Lasso estimator for high-dimensional repeated measurements model

Cite as: AIP Conference Proceedings **2292**, 020002 (2020); https://doi.org/10.1063/5.0030795
Published Online: 27 October 2020

Naser Oda Jassim, and Abdul Hussein Saber Al-Mouel

Lasso Estimator for High-Dimensional Repeated Measurements Model

Naser Oda Jassim[1,a)] and Abdul HusseinSaber Al-Mouel[2,b)]

[1]*Department of Mathematics, Collage of Education for Pure Sciences, University of Basrah, Iraq.*

Corresponding Author: [a)]noj1972.khalef@gmail.com
[b)]abdulhusseinsaber@yahoo.com

Abstract: In this paper, we propose the lasso method for choice of penalty level and investigate the error of the lasso estimator in repeated measurements model. We introduce our repeated measurements model with its expectation error and its derivative. we investigate the choice of level the penalty parameter for the lasso estimator which plays important role in investigation the lasso estimator and also show that with high probability, the random variable is very close to its expectation. Lastly, we investigate the error of lasso estimator and conclude that with high probability the lasso estimator will pick out most of the important variables.

INTRODUCTION

Many scientists and researchers have been given a definition for the repeated measurements in the different periods of time. [1] defined them as term used to describe the data in which observations of response variable are measured repeatedly for each experimental unit under different experimental conditions. While [2] explained that the repeated measurements require two or more independent groups between the most of known experimental designs in the set of different researches type. In the other words, in the repeated measurements , the observations of experimental units are measured repeatedly in the time unit.

The linear regression model is said to be high-dimensional model when the number of explanatory variables exceeds the number of observations. In other hand, the model is called low-dimensional model when the number of explanatory variables is less than the number of observations. Whatever the linear regression model whether low or high-dimensional, we desire to satisfy some important properties which are: estimation, prediction and variables selection.

A least squares estimator can be obtained by minimizing the residual sum of squares. As long as an inverse of the explanatory matrix exists, this leads to have a unique solution to the given problem. But this method cannot apply on the problem which has high-dimensional data. In the sense that, when the number of unknown coefficients which are to be estimated is larger than the number of sample size. In this case, the uniqueness of solution cannot be obtained. Furthermore the traditional methods, like all possible regression, forward regression, backward regression and step-wise regression cannot be used for the variables selection in the high-dimensional data.

As mentioned above, it is cleared that these methods cannot be applied due to increment in the number of coefficients on the account of sample size. In this case, another methods must be utilized instead of conventional method for the estimation and variables selection. It is known that, the common and suitable method which can be employed to treat with the high-dimensional data is called penalized least squares method. The main use of this method is to overcome the computational problems and also improves the prediction accuracy. The penalized method is based on the principle of reduction the residual sum of squares with the some constraints on the unknown parameters. The estimations of penalized method can be obtained by minimizing an objective function which consists of two parts: loss function and penalty function. The general form of the penalized method is given in the following equation,

$$\mathrm{P}_{LS}(\lambda,\beta) = (Y - X\beta)'(Y - X\beta) + \mathrm{P}_i(\lambda,\beta),$$

Proceedings of the 2020 2nd International Conference on Sustainable Manufacturing, Materials and Technologies
AIP Conf. Proc. 2292, 020002-1–020002-10; https://doi.org/10.1063/5.0030795
Published by AIP Publishing. 978-0-7354-4024-1/$30.00

where $\mathrm{P}_i(\lambda, \beta)$ represents the penalty function and λ is penalty parameter. Therefore, the penalized estimator can be attained according to the following form,

$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} = \{(Y - X\beta)'(Y - X\beta) + \mathrm{P}_i(\lambda, \beta)\}$

The penalized method is characterized bythe property thatmaking the penalized estimator and variables selection simultaneously. It is noted that the best penalized function gives an estimator that characterized by three important features which are: unbiasedness, sparsity and continuity. Moreover, As a mentioned in [3] that the estimator from the ideal aspect must have the oracle properties which are consistency and asymptotic normal.

One of the most commonly used penalized methods in the high-dimensional data is called 'lasso' which is proposed by [4]. The word lasso is an abbreviation from 'least absolute shrinkage and selection operator'. The principle of the lasso method is based on the minimizing the residual sum of squares plus the sum of absolute value of coefficients. The lasso estimator of linear regression model can be obtained according to following equation

$$\hat{\beta}_{lasso} = arg\min_{\beta}\left\{\frac{1}{N}\sum_{i=1}^{N}(Y_i - X_i'\beta)^2 + \lambda\sum_{j=1}^{k}\left|\beta_j\right|\right\}$$

where λ is the penalty parameter while $\lambda\sum_{j=1}^{k}\left|\beta_j\right|$ is called penalty function.

The lasso becomes popular and more attractive method through selects a suitable variables and maintains the properties that make some regression coefficients exactly zero and shrinks others to a certain magnitude with reduction the loss function. In [5] introduced a new algorithm for lasso estimator when $\gamma = 1$ in the penalty function $\sum\left|\beta_j\right|^{\gamma}$ bridge function. While in [6] studied the asymptotic behavior of lasso estimator as especial case when $\gamma = 1$ in bridge regression. In [7] Suggested a new algorithm to compute lasso estimator called LARS which is abbreviation to the least angle regression. They showed that this algorithm is simple in application and gives accuracy results. In [8] introduced notes on lasso, LARS and forward stage-wise regression and show that the set of variables selection are not consistently the true set of basic variables when prediction accuracy is used as criterion. [9] studied the lasso properties in terms of the variables selections and estimation in a least absolute deviations. [10] discussed and derived the degrees of freedom of lasso fit without any assumption on the explanatory matrix X. They expressed their results in terms of active set of lasso solution-and developed this results to include the degrees of freedom of design matrix X together with an arbitrary matrix [11] suggested an assumtionless consistency of lasso. He showed that for the lasso considered in Tibshirani original paper, the lasso is consistent under almost no assumption at all. [1] used lasso mothed to demonstrate and examine the feature selection task. [3] explained that the two estimators: the square-root lasso and square-root slope can achieve the optimal minimax prediction rate. [12] introduced generalized lasso problem and studied its uniqueness by using $\mathcal{L}_1$-penalty of matrix D times coefficient vector.In this paper, we will use Lasso method in repeated measurements model to investigate some important and essential principles related with the estimated error of Lasso. Firstly, we introduce our repeated measurements model then discussing the expectation error with its derivative. Secondly, we study and investigate the choice of penalty level that satisfies lasso estimator. Lastly, we investigate the error of lasso estimator and conclude that with high probability the lasso estimator will pick out most of the significant variables [13].

SETTING MODEL

Consider the following repeated measurements linear model

$$Y_{it} = \mu + \sum_{j=1}^{k}\beta_j X_{j(it)} + \upsilon_i + \epsilon_{it} \quad (1)$$

$$i = 1, \dots, N\ , t = 1, \dots, T, where$$

$Y_{it}$the value of response variable for i^{th} unit at time t,

$X_{j(it)}$ the explanatory variables,

$\mu, \beta_j,\ j = 1, \dots, k$ are fixed parameters,

υ_i is the random effect with $\upsilon_i \sim N(0, \sigma_\upsilon^2)$,

ϵ_{it} is an error term with $\epsilon_{it} \sim N(0, \sigma_\epsilon^2)$.

The model (1) can be written as follows:

$$Y_{it} = \mu + \sum_{j=1}^{k}\beta_j X_{j(it)} + \omega_{i_t} \quad (2)$$

Where $\omega_{i_t} = \upsilon_i + \epsilon_{it},\ \omega_{i_t} \sim N(0, \sigma_\omega^2), \sigma_\omega^2 = \sigma_\epsilon^2 + \sigma_\upsilon^2$.

This by using matrix notation the model (2) is

$$Y = G\theta + \omega \,, \tag{3}$$

Where, $G = [e, X], e = [1,1, \ldots ,1]'$ has length NT,
$Y = [Y_{11}, \ldots , Y_{1T}, Y_{21}, \ldots , Y_{N1}, \ldots , Y_{NT}]'$ has length NT,
$X = [X_1, \ldots , X_N]'$ is a $NT \times K$ design matrix of fixed effects,
$\theta = \left[\mu, \beta_1, \beta_{2,} \ldots , \beta_k\right]'$ Has length $k + 1$, and
$\omega = [\omega_{11}, \ldots , \omega_{1T}, \omega_{21}, \ldots , \omega_{N1}, \ldots , \omega_{NT}]'$ has length NT.
From model (3), we have $Y \sim N_{NT}(G\theta, \Sigma)$, where

$$E(\omega\omega') = E(Y - G\theta)(Y - G\theta)' = \Sigma$$

METHODOLOGY

Our target in this paper is to reconstruct the unknown vector $\theta \in R^k$ in high-dimensional case where $k > N$, k is the number of coefficients and N is the sample size. A key assumption of our model is that ℓ_1 −penalized least squares estimator must be thresholding rule that meaning a model with small variables so that it can be interpreted easily. Here we assume that, $S = supp(\theta)$ has $k < N$ elements. The set S of nonzero coefficients is unknown. If $k > N$, the Ordinary least squares is not consistent and it will be used other methods to deal with such case. In recent years, the high dimensional linear model problems have been solved by many new methods. These methods based on ℓ_1 penalization or ℓ_1 norm. It is concluded that for most ℓ_1 penalization methods, the error structure performs a significant role in the estimation of unknown coefficients. One of these methods that relies on error distribution is called lasso (least absolute shrinkage and selection operator) which minimizes the residual sum of squares subject to the sum of absolute value of coefficients being less than constant. From (3) Then the lasso estimator is the solution to

$$\min_{\beta}(y - G\theta)'(y - G\theta), \quad \text{subject to } \textstyle\sum_{j=1}^{p}\left|\theta_j\right| \le t. \tag{4}$$

Here $t \ge 0$ is tuning parameter or equivalently

$$\min_{\beta}(y - G\theta)'(y - G\theta) + \lambda \textstyle\sum_{j=1}^{p}\left|\theta_j\right| \tag{5}$$

The formulation (4) and (5) are equivalent in the sense that, for any given $\lambda \in [0, \infty)$, there exists a $t > 0$ such that the two problems have the same solution and vice versa. In this paper, we introduce analysis for the lasso method and weinvestigate the selection of penalty level, which does not rely on any unknown parameters or the error distribution. It will be discussed important twofold. First, we will propose a rule for setting the penalty level. This choice of penalty will be comprehensive and the mean of errors will be assumed that 0 and $p(\omega_{it} = 0) = 0$ for all i. Second, it will be explained that the estimator with high probability has near oracle performance, i.e. with high probability

$$\left\|\hat{\theta} - \theta\right\|_2 = o\left(\sqrt{\frac{p \log k}{N}}\right), or\ eqivalenty\ \left\|\hat{\theta} - \theta\right\|_2^2 = 0(\frac{p \log k}{N}).$$

It is important to see that there is no any assumption on the distribution or moments of errors, a scale parameter is only needed to control the tail probability of error.We state some lemmas which are used to prove the main theorem.
Lemma 1 Suppose ω_{it} be any continuous random variable, then

$$\frac{dE\left(\left||\omega_{it} + G| - |\omega_{it}|\right|\right)}{dG} = 1 - 2P(\omega_{it} \le -G).$$

Proof: Since$\left||\omega_{it} + G| - |\omega_{it}|\right| \le |G|$ is boundedthe random variable, this implies that its expectation must be existed. Assume the probability of density function of ω_{it} $is\ f(\omega)\ and\ G > 0$. It is easy to see that

$$E(|\omega_{it} + G| - |\omega_{it}|) = \int_0^{\infty} f(t)G\,dt + \int_{-G}^{0} f(t)(2t + G)dt - \int_{-\infty}^{-G} f(t)G\,dt$$

$$= \int_0^{\infty} f(t)G\,dt + \int_{-G}^{0} f(t)Gdt - \int_{-\infty}^{-G} f(t)G\,dt + 2\int_{-G}^{0} t\,f(t)dt$$

$$= G\left(\int_0^{\infty} f(t)dt + \int_{-G}^{0} f(t)dt - \int_{-\infty}^{-G} f(t)dt\right) + 2\int_{-G}^{0} t\,f(t)\,dt$$

$$= G\left(\int_{-G}^{\infty} f(t)dt - \int_{-\infty}^{-G} f(t)\,dt\right) + 2\int_{-G}^{0} t\,f(t)\,dt$$

$$= G[P(\omega_{it} \geq -G) - P(\omega_{it} \leq -G)] + 2\int_{-G}^{0} t\, f(t)\, dt$$

$$= G(1 - 2P(\omega_{it} \leq -G)) + 2\int_{-G}^{0} t\, f(t)\, dt. \tag{6}$$

Hence by taking the derivative with respect to G both sides it is easy to see that

$$\frac{dE(|\omega_{it}+G| - |\omega_{it}|)}{dG} = 1 - 2P(\omega_{it} \leq -G).$$

Lemma 2. Assume that there exists a constant $r > 0$ such that the random variable ω_{it} satisfies the following conditions

$$P(\omega_{it} \geq G) \leq \frac{1}{2+rG} \quad \text{for all } G \geq 0$$

$$P(\omega_{it} \leq G) \leq \frac{1}{2+r|G|} \quad \text{for all } G < 0, \tag{7}$$

where r introduce as a scale parameter of the distribution of ω_{it}.Then

$$E(|\omega_{it} + c| - |\omega_{it}|) \geq \frac{r}{16}|c|\left(|c| \wedge \frac{6}{r}\right). \tag{8}$$

Proof. we have that for any $c > 0$,

$E(|\omega_{it} + c| - |\omega_{it}|) = c - 2\int_0^c P(\omega_{it} < -(G)dG,$ by (4)

$$\geq c - 2\int_0^c \frac{1}{2+rG}\,dG = c - \frac{2}{r}\int_0^c \frac{r}{2+rG}\,dG = \left[c - \frac{2}{r}\log(2+rG\right]_0^c$$

$$= c - \frac{2}{r}[\log(2+2c) - \log(2)]$$

$$= c - \frac{2}{r}\log\left(1 + \frac{r}{2}c\right). \tag{9}$$

i.e.$E(|\omega_{it} + c| - |\omega_{it}|) \geq c - \frac{2}{r}\log\left(1+\frac{r}{2}c\right)$

when $c \geq \frac{6}{r}$, Then

$$c - \frac{2}{r}\log\left(1+\frac{r}{2}c\right) \geq c - \frac{2}{4}\frac{rc}{r} = \frac{c}{2}$$

and when $c \leq \frac{6}{r}$, then

$$c - \frac{2}{r}\log\left(1+\frac{r}{2}c\right) \geq c - \frac{2}{r}\left(\frac{rc}{2} - \frac{1}{8}\left(\frac{rc}{2}\right)^2\right)$$

$$= c - \frac{2}{r}\frac{rc}{2}\left(1 - \frac{rc}{16}\right)$$

$$= c - c + \frac{rc^2}{16} = \frac{rc^2}{16}.$$

In a similar manner, for any real number c when $|c| \geq \frac{6}{r}$ then we have that ,

$$E(|\omega_{it} + c| - |\omega_{it}|) \geq \frac{|c|}{2}$$

and when $|c| \leq \frac{6}{r}$, then

$$E(|\omega_{it} + c| - |\omega_{it}|) \geq \frac{rc^2}{16}.$$

The lemma is proved by Putting the above inequalities together.

Definition 1. (Union Bound)[4].

Let $E_1, E_2, \ldots,$ be a finite or countably infinite set of events, not necessary disjoint then union bound is defined as $P(\bigcup_{i\geq 1} E_i) \leq \sum_{i\geq 1} P(E_i)$. **(Hoeffding's inequality)** [10]**.** Let $Y_1, \ldots, Y_n$ be independent bounded random variables such that $E(Y_i) = \mu$ and $\leq Y_i \leq b$.

Then, for every $> 0, P\left(\left|\overline{Y}_n - \mu\right| \geq \epsilon\right) \leq 2e^{-\frac{2n\epsilon^2}{(b-a)^2}}$.

Choice of Penalty

Assume that for any $\gamma \in R^k$, let

$$Q(\gamma) = \frac{\|Y - G\gamma\|_2^2}{N}. \tag{10}$$

Then the lasso estimator can be expressed as

$$\hat{\theta} \quad \in argmin\{\gamma : Q(\gamma) + \lambda\|\gamma\|_1\}, \tag{11}$$

which means that the lasso estimator is equal or belong to minimize the residual sum of squares plus sum of absolute values of coefficients. Suppose that the measurement errors ω_{it} satisfy $\mathrm{P}(\omega_{it} = 0) = 0$ and $E(\omega_{it}) = 0$ for $i = 1, \dots, N$ and $t = 1, \dots, T$. The penalty level λ can be determined bytaking the sub-differential of (10) with respect to the point of true coefficient $\theta = \gamma$ for all $\omega_{it} \neq 0, i = 1, \dots, N$,which can be written as

$S = \frac{2}{N} G'(sign(\omega_{1t}), sign(\omega_{2t}), \dots, sign(\omega_{Nt}))'$, where $sign(x)$ denotes the sign of x,

i.e. $sign(x) = \begin{cases} 1 & if\ x > 0 \\ -1 & if\ x < 0 \\ 0 & if\ x = 0 \end{cases}$.

Let $I = sign(\omega),\ then\ I = (I_{1t}, I_{2t}, \dots, I_{Nt})'$, where $I_i = sign(\omega_{it}), i = 1, \dots, N$. Since $\omega_{it}'s$ are independent and have mean 0 so that The sub-differential of $Q(\gamma)$at the point of θ, can be written as

$$S = \frac{2}{N} G' I. \tag{12}$$

A penalty parameter λ will be selected such that it dominates the estimation error and it is greater than the maximum absolute value of S with high probability [15]. In other words, we want to select a penalty level λ such that

$$\mathrm{P}(\lambda \geq c\|S\|_\infty) \geq 1 - \alpha, \text{ for some fixed constant } c > 1. \tag{13}$$

It is clear that the distribution of I is known therefore for any given G the distribution of $\|S\|_\infty$ is known and does not rely on any unknown parameter. We have the following lemmas.

Lemma 3. The selection of penalty level $\lambda = c\sqrt{2A(\alpha)N\ log\ k}$ for the $lasso$ estimator of $\theta \in argmin\{\gamma : Q(\gamma) + \lambda\|\gamma\|_1\}$ satisfies $P(\lambda \geq c\|S\|_\infty) \geq 1 - \alpha$.

Proof: by using the definition of the union bound, it can be seen that

$P(c\sqrt{2A(\alpha)N\ logk} \leq c\|S\|_\infty) \leq \sum_{i=1}^{N} P(\sqrt{2A(\alpha)N\ logk} \leq |G_i' I|)$.

For each i, by using Hoefddiding inequality,

$$P(|G_i'\ I| \geq \sqrt{2A(\alpha)N\ logk}\ ) \leq 2exp\left\{-\frac{4A(\alpha)Nlogk}{4\|G_i\|_2^2}\right\} = 2exp\{logk^{-A(\alpha)}\} = 2k^{-A(\alpha)},$$

since $\|G_i\|_2^2 = N\ for\ all\ i$. Therefore,

$P(c\sqrt{2A(\alpha)Nlogk} \leq c\|S\|_\infty) \leq k^2 k^{-A(\alpha)} = k^{-(A(\alpha)-2)} \leq \alpha$

This implies that, $P(\lambda \geq c\|S\|_\infty) \geq 1 - \alpha$.

Lemma 4. the choice of penalty $\lambda = c\sqrt{N}\Phi^{-1}(1 - \frac{\alpha}{2k})$ for the $lasso$ estimator of $\theta \in argmin\{\gamma : Q(\gamma) + \lambda\|\gamma\|_1\}$ satisfies $P(\lambda \geq c\|S\|_\infty) \geq 1 - \alpha(1 + w_N)$, when $w_N \to 0$ as $N \to \infty$, where $B = \underset{N}{sup} \underset{1 \leq j \leq k}{sup} \frac{1}{N}\|G_j\|_q^q < \infty$ for some constant $q > 2$ and $\Phi^{-1}\left(1 - \frac{\alpha}{2k}\right) \leq (q-2)\sqrt{logN}$.

Proof: By using the definition of the union bound, it can be seen that

$$P(c\sqrt{N}\Phi^{-1}\left(1 - \frac{\alpha}{(2k)}\right)) \leq c\|S\|_\infty) \leq \sum_{i=1}^{k} P\left(\sqrt{N}\Phi^{-1}\left(1 - \frac{\alpha}{(2k)}\right)\right) \leq |G_i' I|.$$

For each i, by using Stastinkov, Robin-Sethuraman inequality[7],

$$P\left(\sqrt{N}\Phi^{-1}(1 - \frac{\alpha}{(2k)}) \leq |G_i' I|\right) \leq 2\left(1 - \Phi\left(\Phi^{-1}\left(1 - \frac{\alpha}{(2k)}\right)\right)\right)(1 + w_N)$$

$$= 2(1 - \left(1 - \frac{\alpha}{(2k)}\right)(1 + w_N) = 2(1 - 1 + \frac{\alpha}{(2k)})(1 + w_N) = \frac{\alpha}{k}\ (1 + w_N)$$

where $w_N \to 0$ as $N \to \infty$ infinity, provided that $\Phi^{-1}\left(1 - \frac{\alpha}{2k}\right) \leq (q-2)\sqrt{logN}$.

Therefore $P(c\sqrt{N}\Phi^{-1}\left(1-\frac{\alpha}{(2k)}\right)) \leq c\|S\|_\infty) \leq \alpha((1+w_N)$.

This implies that, $P\left(\lambda \geq c\|S\|_\infty\right) \geq 1-\alpha(1+w_N)$.

Lemma 5. If $\lambda > c\|S\|_\infty$ for some fixed constant $c > 1$, then the estimator error $h = \hat{\theta} - \theta$ belong to set $\Delta_{\overline{C}}$ where $\Delta_{\overline{C}} = \left\{ \begin{matrix} \delta \in R^k: \|\delta_F\|_1 \geq \overline{C}\|\delta_{F^c}\|_1, \; where\ F \subset \{1,2,\ldots,k\} \\ and\ F\ contains\ at\ most\ p\ elements. \end{matrix} \right\}$

Proof. To prove this important feature of the $lasso$ estimator, since $\hat{\theta}$ minimizes $\left\{\frac{\|G\gamma - Y\|_2^2}{N} + \lambda\|\gamma\|_1\right\}$. therefore

$$\|Gh+\omega\|_2^2 + \lambda\|\hat{\theta}\|_1 = \left\|G(\hat{\theta}-\theta) + G\theta - Y\right\|_2^2 + N\,\lambda\|\hat{\theta}\|_1$$

$$= \left\|G\hat{\theta} - G\theta + G\theta - Y\right\|_2^2 + N\lambda\|\hat{\theta}\|_1$$

$$= \left\|G\hat{\theta} - Y\right\|_2^2 + N\lambda\|\hat{\theta}\|_1$$

$$\leq \frac{\|G\theta - Y\|_2^2}{N} + \lambda\|\theta\|_1 = \|\omega\|_2^2 + \lambda\|\theta\|_1$$

i.e. $\|Gh+\omega\|_2^2 + \lambda\|\hat{\theta}\|_1 \leq \|\omega\|_2^2 + \lambda\|\theta\|_1$

Let F denote the set of significant coefficients. Then

$$\|Gh+\omega\|_2^2 - \|\omega\|_2^2 \leq \lambda\left(\|\theta\|_1 - \|\hat{\theta}\|_1\right) = \lambda(\|h_F\|_1 - \|h_{F^c}\|_1). \tag{14}$$

Since the sub-differential of $Q(\gamma)$ at the point θ is $\frac{2}{N}G'I$, where $I = sign(\omega)$.

$$\|Gh+\omega\|_2^2 - \|\omega\|_2^2 \geq (Gh)'I \geq h'G'I \geq -\|h\|_1\|G'I\|_\infty \geq -\frac{N\lambda}{c}(\|h_F\|_1 - \|h_{F^c}\|_1).$$

So $\|h_F\|_1 \geq \overline{C}\|h_{F^c}\|_1$, where $\overline{C} = \frac{c-1}{c+1}$ (15)

SOME IMPORTANT NOTATIONS

Now, some important quantities and assumptions of design matrix G shall be defined and they are very important in proving the following lemma and theorem. Throughout, each vector G_i is assumed to be normalized such that $\|G_i\|_2^2 = N$ for $i = 1, \ldots, N$. Let λ_p^u be the smallest number such that for any p sparse vector $d \in R^k$,

$$\|Gd\|_2^2 \leq N\lambda_p^u\|d\|_2^2.$$

Here $k\ sparse\ vector\ d$ means that the vector d has at most k nonzero coordinate, or $\|d\|_0 \leq p$. Similarly, let λ_p^l be the largest number such that for any $k\ sparse\ vector\ d \in R^k$, $\|Gd\|_2^2 \geq N\lambda_p^l\|d\|_2^2$. Let ψ_{p_1,p_2} be the smallest number such that for any $p_1\ and\ p_2$ sparse vector $c_1\ and\ c_2$ with disjoint support,

$$|\langle Gc_1, Gc_2\rangle| \leq N\psi_{p_1,p_2}\|c_1\|_2\|c_2\|_{2.}$$

The following constrained eigenvalue of design matrix G must be defined. Let

$$p_p^l(\overline{C}) = \min_{h\in\Delta_{\overline{C}}} \frac{\|Gh\|_2}{N\|h_F\|_2}.$$

To show the properties of the $lasso$ high dimensional repeated measurements estimator, we need $p_p^l(\overline{C})$ to be bounded away from 0.. We have the following lemma which plays important role to prove next theorem.

Definition 2. ($\epsilon - Net$)[10]

Consider a subset $K\ of\ \mathbb{R}^n$ and let $\varepsilon > 0$. A subset $N \subseteq K$ is called an $\epsilon - Net$ of K if every point in K is within distance ε of some point of N,

i.e. $\forall x \in K\ \exists x_0 \in N: \|x - x_0\|_2 \leq \varepsilon$.

Equivalently, $N \subseteq K$ is an $\epsilon - Net$ of K if and only if K can be covered by balls with centers in N and radii ε.

Definition 3. (Covering numbers)[10]

The smallest cardinality of an $\epsilon - Net$ of K is called the covering number of K and is denoted $N(K,\varepsilon)$. Equivalently, the $N(K,\varepsilon)$ is the smallest number of closed balls with centers in K and radii ε whose union covers K.

Lemma 6. Suppose $\omega_{it}^{\prime}s$ are independent random variables. Assume that $k > N$ and $k > 3p_p^u$ and also let for any vector $d \in R^k$,

$B(d) = \frac{1}{N}|(\|Gd + \omega\|_2^2 - \|\omega\|_2^2) - E(\|Gd + \omega\|_2^2 - \|\omega\|_2^2)|$ then

$$P\left(\sup_{\|d\|_0 = p, \|d\|_2 = 1} B(d) \geq (1 + 2C_1\sqrt{\lambda_p^u})\sqrt{2p\ logk}\right) \leq 2k^{-4p(C_1^2 - 1),} \qquad (16)$$

where $C_1 > 1$ is a constant.

Proof. for any $1 \leq i \leq N$, $1 \leq t \leq T$, it can be seen

$||(Gd)_{it} - \omega_{it}| - |\omega_{it}|| \leq |(Gd)_{it}|$, where $t = 1,2, \ldots, T$.

So $|(Gd)_{it} - \omega_{it}| - |\omega_{it}|$ is a bounded random variable for any fixed d.

Hence for any fixed p sparse signal $d \in R^k$, by using Hoeffding's inequality, we have that

$P(B(d) \geq \tau) \leq 2\ exp\left(-\frac{N\tau^2}{2\|Gd\|_2^2}\right)$, for all $\tau > 0$.

By the definition of λ_p^u, we have that

$$P(B(d) \geq \tau) \leq 2\ exp\left(-\frac{-N\tau^2}{2N\lambda_p^u\|d\|_2^2}\right) = 2exp\left\{-\frac{\tau^2}{2\lambda_p^u\|d\|_2^2}\right\}. \qquad (17)$$

Let $\tau = C\sqrt{2p\ logk}\|d\|_2^2$ in (15), we have

$P\left(B(d) \geq C\sqrt{2p\ log\ k}\|d\|_2^2\right) \leq 2exp\left(-\frac{2C^2\ p\ log\ k\ \|d\|_2^2}{2\lambda_p^u\ \|d\|_2^2}\right) = 2k^{\frac{-pC^2}{\lambda_p^u}}$, for all $C > 0$.

By using the $\epsilon - Net$ and covering number argument, we will find an upper bound for $\sup_{\|d\|_0 = p, \|d\|_2 = 1}|B(d)|$,

Consider the $\epsilon - Net$ of the set $\{d \in R^k, \|d\|_0 = p, \|d\|_2 = 1\}$.

Since From the standard results of covering number, the covering number of $\{d \in R^p, \|d\|_2 = 1\}\ by\ \epsilon$ balls $(i.e. \{y \in R^p : \|y - x\|_2 \leq \epsilon\})$ is at most $\left(\frac{3}{\epsilon}\right)^p$ for $\epsilon < 1$.

Then the covering number of $\{d \in R^k,\ \|d\|_0 = p,\ \|d\|_2 = 1\}$ by $\epsilon\ balls$ is at most $\left(\frac{3k}{\epsilon}\right)^p$ for $\epsilon < 1$.

Assume that S is such a $\epsilon - Net$ of $\{d \in R^k,\ \|d\|_0 = p,\ \|d\|_2 = 1\}$. By union bound,

$P\left(\sup_{d \in S}|B(d)| \geq C\sqrt{2p \log k}\right) \leq 2\left(\frac{3}{\epsilon}\right)^p k^p\ \ k^{\frac{-pC^2}{\lambda_p^u}}$, for all $C > 0$.

Moreover, it can be seen that,

$$\sup_{d_1, d_2 \in R^k, \|d_1 - d_2\|_0 \leq p, \|d_1 - d_2\|_2 \leq \epsilon}|B(d_1) - B(d_2)| \leq \frac{2}{N}\|G(d_1 - d_2)\|_2^2 \leq 2Np_p^u\epsilon,$$

(since $\|Gd\|_2^2 \leq N\ p_p^u\|d\|_2^2\ and\ \|d\|_2^2 \leq \epsilon$). Therefore

$$\sup_{d \in R^k, \|d\|_0 = p, \|d\|_2 = 1}|B(d)| \leq \sup_{d \in S}|B(d)| + 2N\ p_p^u\epsilon$$

Let $\epsilon = \frac{\sqrt{2p \log k}}{N}\frac{1}{2p_p^u}$, then

$$P\left(\sup_{d \in R^k \|d\|_0 = p, \|d\|_2 = 1}|B(d)| \geq C\sqrt{2p \log k}\right)$$

$$\leq P\left(\sup_{d \in S}|B(d)| \geq (C - 1)\sqrt{2p \log k}\right) \leq 2\left(\frac{3kN\ p_p^u}{K^{\frac{(C-1)^2}{\lambda_p^u}}}\right)^p.$$

Since $k > N$ and $k > 3K_p^u$, let $C = 1 + 2C_1\sqrt{\lambda_p^u}$ for some $C_1 > 1$, then

$$P\left(\sup_{d \in R^k \|d\|_0 = p, \|d\|_2 = 1}|B(d)| \geq (1 + 2C_1\sqrt{\lambda_p^u} - 1)\sqrt{2p \log k}\right) \leq 2\left(\frac{3kN\ K_p^u}{K^{\frac{(C-1)^2}{\lambda_p^u}}}\right)^p$$

$$P\left(\sup_{d \in R^k \|d\|_0 = p, \|d\|_2 = 1}|B(d)| \geq (1 + 2C_1\sqrt{\lambda_p^u})\sqrt{2p \log k}\right) \leq 2K^{-4p(C_1^2 - 1)}.$$

Theorem: Suppose that the measurement errors ω_{it} are independent and identically distributed randomvariables with mean 0. Assume that $\omega_{it}'s$ satisfy condition (7). Moreover, assume that $\lambda_p^l > \psi_{p\,,p}\left(\frac{1}{C}+\frac{1}{4}\right)$ and

$$\frac{3\sqrt{N}}{16}p_p^l > \lambda\sqrt{\frac{p}{N}} + C_1\sqrt{2plog\,k}\left(\frac{5}{4}+\frac{1}{C}\right), \tag{18}$$

For some constant C_1 such that $C_1 > 1 + 2\sqrt{\lambda_p^u}$. Then the $lasso$ estimator $\hat{\theta}$ satisfies with probability at least $1-2k^{-4p(C_2^2-1)+1}$

$$\|\hat{\theta}-\theta\|_2 \;\le \sqrt{\frac{2\,p\log k}{N}}\frac{16(c\sqrt{2}+1.25C_1+\frac{C_1}{\overline{C}}}{\frac{r\left((\lambda_p^l-\psi_{p\,,p}\,(\frac{1}{C}+\frac{1}{4})\right)^2}{\lambda_p^u}}\sqrt{1+\frac{1}{\overline{C}}}\,,$$

where $C_1 = 1+2C_2\sqrt{\lambda_p^u}$ and $C_2 > 1$ is constant.

Proof.

Recall lemma 5, $\Delta_{\overline{C}} = \left\{\begin{matrix}\delta\in R^k\colon \|\delta_F\|_1 \ge \overline{C}\|\delta_{F^c}\|_1,\ where\ F\subset\ \{1,2,\dots,k\} \\ and\ F\ contains\ at\ most\ p\ elements.\end{matrix}\right\}.$

Since $h = \hat{\theta}-\theta$ and $h\in\Delta_{\overline{C}}$.
Assume $|h_1| \ge |h_2| \ge \cdots \ge |h_k|$ and let $S_0 = \{1,2,\dots,p\}$,
therefore $h_{S_0} \ge \overline{C}h_{S_0^c}$.
We write $\{1\,,2\,,\dots,k\}$ in to the following sets:

$$S_0 = \{1\,,2\,,\dots,p\}, S_1 = \{p+1, p+2\,,\dots,2p\}, S_2 = \{2p+1, 2p+2\,,\dots,3p\},\dots\ .$$

From lemma 8 in [22] it follows that

$$\sum_{i\ge1}\|h_{S_i}\|_2 \le \sum_{i\ge1}\frac{\|h_{S_i}\|_1}{\sqrt{p}} + \frac{\sqrt{p}}{4}|h_{p+1}|$$

$$\le \frac{1}{\sqrt{p}}\|h_{S_0^c}\|_1 + \frac{1}{4\sqrt{p}}\|h_{S_0}\|_1$$

$$\le \left(\frac{1}{\sqrt{pC}}+\frac{1}{4p}\right)\|h_{S_0}\|_1$$

$$\le \left(\frac{1}{4}+\frac{1}{\overline{C}}\right)\|h_{S_0}\|_2 \tag{19}$$

Hence,

$$\frac{1}{N}(\|Gh+\omega\|_2^2 - \|\omega\|_2^2) \ge \frac{1}{N}\left(\|Gh_{S_0}+\omega\|_2^2 - \|\omega\|_2^2\right)$$

$$+\sum_{i\ge1}\frac{1}{N}\left(\|G(\textstyle\sum_{j=0}^{i}h_{S_i})+\omega\|_2^2 - \|G(\sum_{j=0}^{i-1}h_{S_i})+\omega\|_2^2\right) \tag{20}$$

for any fixed vector d, Let $M(d) = \frac{1}{N}E(\|Gd+\omega\|_2^2 - \|\omega\|_2^2)$.

By lemma 6, with probability at least $1-2k^{-4p(C_2^2-1)}$,

$$\frac{1}{N}\left(\|Gh_{s_0}+\omega\|_2^2 - \|\omega\|_2^2\right) \ge M(h_{S_0}) - C_1\sqrt{2p\log k}\,\|\boldsymbol{h_{S_0}}\|_{\boldsymbol{2}}, \tag{21}$$

And for $i\ge1$ with probability at least $1-k^{-4p(C_2^2-1)}$,

$$\frac{1}{N}\left(\|G(\textstyle\sum_{j=0}^{i}h_{S_i})+\omega\|_2^2 - \|G(\sum_{j=0}^{i-1}h_{S_i})+\omega\|_2^2\right) \ge M(h_{S_i}) - C_1\sqrt{2\,p\log k}\,\|h_{S_i}\|_2. \tag{22}$$

where $C_1 = 1+2C_2\sqrt{\lambda_p^u}$ $and\ C_2 > 1$ is constant.

Put the above inequalities together, we have that with probability at least $1-k^{-4p(C_2^2-1)+1}$.

$$\frac{1}{N}(\|Gh+\omega\|_2^2 - \|\omega\|_2^2) \ge M(h) - C_1\sqrt{2p\log k}\sum_{i\ge1}\|h_{S_i}\|_2$$

By this and inequalities (14) and (19), we have with probability at least $1-2k^{-4p(C_2^2-1)+1}$,

$$M(h) \leq \frac{1}{N}(\|Gh+\omega\|_2^2 - \|\omega\|_2^2) + C_1\sqrt{2p\log k}\sum_{i\geq 1}\|h_{S_i}\|_2$$

$$\leq \frac{\lambda p}{N}\|h_{S_0}\|_2 + C_1\sqrt{2p\log k}\left(\|h_{S_0}\|_2 + \sum_{i\geq 1}\|h_{s_i}\|_2\right)$$

$$\leq \frac{\lambda p}{N}\|h_{S_0}\|_2 + C_1\sqrt{2p\log k}\left(\|h_{S_0}\|_2 + \left(\frac{1}{4}+\frac{1}{\bar{C}}\right)\|h_{S_0}\|_2\right)$$

$$= \frac{\lambda p}{N}\|h_{S_0}\|_2 + C_1\sqrt{2p\log k}\left(1.25+\frac{1}{\bar{C}}\right)\|h_{S_0}\|_2. \tag{23}$$

Now by condition (7), we consider two cases.

Firstly, if $\|Gh\|_2^2 \geq \frac{3N}{r}$, then by lemma 7 in [22] and inequality (17), we have that

$$\frac{1}{N}E(\|Gh+\omega\|_2^2 - \|\omega\|_2^2) \geq \frac{3}{16N}\|Gh\|_2^2 \geq \frac{3N}{16}p_p^l\|h_{S_0}\|_2 \tag{24}$$

From assumption (14), we must have $\|hS_0\|_2 = 0$ and hence $\hat{\theta} = \theta$.

On the other hand, if $\|Gh\|_2^2 < \frac{3N}{r}$, from lemma 7 in [22] and inequality (8), we have

$$\frac{1}{N}E(\|Gh+\omega\|_2^2 - \|\omega\|_2^2) \geq \frac{r}{16N}\|Gh\|_2^2. \tag{25}$$

We have that,

$$|\langle Gh_{S_0}, Gh\rangle| \geq N\lambda_p^l\|h_{S_0}\|_2^2 - N\psi_{p,p}\|h_{S_0}\|_2\sum_{i\geq 1}\|h_{S_i}\|_2 \geq \lambda_p^u\|h_{S_0}\|_2^2$$

And by (Cauchy-Schwarz inequality)

$$|\langle G\,h_{S_0}, Gh\rangle| \leq \|Gh_{S_0}\|_2^2\|Gh\|_2^2 \leq \|Gh\|_2^2\sqrt{N\lambda_p^u}\|h_{S_0}\|_2$$

Therefore

$$\|Gh\|_2^2 \geq \mathrm{N}\frac{\left(\lambda_p^l - \psi_{p,p}\left(\frac{1}{C}+\frac{1}{4}\right)\right)^2}{\lambda_p^u}\|h_{S_0}\|_2^2 \tag{26}$$

Hence by (23) and (25), with probability at least $1-2k^{-4p(C_2^2-1)+1}$,

$$\|h_{S_0}\|_2 \geq \frac{16N}{r}M(h) \geq \frac{16N}{r}\left(\frac{\lambda p}{N}\|h_{S_0}\|_2\right) + C_1\sqrt{2p\log k}\left(1.25+\frac{1}{\bar{C}}\right)\|h_{S_0}\|_2 \tag{27}$$

This implies

$$\|h_{S_0}\|_2^2 \leq \frac{\lambda_p^u}{N\left(\lambda_p^l - \psi_{p,p}\left(\frac{1}{C}+\frac{1}{4}\right)\right)^2}\frac{16N\lambda}{r}\|h_{S_0}\|_2\left(\frac{p}{N} + C_1\sqrt{2p\log k}\;(1.25+\frac{1}{\bar{C}})\right)$$

i.e.

$$\|h_{S_0}\|_2 \leq \frac{16\lambda p}{Na\frac{\left(\lambda_p^l-\psi_{p,p}\left(\frac{1}{C}+\frac{1}{4}\right)\right)^2}{\lambda_p^u}} + \frac{\frac{\sqrt{2p\log k}}{N}\;16C_1\left(1.25+\frac{1}{\bar{C}}\right)}{Nr\left(\lambda_p^l - \psi_{k,k}\left(\frac{1}{C}+\frac{1}{4}\right)\right)^2}$$

$$= \frac{16\lambda p}{Nr\eta_p^l} + \sqrt{\frac{2p\log k}{N}}\frac{16C_1\left(1.25+\frac{1}{\bar{C}}\right)}{r\eta_p^l}$$

where $\eta_p^l = \frac{\left(\lambda_p^l-\psi_{p,p}\left(\frac{1}{\bar{C}}+\frac{1}{4}\right)\right)^2}{\lambda_p^u}$.

In particular, when $= 2C\sqrt{N\log k}$. we have that,

$$\|h_{S_0}\|_2 \leq \frac{\sqrt{2p\log k}}{N}\frac{16\left(C\sqrt{2}+1.25\,C_1+\frac{C_1}{\bar{C}}\right)}{r\,\eta_p^l}. \tag{28}$$

Since $\sum_{i\geq 1}\|h_{S_i}\|_2^2 \leq |h_{p+1}|\sum_{i\geq 1}\|h_{S_i}\|_1 \leq \frac{1}{C}\|h_{S_0}\|_2^2$, this implies $\sum_{i\geq 1}\|h_{S_i}\|_2 \leq \frac{1}{\sqrt{\bar{C}}}\|h_{S_0}\|_2 \leq \sqrt{\frac{\bar{C}+1}{\bar{C}}}\frac{\sqrt{2p\log k}}{N}\frac{16\left(C\sqrt{2}+1.25\,C_1+\frac{C_1}{\bar{C}}\right)}{r\,\eta_p^l}$

therefore,

$$\left\|\hat{\theta} - \theta\right\|_2 \le \sqrt{\frac{2\,p\log k}{N}}\,\frac{16(c\sqrt{2} + 1.25C_1 + \frac{C_1}{C}}{r\,\eta_p^l}\sqrt{1 + \frac{1}{C}}$$

Where $\eta_p^l = \frac{\left(\lambda_p^l - \psi_{p,p}\left(\frac{1}{C} + \frac{1}{4}\right)\right)^2}{\lambda_p^u}$. , $C_1 = 1 + 2C_2\sqrt{\lambda_p^u}$ and $C_2 > 1$ is a constant.

REFERENCES

[1] M. P. Brown and K. Austin, The New Physique (Publisher Name, Publisher City, 2005), Belitser E. and Fonti V. "Feature Selection Using Lasso", Research Paper in Business Analytics, VU Amsterdam, 2017.

[2] Bühlmann P. and De Geer S. V. "Statistics for High Dimensional data". Springer Heidelberg Dordrecht London New York, 2011.

[3] Derumigny A."Improved Bounds for Square-Root Lasso and Square-Root Slope". Electronic Journal of Statistics, Vol. 12, pp. 741-766, 2018.

[4] A. Derumigny, “Improved bounds for Square-Root Lasso and Square-Root Slope,” Electronic Journal of Statistics, vol. 12, no. 1, pp. 741–766, 2018.

[5] B. Doerr, “Probabilistic Tools for the Analysis of Randomized Optimization Heuristics,” Theory of Evolutionary Computation, pp. 1–87, Nov. 2019.

[6] Gao X. and Huang J. " Asymptotic Analysis of High-dimensional LAD Regression with Lasso". Statistics Sinica, Vol. 20, pp. 1484-1506, 2010.

[7] Lambert-Lacroix S. and Zwald L. "Robust Regression Through the Huber's Criterion and Adaptive Lasso Penalty". Electronic Journal of Statistics, Vol. 5, no. 0, pp. 1015-1053, 2011.

[8] Slastinkov, A. D. "Limit Theorems for Moderate Deviation Probability". Theo. Prob. Appl. Vol. 23, no. 2, pp. 322-340, Mar.1979.

[9] Tibshirani, R. “Regression Analysis and Selection Via the Lasso”. Journal of the Royal Statistical Society Series B Vol. 58, pp. 267-288, 1996.

[10] Van de Geer, S. A. “High-Dimensional Generalized Linear Models and the Lasso”. Ann, Statist. Vol. 36, pp. 614-645, 2008.

[11] R. Vershynin, “High-Dimensional Probability,” Sep. 2018.

[12] E. Vonesh and V. M. Chinchilli, “Linear and Nonlinear Models for the Analysis of Repeated Measurements,” Nov. 1996.

[13] Wang L. "The L_1 Penalized LAD Estimator for High Dimensional Linear Regression". Journal of Multivariate Analysis, Vol. 120, pp. 135-151, Sep. 2013.

[14] S. R. Mohamed and P. Raviraj, “Optimisation of multi-body fishbotundulatory swimming speed based on SOLEIL and BhT simulators,” International Journal of Intelligence and Sustainable Computing, vol. 1, no. 1, p. 19, 2020.

[15] K. V. Kumar, T. Jayasankar, V. Eswaramoorthy, and V. Nivedhitha, “SDARP: Security based Data Aware Routing Protocol for ad hoc sensor networks,” International Journal of Intelligent Networks, vol. 1, pp. 36–42, 2020.