	1.750e-02	1.750e-02	-0.495	2.936e-02	2.938e-02
m=100						
β_{in}	-2.314	9.802e-03	9.991e-03	-2.295	9.791e-03	9.812e-03
β_1	2.496	7.922e-03	7.937e-03	2.498	1.399e-02	1.399e-02
β_2	0.605	1.040e-02	1.042e-02	0.610	1.795e-02	1.706e-02
β_3	-0.502	8.694e-03	8.699e-03	-0.504	1.533e-02	1.534e-02
m=200						
β_{in}	-2.308	4.622e-03	4.693e-03	-2.299	4.649e-03	4.650e-03
β_1	2.502	3.837e-03	3.841e-03	2.501	7.298e-03	7.299e-03
β_2	0.600	5.300e-03	5.300e-03	0.601	9.169e-03	9.170e-03
β_3	-0.500	3.716e-03	3.716e-03	-0.501	6.813e-03	6.814e-03
m=500						
β_{in}	-2.302	1.963e-03	1.968e-03	-2.299	1.968e-03	1.969e-03
β_1	2.499	1.372e-03	1.374e-03	2.500	2.685e-03	2.695e-03
β_2	0.601	1.870e-03	1.871e-03	0.600	3.602e-03	3.602e-03
β_3	-0.501	1.455e-03	1.458e-03	-0.501	2.704e-03	2.704e-03

to 1940), DIS (weighted distances to five Boston employment centres), RAD (index of accessibility to radial highways), TAX (full-value property-tax rate per $10,000$), PTRATIO (pupil-teacher ratio by town), B ($1000(Bk-0.63)^2$ where Bk is the proportion of blacks by town), LSTAT (% lower status of the population) .

We select suitable predictor variables by the correlation coefficient (co-co) between response variable and predictor variables. The calculation results are shown in Table 7. The RM, PTRATIO, LSTAT have relatively strong relations with MEDV. We can also see these relationships intuitively through Figure 1.

Table 7: The correlation coefficient.

MEDV	co-co	MEDV	co-co
CRIM	-0.388	DIS	0.250
ZN	0.360	RAD	-0.382
INDUS	-0.484	TAX	-0.469
CHAS	0.175	PTRATIO	-0.508
NOX	-0.427	B	0.333
RM	0.695	LSTAT	-0.738
AGE	-0.377		

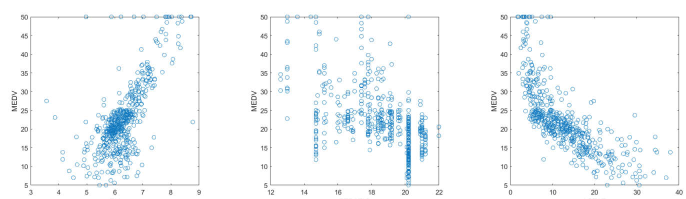

Figure 1: The Scatter plots between MEDV and RM, PTRATIO, LSTAT.

Consider the following linear model to fit the data set:

$$y=\beta_0+x_1\beta_1+x_2\beta_2+x_3\beta_3+\varepsilon,$$

where β_0 is the intercept term. We replicate 100 experiments through the Bootstrap method. The results of OLS estimator and SLS estimator are shown in Table 8. The variance of the SLS estimator is lower than the OLS estimator. In addition, we can get the corresponding residual histograms of both estimators. From Figure 2, we can see that the random error distribution is slightly asymmetric.

$\boxed{6}$ Conclusions

The SLS estimation is the estimation method that makes full use of the second-order moment information of the data and have good statistical theoretical properties. It is asymptotically more efficient than the OLS estimator if the third moment of the random error is nonzero. In this paper, we propose the AU method to calculate SLS estimator in linear regression based on a stronger optimality condition. Numerical experiments show that our method can effectively solve problem (2.2), and also verify the superiority of SLS estimation. This paper provides some basis for the extension of SLS estimation in high-dimensional regression.

Table 8: Real data simulation results.

	SLS	Var	OLS	Var
β_0	18.04	5.99	18.56	6.59
β_1	4.69	1.01e-02	4.52	1.07e-02
β_2	-0.99	2.60e-03	-0.93	3.50e-03
β_3	-0.54	5.997e-04	-0.57	1.10e-03

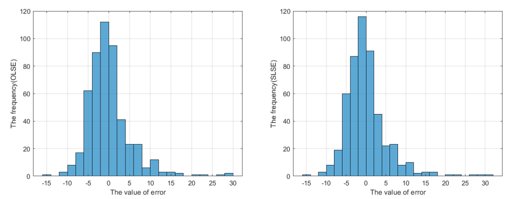

Figure 2: The histograms of residuals of two estimators.

Acknowledgments

We sincerely thank the referees as well as the associate editor for their constructive comments which have improved significantly the quality of the paper.

References

[1] T. Abarin and L. Wang, Comparison of gmm with second-order least squares estimation in nonlinear models, *Far East Journal of Theoretical Statistics* 20 (2006) 179–196.

[2] T. Abarin and L. Wang, Second-order least squares estimation of censored models, *Journal of Statistical Planning and Inference* 139 (2009) 125–135.

[3] A. Beck, *First-Order Methods In Optimization,* Society for Industrial and Applied Mathematics Philadelphia, 2017.

[4] A. Beck and M. Teboulle, A fast iterative shrinkage thresholding algorithm for linear inverse problems, *SIAM Journal on Imaging Sciences* 2 (2009) 183–202.

[5] X. Chen, Z. Lu and T. Pong, Penalty methods for a class of non-lipschitz optimization problems, *SIAM Journal on Optimization* 26 (2016) 1465–1492.

[6] X. Chen, M. Tsao and J. Zhou, Robust second-order least squares estimator for regression, *Statistical Papers* 53 (2012) 371–386.

[7] L. Gao and J. Zhou, New optimal design criteria for regression models with asymmetric errors, *Journal of Statistical Planning and Inference*, 149 (2014) 140–151.

[8] P. Gong, C. Zhang, Z. Lu, J.H. Huang and J. Ye, A general iterative shrinkage and thresholding algorithm for non-convex regularized optimization problems, *Proceedings of the 30th International Conference on Machine Learning* 28 (2013) 37–45.

[9] D. Harrison and D. Rubinfeld, Hedonic housing prices and the demand for clean air, *Journal of Environmental Economics and Management* 5 (1978) 81–102.

[10] D. Peng, N. Xiu and J. Yu, Global optimality and fixed point continuation algorithm for non-lipschitz ℓ_p regularized matrix minimization, *Science China Mathematics* 61 (2018) 1139–1152.

[11] L. Wang, Estimation of nonlinear berkson-type measurement error models, *Statistica Sinica* 13 (2003) 1201–1210.

[12] L. Wang, Estimation of nonlinear models with berkson measurement errors, *The Annals of Statistics* 32 (2004) 2559–2579.

[13] L. Wang and A. Leblanc, Second-order nonlinear least squares estimation, *Annals of the Institute of Statistical Mathematics* 60 (2008) 883–900.

[14] Z. Xu, X. Chang, F. Xu and H. Zhang, $\ell_{1/2}$ regularization: A thresholding representation theory and a fast solver, *IEEE Transactions on Neural Networks and Learning Systems* 23 (2012) 1013–1027.

[15] S. Zhou, L. Pan and N. Xiu, Newton method for ℓ_0-regularized optimization, *Numerical Algorithms* 88 (2021) 1541–1570.

[16] Y. Zhang, An alternating direction algorithm for nonnegative matrix factorization. *IEEE Transactions on Neural Networks and Learning Systems.* 23 (2012), 1013–1027.

[17] Y. Xu, W. Yin and Z. Wen, An alternating direction algorithm for matrix with nonnegative factors. *Frontiers of Mathematics in China.* 7 (2012), 365–384.

[18] L. Yang, T.K. Pong and X. Chen, A non-monotone alternating updating method for a class of matrix factorization problems. *SIAM Journal on Optimization.* 28 (2018), 3402–3430.

Manuscript received 13 September 2021
revised 22 November 2021
accepted for publication 1 December 2021

Xin Wang
Department of Applied Mathematics, Beijing Jiaotong University
No.3 Shangyuancun, Haidian District, Beijing, 100044, P.R. China
E-mail address: 18118020@bjtu.edu.cn

Lingchen Kong
Department of Applied Mathematics, Beijing Jiaotong University
No.3 Shangyuancun, Haidian District, Beijing, 100044, P.R. China
E-mail address: konglchen@126.com

Liqun Wang
Department of Statistics, University of Manitoba
186 Dysart Road, Winnipeg, Manitoba, Canada
E-mail address: Liqun.Wang@umanitoba.ca