VARIABLE SELECTION IN SPARSE REGRESSION WITH QUADRATIC MEASUREMENTS

Jun Fan[1], Lingchen Kong[1], Liqun Wang[2] and Naihua Xiu[1]

[1]*Beijing Jiaotong University and* [2]*University of Manitoba*

Supplementary Material

The supplementary file covers technical lemmas and proofs.

S1 Proofs of Theorem 1 and 2

Without loss of generality, in the following we let $\Gamma^* = \{1, \cdots, s\}$ and $\beta^* = (\beta_1^{*T}, 0^T)^T$. Correspondingly, we partition Z_i and x_i as

$$Z_i = \begin{pmatrix} Z_i^{11} & Z_i^{12} \\ Z_i^{21} & Z_i^{22} \end{pmatrix} \quad \text{and} \quad x_i = (x_i^{1T}, x_i^{2T})^T,$$

where Z_i^{11} is an $s \times s$ symmetric matrix and Z_i^{22} is a $(p-s) \times (p-s)$ symmetric matrix. For convenience, we also denote

$$\tilde{L}_n(\beta_1) := \sum_{i=1}^{n} (y_i - \beta_1^T Z_i^{11} \beta_1 - x_i^{1T} \beta_1)^2 + \lambda_n \|\beta_1\|_q^q$$

and $C_1 = 2\bar{c} + 3\sqrt{(\sigma^2 + 1)/c_1}$.

We first prove some lemmas.

Lemma S1.1. *Let $\{w_n\}$ be a sequence of real numbers and assume that $\{b_n\}$ and $\{B_n\}$ are two sequences of positive numbers tending to infinity. If*

$$B_n \geq \sum_{i=1}^{n} w_i^2 \quad and \quad \frac{b_n}{\sqrt{B_n}} \max_{1\leq i\leq n} |w_i| \to 0,$$

then, for any $\tau > 0$,

$$\limsup_{n\to\infty} b_n^{-2} \log \mathbb{P}\Big(|\sum_{i=1}^{n} w_i\varepsilon_i| > b_n\sqrt{B_n}\tau\Big) \leq -\frac{\tau^2}{2\sigma^2},$$

or

$$\mathbb{P}\big(|\sum_{i=1}^{n} w_i\varepsilon_i| > b_n\sqrt{B_n}\tau\big) \leq \exp\big(-\frac{b_n^2\tau^2}{2\sigma^2} + o(b_n^2)\big).$$

Proof. Based on the similar method proof to that of Lemma 3.2 in Fan, Yan and Xiu (2014), it is easy to show it and so is omitted. □

Lemma S1.2. *Assume that Conditions 1-2 and 4 hold. Let $\{a_n\}$ be a sequence of positive numbers satisfying (3.8) and (3.9). Then for any $\tau > 0$,*

$$\mathbb{P}\Big(\frac{1}{a_n\sqrt{n}} \sup_{u\in S} \|\sum_{i=1}^{n}(Z_i^{11}u + x_i^1)\varepsilon_i\| > \tau\Big) \leq \exp\big(-\frac{a_n^2\tau^2}{2c_2\sigma^2} + o(a_n^2)\big).$$

Proof. Let $A = \{v \in \mathbb{R}^s : \|v\| \leq 1\}$ and denote $r_n = 1/n$. Then by Lemma 14.27 in Bühlmann and Van De Geer (2011), we have

$$A \subseteq \bigcup_{j=1}^{m_n} B(v_j, r_n),$$

where $m_n = (1 + 2n)^s$ and $B(u_j, r_n) = \{v \in \mathbb{R}^s : \|v - v_j\| \leq r_n, v_j \in A\}$ for $j = 1, \cdots, m_n$. By the similar method to the proof the second result of

Lemma 5.1 in Fan, Yan and Xiu (2014), we use Lemma S1.1 with $B_n = nc_2$ and $b_n = a_n$ to obtain that for any $\tau_1 > 0$ and $\epsilon_1 \in (0, \tau_1/2)$,

$$\mathbb{P}\Big(\frac{1}{a_n\sqrt{n}}\|\sum_{i=1}^{n}(Z_i^{11}u + x_i^1)\varepsilon_i\| > \tau_1\Big) \leq m_n \exp\big(-\frac{a_n^2(\tau_1-\epsilon_1)^2}{2c_2\sigma^2} + o(a_n^2)\big) \tag{S1.1}$$

Further, denote $r'_n = C_1\sqrt{s}/n$. Again, by Lemma 14.27 of B*ü*hlmann and van de Geer (2011), we have

$$S \subseteq \bigcup_{j=1}^{m_n} B(u_j, r'_n),$$

where $B(u_j, r'_n) = \{u \in \mathbb{R}^s : \|u - u_j\| \leq r'_n, u_j \in S\}$ for $j = 1, \cdots, m_n$. Analog to (S1.1) we obtain that for any $\epsilon \in (0, \tau/2)$ and $\epsilon_1 \in (0, (\tau-\epsilon)/2)$,

$$\mathbb{P}\Big(\frac{1}{a_n\sqrt{n}}\sup_{u\in S}\|\sum_{i=1}^{n}(Z_i^{11}u + x_i^1)\varepsilon_i\| > \tau\Big) \leq m_n^2 \exp\big(-\frac{a_n^2(\tau-\epsilon-\epsilon_1)^2}{2c_2\sigma^2} + o(a_n^2)\big).$$

From (3.8) we conclude that $a_n^{-2}\log m_n^2 = a_n^{-2}\big(s\log(1+2n)\big) \to 0$, which together with the above inequality implies that

$$\limsup_{n\to\infty} a_n^{-2}\log \mathbb{P}\Big(\frac{1}{a_n\sqrt{n}}\sup_{u\in S}\|\sum_{i=1}^{n}(Z_i^{11}u + x_i^1)\varepsilon_i\| > \tau\Big) \leq -\frac{(\tau-\epsilon-\epsilon_1)^2}{2c_2\sigma^2}.$$

Since ϵ and ϵ_1 are arbitrary, we have for large enough n,

$$\mathbb{P}\Big(\frac{1}{a_n\sqrt{n}}\sup_{u\in S}\|\sum_{i=1}^{n}(Z_i^{11}u + x_i^1)\varepsilon_i\| > \tau\Big) \leq \exp\big(-\frac{a_n^2\tau^2}{2c_2\sigma^2} + o(a_n^2)\big).$$

□

Lemma S1.3. *Under the assumptions of Lemma S1.2, there exists* $\hat\beta_1 = \arg\min_{\beta_1\in\mathbb{R}^s} \tilde L_n(\beta_1)$ *such that*

$$\mathbb{P}\big(\|\hat\beta_1 - \beta_1^*\| \leq r_n\big) \geq 1 - \exp\big(-\frac{(1+c_1^2/4)a_n^2}{2c_2\sigma^2} + o(a_n^2)\big). \tag{S1.2}$$

Proof. To show the existence of minimizer $\hat{\beta}_1$, we consider the level set $\{\beta_1 \in \mathbb{R}^s : \tilde{L}_n(\beta_1) \le \tilde{L}_n(\beta_1^*)\}$. It is apparent that

$$\inf_{\beta_1 \in \mathbb{R}^s} \tilde{L}_n(\beta_1) = \inf_{\beta_1 \in \{\beta_1 \in \mathbb{R}^s : \tilde{L}_n(\beta_1) \le \tilde{L}_n(\beta_1^*)\}} \tilde{L}_n(\beta_1).$$

Since $\tilde{L}_n(\cdot)$ is continuous and the level set is non-empty and closed, $\tilde{L}_n(\cdot)$ has at least one minimizer $\hat{\beta}_1$ in the level set.

Now we prove (S1.2). For notational convenience, we denote $\hat{Z} = (\hat{Z}_1, \cdots, \hat{Z}_n)$, $\hat{\Sigma}_n = \hat{Z}\hat{Z}^T/n$ and $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_n)^T$, where $\hat{Z}_i = Z_i^{11}(\hat{\beta}_1 + \beta_1^*) + x_i^1$. Obviously, Condition 1 implies that $\hat{\Sigma}_n$ is invertible. Then by the definition of $\hat{\beta}_1$ we have $\tilde{L}_n(\hat{\beta}_1) \le \tilde{L}_n(\beta_1)$ for any $\beta_1 \in \mathbb{R}^s$, which implies

$$\begin{aligned}\sum_{i=1}^n \varepsilon_i^2 + \lambda_n \sum_{j=1}^s |\beta_{1j}^*|^q \ge & \sum_{i=1}^n \Big(\varepsilon_i - (\hat{\beta}_1 - \beta_1^*)^T \hat{Z}_i\Big)^2 + \lambda_n \sum_{j=1}^s |e_{s,j}^T \hat{\beta}_1|^q \\ = & \sum_{i=1}^n \varepsilon_i^2 - 2(\hat{\beta}_1 - \beta_1^*)^T \hat{Z}\varepsilon + \lambda_n \sum_{j=1}^s |e_{s,j}^T \hat{\beta}_1|^q \\ & + n(\hat{\beta}_1 - \beta_1^*)^T \hat{\Sigma}_n (\hat{\beta}_1 - \beta_1^*)\end{aligned}$$

and therefore

$$n(\hat{\beta}_1 - \beta_1^*)^T \hat{\Sigma}_n (\hat{\beta}_1 - \beta_1^*) \le 2(\hat{\beta}_1 - \beta_1^*)^T \hat{Z}\varepsilon + \lambda_n \sum_{j=1}^s (|e_{s,j}^T \beta_1^*|^q - |e_{s,j}^T \hat{\beta}_1|^q). \tag{S1.3}$$

By the similar method to the proof of relation (8) in Huang, Horowitz and Ma (2008), we conclude from Condition 1, the second convergence of Condition 4 and the strong law of large number that for large enough n,

$$\|\hat{\beta}_1 - \beta_1^*\| \le C_1\sqrt{s} \ \text{ and } \ \|\hat{\beta}_1 + \beta_1^*\| \le C_1\sqrt{s}, \quad a.s.$$

and

$$\begin{aligned}\|\hat{\beta}_1-\beta_1^*\|^2 &\le \frac{2}{nc_1}\|\hat{\beta}_1-\beta_1^*\|\|\hat{Z}\varepsilon\|+\frac{\eta_n}{nc_1}\\ &\le \frac{2C_1\sqrt{s}}{nc_1}\sup_{u\in S}\|\sum_{i=1}^n(Z_i^{11}u+x_i^1)\varepsilon_i\|+\frac{\lambda_n s\bar{c}^q}{nc_1},\quad a.s., \qquad \text{(S1.4)}\end{aligned}$$

where $\eta_n=\lambda_n\sum_{j=1}^s(|e_{s,j}^T\beta_1^*|^q-|e_{s,j}^T\hat{\beta}_1|^q)$. Therefore,

$$\begin{aligned}1 =&\mathbb{P}\Big(\|\hat{\beta}_1-\beta_1^*\|^2\le\frac{2C_1\sqrt{s}}{nc_1}\sup_{u\in S}\|\sum_{i=1}^n(Z_i^{11}u+x_i^1)\varepsilon_i\|+\frac{\lambda_n s\bar{c}^q}{nc_1}\Big)\\ \le&\mathbb{P}\Big(\|\hat{\beta}_1-\beta_1^*\|^2\le\frac{2C_1\sqrt{s}a_n}{c_1\sqrt{n}}+\frac{\lambda_n s\bar{c}^q}{nc_1}\Big)+\mathbb{P}\Big(\frac{1}{a_n\sqrt{n}}\sup_{u\in S}\|\sum_{i=1}^n(Z_i^{11}u+x_i^1)\varepsilon_i\|>1\Big),\end{aligned}$$

which together with Lemma S1.2 yields that

$$\mathbb{P}(\|\hat{\beta}_1-\beta_1^*\|>r_n')\le\exp\Big(-\frac{a_n^2}{2c_2\sigma^2}+o(a_n^2)\Big), \qquad \text{(S1.5)}$$

where $r_n'=\big(\frac{2C_1a_n\sqrt{s}}{c_1\sqrt{n}}+\frac{\lambda_n s\bar{c}^q}{nc_1}\big)^{1/2}$.

Since $r_n'\to 0$ as $n\to\infty$, it follows that for large enough n,

$$\frac{1}{2}|e_{s,j}^T\beta_1^*|\le|e_{s,j}^T\hat{\beta}_1|\le\frac{3}{2}|e_{s,j}^T\beta_1^*|,\quad j=1,\cdots,s,$$

when $\|\hat{\beta}_1-\beta_1^*\|\le r_n'$. By the mean value theorem and Cauchy-Schwarz inequality, we have, for large enough n,

$$\eta_n\le 2\underline{c}^{q-1}\lambda_n\sqrt{s}\|\hat{\beta}_1-\beta_1^*\|$$

when $\|\hat{\beta}_1-\beta_1^*\|\le r_n'$. Combining the above inequality, (S1.3), Cauchy-Schwarz inequality and Condition 1, we have, for large enough n,

$$\|\hat{\beta}_1-\beta_1^*\| \le \frac{2}{nc_1}\|\hat{Z}\varepsilon\|+\frac{2\underline{c}^{q-1}\lambda_n\sqrt{s}}{nc_1},$$

when $\|\hat{\beta}_1 - \beta_1^*\| \leq r_n'$. Therefore it follows from the first inequality of (S1.4) that for large enough n,

$$\begin{aligned}
1 =& \mathbb{P}\Big(\|\hat{\beta}_1 - \beta_1^*\|^2 \leq \frac{2}{nc_1}\|(\hat{\beta}_1 - \beta_1^*)^T\|\|\hat{Z}\varepsilon\| + \frac{\eta_n}{nc_1}\Big) \\
\leq& \mathbb{P}\Big(\|\hat{\beta}_1 - \beta_1^*\| \leq \frac{2}{nc_1}\|\hat{Z}\varepsilon\| + \frac{2\underline{c}^{q-1}\lambda_n\sqrt{s}}{c_1 n}\Big) + \mathbb{P}\Big(\|\hat{\beta}_1 - \beta_1^*\| > r_n^{'}\Big) \\
\leq& \mathbb{P}\Big(\|\hat{\beta}_1 - \beta_1^*\| \leq \frac{a_n}{\sqrt{n}} + \frac{2\underline{c}^{q-1}\lambda_n\sqrt{s}}{c_1 n}\Big) \\
&+ \mathbb{P}\Big(\frac{1}{a_n\sqrt{n}} \sup_{\|u\|\leq C_1\sqrt{s}} \|\sum_{i=1}^n (Z_i^{11}u + x_i^1)\varepsilon_i\| > c_1/2\Big) + \mathbb{P}\Big(\|\hat{\beta}_1 - \beta_1^*\| > r_n'\Big).
\end{aligned}$$

Then, by Lemma S1.2 and (S1.5) we have

$$\mathbb{P}(\|\hat{\beta}_1 - \beta_1^*\| \geq r_n) \leq \exp\big(- \frac{a_n^2}{2c_2\sigma^2} + o(a_n^2)\big) + \exp\big(- \frac{c_1^2 a_n^2}{8c_2\sigma^2} + o(a_n^2)\big),$$

which yields

$$\limsup_{n\to\infty} a_n^{-2} \log \mathbb{P}(\|\hat{\beta}_1 - \beta_1^*\| \geq r_n) \leq -\frac{1}{2c_2\sigma^2} - \frac{c_1^2}{8c_2\sigma^2}.$$

Thus, we have

$$\mathbb{P}(\|\hat{\beta}_1 - \beta_1^*\| \geq r_n) \leq \exp\big(- \frac{(1 + c_1^2/4)a_n^2}{2c_2\sigma^2} + o(a_n^2)\big),$$

which yields (S1.2). □

Proof of Theorem 1 Denote $b_n = a_n + \frac{2\underline{c}^{q-1}\lambda_n\sqrt{s}}{\sqrt{c_1 n}}$ and $\tilde{r}_n = \big(\frac{a_n}{\sqrt{n}} + \frac{2\underline{c}^{q-1}\lambda_n\sqrt{s}}{c_1 n}\big)\sqrt{s}$. We first show that

$$\lambda_n^{-1}\tilde{r}_n^{1-q} b_n \sqrt{n} s \to 0 \quad \text{and} \quad \lambda_n^{-1}\tilde{r}_n^{2-q}\sqrt{n}s^2 \to 0, \quad \text{as} \quad n \to \infty. \qquad \text{(S1.6)}$$

The first convergence of (S1.6) follows from the second convergence of Condition 4 and (3.10). Since $\lambda_n s^2/n \to 0$, the inequality of Condition 2 implies that

$$\frac{s^{3/2}}{\sqrt{n}} \le \frac{\lambda_n s^2}{n} \cdot \frac{\sqrt{n}}{\lambda_n} \le \frac{\lambda_n s^2}{n} \frac{1}{\sigma \underline{c}^{1-q}\sqrt{\log p}} \to 0, \tag{S1.7}$$

which yields

$$\lambda_n^{-1}\tilde{r}_n^{2-q}\sqrt{n}s^2 = \lambda_n^{-1}\tilde{r}_n^{1-q}b_n\sqrt{n}s \cdot \frac{s^{3/2}}{\sqrt{n}} \to 0.$$

For any $u = (u_1^T, u_2^T) \in \mathbb{R}^p$ and $u_1 \in \mathbb{R}^s$, we show that there exists a sufficiently large constant $\tilde{C}$ such that

$$\mathbb{P}\Big(L_n(\hat{\beta}_1, 0) = \inf_{\|u\|_1 \le \tilde{C}} L_n(\beta_1^* + \tilde{r}_n u_1, \tilde{r}_n u_2)\Big) \ge 1 - \exp\big(-C_0 a_n^2 + o(a_n^2)\big), \tag{S1.8}$$

which implies that with probability $1 - \exp\big(-C_0 a_n^2 + o(a_n^2)\big)$ that $(\hat{\beta}_1^T, 0^T)^T$ is a local minimizer in the ball $\{\beta^* + \tilde{r}_n u : \|u\|_1 \le \tilde{C}\}$, so that both (3.6) and (3.7) hold.

Denote $\zeta_{1i} = Z_i^{11}(2\beta_1^* + \tilde{r}_n u_1) + x_i^1$ and $\zeta_{2i} = 2Z_i^{21}(\beta_1^* + \tilde{r}_n u_1) + x_i^2 + \tilde{r}_n Z_i^{22} u_2$, and define event

$$E_1 := \{\|\sum_{i=1}^n \zeta_{2i}\varepsilon_i\|_\infty \le 4((1+\bar{c}^2))^{1/2} b_n \sqrt{n} s\},$$

where $b_n = a_n + \frac{2\underline{c}^{q-1}\lambda_n\sqrt{s}}{\sqrt{c_1 n}}$. For any $u_2 \in \mathbb{R}^{p-s}$, we show that under event

E_1,

$$L_n(\beta_1^* + \tilde{r}_n u_1, \tilde{r}_n u_2) \geq L_n(\beta_1^* + \tilde{r}_n u_1, 0). \tag{S1.9}$$

Clearly, $L_n(\beta_1^* + \tilde{r}_n u_1, \tilde{r}_n u_2) = L_n(\beta_1^* + \tilde{r}_n u_1, 0)$ when $\|u_2\|_1 = 0$. We proceed to show (S1.9) for $\|u_2\|_1 > 0$. It follows that

$$\begin{aligned}
& L_n(\beta_1^* + \tilde{r}_n u_1, \tilde{r}_n u_2) - L_n(\beta_1^* + \tilde{r}_n u_1, 0) \\
= \;& -2\tilde{r}_n \sum_{i=1}^n u_2^T \zeta_{2i} \varepsilon_i + \tilde{r}_n^2 \sum_{i=1}^n (u_2^T \zeta_{2i})^2 + 2\tilde{r}_n^2 \sum_{i=1}^n u_1^T \zeta_{1i} u_2^T \zeta_{2i} + \lambda_n \tilde{r}_n^q \|u_2\|_q^q \\
\geq \;& -2\tilde{r}_n \sum_{i=1}^n u_2^T \zeta_{2i} \varepsilon_i + 2\tilde{r}_n^2 \sum_{i=1}^n u_1^T \zeta_{1i} u_2^T \zeta_{2i} + \lambda_n \tilde{r}_n^q \|u_2\|_q^q. \qquad (S1.10)
\end{aligned}$$

We now use the fact that $|u^T A v| \leq \|u_1\|_1 \|Av\|_\infty \leq |A|_\infty \|u\|_1 \|v\|_1$ for any $n \times d$ matrix A and vector $u \in \mathbb{R}^n$, $v \in \mathbb{R}^d$ to discuss the bound of $|\sum_{i=1}^n u_1^T \zeta_{1i} u_2^T \zeta_{2i}|$. Noting that

$$|\sum_{i=1}^n u_1^T \zeta_{1i} u_2^T \zeta_{2i}| \leq \|u_1\|_1 \|u_2\|_1 |\sum_{i=1}^n \zeta_{1i} \zeta_{2i}^T|_\infty,$$

we then estimate the upper bound of $|\sum_{i=1}^n \zeta_{1i} \zeta_{2i}^T|_\infty$. Recalling the definition of $|\cdot|_\infty$, we calculate the $e_{s,j}^T \zeta_{1i} \zeta_{2i}^T e_{p-s,k}$ for each $j = 1, \cdots, s$ and $k = 1, \cdots, p-s$. It is easy to check that

$$\begin{aligned}
& e_{s,j}^T \zeta_{1i} \zeta_{2i}^T e_{p-s,k} \\
=& 2(2\beta_1^* + \tilde{r}_n u_1)^T Z_i^{11} e_{s,j} e_{p-s,k}^T Z_i^{21} (\beta_1^* + \tilde{r}_n u_1) + 2 x_i^{1T} e_{s,j} e_{p-s,k}^T Z_i^{21} (\beta_1^* + \tilde{r}_n u_1) \\
& + (2\beta_1^* + \tilde{r}_n u_1)^T Z_i^{11} e_{s,j} e_{p-s,k}^T x_i^2 + x_i^{1T} e_{s,j} e_{p-s,k}^T x_i^2 \\
& + \tilde{r}_n (2\beta_1^* + \tilde{r}_n u_1)^T Z_i^{11} e_{s,j} e_{p-s,k}^T Z_i^{22} u_2 + \tilde{r}_n x_i^{1T} e_{s,j} e_{p-s,k}^T Z_i^{22} u_2.
\end{aligned}$$

So,

$$
\begin{aligned}
&|\sum_{i=1}^{n} e_{s,j}^T \zeta_{1i}\zeta_{2i}^T e_{p-s,k}| \\
\le& 2\|(2\beta_1^* + \tilde{r}_n u_1)\|_1 \|(\beta_1^* + \tilde{r}_n u_1)\|_1 |\sum_{i=1}^{n} Z_i^{11} e_{s,j} e_{p-s,k}^T Z_i^{21}|_\infty \\
&+ 2\|\beta_1^* + \tilde{r}_n u_1\|_1 |\sum_{i=1}^{n} e_{s,j}^T x_i^1 e_{p-s,k}^T Z_i^{21}|_\infty \\
&+ \|(2\beta_1^* + \tilde{r}_n u_1)\|_1 |\sum_{i=1}^{n} Z_i^{11} e_{s,j} e_{p-s,k}^T x_i^2|_\infty + |\sum_{i=1}^{n} x_i^{1T} e_{s,j} e_{p-s,k}^T x_i^2|_\infty \\
&+ \tilde{r}_n \|(2\beta_1^* + \tilde{r}_n u_1)\|_1 |\|u_2\|_1 |\sum_{i=1}^{n} Z_i^{11} e_{s,j} e_{p-s,k}^T Z_i^{22}|_\infty \\
&+ \tilde{r}_n \|u_2\|_1 |\sum_{i=1}^{n} x_i^{1T} e_{s,j} e_{p-s,k}^T Z_i^{22}|_\infty \\
\le& \Big(2(2\bar{c}s + \tilde{r}_n\tilde{C})(\bar{c}s + \tilde{r}_n\tilde{C}) + 2(\bar{c}s + \tilde{r}_n\tilde{C}) + (2\bar{c}s + \tilde{r}_n\tilde{C}) + 1 \\
&+ \tilde{r}_n(2\bar{c}s + \tilde{r}_n\tilde{C})\tilde{C} + \tilde{r}_n\tilde{C}\Big)\sqrt{n}c_0,
\end{aligned}
$$

where the last inequality follows from Condition 3. Since $\tilde{r}_n \to 0$ as $n \to \infty$, we conclude that for large enough n,

$$
|\sum_{i=1}^{n} e_{s,j}^T \zeta_{1i}\zeta_{2i}^T e_{p-s,k}| \le (12\bar{c}^2 s^2 + 12\bar{c}s + 1)\sqrt{n}c_0 \le (12\bar{c}^2 + 1)c_0\sqrt{n}s^2,
$$

and therefore

$$
\begin{aligned}
|\sum_{i=1}^{n} u_1^T \zeta_{1i} u_2^T \zeta_{2i}| &\le \|u_1\|_1 \|u_2\|_1 |\sum_{i=1}^{n} \zeta_{1i}\zeta_{2i}^T|_\infty \\
&\le (12\bar{c}^2 + 1)c_1 C\sqrt{n}s^2 \|u_2\|_1. \qquad \text{(S1.11)}
\end{aligned}
$$

Note that

$$
|\sum_{i=1}^{n} u_2^T \zeta_{2i}\varepsilon_i| \le \|u_2\|_1 \|\sum_{i=1}^{n} \zeta_{2i}\varepsilon_i\|_\infty
$$

and $\|u_2\|_q^q \geq \tilde{C}^{q-1}\|u_2\|_1$. Under the event E_1, it follows from (S1.6), (S1.10) and (S1.11) that

$$\begin{aligned}
& L_n(\beta_1^* + \tilde{r}_n u_1, \tilde{r}_n u_2) - L_n(\beta_1^* + \tilde{r}_n u_1, 0) \\
\geq & - 2\tilde{r}_n \| \sum_{i=1}^n \zeta_{2i}\varepsilon_i \|_\infty \|u_2\|_1 - (12\bar{c}^2 + 1)c_1 C \tilde{r}_n^2 \sqrt{n} s^2 \|u_2\|_1 + \tilde{C}^{q-1}\lambda_n \tilde{r}_n^q \|u_2\|_1 \\
\geq & \lambda_n \tilde{r}_n^q \|u_2\|_1 \Big(- 2(2(1+\bar{c}^2))^{1/2} \lambda_n^{-1} \tilde{r}_n^{1-q} b_n \sqrt{n} s \\
& \qquad - (12\bar{c}^2 + 1)c_1 C \lambda_n^{-1} \tilde{r}_n^{2-q} \sqrt{n} s^2 + \tilde{C}^{q-1} \Big) \\
> & 0,
\end{aligned}$$

when $\|u_2\|_1 > 0$. That is, (S1.9) holds.

On the other hand, under the event $\{\|\hat{\beta}_1 - \beta_1^*\|_1 \leq r_n\}$, we conclude from $\|\hat{\beta}_1 - \beta_1^*\|_1 \leq \sqrt{s}\|\hat{\beta}_1 - \beta_1^*\|$, that $\|\hat{\beta} - \beta^*\|_1 = \|\hat{\beta}_1 - \beta_1^*\|_1 \leq \tilde{r}_n\sqrt{s}$, which yields

$$\inf_{\|u\|_1 \leq \tilde{C}} L_n(\beta_1^* + \tilde{r}_n u_1, \tilde{r}_n u_2) \leq L_n(\hat{\beta}) = L_n(\hat{\beta}_1, 0) \leq L_n(\beta_1^* + \tilde{r}_n u_1, 0).$$

Combining this and (S1.9), we have $L_n(\hat{\beta}) = \inf_{\|u\|_1 \leq \tilde{C}} L_n(\beta_1^* + \tilde{r}_n u_1, \tilde{r}_n u_2)$ under the event $E_1 \cap \{\{\|\hat{\beta}_1 - \beta_1^*\| \leq r_n\}$. That is,

$$E_1 \cap \big\{\{\|\hat{\beta}_1 - \beta_1^*\| \leq r_n\} \subseteq \big\{\hat{\beta} \in \arg \inf_{\|u\|_1 \leq \tilde{C}} L_n(\beta_1^* + \tilde{r}_n u_1, \tilde{r}_n u_2)\big\}. \quad \text{(S1.12)}$$

To complete the proof of (S1.8), we need to verify that

$$\mathbb{P}\big(\| \sum_{i=1}^n \zeta_{2i}\varepsilon_i \|_\infty > 4((1+\bar{c}^2))^{1/2} b_n \sqrt{n} s\big) \leq \exp\big(- \frac{b_n^2}{4\sigma^2} + o(b_n^2)\big). \text{ (S1.13)}$$

Denote the jth element of ζ_{2i} by ζ_{2ij}. Since $\|\beta_1^*+\tilde{r}_n u_1\|_1 \leq \|\beta_1^*\|_1+\tilde{r}_n\|u_1\|_1 \leq \overline{c}s + \tilde{r}_n\tilde{C}$, we use Cauchy-Schwarz inequality to obtain

$$\begin{aligned}
|\zeta_{2ij}| &\leq 2|e_{p-s,j}^T Z_i^{21}(\beta_1^* + \tilde{r}_n u_1)| + |x_{ij}| + \tilde{r}_n|e_{p-s,j}^T Z_i^{22} u_2| \\
&\leq 2\|e_{p-s,j}^T Z_i^{21}\|_\infty \|\beta_1^* + \tilde{r}_n u_1\|_1 + \kappa_{1n} + \tilde{r}_n \|e_{p-s,j}^T Z_i^{22}\|_\infty \|u_2\|_1 \\
&\leq \big(2\overline{c} + 3\tilde{r}_n\tilde{C}\big)\kappa_{2n}s + \kappa_{1n}. \qquad \text{(S1.14)}
\end{aligned}$$

By similar calculation, we have

$$\begin{aligned}
&\sum_{i=1}^n \zeta_{2ij}^2 \\
\leq& 4\sum_{i=1}^n \Big(4\big(e_{p-s,j}^T Z_i^{21}\beta_1^*\big)^2 + 4\tilde{r}_n^2\big(e_{p-s,j}^T Z_i^{21} u_1\big)^2 + x_{ij}^2 + \tilde{r}_n^2\big(e_{p-s,j}^T Z_i^{22} u_2\big)^2\Big) \\
\leq& 4\sum_{i=1}^n \Big(4\|e_{p,j}^T Z_i\|_\infty^2 \big(\|\beta_1^*\|_1^2 + \tilde{r}_n^2\|u_1\|_1^2 + \tilde{r}_n^2\|u_2\|_1^2\big) + x_{ij}^2\Big) \\
\leq& 4\sum_{i=1}^n \Big(4\|Z_i\|_\infty^2 \big(\|\beta_1^*\|_1^2 + \tilde{r}_n^2\|u\|_1^2\big) + x_{ij}^2\Big) \\
\leq& 4(4\overline{c}^2 + 4\tilde{r}_n^2\tilde{C}^2 + 1)ns^2.
\end{aligned}$$

Write $B_n = 4(4\overline{c}^2 + 4\tilde{r}_n^2\tilde{C}^2 + 1)ns^2$. Since the limits (3.5) and (3.9) imply respectively that

$$\frac{\lambda_n\sqrt{s}}{\sqrt{n}}\big(\frac{\kappa_{2n}s + \kappa_{1n}}{\sqrt{n}s}\big) \to 0 \text{ and } a_n\big(\frac{\kappa_{2n}s + \kappa_{1n}}{\sqrt{n}s}\big) \to 0,$$

it follows from (S1.14) and $\tilde{r}_n \to 0$ that

$$\frac{b_n \max_{1\leq i\leq n}|\zeta_{2ij}|}{\sqrt{B_n}} \to 0, \quad \text{as } n \to \infty.$$

We use Lemma S1.1 to obtain that,

$$\mathbb{P}\big(|\sum_{i=1}^{n}\zeta_{2ij}\varepsilon_i| > b_n\sqrt{B_n}\big) \le \exp\Big(-\frac{b_n^2}{2\sigma^2}+o(b_n^2)\Big),$$

which combining the relation $\tilde{r}_n \to 0$ leads to

$$\mathbb{P}\big(\|\sum_{i=1}^{n}\zeta_{2i}\varepsilon_i\|_\infty > 4((1+\overline{c}^2))^{1/2}b_n\sqrt{n}s\big) \le \exp\big(-\frac{b_n^2}{2\sigma^2}+o(b_n^2)\big). \quad \text{(S1.15)}$$

Note that the first relation of Condition 4 implies that

$$b_n > \frac{2\underline{c}^{q-1}\lambda_n\sqrt{s}}{\sqrt{n}} \ge 2\sigma\sqrt{\log p}.$$

Therefore we conclude that

$$\begin{aligned}\mathbb{P}\big(\|\sum_{i=1}^{n}\zeta_{2i}\varepsilon_i\|_\infty > 4((1+\overline{c}^2))^{1/2}b_n\sqrt{n}s\big) &\le \sum_{j=s+1}^{p}\mathbb{P}\big(|\sum_{i=1}^{n}\zeta_{2ij}\varepsilon_i| > 4((1+\overline{c}^2))^{1/2}b_n\sqrt{n}s\big)\\ &\le \exp\Big(-\frac{b_n^2}{4\sigma^2}+o(b_n^2)\Big).\end{aligned}$$

which yields (S1.13). Further by Lemma S1.3, (S1.12) and (S1.13), we have

$$\begin{aligned}\mathbb{P}\big(\hat{\beta}\in\arg\inf_{\|u\|_1\le\tilde{C}}L_n(\beta_1^*+\tilde{r}_nu_1,\tilde{r}_nu_2)\big) &\ge \mathbb{P}\big(E_1\cap\{\|\hat{\beta}_1-\beta_1^*\|\le r_n\}\big)\\ &\ge 1-\exp\big(-C_0a_n^2+o(a_n^2)\big).\end{aligned}$$

□

Proof of Theorem 2 It suffices to show that the sequence $a_n = \sqrt{s}\log n$ satisfies (3.8)-(3.10). First, it is clear that $a_n/\sqrt{s\log n} \to \infty$. Further, it follows from (S1.7) that

$$a_n\sqrt{n} = \frac{\sqrt{s}\log n}{\sqrt{n}} \le \frac{\big(\max(s,\log n)\big)^{3/2}}{\sqrt{n}} \to 0.$$

Moreover, the inequality in Condition 4 and (3.4) imply that

$$\frac{a_n\kappa_{1n}\sqrt{s}}{\sqrt{n}} = \frac{\lambda_n\kappa_{1n}s\log n}{n}\cdot\frac{\sqrt{n}}{\lambda_n} \le \frac{\lambda_n\kappa_{1n}s\log n}{n}\cdot\frac{1}{\sigma\underline{c}^{1-q}\sqrt{\log p}} \to 0$$

and

$$\frac{a_n\kappa_{2n}s^{3/2}}{\sqrt{n}} = \frac{\lambda_n\kappa_{2n}s^2\log n}{n}\cdot\frac{\sqrt{n}}{\lambda_n} \le \frac{\lambda_n\kappa_{2n}s^2\log n}{n}\cdot\frac{1}{\sigma\underline{c}^{1-q}\sqrt{\log p}} \to 0.$$

Therefore by the first convergence of Condition 4, we obtain

$$\frac{a_n^{2-q}n^{\frac{q}{2}}s^{\frac{4-q}{2}}}{\lambda_n} = \frac{\sqrt{n^q}s^{3-q}(\log n)^{2-q}}{\lambda_n} \to 0,$$

which completes the proof. □

S2 Proofs of Theorem 3 and 4

We here also use the notation in Section S1 and provide two lemmas below, i.e., Lemmas S2.1 and S2.2, corresponding to Lemmas S1.2 and S1.3 there.

Lemma S2.1. *For the model (4.1), assume that Conditions* $1'$-$2'$ *and* $4'$ *hold. Let* $\{a_n\}$ *be a sequence of positive numbers satisfying (3.8) and (3.9). Then, for any* $\tau > 0$,

$$\mathbb{P}\Big(\frac{1}{a_n\sqrt{n}}\sup_{u\in S}\|\sum_{i=1}^{n}(Z_i^{11}u)\varepsilon_i\| > \tau\Big) \le \exp\big(-\frac{a_n^2\tau^2}{2c_2\sigma^2}+o(a_n^2)\big),$$

where $S' = S_1 \cap S$.

Proof. First note that

$$\{u \in \mathbb{R}^s : \|u\|_0 \geq s - [\frac{s}{2}]\} \subseteq \bigcup_{k=[\frac{s}{2}]}^{s} \{u \in \mathbb{R}^s : \|u\|_0 = k\},$$

and Lemma 14.27 of Bühlmann and Van De Geer (2011) implies that in the subspace $\mathbb{R}^k$,

$$\begin{aligned}&\{v \in \mathbb{R}^k : \min_{1\leq l\leq k} |e_{k,l}^T v| \geq \frac{\underline{c}}{2}, \|v\| \leq C_1\sqrt{s}\}\\ \subseteq &\bigcup_{j=1}^{(1+2n)^k} \{v \in \mathbb{R}^k : \|v - v_j\| \leq \frac{1}{n}, \min_{1\leq l\leq k} |e_{k,l}^T v_j| \geq \frac{\underline{c}}{2}, \|v_j\| \leq C_1\sqrt{s}\}.\end{aligned}$$

Since

$$\sum_{k=[\frac{s}{2}]}^{s} C_s^k (1+2n)^k \leq (1+2n)^s \sum_{k=[\frac{s}{2}]}^{s} C_s^k \leq (2+4n)^s$$

and

$$\begin{aligned}&\Big\{u \in \mathbb{R}^s : |\{j : |e_{s,j}^T u| \geq \frac{\underline{c}}{2}\}| \geq s - [\frac{s}{2}]\Big\}\\ =&\Big\{u \in \mathbb{R}^s : \|u\|_0 \geq s - [\frac{s}{2}], |e_{s,j}^T u| \geq \underline{c}/2, j \in \mathrm{supp}(u)\Big\},\end{aligned}$$

we have

$$S' \subseteq \bigcup_{j=1}^{(2+4n)^s} \{u \in \mathbb{R}^s : \|u - u_j\| \leq \frac{1}{n}, u_j \in S'\}.$$

Then, we can use the similar method to the proof of Lemma S1.2 to get the desired result. □

Lemma S2.2. *Under the assumptions of Lemma S2.1, $\tilde{L}_n(\beta_1)$ has two minimizers $\hat{\beta}_1$ and $-\hat{\beta}_1$ such that*

$$\begin{aligned}\mathbb{P}\big(\|(-\hat{\beta}_1) - (-\beta_1^*)\| \leq r_n\big) =&\mathbb{P}\big(\|\hat{\beta}_1 - \beta_1^*\| \leq r_n\big)\\ \geq&1 - \exp\big(-\frac{(1+c_1^2/4)a_n^2}{2c_2\sigma^2} + o(a_n^2)\big).\end{aligned}$$

Proof. Define

$$S_1(\beta_1^*) = \Big\{u \in \mathbb{R}^s : \Big|\{j : |e_{s,j}^T u + e_{s,j}^T \beta_1^*| \geq \underline{c}/2\}\Big| \geq s - [\frac{s}{2}]\Big\}$$

and

$$S_1(-\beta_1^*) = \Big\{u \in \mathbb{R}^s : \Big|\{j : |e_{s,j}^T u - e_{s,j}^T \beta_1^*| \geq \underline{c}/2\}\Big| \geq s - [\frac{s}{2}]\Big\}.$$

We first show that

$$S_1(\beta_1^*) \cup S_1(-\beta_1^*) = \mathbb{R}^s. \tag{S2.1}$$

First, it is obvious that $S_1(\beta_1^*) \cup S_1(-\beta_1^*) \subseteq \mathbb{R}^s$. To show the opposite inclusion, we need the following two facts that for any $u \in \mathbb{R}^s$,

$$\big|\{j : |e_{s,j}^T u + e_{s,j}^T \beta_1^*| \geq \underline{c}/2\}\big| \leq [\frac{s}{2}] \Leftrightarrow \big|\{j : |e_{s,j}^T u + e_{s,j}^T \beta_1^*| < \underline{c}/2\}\big| \geq s - [\frac{s}{2}]$$

and

$$\{j : |e_{s,j}^T u + e_{s,j}^T \beta_1^*| < \underline{c}/2\} \subseteq \{j : |e_{s,j}^T u - e_{s,j}^T \beta_1^*| \geq \underline{c}/2\}.$$

It is clear that the first holds. We only need to check the second. Note that for each j with $|e_{s,j}^T u + e_{s,j}^T \beta_1^*| < \underline{c}/2$, it is easy to verify that

$$-2e_{s,j}^T \beta_1^* - \underline{c}/2 < e_{s,j}^T u - e_{s,j}^T \beta_1^* < -2e_{s,j}^T \beta_1^* + \underline{c}/2.$$

Combining this and the assumption $0 < \underline{c} \leq \min\{|e_{p,j}^T \beta^*|, j \in \Gamma^*\}$, we have

$$e_{s,j}^T u - e_{s,j}^T \beta_1^* \begin{cases} < -3\underline{c}/2, & \text{if } e_{s,j}^T \beta_1^* > \underline{c}; \\ > 3\underline{c}/2, & \text{if } e_{s,j}^T \beta_1^* > -\underline{c}. \end{cases}$$

which yields $|e_{s,j}^T u - e_{s,j}^T \beta_1^*| \geq \underline{c}/2$. Therefore the second fact holds. It follows that, for any $\beta_1 \notin S_1(\beta_1^*)$, i.e., $\left|\{j : |e_{s,j}^T u + e_{s,j}^T \beta_1^*| \geq \underline{c}/2\}\right| \leq [\frac{s}{2}]$, the above two facts imply $\beta_1 \in S_1(-\beta_1^*)$, which further implies that (S2.1) holds.

Note that for any $\beta_1 \in S_1(\beta_1^*)$, $-\beta_1 \in S_1(-\beta_1^*)$, and for any $\beta_1 \in S_1(-\beta_1^*)$, $-\beta_1 \in S_1(\beta_1^*)$. That is, the sets $S_1(\beta_1^*)$ and $-S_1(-\beta_1^*)$ are symmetric. Since $\tilde{L}_n(\beta_1)$ is an even function, it follows from (S2.1) that

$$\min_{\beta_1 \in \mathbb{R}^s} \tilde{L}_n(\beta_1) = \min_{\beta_1 \in S_1(\beta_1^*)} \tilde{L}_n(\beta_1) = \min_{\beta_1 \in S_1(-\beta_1^*)} \tilde{L}_n(\beta_1).$$

By the similar method to the proof of Lemma S1.3, we can show that there exists a minimizer $\hat{\beta}_1 = \arg\min_{\beta_1 \in S_1(\beta_1^*)} \tilde{L}_n(\beta_1)$, such that (S1.2) holds. Therefore the desired result follows and the proof is completed. □

Proof of Theorem 3 From Lemmas S2.1 and S2.2, we can use the similar method for model (2.1) to prove that under the event $E_1 \cap \{\|\hat{\beta}_1 - \beta_1^*\| \leq r_n$, $(\hat{\beta}_1^T, 0^T)^T$ is a local minimizer in the ball $\{\beta^* + \tilde{r}_n u : \|u\|_1 \leq \tilde{C}\}$, and $(-\hat{\beta}_1^T, 0^T)^T$ is a local minimizer in the ball $\{-\beta^* + \tilde{r}_n u : \|u\|_1 \leq \tilde{C}\}$. As mentioned before, we identify vectors $\beta, \beta' \in \mathbb{R}^p$ which satisfy $\beta' = \pm\beta$. Then, there exists strict local minimizer $\hat{\beta}$ such that both the results (3.6) and (3.7) remain true. □

Proof of Theorem 4 Proof of Theorem 4 is analogous to that of Theorem 3.

S3 Analysis of the optimization algorithm

Lemma S3.1. *[Chen, Xiu and Peng (2014)] Let* $t \in \mathbb{R}, \lambda > 0, q \in (0,1)$ *be given and* $t^* = (2-q)\big(q(1-q)^{q-1}\lambda\big)^{1/(2-q)}$. *For any* $t_0 > t^*$, *there exists a unique implicit function* $u = \overline{h}_{\lambda,q}(t)$ *on* (t^*, ∞) *such that* $u_0 = \overline{h}_{\lambda,q}(t_0), u = \overline{h}_{\lambda,q}(t) > 0$, $\overline{h}_{\lambda,q}(t) - t + \lambda q \overline{h}_{\lambda,q}(t)^{q-1} = 0$ *and* $u = \overline{h}_{\lambda,q}(t)$ *is continuously differentiable with* $\overline{h}_{\lambda,q}{}'(t) = \frac{1}{1+\lambda q(q-1)\overline{h}_{\lambda,q}(t)^{q-2}} > 0$. *For any* $t_0 < -t^*$, *there exists a unique function* $u = \underline{h}_{\lambda,q}(t)$ *on* $(-\infty, -t^*)$ *such that* $u_0 = \underline{h}_{\lambda,q}(t_0), u = \underline{h}_{\lambda,q}(t) < 0$, $\overline{h}_{\lambda,q}(t) - t - \lambda q|\overline{h}_{\lambda,q}(t)|^{q-1} = 0$ *and* $u = \overline{h}_{\lambda,q}(t)$ *is continuously differentiable with* $\underline{h}_{\lambda,q}{}'(t) = \frac{1}{1+\lambda q(q-1)|\underline{h}_{\lambda,q}(t)|^{q-2}} > 0$.

Furthermore, the global solution $\hat{u}$ *of the problem (5.2) satisfies*

$$\hat{u} = h_{\lambda,q}(t) := \begin{cases} \underline{h}_{\lambda,q}(t), & \text{if } \ t < -t^*; \\ -(2\lambda(1-q))^{\frac{1}{2-q}} \quad \text{or} \ \ 0, & \text{if } \ t = -t^*; \\ 0, & \text{if } \ -t^* < t < t^*; \\ (2\lambda(1-q))^{\frac{1}{2-q}} \quad \text{or} \ \ 0, & \text{if } \ t = t^*; \\ \overline{h}_{\lambda,q}(t), & \text{if } \ t > t^*. \end{cases}$$

Especially, $h_{\lambda,1/2}(t) = \frac{2}{3}t\big(1+\cos\big(\frac{2\pi}{3} - \frac{2}{3}\phi_\lambda(t)\big)\big)$ *with* $\phi_\lambda(t) = \arccos\big(\frac{\lambda}{4}(\frac{|t|}{3})^{-3/2}\big)$.

Lemma S3.2. *For* $q \in (0,1), \lambda > 0$, *let* $\hat{u} = \arg\min_{u \in \mathbb{R}^p} \frac{1}{2}\|u-b\|_2^2 + \lambda\|u\|_q^q$, $\forall\, b \in \mathbb{R}^p$. *Then* $\hat{u} = \mathcal{H}_{\lambda,q}(b)$.

The result is an immediate consequence of Lemma S3.1 and therefore the proof is omitted.

Proof of Theorem 5 For any $\tau > 0$, define the following auxiliary problem

$$\min_{\beta\in\mathbb{R}^p} F_\tau(\beta,u) := \ell(u) + \langle\nabla\ell(u), \beta - u\rangle + \frac{1}{2\tau}\|\beta - u\|_2^2 + \lambda\|\beta\|_q^q, \quad \forall u \in \mathbb{R}^p. \tag{S3.1}$$

It is easy to check that the problem (S3.1) is equivalent to the following minimization problem

$$\min_{\beta\in\mathbb{R}^p} \frac{1}{2}\|\beta - (u - \tau\nabla\ell(u))\|_2^2 + \lambda\tau\|\beta\|_q^q.$$

For any $r > 0$, let $B_r = \{\beta \in \mathbb{R}^p : \|\beta\|_2 \le r\}$ and $G_r = \sup_{\beta\in B_r}\|\nabla^2\ell(\beta)\|_2$. For any $\tau \in (0, G_r^{-1}]$ and $\beta, u \in B_r$, we have

$$\begin{aligned} L(\beta) &= \ell(u) + \langle\nabla\ell(u), \beta - u\rangle + \frac{1}{2}(\beta - u)^T\nabla^2\ell(\xi)(\beta - u) + \lambda\|\beta\|_q^q \\ &= F_\tau(\beta,u) + \frac{1}{2}(\beta - u)^T\nabla^2\ell(\xi)(\beta - u) - \frac{1}{2\tau}\|\beta - u\|_2^2 \\ &\le F_\tau(\beta,u) + \frac{1}{2}\|\nabla^2\ell(\xi)\|_2\|\beta - u\|_2^2 - \frac{1}{2\tau}\|\beta - u\|_2^2 \\ &\le F_\tau(\beta,u) + \frac{L}{2}\|\beta - u\|_2^2 - \frac{1}{2\tau}\|\beta - u\|_2^2 \\ &\le F_\tau(\beta,u), \end{aligned} \tag{S3.2}$$

where $\xi = u + \alpha(\beta - u)$ for some $\alpha \in (0,1)$ and the second inequality follows from $\|\xi\|_2 \le r$.

Further, let $\bar\beta \in \arg\min_{\beta\in\mathbb{R}^p} F_\tau(\beta, \hat\beta)$. Since $L(\beta) \ge 0$ and $\lim_{\|\beta\|_2\to\infty} L(\beta) = \infty$, there exists a positive constant r_1 such that $\|\hat\beta\|_2 \le r_1$. Note that

$$\nabla \ell(\beta) = 2\sum_{i=1}^{m}(\beta^T Z_i \beta + x_i^T \beta - y_i)(2Z_i\beta + x_i) \tag{S3.3}$$

which implies that $\nabla \ell(\beta)$ is continuous differentiable. Then, take

$$r_2 = r_1 + \sup_{\beta \in B_{r_1}} \|\nabla \ell(\beta)\|_2.$$

Hence it follows from Lemma S3.2 that $\|\bar{\beta}\|_2 \leq r_2$ for any $\tau \in (0, 1]$. By the definitions of $\hat{\beta}$ and $\bar{\beta}$, we obtain from the inequality (S3.2) that for any $\tau \in \big(0, \min\{G_{r_2}^{-1}, 1\}\big)$,

$$F_\tau(\bar{\beta}, \hat{\beta}) \leq F_\tau(\hat{\beta}, \hat{\beta}) = L(\hat{\beta}) \leq L(\bar{\beta}) \leq F_\tau(\bar{\beta}, \hat{\beta}),$$

which leads to $F_\tau(\hat{\beta}, \hat{\beta}) = F_\tau(\bar{\beta}, \hat{\beta})$. Therefore $\hat{\beta}$ is also a minimizer of the problem (S3.1) with $u = \hat{\beta}$. The results follows then from Lemma S3.2. □

Lemma S3.3. *Let* $g_k = \|\nabla \ell(\beta^k)\|_2$, $G_k = \sup_{\beta \in B_k} \|\nabla^2 \ell(\beta)\|_2$ *where* $B_k = \{\beta \in \mathbb{R}^p : \|\beta\|_2 \leq \|\beta^k\|_2 + g_k\}$. *For any* $\delta > 0, \gamma, \alpha \in (0, 1)$, *define*

$$j_k = \begin{cases} 0, & \text{if } \gamma(G_k + \delta) \leq 1; \\ -\lceil \log_\alpha \gamma(G_k + \delta) \rceil + 1, & \text{otherwise.} \end{cases}$$

Then (5.4) holds.

Proof. From the definition of τ_k and j_k, it is easy to check that

$$G_k - \frac{1}{\tau_k} \leq -\delta. \tag{S3.4}$$

Indeed, take $\tau_k = \gamma$ which yields to

$$G_k - \frac{1}{\tau_k} = \frac{\gamma G_k - 1}{\gamma} \leq -\delta,$$

when $\gamma(G_k + \delta) \leq 1$. If $\gamma(G_k + \delta) > 1$,

$$\tau_k = \gamma\alpha^{j_k} \leq \gamma\alpha^{-\log_\alpha \gamma(G_k+\delta)} = \frac{1}{G_k + \delta}$$

which also leads to (S3.4).

Note that

$$\beta^{k+1} \in \arg\min_{\beta\in\mathbb{R}^p} G_{\tau_k}(\beta, \beta^k) \tag{S3.5}$$

and

$$\|\beta^{k+1}\|_2 \leq \|\beta^k - \tau_k \nabla\ell(\beta^k)\|_2 \leq \|\beta^k\|_2 + g_k,$$

which yields $\beta^{k+1} \in B_k$. Similar to (S3.2), we obtain from (S3.4) that

$$\begin{aligned} L(\beta^{k+1}) &\leq F_{\tau_k}(\beta^{k+1}, \beta^k) + \frac{1}{2}\|\beta^{k+1} - \beta^k\|_2^2 \big(\|\nabla^2\ell(\xi_k)\|_2 - \frac{1}{\tau_k}\big) \\ &\leq F_{\tau_k}(\beta^{k+1}, \beta^k) + \frac{1}{2}\|\beta^{k+1} - \beta^k\|_2^2 (G_k - \frac{1}{\tau_k}) \\ &\leq F_{\tau_k}(\beta^{k+1}, \beta^k) - \frac{\delta}{2}\|\beta^{k+1} - \beta^k\|_2^2, \end{aligned}$$

where $\xi_k = \beta^k + \varrho(\beta^{k+1} - \beta^k)$ for some $\varrho \in (0, 1)$ and then $\xi_k \in B_k$ leads to the second inequality. Combining this and (S3.5), we have

$$\begin{aligned} L(\beta^k) - L(\beta^{k+1}) &= F_{\tau_k}(\beta^k, \beta^k) - L(\beta^{k+1}) \geq F_{\tau_k}(\beta^{k+1}, \beta^k) - L(\beta^{k+1}) \\ &\geq \frac{\delta}{2}\|\beta^{k+1} - \beta^k\|_2^2, \end{aligned}$$

which completes the proof. □

Lemma S3.4. *Let $\{\beta^k\}$ and $\{\tau_k\}$ be generated by FPIA. Then,*

(i) $\{\beta^k\}$ is bounded; and

(ii) there is a nonnegative integer $\bar{j}$ such that $\tau_k \in [\gamma\alpha^{\bar{j}}, \gamma]$.

Proof. Lemma S3.3 implies that $\{L(\beta^k)\}$ is strictly decreasing. From this, $\ell(\cdot) \geq 0$ and the definition of $L(\cdot)$, it is easy to check that $\{\beta^k\}$ is bounded. Since $\ell(\cdot)$ is a twice continuous differentiable function, it then follows from the bound of $\{\beta^k\}$ that there exist two positive constants $\bar{g}$ and $\bar{G}$ such that $\sup_{k\geq 0}\{g_k\} \leq \bar{g}$ and $\sup_{k\geq 0}\{G_k\} \leq \bar{G}$. Define $\bar{j} = \max(0, [-\log_\alpha \gamma(\bar{G} + \delta)] + 1)$. Then, $0 \leq j_k \leq \bar{j}$ which combining the definition of τ_k imply that $\tau_k \in [\gamma\alpha^{\bar{j}}, \gamma]$. □

Now we consider the convergence of the sequence $\{\beta^k\}$. To this end we slightly modify $h_{\lambda,q}(\cdot)$ as follows

$$h_{\lambda,q}(t) := \begin{cases} \underline{h}_{\lambda,q}(t), & \text{if } \ t < -t^*; \\ 0, & \text{if } \ |t| \leq t^*; \\ \overline{h}_{\lambda,q}(t), & \text{if } \ t > t^*. \end{cases} \tag{S3.6}$$

Then we have the following result.

Theorem S3.1. *Let $\{\beta^k\}$ be the sequence generated by FPIA. Then,*

(i) $\{L(\beta^k)\}$ converges to $L(\widetilde{\beta})$, where $\widetilde{\beta}$ is any accumulation point of

$\{\beta^k\}$;

(ii) $\lim_{k\to\infty} \frac{\|\beta^{k+1}-\beta^k\|_2}{\tau_k} = 0$;

(iii) any accumulation point of $\{\beta^k\}$ *is a stationary point of the minimization problem (5.1) when* $\gamma \leq \left(\frac{q}{16(1-q)}\bar{g}^{-1}\right)^{\frac{2-q}{1-q}}\left(\lambda(1-q)\right)^{\frac{1}{1-q}}$ *and* $\bar{g} = \sup_{k\geq 0}\|\nabla\ell(\beta^k)\|_2$.

Proof. (i) Since $\{\beta^k\}$ is bounded, it has at least one accumulation point. Since $\{L(\beta^k)\}$ is monotonically decreasing and $L(\cdot) \geq 0$, $\{L(\beta^k)\}$ converges to a constant $\widetilde{L}(\geq 0)$. Since $L(\beta)$ is continuous, we have $\{L(\beta^k)\} \to \widetilde{L} = L(\widetilde{\beta})$, where $\widetilde{\beta}$ is an accumulation point of $\{\beta^k\}$ as $k \to \infty$.

(ii) From the definition of β^{k+1} and (5.4), we have

$$\sum_{k=0}^{n}\|\beta^{k+1}-\beta^k\|_2^2 \leq \frac{2}{\delta}\sum_{k=0}^{n}[L(\beta^k)-L(\beta^{k+1})] = \frac{2}{\delta}[L(\beta^0)-L(\beta^{n+1})] \leq \frac{2}{\delta}L(\beta^0).$$

Hence, $\sum_{k=0}^{\infty}\|\beta^{k+1}-\beta^k\|_2^2 < \infty$ and $\|\beta^{k+1}-\beta^k\|_2 \to 0$ as $k\to\infty$. Then the second result of Lemma S3.4 leads to the result (ii).

(iii) Since $\{\beta^k\}$ and $\{\tau_k\}$ have convergent sequences, without loss of generality, assume that

$$\beta^k \to \widetilde{\beta} \text{ and } \tau_k \to \widetilde{\tau}, \quad \text{as} \quad k\to\infty. \tag{S3.7}$$

It suffices to prove that $\hat{\beta}$ and $\widetilde{\tau}$ satisfy (5.3). Note that

$$\begin{aligned} & \|\widetilde{\beta} - \mathcal{H}_{\lambda\widetilde{\tau},q}\big(\widetilde{\beta} - \widetilde{\tau}\nabla\ell(\widetilde{\beta})\big)\|_2 \\ \leq\ & \|\widetilde{\beta} - \beta^{k+1}\|_2 + \|\mathcal{H}_{\lambda\tau_k,q}\big(\beta^k - \tau_k\nabla\ell(\beta^k)\big) - \mathcal{H}_{\lambda\widetilde{\tau},q}\big(\widetilde{\beta} - \widetilde{\tau}\nabla\ell(\widetilde{\beta})\big)\|_2 \\ =\ & I_1 + I_2. \end{aligned} \tag{S3.8}$$

The result (ii) and (S3.7) imply that $I_1 \to 0$ as $k \to \infty$.

To complete the proof, we need show $I_2 \to 0$ for $q \in (0,1)$. For $i = 1, \cdots, p$, denote

$$v_i^k = e_{p,i}^T\big(\beta^k - \tau_k\nabla\ell(\beta^k)\big), \widetilde{v}_i = e_{p,i}^T\big(\widetilde{\beta} - \widetilde{\tau}\nabla\ell(\widetilde{\beta})\big), t_i^* = \frac{2-q}{2(1-q)}[2\lambda\widetilde{\tau}(1-q)]^{1/(2-q)}$$

and $\widetilde{\beta}_i = \big(2\lambda\widetilde{\tau}(1-q)\big)^{1/(2-q)}$. Then it suffices to prove that

$$h_{\lambda\tau_k,q}(v_i^k) \to h_{\lambda\tau_k,q}(\widetilde{v}_i) \tag{S3.9}$$

when $v_i^k \to \widetilde{v}_i$ as $k \to \infty$. We only give the proof of (S3.9) as $\widetilde{v}_i > 0$ because the case of $\widetilde{v}_i < 0$ can be similarly proved.

For $\widetilde{v}_i < t_i^*$, the limit (S3.7) and the definition of $h_{\lambda\tau,q}$ imply that $h_{\lambda\tau_k,q}(v_i^k) = 0 = h_{\lambda\widetilde{\tau},q}(\widetilde{v}_i)$. For $\widetilde{v}_i > t_i^*$, one can conclude from (S3.7) and the continuity of $h_{\lambda\tau,q}$ on (t_i^*, ∞) that $h_{\lambda\tau_k,q}(v_i^k) \to h_{\lambda\widetilde{\tau},q}(\widetilde{v}_i)$. For $\widetilde{v}_i = t_i^*$, we show that any subsequence of $\{v_i^k\}$ converging to $\widetilde{v}_i$, without loss of generality, say $\{v_i^k\}$, must satisfy

$$v_i^k \leq t_i^*, \quad \text{for large enough } k. \tag{S3.10}$$

We prove the above inequality by contradiction. Denote $\Delta = \frac{q}{16(1-q)}\big(\lambda(1-q)\big)^{\frac{1}{2-q}}$ and $\delta_i = \frac{t_i^* - \widetilde{\beta}_i}{4}$. Note that $t_i^* > \widetilde{\beta}_i$ implies that $\delta_i = \frac{p}{16(1-q)}\Delta(2\widetilde{\tau})^{\frac{1}{2-q}} > 0$. The second limit of (S3.7) implies $\widetilde{\tau} \geq \frac{1}{2}\tau_k$ and hence $\delta_i \geq 2\Delta(\tau_k)^{\frac{1}{2-q}}$ for large enough k. Since $\tau_k^{\frac{1-q}{2-q}}\Delta^{-1} \leq \gamma^{\frac{1-q}{2-q}}\Delta^{-1} \leq \bar{\ell}^{-1}$, for large enough k, we have

$$\tau_k \|\nabla \ell(\beta^k)\|_2 \leq \Delta \tau_k \bar{\ell} \Delta^{-1} \leq \frac{\delta_i}{2}\tau_k^{-\frac{1}{2-q}}\tau_k \bar{\ell}\Delta^{-1} \leq \frac{\delta_i}{2}$$

and therefore

$$e_{p,i}^T \beta^k = v_i^k + \tau_k[\nabla \ell(\beta^k)]_i \geq v_i^k - \tau_k \|[\nabla \ell\big(e_{p,i}^T \beta^k\big)\big)\|_2 \geq v_i^k - \frac{1}{2}\delta_i.$$

Combining this, the result (ii) and $v_i^k \to t_i^*$, we have

$$e_{p,i}^T \beta^{k+1} \geq e_{p,i}^T \beta^k - \frac{1}{2}\delta_i \geq v_i^k - \delta_i \geq t_i^* - 2\delta_i = \widetilde{\beta}_i + 2\delta_i, \quad \text{for large enough } k. \tag{S3.11}$$

Note that $h_{\lambda\tau,q}$ is continuous on (t_i^*, ∞) and $\lim_{n\to\infty} h_{\lambda\tau_k,q}(v_i^k) = \widetilde{\beta}_i$. For large enough k, we have $e_{p,i}^T\beta^{k+1} = h_{\lambda\tau_k,q}(v_i^k) \in [\widetilde{\beta}_i - \delta_i, \widetilde{\beta}_i + \delta_i]$, which is in contradiction with $(S3.11)$. So (S3.10) holds. By the definition of $h_{\lambda,q}(\cdot)$, we have $h_{\lambda\tau_k,q}(v_i^k) = 0 = h_{\lambda\widetilde{\tau},q}(\widetilde{v}_i)$. □

References

Bühlmann, P. and Van De Geer. S., (2011). *Statistics for high-dimensional data: methods, theory and applications*. Springer, Heidelberg.

Chen, Y., Xiu, N. and Peng, D. (2014). Global solutions of non-Lipschitz $S_2 - S_p$ minimization over positive semidefinite cone. *Optimization Letters* **8**, 2053-2064.

Fan, Jun, Yan, Ailing and Xiu, Naihua. (2014). Asymptotic Properties for M-estimators in Linear Models with Dependent Random Errors. *J. Stat. Plan. Infer.* **148**, 49-66.

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.