

ASYMPTOTICS FOR REGULARIZED MATRIX REGRESSIONS

CHEN Wei

(School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China)

Abstract: In this paper, we study the asymptotics properties of regularized matrix regressions estimators. We use Knight, Fu and Chatterjee, Lahiri's asymptotics method on vector regressions, and extend it to matrix regressions to study the asymptotics properties of nuclear norm regularized matrix regressions estimators. We have obtained the weak consistency and limiting distribution of nuclear norm estimators when the second order moments of random errors exist, i.e. $\mathbb{E}|\epsilon_i|^2 < \infty$. And under the condition of the lower order moments of random errors existing, i.e. $\mathbb{E}|\epsilon_i|^\alpha < \infty, 1 < \alpha < 2$, we also have obtained the strong consistency and convergence rates of nuclear norm estimators.

Keywords: asymptotic theory; linear regression; nuclear norm

2010 MR Subject Classification: 62E20; 62F12; 62H12

Document code: A **Article ID:** 0255-7797(2022)02-0146-07

1 Introduction

For a rectangular matrix $\mathbf{B} \in \mathbb{R}^{p \times q}$, $\sigma_j(\mathbf{B})$ denotes the largest singular value of it and is equal to the square root of the i th largest eigenvalue of $\mathbf{B}\mathbf{B}^T$. The rank of \mathbf{B} will usually be denoted by \mathbf{r} and is equal to the number of nonzero singular values. For matrices \mathbf{B} and \mathbf{X} of the same dimensions, we define the inner product in $\mathbb{R}^{p \times q}$ as $\langle \mathbf{B}, \mathbf{X} \rangle = \text{tr}(\mathbf{X}^T \mathbf{B}) = \langle \mathbf{B}, \mathbf{X} \rangle = \langle \text{vec}(\mathbf{B}), \text{vec}(\mathbf{X}) \rangle$, where $\text{vec}(\cdot)$ is the vectorization operator that stacks the columns of a matrix into a vector. The norm associated with this inner product is called the Frobenius (or Hilbert - Schmidt) norm $\|\cdot\|_F$. The Frobenius norm is also equal to the Euclidean, or l_1 norm of the vector of singular values, i.e.

$$\|\mathbf{B}\|_F = \sqrt{\langle \mathbf{B}, \mathbf{B} \rangle} = \sqrt{\text{tr}(\mathbf{B}^T \mathbf{B})} = \left(\sum_{i=1}^p \sum_{i=1}^q B_{ij}^2 \right)^{1/2} = \left(\sum_{j=1}^r \sigma_j^2 \right)^{1/2},$$

The nuclear norm of a matrix is equal to the sum of its singular values, i.e.

$$\|\mathbf{B}\|_* := \sum_{i=1}^r \sigma_i(\mathbf{B}),$$

* Received date: 2021-03-27

Accepted date: 2021-04-27

Biography: Chen Wei(1993-), female, born at Xianning, Hubei, postgraduate, major in probability theory and mathematical statistics.

and is alternatively known by several others including the Scatten-1-norm, the Ky Fan r-norm, and the trace class norm. Since the singular values are all positive, the nuclear norm is also equal to the l_1 norm of the vector of singular values. These two norms are related by the following inequalities which hold for any matrix \mathbf{B} of rank at most r (see [1]): $\|\mathbf{B}\|_F \leq \|\mathbf{B}\|_* \leq \sqrt{r}\|\mathbf{B}\|_F$.

Consider the regularized matrix regression model (see [2])

$$Y_i = \gamma^T \mathbf{T}_i + \langle \mathbf{B}, \mathbf{X}_i \rangle + \epsilon_i \quad (1.1)$$

where $\epsilon_1, \dots, \epsilon_n$ are i.i.d. random variables with mean 0 and variance σ^2 , $\gamma \in \mathbb{R}^m$, $\mathbf{T}_i \in \mathbb{R}^m$, $\mathbf{B} \in \mathbb{R}^{p \times q}$, $\mathbf{X}_i \in \mathbb{R}^{p \times q}$. This model is no longer limited to rank 1 matrices or low rank matrices, and is a generalization of the fixed rank matrix regression. Without loss of generality, we drop the vector covariate \mathbf{T} and its associated parameter γ in subsequent discussion, that is

$$Y_i = \langle \mathbf{B}, \mathbf{X}_i \rangle + \epsilon_i. \quad (1.2)$$

Then we estimate \mathbf{B} by minimizing the penalized least squares criterion, where the penalty is the nuclear norm of \mathbf{B} (see [2]), i.e.

$$\sum_{i=1}^n (Y_i - \langle \mathbf{B}, \mathbf{X}_i \rangle)^2 + \lambda \|\mathbf{B}\|_*. \quad (1.3)$$

In order to facilitate the next work, we first introduce some symbols. For a given λ_n , we will denote the estimator minimizing (1.3) by $\hat{\mathbf{B}}_n$. In particular, $\lambda_n = 0$ corresponds to the ordinary LS. We denote this estimator by $\hat{\mathbf{B}}_n^{(0)}$. We will assume the following regularity conditions for the design (see [3]),

$$C_n = \frac{1}{n} \sum_{i=1}^n \text{vec}(\mathbf{X}_i) \text{vec}(\mathbf{X}_i)^T \longrightarrow C \quad (1.4)$$

where C is a nonnegative definite matrix and

$$\frac{1}{n} \max \text{vec}(\mathbf{X}_i)^T \text{vec}(\mathbf{X}_i) = \frac{1}{n} \max \|\mathbf{X}_i\|_F^2 \longrightarrow 0 \quad (1.5)$$

where $\mathbf{X}_i \in \mathbb{R}^{p \times q}$, and \mathbf{X}_i is the i th sample observation, then $\text{vec}(\mathbf{X}_i) = (x_{11}^i, \dots, x_{1q}^i, x_{21}^i, \dots, x_{2q}^i, \dots, x_{p1}^i, \dots, x_{pq}^i)^T$. In this paper, we assume that C_n is nonsingular for all n .

The next of the paper is organized as follow. We formulate the limiting distribution about nuclear norm regularized matrix regression in Section 2. In Section 3, we discuss strong consistency about this regression estimators. While in Section 4, we give a discussion about the outlook of this paper.

2 Limiting Distribution

In this section, first we show the study of the limiting distribution of matrix regression based on the method of Knight and Fu. By studying the asymptotics behavior of the objective function (1.3) can determine the limiting behavior of the estimators $\hat{\mathbf{B}}_n$. For instance, we will define the function

$$Z_n(\Phi) = \frac{1}{n} \sum_{i=1}^n (Y_i - \langle \mathbf{B}, \mathbf{X}_i \rangle)^2 + \frac{\lambda_n}{n} \|\Phi\|_*, \tag{2.1}$$

to consider consistency of $\hat{\mathbf{B}}_n$, when $\Phi = \hat{\mathbf{B}}_n$, the function (2.1) gets the smallest value. The following result shows that is consistent under the condition of $\lambda_n = o(n)$.

Theorem 2.1 If C in (1.4) is nonsingular and $\lambda_n/n \rightarrow \lambda_0 \geq n$, then $\hat{\mathbf{B}}_n \rightarrow_p \arg \min(Z)$ where $Z(\Phi) = \text{vec}(\Phi - \mathbf{B})^T C \text{vec}(\Phi - \mathbf{B}) + \lambda_0 \|\Phi\|_*$.

Thus if $\lambda_n = o(n)$, $\arg \min(Z) = \mathbf{B}$, and so $\hat{\mathbf{B}}_n$ is consistent.

Proof Define Z_n as in (2.1). Then let $Y_i = \langle \mathbf{B}, \mathbf{X}_i \rangle + \epsilon_i$, so we have

$$Z_n(\Phi) = \frac{1}{n} \sum_{i=1}^n (\epsilon_i - \text{vec}(\mathbf{X}_i) \text{vec}(\Phi - \mathbf{B})^T)^2 + \frac{\lambda_n}{n} \|\Phi\|_*.$$

Now we need to show that

$$\sup_{\Phi \in K} |Z_n(\Phi) - Z(\Phi) - \sigma^2| \rightarrow_p 0 \tag{2.2}$$

for any compact set K and that

$$\hat{\mathbf{B}}_n = O_p(1). \tag{2.3}$$

Under (2.2) and (2.3), we have $\arg \min Z_n \rightarrow_p \arg \min Z$. Note that Z_n is convex; Thus (2.2) and (2.3) follow from the pointwise convergence in probability of $Z_n(\Phi)$ to $Z(\Phi) + \sigma^2$ by applying standard results(see [4]).

Theorem 2.2 If C in (1.4) is nonsingular and $\lambda_n/\sqrt{n} \rightarrow \lambda_0 \geq 0$, then

$$\sqrt{n}(\hat{\mathbf{B}}_n - \mathbf{B}) \rightarrow_d \arg \min(\mathbf{V}),$$

where

$$V(U) = -2\text{vec}(\mathbf{U})^T W + \text{vec}(\mathbf{U})^T C \text{vec}(\mathbf{U}) + \lambda_0 \max_{d \in \partial \|\mathbf{B}\|_*} \sum_{i=1}^{\min(p,q)} d_i \alpha_i^T \mathbf{U} \beta_i$$

and W has a $N(\mathbf{0}, \sigma_2 C)$ distribution.

Proof Define $V_n(\mathbf{U})$ by

$$V_n(U) = \sum_{i=1}^n [(\epsilon_i - \text{vec}(\mathbf{U})^T \mathbf{X}_i / \sqrt{n})^2 - \epsilon_i^2] + \lambda_n [\|\mathbf{B} + 1/\sqrt{n} \mathbf{U}\|_* - \|\mathbf{B}\|_*],$$

where $\mathbf{U} \in \mathbb{R}^{p \times q}$, note that V_n is minimized at $\sqrt{n}(\hat{\mathbf{B}}_n - \mathbf{B})$. Because adding a constant does not change the position of the minimum point of the objection function. We rewrite the $V_n(\mathbf{U})$, that is

$$V_n(U) = \left[\sum_{i=1}^n [(\epsilon_i - \text{vec}(\mathbf{U})^T \mathbf{X}_i / \sqrt{n})^2 - \epsilon_i^2] + \lambda_n \|\mathbf{B} + 1/\sqrt{n} \mathbf{U}\|_* \right] - \lambda_n \|\mathbf{B}\|_*.$$

First we have

$$\sum_{i=1}^n [(\epsilon_i - \text{vec}(\mathbf{U})^T \mathbf{X}_i / \sqrt{n})^2 - \epsilon_i^2] \rightarrow -2\text{vec}(\mathbf{U})^T W + \text{vec}(\mathbf{U})^T C \text{vec}(\mathbf{U}).$$

Next using the theorem 2 (see [5]), let $t = 1/\sqrt{n}$, then the second term

$$\lambda_n [\|\mathbf{B} + 1/\sqrt{n} \mathbf{U}\|_* - \lambda_n \|\mathbf{B}\|_*] \rightarrow \lambda_0 \max_{d \in \partial \|\mathbf{B}\|_*} \sum_{i=1}^{\min(p,q)} d_i \alpha_i^T \mathbf{U} \beta_i,$$

where α_i and β_i are singular vectors of \mathbf{B} corresponding to the i th largest singular value, $\partial \|\mathbf{B}\|_*$ denotes the subdifferential of $\|\mathbf{B}\|_*$. Therefore $V_n(\mathbf{U}) \rightarrow_d V(\mathbf{U})$ (as defined above) with the finite dimensional convergence holding trivially. Since V_n is convex and V has a unique minimum, it follows (see [6]) that $\arg \min V_n(\mathbf{U}) = \sqrt{n}(\hat{\mathbf{B}}_n - \mathbf{B}) \rightarrow_d \arg \min(V)$. In particular, when $\lambda_n = 0$, $V = -2\text{vec}(\mathbf{U})^T W + \text{vec}(\mathbf{U})^T C \text{vec}(\mathbf{U})$, we have

$$\sqrt{n}(\hat{\mathbf{B}}_n - \mathbf{B}) = \arg \min(V) = C^{-1}W \sim N(\mathbf{0}, \sigma^2 C^{-1}).$$

3 Strong Consistency

Under the matrix regression model (1.2) with i.i.d. error variables ϵ_i , where $\mathbb{E}|\epsilon_i| < \infty$ and $\mathbb{E}\epsilon_i = 0$. Now we consider the problem of strong consistency of the Lasso estimator assuming only finiteness of the first moment and some mild regularity conditions on the design matrix \mathbf{X}_i 's.

Theorem 3.1 Let ϵ_i be i.i.d. random variables with $\mathbb{E}|\epsilon_i| < \infty$ and $\mathbb{E}\epsilon_i = 0$. If C in (1.4) is nonsingular and if $\frac{\lambda_n}{n} \rightarrow 0$, then $\hat{\mathbf{B}}_n \rightarrow \mathbf{B}$ w.p.1.

Proof Note that

$$\begin{aligned} \hat{\mathbf{B}}_n &= \arg \min \left\{ \sum_{i=1}^n (Y_i - \langle \Phi, \mathbf{X}_i \rangle)^2 + \lambda_n \|\Phi\|_* \right\} \\ &= \arg \min \left\{ \sum_{i=1}^n (\epsilon_i - \langle \Phi - \mathbf{B}, \mathbf{X}_i \rangle)^2 + \lambda_n \|\mathbf{B} + \Phi - \mathbf{B}\|_* \right\}. \end{aligned}$$

Then $\hat{\mathbf{B}}_n - \mathbf{B} = \arg \min \sum_{i=1}^n (\epsilon_i - \langle \mathbf{U}, \mathbf{X}_i \rangle)^2 + \lambda_n \|\mathbf{B} + \mathbf{U}\|_*$. Recall that $C_n = \frac{1}{n} \sum_{i=1}^n \text{vec}(\mathbf{X}_i) \cdot \text{vec}(\mathbf{X}_i)^T \rightarrow C$, let $\gamma_{0,n}$ = the smallest eigenvalue of C_n and γ_0 = the smallest eigenvalue of C and let $W_n = \frac{1}{n} \sum_{i=1}^n \text{vec}(\mathbf{X}_i) \epsilon_i$. With lemma 3.1 (see [7]),

$$W_n = \frac{1}{n} \sum_{i=1}^n \text{vec}(\mathbf{X}_i) \epsilon_i \rightarrow 0 \quad \text{as } n \rightarrow \infty \quad \text{w.p.1.} \quad (3.1)$$

Since $\sum_{i=1}^n \epsilon_i^2$ does not involve \mathbf{U} , discarding this term from the criterion function above and dividing the resulting expression by \mathbf{B} , we have

$$\begin{aligned} \hat{\mathbf{B}}_n - \mathbf{B} &= \arg \min \left\{ \text{vec}(\mathbf{U})^T C_n \text{vec}(\mathbf{U}) - 2W_n^T \text{vec}(\mathbf{U}) + \frac{\lambda_n}{n} [\|\mathbf{B} + \mathbf{U}\|_* - \|\mathbf{B}\|_*] \right\} \\ &= \arg \min V_n(\mathbf{U}). \end{aligned} \quad (3.2)$$

Note that for any $\mathbf{U} \in \mathbb{R}^{p \times q}$,

$$V_n(\mathbf{U}) \geq \frac{\gamma_{0,n}}{\min\{p, q\}} \|\mathbf{U}\|_*^2 - 2\|W_n\|_* \|\mathbf{U}\|_* - \frac{\lambda_n}{n} \|\mathbf{U}\|_*. \quad (3.3)$$

Next fix $\eta \in (0, 1)$. Since $\frac{\lambda_n}{n} \rightarrow o(1)$, there exists a $n_0 \in (0, \infty)$ such that $\frac{\lambda_n}{n} \leq \eta$ and $\gamma_{0,n} > \gamma_0/2$ for all $n \geq n_0$. On the set $\{\|W_n\|_* \leq \eta\}$, by (3.3), for all $\mathbf{U} \in \mathbb{R}^{p \times q}$, with $\|\mathbf{U}\|_* > 6 \min\{p, q\} \eta / \gamma_{0,n}$,

$$V_n(\mathbf{U}) \geq \|\mathbf{U}\|_* \left(\frac{\gamma_{0,n}}{\min\{p, q\}} \|\mathbf{U}\|_* - 3\eta \right) \geq \frac{\gamma_{0,n}}{2 \min\{p, q\}} \|\mathbf{U}\|_*^2 > 0.$$

Since $V_n(0) = 0$, it follows that for $n \geq n_0$, the minimum of it cannot be attained in the set $\{\mathbf{U} : \|\mathbf{U}\|_* > 6 \min\{p, q\} \eta / \gamma_{0,n}\}$, whenever $\{\|W_n\|_* \leq \eta\}$. Hence, it follows that for $n \geq n_0$, $\{\|W_n\|_* \leq \eta\}$, implies $\hat{\mathbf{B}}_n - \mathbf{B} = \arg \min V_n(\mathbf{U}) \in \{\mathbf{U} : \|\mathbf{U}\|_* \leq 6 \min\{p, q\} \eta / \gamma_{0,n}\}$.

In particular,

$$\begin{aligned} P \left(\|\hat{\mathbf{B}}_n - \mathbf{B}\|_* > \frac{12 \min\{p, q\} \eta}{\gamma_0}, i.o. \right) &\leq P \left(\|\hat{\mathbf{B}}_n - \mathbf{B}\|_* > \frac{6 \min\{p, q\} \eta}{\gamma_{0,n}}, i.o. \right) \\ &\leq P(\|W_n\|_* > \eta, i.o.) = 0, \end{aligned}$$

which follows from (3.1). Since $\eta \in (0, \infty)$ is arbitrary, this completes the proof.

Theorem 3.2 Let ϵ_i be i.i.d. random variables with $\mathbb{E}|\epsilon_i| < \infty$ and $\mathbb{E}|\epsilon_i| = 0$. Assume that (2.1) holds as $n \rightarrow \infty$.

(a) if $\frac{\lambda_n}{n} \rightarrow a \in (0, \infty)$ then

$$\hat{\mathbf{B}}_n - \mathbf{B} \rightarrow \arg \min V_\infty(\mathbf{U}, a),$$

where $V_\infty(\mathbf{U}, a) = \text{vec}(\mathbf{U})^T C_n \text{vec}(\mathbf{U}) + a [\|\mathbf{B} + \mathbf{U}\|_* - \|\mathbf{U}\|_*]$.

(b) if $\frac{\lambda_n}{n} \rightarrow \infty$ then $\hat{\mathbf{B}}_n \rightarrow \mathbf{0}$, w.p.1.

Proof First consider part (a), let $V_n(\cdot)$ be as in (3.2). Note that $|\|\mathbf{B} + \mathbf{U}\|_* - \|\mathbf{U}\|_*| \leq \|\mathbf{U}\|_*$, since $\frac{\lambda_n}{n} \rightarrow a \in (0, \infty)$, for any compact set $K \in \mathbb{R}^{p \times q}$,

$$\begin{aligned} &\sup_{U \in K} |V_n(\mathbf{U}) - V_\infty(\mathbf{U}, a)| \\ &\leq \sup_{U \in K} \left[\|\mathbf{U}\|_*^2 \|C_n - C\|_* + 2\|W_n\|_* \|\mathbf{U}\|_* + \left| \frac{\lambda_n}{n} - a \right| \|\mathbf{U}\|_* \right] \\ &= o(1) \quad \text{as } n \rightarrow \infty \quad w.p.1. \end{aligned} \quad (3.4)$$

Let $n_0 \geq 1$ such that for all $n \geq n_0$, $\lambda_n/n < 2a$, and $\gamma_{0,n} > \gamma_0/2$. From (3.3), for all $n \geq n_0$, on the set $\{\|W_n\|_* \leq a\}$, we have

$$V_n(\mathbf{U}) \geq \|\mathbf{U}\|_* \left(\frac{\gamma_{0,n}}{\min\{p, q\}} \|\mathbf{U}\|_* - 2\|W_n\|_* - \lambda_n/n \right) \geq \frac{\|\mathbf{U}\|_*}{2}$$

for all $\|\mathbf{U}\|_* > (1 + 8a)/\gamma_0 \equiv c_0$. Since $V_n(\mathbf{0}) = 0$, this implies $\|\hat{\mathbf{B}}_n - \mathbf{B}\|_* \leq c_0$, whenever $n \geq n_0$ and $\{\|W_n\|_* \leq a\}$. Thus, the minimizer of $V_n(\mathbf{U})$ lies in a compact set for all $n \geq n_0$, provided $\{\|W_n\|_* \leq a\}$. Since $V_\infty(\cdot, a)$ is a convex function, by (3.4) and with (3.1), part (a) follows.

Next consider part (b). Let $a_n^2 = \lambda_n/n$, then $a_n \rightarrow \infty$. Also, let

$$D_n = \{\mathbf{U} : \|\mathbf{U}\|_* \leq a_n, \|\mathbf{B} + \mathbf{U}\|_* \geq a_n^{-1}\}.$$

With (3.1),

$$\begin{aligned} & \inf_{\mathbf{U} \in D_n} [\text{vec}(\mathbf{U})^T C_n \text{vec}(\mathbf{U}) - 2W_n^T \text{vec}(\mathbf{U}) + a_n^2 \|\mathbf{B} + \mathbf{U}\|_*] \\ & \geq \inf_{\mathbf{U} \in D_n} [a_n^2 \|\mathbf{B} + \mathbf{U}\|_* - 2\|W_n\|_* \sup\{\|\mathbf{U}\|_* : \mathbf{U} \in D_n\}] \\ & \geq a_n [1 - 2\|W_n\|_*] \rightarrow \infty \text{ as } n \rightarrow \infty \text{ w.p.1.} \end{aligned} \quad (3.5)$$

Also by (3.1),

$$\begin{aligned} & \inf_{\|\mathbf{U}\|_* \geq a_n} [\text{vec}(\mathbf{U})^T C_n \text{vec}(\mathbf{U}) - 2W_n^T \text{vec}(\mathbf{U}) + a_n^2 \|\mathbf{B} + \mathbf{U}\|_*] \\ & \geq \inf_{\|\mathbf{U}\|_* \geq a_n} [\gamma_{0,n} \|\mathbf{U}\|_*^2 - 2\|W_n\|_* \|\mathbf{U}\|_*] \\ & \geq a_n [a_n \gamma_{0,n} - 2\|W_n\|_*] \rightarrow \infty \text{ as } n \rightarrow \infty \text{ w.p.1.} \end{aligned} \quad (3.6)$$

Finally, with $D_{1,n} = \{\mathbf{U} : \|\mathbf{B} + \mathbf{U}\|_* \leq a_n^{-1}\}$

$$\begin{aligned} & \inf_{\mathbf{U} \in D_{1,n}} [\text{vec}(\mathbf{U})^T C_n \text{vec}(\mathbf{U}) - 2W_n^T \text{vec}(\mathbf{U}) + a_n^2 \|\mathbf{B} + \mathbf{U}\|_*] \\ & \leq \mathbf{U}_0^T C_n \mathbf{U}_0 - 2W_n^T \mathbf{U}_0 + a_n^2 \|\mathbf{B} + \mathbf{U}_0\| \\ & = \mathbf{U}_0^T C_n \mathbf{U}_0 - 2W_n^T \mathbf{U}_0 \\ & \rightarrow \mathbf{B}^T C \mathbf{B} \rightarrow \infty \text{ as } n \rightarrow \infty \text{ w.p.1} \end{aligned} \quad (3.7)$$

where $\mathbf{U}_0 = -\mathbf{B} \in D_{1,n}$.

Note that for any sequence $\{\mathbf{U}_n\}_{n \geq 1}$, with $\mathbf{U}_n \in D_{1,n}$, $\|\mathbf{B} + \mathbf{U}_n\|_* \leq a_n^{-1} \rightarrow 0$ as $n \rightarrow \infty$. Hence, from (3.5)–(3.7) and (3.2), it follows that there exists a set A with $p(A) = 1$ and for all $\omega \in A$, there exists a $n_\omega \geq 1$ such that for all $n \geq n_\omega$,

$$\begin{aligned} \hat{\mathbf{B}}_n - \mathbf{B} &= \arg \min_{\mathbf{U}} V_n(\mathbf{U}) \\ &= \arg \min_{\mathbf{U} \in D_{1,\infty}} [\text{vec}(\mathbf{U})^T C_n \text{vec}(\mathbf{U}) - 2W_n^T \text{vec}(\mathbf{U}) + a_n^2 \|\mathbf{B} + \mathbf{U}\|_*] \\ &\rightarrow -\mathbf{B} \text{ as } n \rightarrow \infty. \end{aligned}$$

This completes the proof of part (b).

4 Discussion

An important contribution of this study is that, based on the regularized matrix regression of Zhou and Li (see [2]), we provide the corresponding statistical justification. Such that we derive the asymptotic normality of the proposed estimators. But there still have other issues. For example, when C_n [defined in (1.4)] is not nonsingular or nearly singular for each n , then the parametrization in (1.2) is not unique. So a new singular design need to be brought to solve this problem.

References

- [1] Recht B, Fazel M, Parrilo P A. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization[J]. SIAM Review, 2010, 52(3): 471–501.
- [2] Zhou H, Li L. Regularized matrix regression[J]. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2014, 76(2): 463–483.
- [3] Knight K, Fu W. Asymptotics for lasso-type estimators[J]. Annals of Statistics, 2000, 28(2): 1356–1378.
- [4] Anderson P K, Gill R D. Cox' s regression model for counting processes: a large sample study[J]. Ann. Statist, 1982, 10(3): 1100–1120.
- [5] Watson G A. Characterization of the subdifferential of some matrix norms[J]. Linear Algebra and Its Applications, 1992, 170(3): 33–45.
- [6] Geyer C J. On the asymptotics of convex stochastic optimization[R]. Technical Report, School of Statistics, University of Minnesota, 1996.
- [7] Chatterjee A, Lahiri S N. Strong consistency of Lasso estimators[J]. Sankhya A, 2011, 73(1): 55–78.

正则矩阵回归的渐近性质

陈 伟

(武汉大学数学与统计学院, 湖北 武汉 430072)

摘要: 本文研究了正则化矩阵回归估计量的渐近性质等问题. 利用Knight, Fu 和Chatterjee, Lahiri 分别关于向量回归的Lasso估计量渐近性研究方法, 推广到矩阵回归, 研究核范数正则化矩阵回归估计量对应的渐近性质. 从而得到了核范数矩阵估计量在随机误差二阶矩存在即 $\mathbb{E}|\epsilon_i|^2 < \infty$ 的条件下的弱相合性和极限分布, 以及在随机误差的低阶矩存在即 $\mathbb{E}|\epsilon_i|^\alpha < \infty, 1 < \alpha < 2$ 的条件下, 核范数矩阵参数估计量的强相合性以及对应的收敛速度.

关键词: 渐近理论; 线性回归; 核范数

MR(2010)主题分类号: 62E20; 62F12; 62H12

中图分类号: O211.4