Sankhyā : The Indian Journal of Statistics
2011, Volume 73-A, Part 1, pp. 55-78
© 2011, Indian Statistical Institute

Strong Consistency of Lasso Estimators

A. Chatterjee
Indian Statistical Institute, New Delhi, India
S. N. Lahiri
Texas A&M University, College Station, USA

Abstract

In this paper, we study the strong consistency and rates of convergence of the Lasso estimator. It is shown that when the error variables have a finite mean, the Lasso estimator is strongly consistent, provided the penalty parameter (say, λ_n) is of *smaller* order than the sample size (say n). We also show that this condition on λ_n cannot be relaxed. More specifically, we show that consistency of the Lasso estimators fail in the cases where $\lambda_n/n \to a$ for some $a \in (0, \infty]$. For error variables with a finite αth moment, $1 < \alpha < 2$, we also obtain convergence rates of the Lasso estimator to the true parameter. It is noted that the convergence rates of the Lasso estimators of the non-zero components of the regression parameter vector can be *worse* than the corresponding least squares estimators. However, when the design matrix satisfies some orthogonality conditions, the Lasso estimators of the zero components are surprisingly accurate; The Lasso recovers the zero components *exactly*, for large n, almost surely.

AMS (2000) *subject classification.* Primary 62J07; secondary 60F15, 62E20.
Keywords and phrases. Penalized regression, strong law, convergence rates.

1 Introduction

Consider the following regression model

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \epsilon_i, \quad i = 1, \ldots, n, \tag{1.1}$$

where, y_i is the response, $\mathbf{x}_i' = (x_{i,1}, \ldots, x_{i,p})$ is a $p \times 1$ covariate vector, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is the regression parameter and $\{\epsilon_i\}$ are *iid* errors. We assume that p is fixed. The Lasso estimator of $\boldsymbol{\beta}$ is defined as the minimizer of the l_1-penalized least square criterion function,

$$\widehat{\boldsymbol{\beta}}_n := \operatorname*{argmin}_{\mathbf{u}\in\mathbb{R}^p} \sum_{i=1}^{n} \left(y_i - \mathbf{x}_i'\mathbf{u}\right)^2 + \lambda_n \sum_{j=1}^{p} |u_j|, \tag{1.2}$$

where, λ_n is a penalty or regularization parameter. The Lasso estimate was introduced by Tibshirani (1996) as an estimation and variable selection method. Recently the Lasso has emerged as a very popular method for both estimation as well as model selection. Two main benefits of the Lasso are: (i) the nature of regularization used in the Lasso leads to sparse solutions, which automatically leads to parsimonious model selection (see Zhao and Yu (2006), Wainwright (2006), Zou (2006)) and (ii) it is computationally feasible (see Efron et. al (2004), Osborne et al. (2000), Fu (1998)), even in high dimensional settings.

The asymptotic properties of the Lasso was first studied by Knight and Fu (2000) for the finite dimensional regression model (1.1). In addition to finding the asymptotic distribution of the Lasso estimator, Knight and Fu (2000) also showed that the Lasso was weakly consistent under some mild regularity conditions. In this paper, we investigate the problem of strong consistency of the Lasso estimator under different moment conditions on the error variables ϵ_i's in (1.1). It is shown that when $\mathbf{E}|\epsilon_1| < \infty$ and the regularization parameter $\lambda_n = o(n)$ as $n \to \infty$, the Lasso estimator of the regression parameter $\boldsymbol{\beta}$ is strongly consistent. However, if $\lim_{n\to\infty} \lambda_n/n \to a \in (0, \infty]$, then the Lasso fails to be strongly consistent, and converges to a different limiting quantity. Thus, when $\mathbf{E}|\epsilon_1| < \infty$, one needs to choose $\lambda_n = o(n)$ to guarantee consistency of the Lasso estimator.

Next we consider the rate of almost sure convergence of $\|\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\|$ when the error variables have a finite αth absolute moment for some $\alpha \in (1, 2)$. Theorem 2.3 below shows that for $\mathbf{E}|\epsilon_1|^\alpha < \infty$, $\alpha \in (1, 2)$,

$$\|\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\| = O\left(n^{-(\alpha-1)/\alpha}\right) \quad \text{with probability 1.}$$

Further, for the Lasso estimators of the non-zero components of $\boldsymbol{\beta}$, we also obtain a lower bound on the rate of convergence (cf. Theorem 2.4), which shows that the *exact* convergence rate of the Lasso estimators of the non-zero parameter components is $n^{-(\alpha-1)/\alpha}$ as $n \to \infty$. This is an interesting finding as it allows one to compare the relative performances of the Lasso estimator and the ordinary least squares (OLS) estimator of $\boldsymbol{\beta}$. It can be shown that under the regularity conditions of Theorem 2.4 on the design matrix, for $\mathbf{E}|\epsilon_1|^\alpha < \infty$, $\alpha \in (1,2)$, the OLS estimator $\widehat{\boldsymbol{\beta}}_n^{\text{OLS}}$ of $\boldsymbol{\beta}$ satisfies

$$\|\widehat{\boldsymbol{\beta}}_n^{\text{OLS}} - \boldsymbol{\beta}\| = o\Big(n^{-(\alpha-1)/\alpha}\Big) \quad \text{with probability 1.}$$

Thus, the penalization used in the definition of the Lasso estimator results in a loss of accuracy for the non-zero components of $\boldsymbol{\beta}$ compared to the OLS estimator, which uses no penalty. For the zero-components of $\boldsymbol{\beta}$, however, this is not necessarily true. When the design matrix satisfies an orthogonality condition and $\mathbf{E}|\epsilon_1|^\alpha < \infty$, $\alpha \in (1,2)$, Theorem 2.5 shows that the Lasso estimator of the zero-components of $\boldsymbol{\beta}$ recovers the true values *exactly* for large n, almost surely. Thus, in this case, the rate of convergence of the Lasso estimator of the zero-components is $O(b_n)$ for any $b_n \to 0$, with probability (w.p.) 1. On the other hand, when the orthogonality condition on the design matrix fails, the convergence rate of the Lasso estimator of the zero-components also can be $n^{-(\alpha-1)/\alpha}$ exactly (like the non-zero components), making it worse than the OLS.

Finally, we also consider the case where $\mathbf{E}|\epsilon_1|^\alpha < \infty$ for some $\alpha \in (0,1)$. In this case, the mean of the error variables may not even exist. Theorem 2.6 shows that under suitable regularity conditions, $\widehat{\boldsymbol{\beta}}_n$ converges to the zero-vector, almost surely. Hence, it follows that for the (strong) consistency of the Lasso estimator, finiteness of $\mathbf{E}|\epsilon_1|$ cannot be dispensed with.

We now conclude this section with a brief literature review. For the case of ordinary least squares estimators, results on strong consistency were studied by Lai et al. (1978) and Drygas (1976), Knight and Fu (2000) proved consistency of the Lasso estimator and derived its asymptotic distribution under the moment condition $\mathbf{E}\left(\epsilon_1^2\right) < \infty$, in the finite dimensional case. Recently, Lounici (2008) showed that the ℓ_∞ distance $\|\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\|_\infty$ converges *weakly* to zero, and also derived the rate of convergence. There has been a large amount work on asymptotic properties of the Lasso in high dimensional settings in the context of model-selection. For further details, see the works of Huang et al. (2008), Meinshausen and Yu (2009), Zhang and Huang (2008), Meinshausen and Bühhnann (2006), Bickel et al. (2009) and references therein.

The rest of the paper is organized as follows. The main results on strong consistency and rates of convergence are stated in Section 2. The proofs of our results are given in Section 3.

2 Main Results

2.1. Strong Consistency. Consider the regression model (1.1) with iid error variables ϵ_i's where $\mathbf{E}|\epsilon_1| < \infty$ and $\mathbf{E}(\epsilon_1) = 0$. Although the Lasso criterion function is used mainly for the case where the second moment of ϵ_1 is finite, the Lasso estimators (and also the least squares estimators) are well-defined even when the second moment of ϵ_1 does not exist. Here 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 vectors $\mathbf{x}_i$'s. The first result asserts strong consistency of the Lasso estimator when the penalty λ_n is $o(n)$ as $n \to \infty$.

THEOREM 2.1. *Let $\{\epsilon_i\}$ be iid random variables with $\mathbf{E}|\epsilon_1| < \infty$ and $\mathbf{E}(\epsilon_1) = 0$. Suppose that there exists a nonsingular matrix $\boldsymbol{C}$ such that*

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i\mathbf{x}_i' \to \boldsymbol{C}, \quad \text{as} \quad n \to \infty. \tag{2.1}$$

If $\frac{\lambda_n}{n} \to 0$, then $\widehat{\boldsymbol{\beta}}_n \to \boldsymbol{\beta}$, w.p. 1.

Theorem 2.1 extends the weak consistency results of Knight and Fu (2000) who established the convergence in probability of $\widehat{\boldsymbol{\beta}}_n$, under the assumption that $\mathbf{E}\left(\epsilon_1^2\right) < \infty$. It also shows that strong consistency of the Lasso estimator holds merely under the finiteness of the first moment of ϵ_1, provided $\frac{\lambda_n}{n} \to 0$. When the regularization parameter λ_n grows at a faster rate, the strong consistency of $\widehat{\boldsymbol{\beta}}_n$ may fail, as shown by the following result.

THEOREM 2.2. *Let $\{\epsilon_i\}$ be iid random variables with $\mathbf{E}|\epsilon_1| < \infty$ and $\mathbf{E}(\epsilon_1) = 0$. Assume that (2.1) holds as $n \to \infty$.*

(a) If $\frac{\lambda_n}{n} \to a \in (0, \infty)$, then

$$\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta} \longrightarrow \operatorname*{argmin}_{\mathbf{u}} V_\infty(\mathbf{u}, a),$$

where $V_\infty(\mathbf{u}, a) = \mathbf{u}'\boldsymbol{C}\mathbf{u} + a\sum_{i=1}^{p}\left(|\beta_i + u_i| - |\beta_i|\right)$.

(b) If $\frac{\lambda_n}{n} \to \infty$, then $\widehat{\boldsymbol{\beta}}_n \to \mathbf{0}$, w.p. 1.

Theorem 2.2 shows that, in general, the Lasso estimator is inconsistent whenever λ_n grows precisely at the rate n or faster. Under part (a), consider the special case where all $\beta_j = 0$. In this case, it is easy to check that $\operatorname{argmin}_{\mathbf{u}} V_\infty(\mathbf{u}, a) = \mathbf{0}$ and therefore, $\widehat{\boldsymbol{\beta}}_n$ is consistent for the zero components of $\boldsymbol{\beta}$. Part (b) says that for $\lambda_n \gg n$, the Lasso estimators are consistent for the zero components, but not for the non-zero components. Thus, to ensure strong consistency of the Lasso estimators for all components, the regularization parameter λ_n should be chosen in a such way that it grows at a rate slower than the sample size n.

2.2. Rates of convergence. In this section, we consider the rate of almost sure convergence of the Lasso estimator under a stronger moment condition on the error variables, where we assume that $\mathbf{E}|\epsilon_1|^\alpha < \infty$ for some $1 < \alpha < 2$. In this case, we have the following rate bound:

Theorem 2.3. *Suppose that* $\mathbf{E}|\epsilon_1|^\alpha < \infty$ *for some* $1 < \alpha < 2$ *and* $\mathbf{E}(\epsilon_1) = 0$. *Also suppose that* (2.1) *holds and that*

$$\max\Big\{\|\mathbf{x}_i\| : 1 \leqslant i \leqslant n\Big\} = O(1), \quad \text{as } n \to \infty. \tag{2.2}$$

If, in addition, $\lambda_n/n^{1/\alpha} \to a \in (0,\infty)$ *as* $n \to \infty$*, then*

$$\|\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\| = O\Big(n^{-(\alpha-1)/\alpha}\Big), \quad \text{w.p. 1}$$

Thus for $1 < \alpha < 2$, $\widehat{\boldsymbol{\beta}}_n$ converges to $\boldsymbol{\beta}$ at the rate of $O\left(n^{(\alpha-1)/\alpha}\right)$, w.p. 1. For $\alpha \geqslant 2$, the results in Knight and Fu (2000) shows that $n^{1/2}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ has a non-degenerate limit distribution and therefore, an almost sure bound on the difference $\|\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\|$ similar to that in Theorem 2.3 for values of $\alpha \in [2,\infty)$ is, in general, not possible. Also, note that in Theorem 2.3, we set the regularization parameter λ_n to grow at the rate $n^{1/\alpha}$ which, in particular, satisfies the requirement of Theorem 2.1.

In general, the rate bound given in Theorem 2.3 cannot be improved upon for the Lasso estimators of the non-zero components of $\boldsymbol{\beta}$. The next theorem gives a lower bound on the almost sure rate of convergence for the non-zero components of $\boldsymbol{\beta}$ which shows that under some additional conditions, the rate $n^{-(\alpha-1)/\alpha}$ is *optimal.* To state the result, let γ_0 denote the smallest eigen-value of the matrix $\boldsymbol{C}$ (cf. (2.1)) and let γ^* denote the largest eigen-value of the submatrix of $\boldsymbol{C}$ corresponding to the non-zero components of $\boldsymbol{\beta}$. Also, define

$$y_0 = \left(1 - p_0^{-1}\right), \tag{2.3}$$

where p_0 is the number of nonzero components of $\boldsymbol{\beta}$, i.e., $p_0 = |\{j : 1 \leqslant j \leqslant p, \beta_j \neq 0\}|$. Note that $y_0 \in [0,1)$ for $p_0 \geqslant 1$. *Without loss of generality, for* *the rest of the paper, we will suppose that* $\beta_j \neq 0$ *for all* $j = 1, \ldots, p_0$ *(if* $p_0 > 0$*) and* $\beta_j = 0$ *for* $j = p_0 + 1, \ldots, p$. With this notation, we are now ready to state the lower bound result.

THEOREM 2.4. *Suppose that* $\mathbf{E}|\epsilon_1|^\alpha < \infty$ *for some* $1 < \alpha < 2$, $\mathbf{E}(\epsilon_1) = 0$, *and* $\lambda_n / n^{1/\alpha} \to a \in (0, \infty)$ *as* $n \to \infty$. *Also, suppose that* (2.1) *and* (2.2) *hold, and that* $p_0 > 0$ *and* $\gamma_0/\gamma^* > y_0$. *Then, there exists a constant* $K_1 \in (0, \infty)$, *such that for all* $1 \leqslant j \leqslant p_0$,

$$\liminf_{n\to\infty} \left|\widehat{\beta}_{n,j} - \beta_j\right| n^{(\alpha-1)/\alpha} > K_1 > 0, \quad \text{w.p. } 1 \tag{2.4}$$

Thus, for nonzero components of $\boldsymbol{\beta}$, its Lasso estimator $\widehat{\beta}_{n,j}$ has an almost sure convergence rate that is precisely $n^{(\alpha-1)/\alpha}$ when the conditions of Theorem 2.4 are satisfied. Note that for $p_0 = 1$, $y_0 = 0$ and hence, $\gamma_0/\gamma^* > y_0$. That is, if $\boldsymbol{\beta}$ has a single non-zero component, then the corresponding Lasso estimator cannot converge at an almost sure rate faster than $n^{(\alpha-1)/\alpha}$. More generally, the condition $\gamma_0/\gamma^* > y_0$ holds in the case of 'balanced' designs where the eigen-values of $\boldsymbol{C}$ are equal. In particular, when the covariates $\mathbf{x}_i$'s are given by realizations of a collection of normalized iid random vectors with the identity covariance matrix, (2.1) holds with $\boldsymbol{C} = \boldsymbol{I}_p$, the identity matrix of order p. In this case, $\gamma_0/\gamma^* = 1$, which is greater than y_0, and the Lasso estimator of the non-zero components of $\boldsymbol{\beta}$ has the exact rate $n^{(\alpha-1)/\alpha}$ of almost sure convergence.

The convergence rate of the Lasso estimator from Theorems 2.3, 2.4, may be compared with the corresponding rate for the (ordinary) least squares estimator $\widehat{\beta}_n^{\mathrm{OLS}}$ of $\boldsymbol{\beta}$. For iid, zero-mean error variables $\epsilon_1, \epsilon_2, \ldots$, with $\mathbf{E}|\epsilon_1|^\alpha < \infty$, $1 < \alpha < 2$ in (1.1), by a weighted version of the Marcinkiewz-Zygmund strong law of large numbers (SLLN) (cf. Lemma 3.2 below), it follows that

$$\|\widehat{\boldsymbol{\beta}}_n^{\mathrm{OLS}} - \boldsymbol{\beta}\| = o\left(n^{(\alpha-1)/\alpha}\right), \quad \text{w.p. } 1. \tag{2.5}$$

Hence, under the conditions of Theorem 2.4, the Lasso estimator of the nonzero components of $\boldsymbol{\beta}$ has a *slower rate* of convergence than the OLS estimator of $\boldsymbol{\beta}$. The penalization leads to a loss of accuracy of the Lasso estimator of nonzero components, compared to the ordinary least squares estimation with no penalty.

Next consider the zero components of $\boldsymbol{\beta}$. It turns out that the scenario can be drastically different for the Lasso estimators of the zero-components of $\boldsymbol{\beta}$. Under some structural conditions on the design matrix $\boldsymbol{C}$, the Lasso estimators of the zero component can capture the true parameter value *exactly*, as shown in Theorem 2.5 below :

Theorem 2.5. *Suppose that the conditions of Theorem 2.4 hold. Let* $\boldsymbol{C}_{12} = \{((c_{i,j}))\,,\ 1 \leqslant i \leqslant p_0,\ (p_0+1) \leqslant j \leqslant p\}$, *denote the upper right submatrix of* $\boldsymbol{C}$ *of order* $p_0 \times p_1$, *where* $p_1 = p - p_0$. *Suppose that* $\boldsymbol{C}_{12} = \mathbf{O}$. *Then,*

$$\widehat{\beta}_{n,j} = 0, \quad \textit{eventually w.p. 1 for all } j = (p_0+1)\,, \ldots, p,$$

i.e., there exists a set A *with* $\mathbf{P}(A) = 1$ *such that for all* $\omega \in A$, *there exists* $n_\omega \geqslant 1$ *such that* $\widehat{\beta}_{n,j}(\omega) = 0$ *for all* $j = (p_0+1)\,, \ldots, p$, *whenever* $n \geqslant n_\omega$.

The condition '$\boldsymbol{C}_{12} = \mathbf{O}$' can be thought of as some sort of an orthogonality condition and can be achieved by suitably choosing the matrix $\boldsymbol{X}_n$ in applications that allow design of experiments. Thus, unlike the Lasso estimators of the non-zero components of $\boldsymbol{\beta}$, the Lasso estimator is more accurate than the OLS of the zero components, and reproduce the exact true value of the unknown parameter when this orthogonality condition is satisfied.

It is worth pointing out that the remarkable property of the Lasso estimators of the zero components may fail for a general $\boldsymbol{C}$ matrix when $\boldsymbol{C}_{12} \neq \mathbf{O}$, as shown by the following example.

Example 2.1. Let $\{\epsilon_i\}_{i\geqslant 1}$ be *iid* with $\mathbf{E}|\epsilon_1|^\alpha < \infty$ for some $\alpha \in (1,2)$, $\mathbf{E}(\epsilon_1) = 0$. Suppose that $\lim_{n\to\infty} \lambda_n n^{-1/\alpha} = a \in (0,\infty)$, and that $\beta_j > 0$ for all $j = 1, \ldots, p_0$ and $\beta_p = 0$, where $p = (p_0+1)$. Also, suppose that (2.1) and (2.2) hold and $\boldsymbol{C}$ is of the form

$$\boldsymbol{C} = \begin{bmatrix} M\boldsymbol{I}_{p_0} & \gamma\mathbf{1} \\ \gamma\mathbf{1}' & m \end{bmatrix},$$

where $\boldsymbol{I}_{p_0}$ is the identity matrix of order p_0, $\mathbf{1} \in \mathbb{R}^{p_0}$ is a vector of 1's, and $m, M, \gamma \in \mathbb{R}$. Then, there exists a choice of m, γ and M satisfying $0 < m < \gamma \leqslant 1$ and $M > \gamma p_0$ and a constant $K_0 = K_0(m,\gamma,M,p_0) \in (0,\infty)$ such that

$$\liminf_{n\to\infty} n^{(2-\alpha)/\alpha} \left|\widehat{\beta}_{n,j} - \beta_j\right| \geqslant K_0 > 0, \quad \text{for all } j = 1, \ldots, p. \tag{2.6}$$

A proof of (2.6) is given in Section 3. This example shows that in general, the Lasso estimators of the zero components also cannot converge at a rate faster than $n^{-(2-\alpha)/\alpha}$.

Thus, for a general design matrix $\boldsymbol{C}$, the rate bound given in Theorem 2.3 is optimal.

2.3. The infinite mean case. For the sake of completeness, we also investigate almost sure behavior of $\widehat{\boldsymbol{\beta}}_n$ when the error variables have a finite αth absolute moment for some $0 < \alpha < 1$. Note that in this case, the mean of the ϵ_i's is not necessarily well defined. As a result, for the regression model (1.1) to make sense, some symmetry conditions (such as the median of ϵ_1 is zero) on the distribution of ϵ_1 is needed. The next result, is however, valid without such symmetry assumptions.

THEOREM 2.6. *Suppose that* $\mathbf{E}|\epsilon_1|^{\alpha} < \infty$ *for some* $\alpha \in (0,1)$ *and that* (2.1) *hold. Suppose* $\lambda_n/n^{1/\alpha} \to a \in (0,\infty]$. *Then,*

$$\widehat{\boldsymbol{\beta}}_n \to \mathbf{0}, \quad \textit{w.p. 1.}$$

Theorem 2.6 shows that even for such heavy-tailed error distributions, the Lasso estimators of the zero components of $\boldsymbol{\beta}$ are strongly consistent, but those for the non-zero components are not.

3 Proofs

Let $\boldsymbol{C}_n = n^{-1}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i'$, $n \geqslant 1$. Let $\gamma_{0,n}$ = the smallest eigenvalue of $\boldsymbol{C}_n$ and γ_n^* = the largest eigenvalue of the the $p_0 \times p_0$ submatrix of $\boldsymbol{C}_n$ consisting of the first p_0 rows and p_0 columns. In the proofs below, we write $C, C(\cdot)$ to denote generic constants that depends on its arguments, but not on n. Let *i.o.* stand for 'infinitely often'. Also, let $\text{sgn}(\cdot)$ denote the sign function, i.e., $\text{sgn}(x) = -1, 0, 1$ according as $x < 0$, $x = 0$, $x > 0$. Let $\mathbb{1}(\cdot)$ denotes the indicator function. Unless otherwise specified, the limits in the order symbols and elsewhere are taken by letting n tend to infinity.

LEMMA 3.1. *Suppose* $\mathbf{E}|\epsilon_1| < \infty$ *and* $\mathbf{E}(\epsilon_1) = 0$. *Also, suppose that*

$$\frac{1}{n}\sum_{i=1}^{n}\|\mathbf{x}_i\|^2 = O(1). \tag{3.1}$$

Then, $\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\epsilon_i \longrightarrow \mathbf{0}$ *as* $n \to \infty$, *w.p. 1.*

PROOF OF LEMMA 3.1. Since it is enough prove the almost sure convergence componentwise, for notational simplicity, w.l.g., we assume that the $\mathbf{x}_i$'s are scalars. Let

$$\breve{\epsilon}_i = \epsilon_i \mathbb{1}(|\epsilon_i| \leqslant i), \quad i \geqslant 1.$$

Since, $\mathbf{E}|\epsilon_1| < \infty$,

$$\mathbf{P}(\epsilon_i \neq \breve{\epsilon}_i,\ i.o.) = \mathbf{P}(|\epsilon_i| > i,\ i.o.) = 0.$$

Note that $\mathbf{E}\big(\epsilon_1 \mathbb{1}\,(|\epsilon_1| \leqslant i)\,\big) \to \mathbf{E}\,(\epsilon_1) = 0$ as $i \to \infty$. Hence,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} x_i \mathbf{E}\,(\breve{\epsilon}_i) &= \frac{1}{n}\sum_{i=1}^{n} x_i \mathbf{E}\Big\{\epsilon_i \mathbb{1}\,(|\epsilon_i| \leqslant i)\,\Big\} \\
&\leqslant \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right)^{1/2} \left(\frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{E}\Big\{\epsilon_1 \mathbb{1}\,(|\epsilon_1| \leqslant i)\,\Big\}\right)^2\right)^{1/2} \\
&\to 0, \quad \text{as } n \to \infty.
\end{aligned}$$

By Kronecker's Lemma, it is now enough to show that

$$\sum_{i=1}^{\infty} \frac{x_i}{i}\big(\breve{\epsilon}_i - \mathbf{E}\,(\breve{\epsilon}_i)\,\big) \quad \text{converges w.p. 1.} \tag{3.2}$$

To that end, write $s_n^2 = \sum_{j=1}^{n} x_j^2$, $n \geqslant 1$. By condition (3.1), $s_n^2 \leqslant Cn$ for all $n \geqslant 1$. Hence, for any $m \geqslant 1, j \geqslant 1$,

$$\begin{aligned}
\sum_{i=j}^{j+m} x_i^2 i^{-2} &= \sum_{i=j}^{j+m} s_i^2 i^{-2} - \sum_{i=j-1}^{j+m-1} s_i^2 (i+1)^{-2} \\
&= \sum_{i=j}^{j+m} s_i^2 \left(i^{-2} - (i+1)^{-2}\right) + s_{j+m}^2 (j+m+1)^{-2} - s_{j-1}^2 j^{-2} \\
&\leqslant C \sum_{i=j}^{j+m} i\,(2i+1)\,\big[i\,(i+1)\,\big]^{-2} + C\,(j+m)\,(j+m+1)^{-2} \\
&\leqslant c \sum_{i=j}^{\infty} i^{-2} + c(j+m)^{-1} \;\leqslant\; C j^{-1}.
\end{aligned} \tag{3.3}$$

Using (3.3) we can write

$$\begin{aligned}
\sum_{i=1}^{\infty} \mathbf{Var}\Big(i^{-1} x_i (\breve{\epsilon}_i)^2\Big) &\leqslant \sum_{i=1}^{\infty} i^{-2} x_i^2 \mathbf{E}\Big\{\epsilon_1^2 \mathbb{1}\,(|\epsilon_1| \leqslant i)\,\Big\} \\
&= \sum_{j=1}^{\infty} \mathbf{E}\Big\{\epsilon_1^2 \mathbb{1}\big(\,(j-1) < |\epsilon_1| \leqslant j\big)\Big\} \sum_{i=j}^{\infty} i^{-2} x_i^2 \\
&\leqslant C \sum_{j=1}^{\infty} \mathbf{E}\Big\{|\epsilon_1| \mathbb{1}\big(\,(j-1) < |\epsilon_1| \leqslant j\big)\Big\} = C \cdot \mathbf{E}|\epsilon_1| \;<\; \infty.
\end{aligned}$$

Hence, by Theorem 8.34 of Athreya and Lahiri (2006), (3.2) follows. □

Lemma 3.2. *Suppose that ϵ_i's are iid with $\mathbf{E}|\epsilon_1|^\alpha < \infty$ for some $\alpha \in (0,2)$, and $\mathbf{E}(\epsilon_1) = 0$, if $\alpha \geqslant 1$. Suppose that* (2.1) *and* (2.2) *hold. Then,*

$$n^{-1/\alpha} \sum_{i=1}^{n} \mathbf{x}_i \epsilon_i \to \mathbf{0}, \quad \textit{w.p. 1.} \tag{3.4}$$

Proof of Lemma 3.2. Under (2.2), one can easily modify the steps in the proof of Theorem 8.44 of Athreya and Lahiri (2006) (Marcinkiewicz-Zygmund SLLN) to prove (3.4). We omit the details. □

Proof of Theorem 2.1. Note that

$$\begin{aligned} \hat{\boldsymbol{\beta}}_n &= \operatorname*{argmin}_{\mathbf{t}} \sum_{i=1}^{n} \left(y_i - \mathbf{x}_i'\mathbf{t}\right)^2 + \lambda_n \sum_{i=1}^{p} |t_i| \\ &= \operatorname*{argmin}_{\mathbf{t}} \sum_{i=1}^{n} \Big[\epsilon_i - \mathbf{x}_i'\left(\mathbf{t} - \boldsymbol{\beta}\right)\Big]^2 + \lambda_n \sum_{i=1}^{p} |\beta_i + t_i - \beta_i| \\ \Rightarrow \left(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\right) &= \operatorname*{argmin}_{\mathbf{u}} \sum_{i=1}^{n} \left(\epsilon_i - \mathbf{x}_i'\mathbf{u}\right)^2 + \lambda_n \sum_{i=1}^{p} |\beta_i + u_i|. \end{aligned}$$

Recall that $\boldsymbol{C}_n = n^{-1}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i'$, $\gamma_{0,n}$ = is the smalles eigenvalue of $\boldsymbol{C}_n$ and γ_0 = the smallest eigenvalue of $\boldsymbol{C}$. Let $\overline{\boldsymbol{W}}_n = n^{-1}\sum_{i=1}^n \mathbf{x}_i\epsilon_i$. 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 $\mathbf{u}$, we have

$$\begin{aligned} \left(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\right) &= \operatorname*{argmin}_{\mathbf{u}} \left\{\mathbf{u}'\boldsymbol{C}_n\mathbf{u} - 2\overline{\boldsymbol{W}}_n'\mathbf{u} + \frac{\lambda_n}{n}\sum_{i=1}^{p}\Big[|\beta_i + u_i| - |\beta_i|\Big]\right\} \\ &\equiv \operatorname*{argmin}_{\mathbf{u}} V_n(\mathbf{u}), \quad \text{(say)}. \end{aligned} \tag{3.5}$$

Note that for any $\mathbf{u} \in \mathbb{R}^p$,

$$\begin{aligned} V_n(\mathbf{u}) &\geqslant \gamma_{0,n}\|\mathbf{u}\|^2 - 2\|\overline{\boldsymbol{W}}_n\|\|\mathbf{u}\| - \frac{\lambda_n}{n}\sum_{i=1}^{p}|u_i| \\ &\geqslant \gamma_{0,n}\|\mathbf{u}\|^2 - 2\|\overline{\boldsymbol{W}}_n\|\|\mathbf{u}\| - \frac{\lambda_n}{n}\sqrt{p}\|\mathbf{u}\|. \end{aligned} \tag{3.6}$$

Next fix $\eta \in (0,1)$. Since $\lambda_n/n = o(1)$, there exists a $n_0 \in (0,\infty)$ such that $\lambda_n/n \leqslant \eta$ and $\gamma_{0,n} > \gamma_0/2$ for all $n \geqslant n_0$. On the set $\left\{\|\overline{\boldsymbol{W}}_n\| \leqslant \eta\right\}$, by (3.6), for any $\mathbf{u} \in \mathbb{R}^p$, with $\|\mathbf{u}\| > \eta\left(4 + 2\sqrt{p}\right)/\gamma_{0,n}$,

$$V_n(\mathbf{u}) \geqslant \|\mathbf{u}\|\Big(\gamma_{0,n}\|\mathbf{u}\| - 2\eta - \sqrt{p}\eta\Big) \geqslant \gamma_{0,n}\frac{\|\mathbf{u}\|^2}{2} \; > \; 0.$$

Since $V_n(\mathbf{0}) = 0$, it follows that for $n \geqslant n_0$, the minimum of $V_n(\mathbf{u})$ cannot be attained in the set $\left\{\mathbf{u} : \|\mathbf{u}\| > \eta\left(4 + 2\sqrt{p}\right)/\gamma_{0,n}\right\}$, whenever $\{\|\overline{\boldsymbol{W}}_n\| \leqslant \eta\}$. Hence, it follows that for $n \geqslant n_0$, $\{\|\overline{\boldsymbol{W}}_n\| \leqslant \eta\}$ implies

$$\left(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\right) = \underset{\mathbf{u}}{\operatorname{argmin}}\ V_n(\mathbf{u}) \in \left\{\mathbf{u} : \|\mathbf{u}\| \leqslant \frac{\eta\left(4 + 2\sqrt{p}\right)}{\gamma_{0,n}}\right\}.$$

In particular,

$$\begin{aligned}\mathbf{P}\left(\|\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\| > \frac{2\eta\left(4 + 2\sqrt{p}\right)}{\gamma_0},\ i.o.\right) &\leqslant \mathbf{P}\left(\|\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\| > \frac{\eta\left(4 + 2\sqrt{p}\right)}{\gamma_{0,n}},\ i.o.\right) \\ &\leqslant \mathbf{P}\left(\|\overline{\boldsymbol{W}}_n\| > \eta,\ i.o.\right) \ = \ 0,\end{aligned}$$

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

PROOF OF THEOREM 2.2. First, consider part (a). Let $V_n(\cdot)$ be as in (3.5). Note that for each i,

$$\Big||\beta_i + u_i| - |\beta_i|\Big| \leqslant |u_i|.$$

Since $\frac{\lambda_n}{n} \to a \in (0, \infty)$, for any compact set $\mathbf{K} \subset \mathbb{R}^p$,

$$\begin{aligned}&\sup_{\mathbf{u}\in\mathbf{K}} |V_n(\mathbf{u}) - V_\infty(\mathbf{u}, a)| \\ &\leqslant \sup_{\mathbf{u}\in\mathbf{K}} \left[\|u\|^2\|\boldsymbol{C}_n - \boldsymbol{C}\| + 2\|\overline{\boldsymbol{W}}_n\|\|\mathbf{u}\| + \left|n^{-1}\lambda_n - a\right|\sum_{i=1}^p |u_i|\right] \\ &= o(1) \quad \text{as } n \to \infty, \quad \text{w.p. } 1. \end{aligned} \tag{3.7}$$

Let $n_0 \geqslant 1$ be such that for all $n \geqslant n_0$, $\lambda_n/n < 2a$ and $\gamma_{0,n} > \gamma_0/2$. From (3.6), for all $n \geqslant n_0$, on the set $\left\{\|\overline{\boldsymbol{W}}_n\| \leqslant a\right\}$, we have

$$\begin{aligned}V_n(\mathbf{u}) &\geqslant \|\mathbf{u}\|\left[\gamma_{0,n}\|\mathbf{u}\| - 2\|\overline{\boldsymbol{W}}_n\| - \frac{\lambda_n}{n}\sqrt{p}\right] \\ &\geqslant \|\mathbf{u}\|\left[\frac{\gamma_0}{2}\|\mathbf{u}\| - 2a - 2a\sqrt{p}\right] \geqslant \frac{\|\mathbf{u}\|}{2},\end{aligned}$$

for all $\|\mathbf{u}\| > \left[1 + 4a\left(1 + \sqrt{p}\right)\right]/\gamma_0 \equiv c_0$. Since, $V_n(\mathbf{0}) = 0$, this implies $\|\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\| \leqslant c_0$, whenever $n \geqslant n_0$ and $\left\{\|\overline{\boldsymbol{W}}_n\| \leqslant a\right\}$. Thus, the minimizer of $V_n(\mathbf{u})$ lies in a compact set for all $n \geqslant n_0$, provided $\Big\{\|\overline{\boldsymbol{W}}_n\| \leqslant a\Big\}$. Since $V_\infty(\cdot; a)$ is a convex function, by (3.7) and Lemma 3.1, part (a) follows.

Next consider part (b). Let $a_n^2 = \lambda_n/n$. Then, $a_n \to \infty$. Also, let $\mathbf{B}_n = \Big\{\mathbf{u} : \|\mathbf{u}\| \leqslant a_n,\ |\beta_i + u_i| \geqslant a_n^{-1}$ for atleast one $i = 1, \ldots, p\Big\}$. By Lemma 3.1,

$$
\begin{aligned}
&\inf_{\mathbf{u}\in\mathbf{B}_n}\Big[\mathbf{u}'\boldsymbol{C}_n\mathbf{u} - 2\overline{\boldsymbol{W}}_n'\mathbf{u} + a_n^2\sum_{i=1}^p |\beta_i + u_i|\Big] \\
&\geqslant \inf_{\mathbf{u}\in\mathbf{B}_n}\Big[a_n^2\sum_{i=1}^p |\beta_i + u_i| - 2\|\overline{\boldsymbol{W}}_n\|\|\mathbf{u}\|\Big] \\
&\geqslant \inf_{\mathbf{u}\in\mathbf{B}_n}\Big[a_n^2\sum_{i=1}^p |\beta_i + u_i| - 2\|\overline{\boldsymbol{W}}_n\|\ \sup\big\{\|\mathbf{u}\| : \mathbf{u}\in\mathbb{B}_n\big\}\Big] \\
&= \inf_{\mathbf{u}\in\mathbf{B}_n}\Big[a_n^2\sum_{i=1}^p |\beta_i + u_i| - 2\|\overline{\boldsymbol{W}}_n\| a_n\Big] \\
&\geqslant a_n\Big[1 - 2\|\overline{\boldsymbol{W}}_n\|\Big] \ \to\ \infty, \quad \text{as } n\to\infty,\ \text{w.p. } 1.
\end{aligned} \tag{3.8}
$$

Also, by Lemma 3.1,

$$
\begin{aligned}
&\inf_{\|\mathbf{u}\|>a_n}\Big[\mathbf{u}'\boldsymbol{C}_n\mathbf{u} - 2\overline{\boldsymbol{W}}_n'\mathbf{u} + a_n^2\sum_{i=1}^p |\beta_i + u_i|\Big] \\
&\geqslant \inf_{\|\mathbf{u}\|>a_n}\Big[\mathbf{u}'\boldsymbol{C}_n\mathbf{u} - 2\|\overline{\boldsymbol{W}}_n\|\|\mathbf{u}\|\Big] \\
&\geqslant \inf_{\|\mathbf{u}\|>a_n}\Big[\gamma_{0,n}\|\mathbf{u}\|^2 - 2\|\overline{\boldsymbol{W}}_n\|\|\mathbf{u}\|\Big] \\
&= \inf_{\|\mathbf{u}\|>a_n}\|\mathbf{u}\|\Big[\gamma_{0,n}\|\mathbf{u}\| - 2\|\overline{\boldsymbol{W}}_n\|\Big] \\
&\geqslant a_n\Big[a_n\gamma_{0,n} - 2\|\overline{\boldsymbol{W}}_n\|\Big] \to \infty, \quad \text{as } n\to\infty,\ \text{w.p. } 1.
\end{aligned} \tag{3.9}
$$

Finally, with $\mathbf{B}_{1,n} = \Big\{\mathbf{u} : |\beta_i + u_i| \leqslant a_n^{-1},\ \text{for all } i = 1, \ldots, p\Big\}$,

$$
\begin{aligned}
&\inf_{\mathbf{u}\in\mathbf{B}_{1,n}}\Big[\mathbf{u}'\boldsymbol{C}_n\mathbf{u} - 2\overline{\boldsymbol{W}}_n'\mathbf{u} + a_n^2\sum_{i=1}^p |\beta_i + u_i|\Big] \\
&\leqslant \mathbf{u}_0'\boldsymbol{C}_n\mathbf{u}_0 - 2\overline{\boldsymbol{W}}_n'\mathbf{u}_0 + a_n^2\sum_{i=1}^p |\beta_i + u_{0,i}|,
\end{aligned}
$$

$$
\begin{aligned}
&= \mathbf{u}_0' \boldsymbol{C}_n \mathbf{u}_0 - 2\overline{\boldsymbol{W}}_n' \mathbf{u}_0 \quad (\text{where, } \mathbf{u}_0 = (u_{0,1}, \ldots, u_{0,p})' = -\boldsymbol{\beta} \in \mathbf{B}_{1,n}), \\
&\to \boldsymbol{\beta}' \boldsymbol{C} \boldsymbol{\beta} \in [0, \infty), \quad \text{as } n \to \infty, \text{ w.p. 1.} \qquad (3.10)
\end{aligned}
$$

Note that for any sequence $\{\mathbf{u}_n\}_{n \geqslant 1}$, with $\mathbf{u}_n \in \mathbf{B}_n$, $\|\mathbf{u}_n + \boldsymbol{\beta}\| \leqslant a_n^{-1}\sqrt{p} \to 0$, as $n \to \infty$. Hence, from (3.8)- (3.10) and (3.5), it follows that there exists a set A with $\mathbf{P}(A) = 1$ and for all $\omega \in A$, there exists a $n_\omega \geqslant 1$ such that for all $n \geqslant n_\omega$,

$$
\begin{aligned}
\left(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\right) &= \underset{\mathbf{u}}{\operatorname{argmin}}\ V_n(\mathbf{u}) \\
&= \underset{\mathbf{u}}{\operatorname{argmin}} \left[\mathbf{u}' \boldsymbol{C}_n \mathbf{u} - 2\overline{\boldsymbol{W}}_n' \mathbf{u} + a_n^2 \sum_{i=1}^{p} |\beta_i + u_i|\right] \\
&= \underset{\mathbf{u} \in \mathbf{B}_{1,n}}{\operatorname{argmin}} \left[\mathbf{u}' \boldsymbol{C}_n \mathbf{u} - 2\overline{\boldsymbol{W}}_n' \mathbf{u} + a_n^2 \sum_{i=1}^{p} |\beta_i + u_i|\right] \\
&\to -\boldsymbol{\beta}, \quad \text{as } n \to \infty,
\end{aligned}
$$

This completes the proof of part (b). □

Proof of Theorem 2.3. Since

$$
\widehat{\boldsymbol{\beta}}_n = \underset{\mathbf{t}}{\operatorname{argmin}} \sum_{i=1}^{n} \left[\epsilon_i - \mathbf{x}_i' (\mathbf{t} - \boldsymbol{\beta})\right]^2 + \lambda_n \sum_{i=1}^{p} |\beta_i + t_i - \beta_i|,
$$

it follows that

$$
\begin{aligned}
&n^{1/\alpha}\left(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\right) \\
&= \underset{\mathbf{u}}{\operatorname{argmin}} \sum_{i=1}^{n} \left(\epsilon_i - n^{-1/\alpha} \mathbf{x}_i' \mathbf{u}\right)^2 + \lambda_n \sum_{i=1}^{p} \left|\beta_i + n^{-1/\alpha} u_i\right| \\
&= \underset{\mathbf{u}}{\operatorname{argmin}} \left[n^{-2/\alpha} \mathbf{u}' \left(\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i'\right) \mathbf{u} - 2n^{-1/\alpha} \mathbf{u}' \sum_{i=1}^{n} \mathbf{x}_i \epsilon_i \right. \\
&\qquad \left. + \lambda_n \sum_{i=1}^{p} \left(\left|\beta_i + n^{-1/\alpha} u_i\right| - |\beta_i|\right)\right] \\
&\equiv \underset{\mathbf{u}}{\operatorname{argmin}}\ \check{V}_n(\mathbf{u}), \quad (\text{say}).
\end{aligned}
$$

Write $\breve{\boldsymbol{W}}_n = n^{-1/\alpha}\sum_{i=1}^n \mathbf{x}_i\epsilon_i$. Then, by Lemma 3.2, $\|\breve{\boldsymbol{W}}_n\| = o(1)$, almost surely. Define the sets

$$\begin{aligned}
\mathbf{B}_{1,n} &= \left\{\mathbf{u} : \|\mathbf{u}\| > K_1 n^{(3-\alpha)/2\alpha}\right\}\\
\mathbf{B}_{2,n} &= \left\{\mathbf{u} : \mathbf{u} \in \mathbf{B}_{1,n}^c,\ \max_{1\leqslant i\leqslant p} |u_i| > K_2 n^{(2-\alpha)/\alpha}\right\}\\
\mathbf{B}_{3,n} &= \Big\{\mathbf{u} : K_3 n^{(2-\alpha)/\alpha} < |u_i| \leqslant K_2 n^{(2-\alpha)/\alpha}\ \forall\ i = 1,\ldots,p_0;\\
&\qquad |u_i| \leqslant K_2 n^{(2-\alpha)/\alpha}\ \forall\ i = (p_0+1),\ldots,p\Big\} \text{ and,}\\
\mathbf{B}_{4,n} &= \Big\{\mathbf{u} :\ |u_i| \leqslant K_3 n^{(2-\alpha)/\alpha} \text{ for some } i = 1,\ldots,p_0;\\
&\qquad |u_j| \leqslant K_2 n^{(2-\alpha)/\alpha} \text{ for all } j = 1,\ldots,p\Big\},
\end{aligned} \tag{3.11}$$

where,

$$K_1 = \left[8a\left(\sum_{i=1}^p |\beta_i| + 1\right)\gamma_0^{-1}\right]^{1/2}$$

and K_2, K_3 are any given real numbers (not depending on n) satisfying

$$K_2 \in \left[4a\,(p+1)\,\gamma_0^{-1}, \infty\right) \quad \text{and} \quad K_3 \in \left(0, 1 - \sqrt{p_0\left(1 - \frac{\gamma_0}{\gamma^*}\right)\frac{a}{2\gamma_0}}\right).$$

The sets $\mathbf{B}_{3,n}$ and $\mathbf{B}_{4,n}$ would be used in the proof of the next result. Let

$$A = \left\{\|\breve{\boldsymbol{W}}_n\| \to 0 \quad \text{as } n \to \infty\right\}.$$

Then for all $\omega \in A$, there exists $n_\omega \geqslant 1$ such that for all $n \geqslant n_\omega$,

$$\begin{aligned}
&\inf_{\mathbf{u}\in\mathbf{B}_{1,n}} \breve{V}_n(\mathbf{u})\\
\geqslant\ &\inf_{\mathbf{u}\in\mathbf{B}_{1,n}} \left[n^{(1-2/\alpha)}\gamma_{0,n}\|\mathbf{u}\|^2 - 2\|\breve{\boldsymbol{W}}_n\|\|\mathbf{u}\| + \lambda_n \sum_{i=1}^p \left(n^{-1/\alpha}|u_i| - 2|\beta_i|\right)\right]\\
\geqslant\ &\inf_{\mathbf{u}\in\mathbf{B}_{1,n}} \|\mathbf{u}\|\left[n^{(1-2/\alpha)}\gamma_{0,n}\|\mathbf{u}\| - 2\|\breve{\boldsymbol{W}}_n\| + \lambda_n n^{-1/\alpha}\right] - 2\lambda_n \sum_{i=1}^p |\beta_i|\\
\geqslant\ &\inf_{\mathbf{u}\in\mathbf{B}_{1,n}} \|\mathbf{u}\|\left[n^{(1-2/\alpha)}\gamma_{0,n}\|\mathbf{u}\| + \frac{a}{2}\right] - 2\lambda_n \sum_{i=1}^p |\beta_i|
\end{aligned}$$

$$\geqslant \inf_{\mathbf{u}\in\mathbf{B}_{1,n}} \left[n^{(1-2/\alpha)} \|\mathbf{u}\|^2 \frac{\gamma_0}{2} \right] - 2\lambda_n \sum_{i=1}^{p} |\beta_i|$$
$$= \left[\gamma_0/2\right] K_1^2 n^{(1-2/\alpha)} n^{(3-\alpha)/\alpha} - 2\lambda_n \sum_{i=1}^{p} |\beta_i|$$
$$\geqslant C n^{1/\alpha}. \tag{3.12}$$

Note that for $\alpha > 1$, $(3-\alpha)/[2\alpha] < \alpha^{-1}$, and therefore

$$\sup \left\{ \|\mathbf{u}\| n^{-1/\alpha} : \mathbf{u} \in \mathbf{B}_{1,n}^c \right\} = o(1).$$

Hence for $\mathbf{u} \in \mathbf{B}_{1,n}^c$,

$$\breve{V}_n(\mathbf{u}) = n^{(1-2/\alpha)} \mathbf{u}' \boldsymbol{C}_n \mathbf{u} - 2\mathbf{u}' \breve{\boldsymbol{W}}_n + \frac{\lambda_n}{n^{1/\alpha}} \Big[\sum_{j=1}^{p_0} \operatorname{sgn}(\beta_j) u_j + \sum_{j=p_0+1}^{p} |u_j| \Big]$$
$$= n^{(1-2/\alpha)} \mathbf{u}' \boldsymbol{C}_n \mathbf{u} + \sum_{j=1}^{p_0} u_j \left(\operatorname{sgn}(\beta_j) n^{-1/\alpha} \lambda_n - 2\breve{W}_{j,n} \right)$$
$$+ \sum_{j=p_0+1}^{p} |u_j| \left(n^{-1/\alpha} \lambda_n - 2\breve{W}_{j,n} \operatorname{sgn}(u_j) \right). \tag{3.13}$$

It is easy to check that for any $\omega \in A$, there exists $n_\omega \geqslant 1$, such that for all $n \geqslant n_\omega$,

$$\inf_{\mathbf{u}\in\mathbf{B}_{2,n}} \breve{V}_n(\mathbf{u})$$
$$\geqslant \inf_{\mathbf{u}\in\mathbf{B}_{2,n}} \left[n^{(1-2/\alpha)} \gamma_{0,n} \|\mathbf{u}\|^2 - \sum_{j=1}^{p} |u_j| \left(n^{-1/\alpha} \lambda_n + 2 |\breve{W}_{j,n}| \right) \right]$$
$$\geqslant \inf_{\mathbf{u}\in\mathbf{B}_{2,n}} \left[n^{(1-2/\alpha)} \gamma_{0,n} \|\mathbf{u}\|^2 - n^{-1/\alpha} \lambda_n \|\mathbf{u}\| (p+1)^{1/2} \right]$$
$$= \inf_{\|\mathbf{u}\|\in\mathbf{B}_{2,n}} \|\mathbf{u}\| \left[n^{(1-2/\alpha)} \gamma_{0,n} \|\mathbf{u}\| - n^{-1/\alpha} \lambda_n (p+1)^{1/2} \right]$$
$$\geqslant K_2 n^{(2-\alpha)/\alpha} \left[n^{(1-2/\alpha)} \gamma_{0,n} K_2 n^{(2-\alpha)/\alpha} - n^{-1/\alpha} \lambda_n (p+1)^{1/2} \right]$$
$$\geqslant C n^{(2-\alpha)/\alpha}. \tag{3.14}$$

Since, by (3.13), $\breve{V}_n(\mathbf{0}) = 0$,

$$\inf_{\mathbf{u}\in\mathbf{B}_{2,n}^c} \breve{V}_n(\mathbf{u}) \leqslant \breve{V}_n(\mathbf{0}) = 0 < \min \left\{ C n^{(2-\alpha)/\alpha}, C n^{1/a} \right\}.$$

Hence by (3.12) and (3.14), $\inf_{\mathbf{u}} \check{V}_n(\mathbf{u}) = \inf_{\mathbf{u}\in\mathbf{B}_{2,n}^c} \check{V}_n(\mathbf{u})$ for $n \geqslant n_\omega$ for all $\omega \in A$. This completes the proof. □

Proof of Theorem 2.4. Next consider Theorem 2.4 and let A be defined as before, *i.e.*

$$A = \left\{ \|\breve{\boldsymbol{W}}_n\| \to 0 \quad \text{as } n \to \infty \right\}.$$

Then, for each $\omega \in A$, there exists $n_\omega \geqslant 1$, such that for all $n \geqslant n_\omega$, by (3.11),

$$\begin{aligned}
\inf_{\mathbf{u}\in\mathbf{B}_{4,n}} \check{V}_n(\mathbf{u}) &\geqslant \inf_{\mathbf{u}\in\mathbf{B}_{4,n}} \Bigg[n^{1-2/\alpha}\gamma_{0,n}\|\mathbf{u}\|^2 - \sum_{j=1}^{p_0} |u_j| \left(n^{-1/\alpha}\lambda_n + 2|\breve{W}_{j,n}| \right) \\
&\qquad + \sum_{j=p_0+1}^{p} |u_j| \left(\frac{\lambda_n}{\sqrt{n}} - 2\breve{W}_{j,n} \right) \Bigg] \\
&\geqslant \inf_{\mathbf{u}\in\mathbf{B}_{4,n}} \left[n^{1-2/\alpha}\gamma_{0,n} \sum_{j=1}^{p_0} u_j^2 - \sum_{j=1}^{p_0} |u_j| \left(n^{-1/\alpha}\lambda_n + 2\|\breve{\boldsymbol{W}}_n\| \right) \right] \\
&\geqslant \inf_{\mathbf{u}\in\mathbf{B}_{4,n}} \sum_{j=1}^{p_0} \left\{ n^{1-2/\alpha}\gamma_{0,n} u_j^2 - |u_j| \left(n^{-1/\alpha}\lambda_n + 2\|\breve{\boldsymbol{W}}_n\| \right) \right\}.
\end{aligned}$$

Now, consider the function $f(x) = c_1 x^2 - c_2 x$, $x \geqslant 0$, $c_1, c_2 \geqslant 0$. This function is strictly decreasing on $\left(0, \frac{c_2}{2c_1}\right)$, strictly increasing on $\left(\frac{c_2}{2c_1}, \infty\right)$, and attains its minimum at $x = \frac{c_2}{2c_1}$. The minimum value of $f(\cdot)$ is given by $f\left(\frac{c_2}{2c_1}\right) = -\frac{c_2^2}{4c_1}$, and $\min_{0\leqslant x\leqslant x_0} f(x) = f(x_0)$ for all $x_0 \in \left[0, \frac{c_2}{2c_1}\right]$.

Now, apply this to each of the p_0 terms and use the definition of $\mathbf{B}_{4,n}$ to conclude that for all $\omega \in A$ and $\eta \in (0,1)$, there exists $n_\omega \geqslant 1$ such that for all $n \geqslant n_\omega$,

$$\begin{aligned}
\inf_{\mathbf{u}\in\mathbf{B}_{4,n}} \check{V}_n(\mathbf{u}) \geqslant\ & (p_0 - 1) \left[-\frac{\left(n^{-1/\alpha}\lambda_n + 2\|\breve{\boldsymbol{W}}_n\| \right)^2}{4\gamma_{0,n} n^{1-2/\alpha}} \right] \\
&+ \left[n^{1-2/\alpha}\gamma_{0,n} K_3^2 n^{2(2-\alpha)/\alpha} - K_3 n^{(2-\alpha)/\alpha} \left(\frac{\lambda_n}{n^{1/\alpha}} + 2\|\breve{\boldsymbol{W}}_n\| \right) \right] \\
= n^{(2-\alpha)/\alpha} & \left[\gamma_{0,n} K_3^2 - K_3 \left(\frac{\lambda_n}{n^{1/\alpha}} + 2\|\breve{\boldsymbol{W}}_n\| \right) \right.
\end{aligned}$$

$$- (p_0 - 1) \frac{\left(n^{-1/\alpha}\lambda_n + 2\|\breve{\boldsymbol{W}}_n\|\right)^2}{4\gamma_{0,n}} \Bigg]$$

$$\geqslant \; - n^{(2-\alpha)/\alpha} \left[-\gamma_0 K_3^2 + K_3 a + (p_0 - 1) \frac{a^2}{4\gamma_0} \right] (1 + \eta) \, ; \tag{3.15}$$

Finally, consider $\breve{V}_n(\mathbf{u})$ for $\mathbf{u} \in \mathbf{B}_{3,n}$. Let,

$$\mathbf{u}_0 = \Big(- \operatorname{sgn}(\beta_1), \ldots, -\operatorname{sgn}(\beta_{p_0}), 0, \ldots, 0 \Big)' K_4 n^{(2-\alpha)/\alpha},$$

with $K_4 = a/2\gamma^*$.

Then, for all $\omega \in A$ and $\eta \in (0,1)$, there exists $n_\omega \geqslant 1$ such that for all $n \geqslant n_\omega$,

$$\inf_{\mathbf{u} \in \mathbf{B}_{3,n}} \breve{V}_n(\mathbf{u}) \leqslant \; \breve{V}_n(\mathbf{u}_0) \quad (\text{as } \mathbf{u}_0 \in \mathbf{B}_{3,n})$$

$$\leqslant n^{(1-2/\alpha)} \gamma_n^* \|\mathbf{u}_0\|^2 + \sum_{j=1}^{p_0} u_{j,0} \left(\operatorname{sgn}(\beta_j) n^{-1/\alpha} \lambda_n - 2\breve{W}_{j,n} \right)$$

$$\leqslant n^{1-2/\alpha} \gamma_n^* p_0 K_4^2 n^{2(2-\alpha)/\alpha} - p_0 n^{-1/\alpha} \lambda_n K_4 n^{(2-\alpha)/\alpha} + 2K_4 n^{(2-\alpha)/\alpha} \sum_{j=1}^{p_0} |\breve{W}_{j,n}|$$

$$= \; - n^{(2-\alpha)/\alpha} \left[K_4 p_0 n^{-1/\alpha} \lambda_n - K_4^2 p_0 \gamma_n^* - 2K_4 \sum_{j=1}^{p_0} |\breve{W}_{j,n}| \right]$$

$$\leqslant \; - n^{-(2-\alpha)/\alpha} \left[K_4 p_0 a - K_4^2 p_0 \gamma^* \right] (1 - \eta) \tag{3.16}$$

Now check that by the condition, $\gamma_0 > \gamma^* \left(1 - p_0^{-1}\right)$, we have

$$\left[K_4 p_0 a - K_4^2 p_0 \gamma^* \right] > \left[-\gamma_0 K_3^2 + K_3 a + \frac{(p_0 - 1)\, a^2}{4\gamma_0} \right].$$

Hence, there exists $\eta_0 \in (0, 1$ such that for all $\eta \in (0, \eta_0)$,

$$\left[K_4 p_0 a - K_4^2 p_0 \gamma^* \right] (1 - \eta) > \left[-\gamma_0 K_3^2 + K_3 a + \frac{(p_0 - 1)\, a^2}{4\gamma_0} \right] (1 + \eta) \, . \tag{3.17}$$

By choosing $\eta < \eta_0$, and using (3.12)-(3.17), it follows that for all $\omega \in A$, there exists $n_\omega \geqslant 1$ such that for all $n \geqslant n_\omega$,

$$\inf_{\mathbf{u}} \breve{V}_n(\mathbf{u}) = \inf_{\mathbf{u} \in \mathbf{B}_{3,n}} \breve{V}_n(\mathbf{u}) \, . \tag{3.18}$$

Theorem 2.4 follows from this. □

Proof of Theorem 2.5. Write

$$\boldsymbol{C} = \begin{pmatrix} \boldsymbol{C}_{11} & \boldsymbol{C}_{12} \\ \boldsymbol{C}_{21} & \boldsymbol{C}_{22} \end{pmatrix},$$

where $\boldsymbol{C}_{11}$ is $p_0 \times p_0$. From the proof of Theorem 2.4 (cf. (3.18)), it follows that

$$n^{1/\alpha}\left(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\right) = \underset{\mathbf{u}}{\text{argmin}}\ \breve{V}_n(\mathbf{u}) = \underset{\mathbf{u}\in\mathbf{B}_{3,n}}{\text{argmin}}\ \breve{V}_n(\mathbf{u}),$$

where $\breve{V}_n(\mathbf{u})$ has the representation (3.13), and

$$\mathbf{B}_{3,n} = \Big\{\mathbf{u} : K_3 n^{(2-\alpha)/\alpha} < |u_i|\ \forall\ i = 1,\dots,p_0; \\ |u_j| \leqslant K_2 n^{(2-\alpha)/\alpha}\ \forall\ j = 1,\dots,p\Big\}$$

with K_2 and K_3 as defined in (3.11). Fix

$$\mathbf{u}^{(1)} = (u_1,\dots,u_{p_0})' \in \left[K_3 n^{(2-\alpha)/\alpha}, K_2 n^{(2-\alpha)/\alpha}\right],$$

and let

$$\mathbf{B}_{3,n}^{(1)} = \Big\{\mathbf{u}^{(2)} : |u_j| \leqslant K_2 n^{(2-\alpha)/\alpha},\ (p_0+1) \leqslant j \leqslant p\Big\}.$$

Note that, as $\boldsymbol{C}_{12} = 0$,

$$\begin{aligned}
&\underset{\mathbf{u}^{(2)}\in\mathbf{B}_{2,n}^{(1)}}{\text{argmin}}\ \breve{V}_n\left(\mathbf{u}^{(1)},\mathbf{u}^{(2)}\right) \\
&= \underset{\mathbf{u}^{(2)}\in\mathbf{B}_{2,n}^{(1)}}{\text{argmin}}\ n^{(1-2/\alpha)}\Big\{\left(\mathbf{u}^{(2)}\right)'\boldsymbol{C}_{22}\mathbf{u}^{(2)} + 2\left(\mathbf{u}^{(1)}\right)'\boldsymbol{C}_{12}\mathbf{u}^{(2)}\Big\} \\
&\qquad + \sum_{j=p_0+1}^{p} |u_j|\Big[n^{-1/\alpha}\lambda_n - 2\breve{W}_{j,n}\text{sgn}(u_j)\Big] \\
&= \underset{\mathbf{u}^{(2)}\in\mathbf{B}_{2,n}^{(1)}}{\text{argmin}}\ n^{(1-2/\alpha)}\left(\mathbf{u}^{(2)}\right)'\boldsymbol{C}_{22}\mathbf{u}^{(2)} + \sum_{j=p_0+1}^{p} |u_j|\Big[n^{-1/\alpha}\lambda_n - 2\breve{W}_{j,n}\text{sgn}(u_j)\Big].
\end{aligned}$$

Since $\|\breve{\boldsymbol{W}}_n\| = o(1)$ w.p. 1, for every $\omega \in A \equiv \Big\{\|\breve{\boldsymbol{W}}_n\| = o(1)\Big\}$, there exists $n_\omega \geqslant 1$, such that for all $n \geqslant n_\omega$,

$$n^{-1/\alpha}\lambda_n - 2|\breve{W}_{j,n}| > \frac{a}{2} > 0, \quad \text{for all } j = (p_0+1),\dots,p.$$

Hence, by the positive definiteness of $\boldsymbol{C}_{22}$, the minimizer of the expression above is $\mathbf{u}^{(2)} = \mathbf{0}$. This proves the theorem. $\square$

Proof of Example 2.1. Let $\mathbf{y} = n^{-(2-\alpha)/\alpha}\mathbf{u}$, *i.e.*, $\mathbf{u} = n^{(2-\alpha)/\alpha}\mathbf{y}$, and,

$$\begin{aligned}\mathbf{B}_3 &= n^{-(2-\alpha)/\alpha}\mathbf{B}_{3,n}\\ &= \{\mathbf{y} : K_3 < |y_i| \leqslant K_2,\ \text{for } i = 1, \dots, p_0,\ \text{and}\\ &\qquad |y_i| \leqslant K_2 \text{ for } i = (p_0+1), \dots, p\}.\end{aligned}$$

Note that

$$\begin{aligned}\breve{V}_n(\mathbf{u}) &= \breve{V}_n\left(n^{(2-\alpha)/\alpha}\mathbf{y}\right)\\ &= n^{(2-\alpha)/\alpha}\left[\mathbf{y}'\boldsymbol{C}_n\mathbf{y} + \sum_{j=1}^{p_0} y_j\left\{n^{-1/\alpha}\lambda_n\operatorname{sgn}(\beta_j) - 2\breve{W}_{j,n}\right\}\right.\\ &\qquad\left. + \sum_{j=p_0+1}^{p}|y_j|\left\{n^{-1/\alpha}\lambda_n - 2\breve{W}_{j,n}\operatorname{sgn}(y_j)\right\}\right],\end{aligned}$$

and $\mathbf{u} = n^{(2-\alpha)/\alpha}\mathbf{y} \in \mathbf{B}_{3,n} \Leftrightarrow y \in \mathbf{B}_3$.

Hence,

$$\begin{aligned}&\operatorname*{argmin}_{\mathbf{u}} \breve{V}_n(\mathbf{u}) = \operatorname*{argmin}_{\mathbf{u}\in\mathbf{B}_{3,n}} \breve{V}_n(\mathbf{u}) = \operatorname*{argmin}_{\mathbf{u}\in\mathbf{B}_{3,n}} n^{-(2-\alpha)/\alpha}\breve{V}_n(\mathbf{u})\\ =\ & n^{(2-\alpha)/\alpha}\operatorname*{argmin}_{\mathbf{y}\in\mathbf{B}_3}\left[\mathbf{y}'\boldsymbol{C}_n\mathbf{y} + \sum_{j=1}^{p_0} y_j\left\{n^{-1/\alpha}\lambda_n\operatorname{sgn}(\beta_j) - 2\breve{W}_{j,n}\right\}\right.\\ &\left. + \sum_{j=p_0+1}^{p}|y_j|\left\{n^{-1/\alpha}\lambda_n - 2\breve{W}_{j,n}\operatorname{sgn}(y_j)\right\}\right]\\ =\ & n^{(2-\alpha)/\alpha}\big(1 + o(1)\big)\\ &\times \operatorname*{argmin}_{\mathbf{y}\in\mathbf{B}_3}\left[\mathbf{y}'\boldsymbol{C}\mathbf{y} + a\left\{\sum_{j=1}^{p_0}\operatorname{sgn}(\beta_j)\,y_j + \sum_{j>p_0}|y_j|\right\}\right],\ \text{w.p. } 1. \quad (3.19)\end{aligned}$$

Since $\operatorname{sgn}(\beta_j) = 1$ for $1 \leqslant j \leqslant p_0$, for any $\mathbf{y} \in \mathbb{R}^p$, we have

$$Q(y) \equiv \mathbf{y}'\boldsymbol{C}\mathbf{y} + a\left(\sum_{j=1}^{p_0} y_j + \sum_{j=p_0+1}^{p}|y_j|\right)$$

$$
\begin{aligned}
&= \sum_{i=1}^{p} c_{i,i} y_i^2 + 2 \sum_{i=1}^{p_0} \sum_{j=p_0+1}^{p} c_{i,j} y_i y_j + a \left(\sum_{j=1}^{p_0} y_j + \sum_{j=p_0+1}^{p} |y_j| \right) \\
&\geqslant \sum_{i=1}^{p} c_{i,i} y_i^2 - 2 \sum_{i=1}^{p_0} \sum_{j=p_0+1}^{p} c_{i,j} |y_i| |y_j| - a \sum_{j=1}^{p_0} |y_j| + a \sum_{j=p_0+1}^{p} |y_j| \\
&= Q\left(\mathbf{g}\left(\mathbf{y}\right)\right),
\end{aligned}
$$

where, $g_j\left(\mathbf{y}\right) = -|y_j|$, $1 \leqslant j \leqslant p_0$ and $g_j\left(\mathbf{y}\right) = |y_j|$, $(p_0+1) \leqslant j \leqslant p$. Hence, it follows that

$$
\begin{aligned}
&\operatorname*{argmin}_{\mathbf{y} \in \mathbf{B}_3} \mathbf{y}' \boldsymbol{C} \mathbf{y} + a \left(\sum_{j=1}^{p_0} y_j + \sum_{j=p_0+1}^{p} |y_j| \right) \\
= &\operatorname*{argmin}_{\substack{K_3 \leqslant y_i \leqslant K_2,\ i=1,\ldots,p_0 \\ 0 \leqslant y_j \leqslant K_2,\ j=p_0+1,\ldots,p}} \sum_{i=1}^{p} c_{i,i} y_i^2 - 2 \sum_{i=1}^{p_0} \sum_{j=p_0+1}^{p} c_{i,j} y_i y_j - a \sum_{i=1}^{p_0} y_i + a \sum_{j=p_0+1}^{p} y_j.
\end{aligned}
$$

Next consider the quadratic form

$$
\begin{aligned}
Q_1\left(\mathbf{y}\right) &\equiv \sum_{i=1}^{p} c_{i,i} y_i^2 - 2 \sum_{i=1}^{p_0} \sum_{j=p_0+1}^{p} c_{i,j} y_i y_j - a \sum_{i=1}^{p_0} y_i + a \sum_{j=p_0+1}^{p} y_j \\
&= \mathbf{y}' \boldsymbol{A} \mathbf{y} - 2\mathbf{b}'\mathbf{y},
\end{aligned}
$$

where

$$
\boldsymbol{A} = \begin{bmatrix} \boldsymbol{C}_{11} & -\boldsymbol{C}_{12} \\ -\boldsymbol{C}_{12}' & \boldsymbol{C}_{22} \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \Big(\underbrace{-a, \ldots, -a}_{p_0}, \underbrace{a, \ldots, a}_{(p-p_0)} \Big)'.
$$

It is easy to check that $Q_1\left(\mathbf{y}\right)$ attains it minimum (over $\mathbb{R}^p$) at $\mathbf{y}_0 = \boldsymbol{A}^{-1}\mathbf{b}$. Now from Rao (1973) (pp. 33),

$$
\boldsymbol{A}^{-1} = \begin{bmatrix} \boldsymbol{C}_{11}^{-1} + \boldsymbol{F} E^{-1} \boldsymbol{F}' & -\boldsymbol{F} E^{-1} \\ -E^{-1} \boldsymbol{F}' & -E^{-1} \end{bmatrix},
$$

where,

$$
\begin{aligned}
\boldsymbol{F} &= -\boldsymbol{C}_{11}^{-1} \boldsymbol{C}_{12} = -\frac{\gamma}{M} \mathbf{1}, \quad \text{and,} \\
E &= \boldsymbol{C}_{22} - \boldsymbol{C}_{12}' \boldsymbol{C}_{11}^{-1} \boldsymbol{C}_{12} = m - \frac{\gamma^2}{M} \mathbf{1}'\mathbf{1} = \left(m - \frac{p_0 \gamma}{M} \right) \\
\Longrightarrow \boldsymbol{A}^{-1} &= \begin{bmatrix} M^{-1} \mathbf{I}_{p_0} + \left(\frac{\gamma}{M}\right)^2 \frac{\mathbf{1}\mathbf{1}'}{E} & \frac{\gamma}{ME} \mathbf{1} \\ \frac{\gamma}{ME} \mathbf{1}' & E^{-1} \end{bmatrix}.
\end{aligned}
$$

Thus for all $1 \leqslant i \leqslant p_0$,

$$\begin{aligned}\left(\boldsymbol{A}^{-1}\mathbf{b}\right)_i &= \text{the } i\text{-th coordinate of } \boldsymbol{A}^{-1}\mathbf{b} = -a\left[M^{-1} + \frac{\gamma^2 p_0}{M^2 E}\right] + \frac{a\gamma}{ME} \\ &= \frac{a}{ME}\left[\gamma + \frac{p_0\gamma}{M} - m - \frac{\gamma^2 p_0}{M}\right] \\ &= \frac{a}{ME}\left[\gamma - m + \frac{p_0\gamma}{M}(1-\gamma)\right].\end{aligned}$$

and for $i = p_0 + 1$,

$$\left(\boldsymbol{A}^{-1}\mathbf{b}\right)_{p_0+1} = \frac{a}{E}\left[1 - \frac{\gamma p_0}{M}\right].$$

Clearly

$$m < \gamma \leqslant 1 \quad \text{and} \quad M > \gamma p_0, \tag{3.20}$$

makes all co-ordinates of $\boldsymbol{A}^{-1}\mathbf{b}$ positive. Also, note that the eigenvalues of $\boldsymbol{C}$ are given by (cf. pp. 32, Rao (1973)):

$$\begin{aligned}&\det(\boldsymbol{C} - \lambda\mathbf{I}) = 0 \\ \Rightarrow\ &\begin{vmatrix}(M-\lambda)\mathbf{I} & \gamma\mathbf{1} \\ \gamma\mathbf{1}' & (m-\lambda)\end{vmatrix} = 0 \\ \Rightarrow\ &(M-\lambda)^{p_0-1}\left[(m-\lambda)(M-\lambda) - \gamma^2 p_0\right] = 0 \\ \Rightarrow\ &\lambda_i = M, \quad \text{for } (p_0 - 1) \text{ many } i\text{'s and}\end{aligned}$$

$$\lambda = \frac{1}{2}\left[(m+M) \pm \left\{(M-m)^2 + 4\gamma^2 p_0\right\}^{1/2}\right].$$

Note that for M large, $(M+m) > \left[(M-m)^2 + 4\gamma^2 p_0\right]^{1/2}$, and hence,

$$\begin{aligned}\gamma_0 &= \frac{1}{2}\left[(M+m) - \left\{(M-m)^2 + 4\gamma^2 p_0\right\}^{1/2}\right] \quad \text{and,} \\ \gamma_p &= \frac{1}{2}\left[(M+m) + \left\{(M-m)^2 + 4\gamma^2 p_0\right\}^{1/2}\right],\end{aligned}$$

where γ_p is the largest eigenvalue of $\boldsymbol{C}$. Since K_2^{-1} and K_3 can be chosen arbitrarily small, it is easy to find a set of m, γ, M such that (3.20) holds, $\gamma_0 > 0$ and $\mathbf{y}_0 \equiv \boldsymbol{A}^{-1}\mathbf{b} \in \mathbf{B}_3$. For any such choice of m, γ_0, M, $y_{0,p_0+1} = \left(\boldsymbol{A}^{-1}\mathbf{b}\right)_{p_0+1} \in (0, \infty)$, and therefore, by (3.19), (2.6) holds. □

Proof of Theorem 2.6. Note that,

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_n &= \underset{\mathbf{t}}{\operatorname{argmin}} \frac{1}{n^{1/\alpha}} \sum_{i=1}^{n} \left(y_i - \mathbf{x}_i'\mathbf{t}\right)^2 + \frac{\lambda_n}{n^{1/\alpha}} \sum_{i=1}^{n} |t_i| \\
\Rightarrow \left(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\right) &= \underset{\mathbf{u}}{\operatorname{argmin}}\ n^{-1/\alpha}\left(\mathbf{u}'\boldsymbol{X}_n'\boldsymbol{X}_n\mathbf{u}\right) - 2\widetilde{\boldsymbol{W}}_n'\mathbf{u} + n^{-1/\alpha}\lambda_n \sum_{i=1}^{p} |\beta_i + u_i| \\
&= \underset{\mathbf{u}}{\operatorname{argmin}}\ \widetilde{V}_n(\mathbf{u}) \quad \text{(say)}.
\end{aligned}$$

Fix $\eta \in (0, \max\{1, \delta\})$, where $\delta = \max\{|\beta_i| : 1 \leqslant i \leqslant p_0\}$. (Set $\delta = 0$ if $p_0 = 0$). Let

$$\begin{aligned}
\mathbf{D}_1 &= \{\mathbf{u} : \|\mathbf{u}\| > (4p\delta + 1)\}, \\
\mathbf{D}_2 &= \Big\{\mathbf{u} : |\beta_i + u_i| > \eta,\ \text{for some } i,\ 1 \leqslant i \leqslant p\ ;\ \mathbf{u} \in \mathbf{D}_1^c\Big\}, \\
\mathbf{D}_3 &= \Big\{\mathbf{u} : |\beta_i + u_i| \leqslant \eta,\ \text{for all } i = 1, \ldots, p\Big\}.
\end{aligned}$$

Then, there exists a set A with $P(A) = 1$ such that for all $\omega \in A$, there exists a $n_\omega \geqslant 1$ such that for all $n \geqslant n_\omega$, the following inequalities hold:

$$\begin{aligned}
\inf_{\mathbf{u} \in \mathbf{D}_1} \widetilde{V}_n(\mathbf{u}) &\geqslant \inf_{\mathbf{u}} \left[-2\widetilde{\boldsymbol{W}}_n'\mathbf{u} + n^{-1/\alpha}\lambda_n \sum_{i=1}^{p} |\beta_i + u_i| \right] \\
&\geqslant \inf_{\mathbf{u}} \left[-2\|\widetilde{\boldsymbol{W}}_n\|\|\mathbf{u}\| + n^{-1/\alpha}\lambda_n \Big(\|\mathbf{u}\| - \sum_{i=1}^{p} |\beta_i| \Big) \right] \\
&\geqslant \inf_{\mathbf{u}} \left[\|\mathbf{u}\| \Big(-2\|\widetilde{\boldsymbol{W}}_n\| + n^{-1/\alpha}\lambda_n \Big) - n^{-1/\alpha}\lambda_n \sum_{i=1}^{p} |\beta_i| \right] \\
&= (4p\delta + 1)\frac{a}{2} - 2a \sum_{i=1}^{p} |\beta_i| \geqslant \frac{a}{2}\ ;
\end{aligned}$$

Similarly,

$$\begin{aligned}
\inf_{\mathbf{u} \in \mathbf{D}_2} \widetilde{V}_n(\mathbf{u}) &\geqslant \inf_{\mathbf{u} \in \mathbf{D}_2} \left[-2\|\widetilde{\boldsymbol{W}}_n\|\|\mathbf{u}\| + n^{-1/\alpha}\lambda_n \sum_{i=1}^{p} |\beta_i + u_i| \right] \\
&\geqslant -2\|\widetilde{\boldsymbol{W}}_n\|(4p\delta + 1) + n^{-1/\alpha}\lambda_n\eta \\
&\geqslant \frac{a\eta}{2}, \quad \text{w.p. 1. ;}
\end{aligned}$$

and, using the fact that $0 < \alpha < 1$,

$$\inf_{\mathbf{u}\in\mathbf{D}_3} \widetilde{V}_n(\mathbf{u}) \leqslant V_n(-\boldsymbol{\beta}) \leqslant n^{(1-1/\alpha)}\boldsymbol{\beta}'\boldsymbol{C}_n\boldsymbol{\beta} + 2\widetilde{\boldsymbol{W}}'_n\boldsymbol{\beta} \leqslant \frac{a\eta}{4}.$$

Hence, for all $\omega \in A$ (where $P(A) = 1$), there exists $n_\omega \geqslant 1$ and that for all $n \geqslant n_\omega$,

$$\operatorname*{argmin}_{\mathbf{u}} \widetilde{V}_n(\mathbf{u}) = \operatorname*{argmin}_{\mathbf{u}\in\mathbf{D}_3} \widetilde{V}_n(\mathbf{u}) \Rightarrow \left(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\right) \in \mathbf{D}_3$$

Hence, it follows that $\left(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\right) \to -\boldsymbol{\beta}$ almost surely, which implies $\widehat{\boldsymbol{\beta}}_n \to \mathbf{0}$ almost surely. □

References

Athreya, K. B. and Lahiri, S. N. (2006). *Measure theory and Probability theory.* Springer Texts in Statistics, Springer, New York.

Bickel, P. J., Ritov, Y. and Tsybakov, A. B. (2009). Simultaneous analysis of lasso and dantzig selector. *Ann. Statist.*, **37**, 1705–1732.

Drygas, H. (1976). Weak and strong consistency of the least squares estimators in regression models. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, **34**, 119–127.

Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004). Least angle regression. *Ann. Statist.*, **32**, 407–499. With discussion, and a rejoinder by the authors.

Fu, W. J. (1998). Penalized regressions: the bridge versus the lasso. *J. Comput. Graph. Statist.*, **7**, 397–416.

Huang, J., Horowitz, J. L. and Ma, S. (2008). Asymptotic properties of bridge estimators in sparse high-dimensional regression models. *Ann. Statist.*, **36**, 587–613.

Knight, K. and Fu, W. (2000). Asymptotics for lasso-type estimators. *Ann. Statist.*, **28**, 1356–1378.

Lai, T. L., Robbins, H. and Wei, C. Z. (1978). Strong consistency of least squares estimates in multiple regression. *Proc. Nat. Acad. Sci. U.S.A.*, **75**, 3034–3036.

Lounici, K. (2008). Sup-norm convergence rate and sign concentration property of Lasso and Dantzig estimators. *Electron. J. Stat.*, **2**, 90–102.

Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.*, **34**, 1436–1462.

Meinshausen, N. and Yu, B. (2009). Lasso-type recovery of sparse representations for high-dimensional data. *Ann. Statist.*, **37**, 246–270.

Osborne, M. R., Presnell, B. and Turlach, B. A. (2000). On the LASSO and its dual. *J. Comput. Graph. Statist.,* **9**, 319–337.

Rao, C. R. (1973). *Linear statistical inference and its applications.* John Wiley & Sons, New York London-Sydney, second edition. Wiley Series in Probability and Mathematical Statistics.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc.*, **Ser. B, 58**, 267–288.

Wainwright, M. J. (2006). Sharp thresholds for high-dimensional and noisy recovery of sparsity. Technical report, Dept. of Statistics, UC Berkeley. Preprint available at `http://arxiv.org/abs/math/0605740`.

Zhang, C.-H. and Huang, J. (2008). The sparsity and bias of the lasso selection in high-dimensional linear regression. *Ann. Statist.*, **36**, 1567–1594.

Zhao, P. and Yu, B. (2006). On model selection consistency of Lasso. *J. Mach. Learn. Res.*, **7**, 2541–2563.

Zou, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.*, **101**, 1418–1429.

A. Chatterjee
Stat-Math Unit
Indian Statistical Institute
7, SJSS Marg
New Delhi 110016
India
E-mail: cha@isid.ac.in

S. N. Lahiri
Department of Statistics
Texas A&M University
3143 TAMU
College Station, TX, 77843
USA
E-mail: snlahiri@stat.tamu.edu

Paper received August 2009; revised February 2010.