Web Supplementary Materials for "A Modified Adaptive Lasso for Identifying Interactions in the Cox Model with the Heredity Constraint"

Lu Wang[a,*], Jincheng Shen[a], Peter F. Thall[b]

[a] *Department of Biostatistics, University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48109, USA*
[b] *Department of Biostatistics, M.D. Anderson Cancer Center, Houston, Texas 77030, USA*

Appendix 1: Regularity conditions.

Following the notation in Andersen & Gill (1982), we consider a finite time interval $[0, \tau]$ with $\tau < \infty$. To facilitate the notation, let $N_i(t) = I\{T_i \le t, T_i \le C_i\}$ and $Y_i(t) = I\{T_i \ge t, C_i \ge t\}$. Define $H_i(\theta, t) = Y_i(t)\exp\{g(X_i, \theta)\}$, $S^{(0)}(\theta, t) = n^{-1}\sum_{i=1}^n H_i(\theta, t)$, $S^{(1)}(\theta, t) = n^{-1}\sum_{i=1}^n \nabla_\theta H_i(\theta, t)$, $S^{(2)}(\theta, t) = n^{-1}\sum_{i=1}^n \nabla_\theta^2 H_i(\theta, t)$, and $S^{(3)}(\theta, t) = n^{-1}\sum_{i=1}^n \nabla_\theta^3 H_i(\theta, t)$, where $\nabla_\theta(\cdot)$ denotes the first derivative with respect of θ, $\nabla_\theta^2(\cdot)$ and $\nabla_\theta^3(\cdot)$ denote the second and third order derivatives respectively.

We assume the following regularity conditions hold for Theorem 1:

(1) $\int_0^\tau \lambda_0(t)dt < \infty$

(2) There exists a neighbourhood Θ of θ_0 and $s^{(0)}(\theta, t)$, $s^{(1)}(\theta, t)$, $s^{(2)}(\theta, t)$ and $s^{(3)}(\theta, t)$ defined on $\Theta \times [0, \tau]$ such that for j = 0,1,2 and 3.

$$\sup_{t\in[0,\tau],\theta\in\Theta} \left\| S^{(j)}(\theta, t) - s^{(j)}(\theta, t) \right\| \longrightarrow_{\mathscr{P}} 0$$

where $\|\cdot\|$ is the L_1-norm.

(3) Let Θ, $s^{(0)}(\cdot,\cdot)$, $s^{(1)}(\cdot,\cdot)$, $s^{(2)}(\cdot,\cdot)$ and $s^{(3)}(\cdot,\cdot)$ be as in Condition (2) and define $e = s^{(1)}/s^{(0)}$ and $v = s^{(2)}/s^{(0)} - e \otimes e$. For all $\theta \in \Theta$, $t \in [0, \tau]$, $s^{(0)}(\cdot, t)$, $s^{(1)}(\cdot, t)$ and $s^{(2)}(\cdot, t)$ are continuous functions of $\theta \in \Theta$, uniformly in $t \in [0, \tau]$, $s^{(0)}(\theta, t)$, $s^{(1)}(\theta, t)$, $s^{(2)}(\theta, t)$, and $s^{(3)}(\theta, t)$ are bounded on $\Theta \times [0, \tau]$; and $s^{(0)}(\theta, t)$ is bounded away from zero on $\Theta \times [0, \tau]$. Let

*Corresponding author.
Email address: `luwang@umich.edu` (Lu Wang)

$u^{(2)}(\theta) = \nabla^2_\theta g(X, \theta)$,

$$I(\theta_0) = \int_0^\tau \left\{\nu(\theta_0, t) - u^{(2)}(\theta_0)\right\} s^{(0)}(\theta_0, t)\lambda_0(t)dt,$$

and we require the submatrix $\mathcal{I}_a(\beta_{a0}, \gamma_{a0})$ from $I(\theta_0)$ that corresponds to the non-zero $(\beta_{a0}, \gamma_{a0})$ is positive definite.

For Theorem 2, we denote $H_i(\theta_n, t) = Y_i(t)\exp\{g(\theta_n, X_{n,i})\}$. Define $\phi_{i,j}(\theta_n, t) = \{\partial H_i(\theta_n, t)/\partial\theta_{n,j}\} / \{n^{-1}\sum_{i=1}^n H_i(\theta_n, t)\}$, $W_j^{(1)}(\theta_n, t) = n^{-1}\sum_{i=1}^n \phi_{i,j}(\theta_n, t)$, $W_{jk}^{(2)}(\theta_n, t) = n^{-1}\sum_{i=1}^n \partial\phi_{i,j}(\theta_n, t)/\partial\theta_{n,k}$, and $W_{jkl}^{(3)}(\theta_n, t) = n^{-1}\sum_{i=1}^n \partial\phi_{i,j}(\theta_n, t)/(\partial\theta_{n,k}\partial\theta_{n,l})$, for any $j, k, l = 1, \cdots, p_n(p_n+1)/2$. We further denote $W_{jk}^{(1,2)}(\theta_n, t) = n^{-1}\sum_{i=1}^n \{\phi_{i,j}(\theta_n, t)\phi_{i,k}(\theta_n, t)\}^2$, $W_{jk}^{(2,2)}(\theta_n, t) = n^{-1}\sum_{i=1}^n \{\partial\phi_{i,j}(\theta_n, t)/\partial\theta_{n,k}\}^2$, and $W_{jkl}^{(3,2)}(\theta_n, t) = n^{-1}\sum_{i=1}^n \{\partial^2\phi_{i,j}(\theta_n, t)/\partial\theta_{n,k}\partial\theta_{n,l}\}^2$. We assume the following regularity conditions in Theorem 2.

(4) $\int_0^\tau \lambda_0(t)dt < \infty$

(5) There exists a neighbourhood Θ_n of $\theta_{n,0}$ and $w_j^{(1)}(\theta_n, t)$, $w_{jk}^{(2)}(\theta_n, t)$, $w_{jkl}^{(3)}(\theta_n, t)$, $w_{jk}^{(1,2)}(\theta_n, t)$, $w_{jk}^{(2,2)}(\theta_n, t)$, $w_{jkl}^{(3,2)}(\theta_n, t)$, defined on $\Theta_n \times [0, \tau]$ such that for $m = 1, 2, 3$,

$$\sup_{t\in[0,\tau],\theta_n\in\Theta_n} \left\|W_\cdot^{(m)}(\theta_n, t) - w_\cdot^{(m)}(\theta_n, t)\right\| \longrightarrow_{\mathscr{P}} 0,$$

and moreover,

$$\sup_{t\in[0,\tau],\theta_n\in\Theta_n} \left\|W_\cdot^{(1,2)}(\theta_n, t) - w_\cdot^{(1,2)}(\theta_n, t)\right\| \longrightarrow_{\mathscr{P}} 0$$

$$\sup_{t\in[0,\tau],\theta_n\in\Theta_n} \left\|W_\cdot^{(2,2)}(\theta_n, t) - w_\cdot^{(2,2)}(\theta_n, t)\right\| \longrightarrow_{\mathscr{P}} 0$$

$$\sup_{t\in[0,\tau],\theta_n\in\Theta_n} \left\|W_\cdot^{(3,2)}(\theta_n, t) - w_\cdot^{(3,2)}(\theta_n, t)\right\| \longrightarrow_{\mathscr{P}} 0.$$

(6) For all $\theta_n \in \Theta_n$, $t \in [0, \tau]$, $w_\cdot^{(1)}(\cdot, t)$, $w_\cdot^{(2)}(\cdot, t)$, $w_\cdot^{(3)}(\cdot, t)$, $w_\cdot^{(1,2)}(\cdot, t)$, $w_\cdot^{(2,2)}(\cdot, t)$, $w_\cdot^{(3,2)}(\cdot, t)$ are continuous functions of $\theta_n \in \Theta_n$, uniformly in $t \in [0, \tau]$, and $w_\cdot^{(1)}(\theta_n, t)$, $w_\cdot^{(2)}(\theta_n, t)$, $w_\cdot^{(3)}(\theta_n, t)$ and $w_\cdot^{(1,2)}(\theta_n, t)$, $w_\cdot^{(2,2)}(\theta_n, t)$, $w_\cdot^{(3,2)}(\theta_n, t)$ are bounded on $\Theta_n \times [0, \tau]$. Let $u^{(2)}(\theta_n)$ denote $\nabla^2_{\theta_n} g(X, \theta_n)$ and $w^{(2)}(\theta_n, t)$ denote the matrix with $\{w^{(2)}(\theta_n, t)\}_{jk} = w_{jk}^{(2)}(\theta_n, t)$ for all

$j, k = 1, \cdots, p_n(p_n+1)/2$. Then define

$$I(\theta_{n,0}) = \int_0^\tau \left\{ w^{(2)}(\theta_{n,0}, t) - u^{(2)}(\theta_{n,0}) \right\} s^{(0)}(\theta_{n,0}, t) \lambda_0(t) dt,$$

and let $\mathcal{I}_{an}(\beta_{a0}, \gamma_{a0})$ denote the submatrix of $I(\theta_{n,0})$ with respect to the non-zero $(\beta_{a0}, \gamma_{a0})$. It satisfies $0 < C_1 < \lambda_{\min}\{\mathcal{I}_{an}(\beta_{a0}, \gamma_{a0})\} \leq \lambda_{\max}\{\mathcal{I}_{an}(\beta_{a0}, \gamma_{a0})\} < C_2 < \infty$ for all n, where $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ represent the smallest and largest eigenvalues of a matrix respectively.

Appendix 2: Proof of Theorem 1.

The log partial likelihood $l_n(\theta)$ can be written as

$$l_n(\theta) = \sum_{i=1}^{n} \int_0^\tau g(X_i, \theta) dN_i(s) - \int_0^\tau log \left[\sum_{i=1}^{n} Y_i(s) exp\left\{ g(X_i, \theta) \right\} \right] d\widetilde{N}(s)$$

where $\widetilde{N}(\cdot) = \sum_{i=1}^{n} N_i(\cdot)$. By Theorem 4.1 and Lemma 3.1 of **?**, it follows that, for each θ in a neighbourhood of θ_0:

$$\frac{1}{n}\left\{ l_n(\theta) - l_n(\theta_0) \right\} = \int_0^\tau \left[(\theta - \theta_0)^T s^{(1)}(\theta_0, t) - log \left\{ \frac{s^{(0)}(\theta, t)}{s^{(0)}(\theta_0, t)} \right\} s^{(0)}(\theta_0, t) \right] \lambda_0(t) dt + O_p\left(\frac{\|\theta - \theta_0\|}{\sqrt{n}} \right).$$

Let $\eta_n = n^{-1/2} + \xi_n$, consider the C-ball $B_n(C) = \left\{ \theta : \theta = \theta_0 + \eta_n \delta, \|\delta\| \leq C \right\}, C > 0$. For any $\theta \in B_n(C)$, by the second-order Taylor expansion of the log partial likelihood, and by the weak law of large numbers, we have

$$\frac{1}{n}\left\{ l_n(\theta_0 + \eta_n \delta) - l_n(\theta_0) \right\} = \frac{1}{n} \nabla_\theta^T l_n(\theta_0) \eta_n \delta - \frac{1}{2} \eta_n^2 \delta^T \left\{ I(\theta_0) + o_p(1) \right\} \delta$$

where $\|\delta\| \leq C$. We further write $\delta = (u_1, ..., u_p, v_{12}, ..., v_{p-1,p})^T = (u^T, v^T)^T$. Then let

$$D_n(\delta) \equiv \frac{1}{n}\left\{ Q_n(\theta_0 + \eta_n \delta) - Q_n(\theta_0) \right\}$$

$$= -\frac{1}{n}\left\{ l_n(\theta_0 + \eta_n \delta) - l_n(\theta_0) \right\} + \sum_{j=1}^{p} \lambda_{j,n}^{\beta} \left(|\beta_{0j} + \eta_n u_j| - |\beta_{0j}| \right) + \sum_{j<j'} \lambda_{j,j',n}^{\gamma} \left(|\gamma_{0j,j'} + \eta_n v_{j,j'}| - |\gamma_{0j,j'}| \right)$$

$$\geq -\frac{1}{n}\left\{ l_n(\theta_0 + \eta_n \delta) - l_n(\theta_0) \right\} - \eta_n^2 \left(\sum_{\{j:\beta_{0j} \in \beta_{a0}\}} |u_j| + \sum_{\{(j,j'):\gamma_{0j,j'} \in \gamma_{a0}\}} |v_{j,j'}| \right)$$

$$\geq -\frac{1}{n}\left\{l_n(\theta_0+\eta_n\delta)-l_n(\theta_0)\right\} - \eta_n^2\left(|\beta_{a0}|+|\gamma_{a0}|\right)C \equiv A_1+A_2+A_3$$

where $|\cdot|$ measures the number of elements of the vector inside,

$$A_1 = -\frac{1}{n}\nabla_\theta l_n(\theta_0)\left(\eta_n\delta\right) = O_p(n^{-1/2})\left(\eta_n\delta\right)$$

$$A_2 = \frac{1}{2}\left(\eta_n\delta\right)^T\left\{I(\theta_0)+o_p(1)\right\}\left(\eta_n\delta\right) = \frac{1}{2}\left(\eta_n\delta_a\right)^T\left\{\mathcal{I}_a(\beta_{a0},\gamma_{a0})+o_p(1)\right\}\left(\eta_n\delta_a\right)$$

$$A_3 = -\eta_n^2\left(|\beta_{a0}|+|\gamma_{a0}|\right)C,$$

and δ_a is the sub-vector of δ correspond to non-zero (β_{a0},γ_{a0}). Notice that A_2 dominates A_1 and A_3 and is positive since $\mathcal{I}_a(\beta_{a0},\gamma_{a0})$ is positive definite. Therefore, for any given $\epsilon > 0$, there exists a large enough constant d such that

$$P\left\{\inf_{\theta\in B_n(d)} Q_n(\theta) > Q_n(\theta_0)\right\} \geq 1-\epsilon.$$

This implies that with probability at least $1-\epsilon$, there exists a local minimizer in the ball $B_n(C)$ such that $\left\|\hat{\theta}_n-\theta_0\right\| = O_p(\eta_n) = O_p(n^{-1/2})$.

Now for the sparsity, we first show $P(\hat{\beta}_{bn}=0)\to 1$. It is sufficient to show for any $\{j:\beta_{0j}\in\beta_{b0}\}$,

$$\frac{\partial Q_n(\hat{\theta}_n)}{\partial\beta_j} > 0 \text{ for } 0<\hat{\beta}_j<\epsilon_n \tag{1}$$

and

$$\frac{\partial Q_n(\hat{\theta}_n)}{\partial\beta_j} < 0 \text{ for } -\epsilon_n<\hat{\beta}_j<0 \tag{2}$$

with probability tending to 1, where $\epsilon_n = Cn^{-1/2}$ and $C>0$ is any constant. To show (1), notice

$$\frac{\partial Q_n(\hat{\theta}_n)}{\partial\beta_j} = -\frac{\partial l_n(\hat{\theta}_n)}{\partial\beta_j} + \lambda_{j,n}^\beta\text{sign}(\beta_j) = -\frac{\partial l_n(\theta_0)}{\partial\beta_j} - \sum_{k=1}^{p(p+1)/2}\frac{\partial^2 l_n(\theta_0)}{\partial\beta_j\partial\theta_k}\left(\hat{\theta}_k-\theta_{0k}\right)$$

$$-\sum_{k=1}^{\frac{p(p+1)}{2}}\sum_{l=1}^{\frac{p(p+1)}{2}}\frac{\partial^3 l_n(\tilde{\theta})}{\partial\beta_j\partial\theta_k\partial\theta_l}\left(\hat{\theta}_k-\theta_{0k}\right)\left(\hat{\theta}_l-\theta_{0l}\right) + \lambda_{j,n}^\beta\text{sign}(\beta_j),$$

where $\tilde{\theta}$ lies between $\hat{\theta}_n$ and θ_0. By the regularity conditions and $\left\|\hat{\theta}_n - \theta_0\right\| = O_p(n^{-1/2})$,

$$\frac{\partial Q_n(\hat{\theta}_n)}{\partial \beta_j} = \sqrt{n}\left\{O_p(1) + \sqrt{n}\lambda_{j,n}^{\beta}\text{sign}(\hat{\beta}_j)\right\}.$$

As $\sqrt{n}\lambda_j^{\beta} \to \infty$ for $j \in \{j : \beta_{0j} \in \beta_{b0}\}$, the sign of $\partial Q_n(\hat{\theta}_n)/\partial\beta_j$ is dominated by sign$(\hat{\beta}_j)$. Therefore,

$$P\left[\frac{\partial Q_n(\hat{\theta}_n)}{\partial \beta_j} > 0 \; for \; 0 < \hat{\beta}_j < \epsilon_n\right] \to 1 \; as \; n \to \infty$$

Similarly, we can show (2), and $P(\hat{\beta}_{bn} = 0) \to 1$ follows. We can similarly prove that $P(\hat{\gamma}_{bn} = 0) \to 1$.

For $(j, j') \in \{(j, j') : \gamma_{0j,j'} \in \gamma_{c0}\}$, without loss of generality, assume that $\beta_{0j} = 0$. Notice that $\hat{\beta}_j = 0$ implies $\hat{\gamma}_{j,j'} = 0$. Since we already have $P(\hat{\beta}_j = 0) \to 1$, we can conclude $P(\hat{\gamma}_{j,j'} = 0) \to 1$ as well, i.e. $P(\hat{\gamma}_{cn} = 0) \to 1$ as $n \to \infty$. Thus, we finish the proof for Part (i) of Theorem 1.

Next we show the asymptotic normality. Let $\widetilde{Q}_n(\theta_a)$ denote the objective function Q_n only on the nonzero component of θ, i.e. $\theta_a = (\beta_a^T, \gamma_a^T)^T$. We define $\theta_b = (\beta_b^T, \gamma_b^T, \gamma_c^T)^T$, and from the above derivation, we have $P\left(\hat{\theta}_b = 0\right) \to 1$. Thus,

$$P\left[\arg\min_{\theta_a} \widetilde{Q}_n(\theta_a) = \left(\theta_a - \text{component of } \arg\min_{\theta} Q_n(\theta)\right)\right] \to 1.$$

It means that $\hat{\theta}_a$ should satisfy

$$\frac{\partial \widetilde{Q}_n(\theta_a)}{\partial \theta_j}|_{\theta_a = \hat{\theta}_a} = 0, \quad \forall j \in \{j : \theta_j \in \theta_a\}$$

with probability tending to 1.

Let $\widetilde{l}_n(\theta_a)$ and $\widetilde{P}_\lambda(\theta_a)$ denote the log-likelihood function of θ_a and the penalty function of θ_a respectively so that we have

$$\widetilde{Q}_n(\theta_a) = -\widetilde{l}_n(\theta_a) + \widetilde{P}_\lambda(\theta_a)$$

Then

$$\nabla_{\theta_a}\widetilde{Q}_n(\hat{\theta}_a) = -\nabla_{\theta_a}\widetilde{l}_n(\hat{\theta}_a) + \nabla_{\theta_a}\widetilde{P}_\lambda(\hat{\theta}_a) = 0 \tag{3}$$

with probability tending to 1.

By Taylor expansion, it is easy to show that

$$B_1 = \nabla_{\theta_a} \widetilde{l}_n(\hat{\theta}_a) = \sqrt{n}\left[\frac{1}{\sqrt{n}}\nabla_{\theta_a}\widetilde{l}_n(\theta_{a0}) - \mathcal{I}_a(\beta_{a0},\gamma_{a0})\sqrt{n}(\hat{\theta}_a - \theta_{a0}) + o_p(1)\right]$$

and

$$B_2 = \nabla_{\theta_a}\widetilde{P}_\lambda(\hat{\theta}_a) = \left\{\begin{bmatrix} \lambda^{\beta}_{j,n}\text{sign}(\beta_j) \\ \lambda^{\gamma}_{j,j',n}\text{sign}(\gamma_{j,j'}) \end{bmatrix}_{\beta_j\in\beta_a,\gamma_{j,j'}\in\gamma_a} + o_p(1)(\hat{\theta}_a - \theta_{a0})\right\}.$$

Since we have $\left\|\hat{\theta}_a - \theta_{a0}\right\| = O_p(n^{-1/2})$, together with (3), we have

$$\sqrt{n}(\hat{\theta}_a - \theta_{a0}) = \mathcal{I}_a(\beta_{a0},\gamma_{a0})^{-1}\frac{1}{\sqrt{n}}\nabla_{\theta_a}\widetilde{l}_n(\theta_{a0}) + o_p(1).$$

Part (ii) of Theorem 1 then follows by applying the central limit theorem.

Appendix 3: Proof of Theorem 2.

Similarly, under the regularity conditions in Appendix 1, we argue that there exists a local minimizer $\hat{\theta}_n$ of $Q_n(\theta)$ such that $\left\|\hat{\theta}_n - \theta_{n,0}\right\| = O_p(\sqrt{q_n}(n^{1/2}+\xi_n))$. Let $\eta_n = \sqrt{q_n}(n^{-1/2}+\xi_n)$, consider the C-ball $\{\theta_n = \theta_{n,0} + \eta_n\delta, \|\delta\| \le C\}, C > 0$. We define $D_n(\delta) \equiv \{Q_n(\theta_{n,0}+\eta_n\delta) - Q_n(\theta_{n,0})\}/n$, then for any $\delta = (u_1, ..., u_p, v_{12}, ..., v_{p-1,p})^T = (u^T, v^T)^T$ that satisfies $\|\delta\| \le C$, similarly as in Appendix 2, we have

$$D_n(\delta) \equiv \frac{1}{n}\{Q_n(\theta_{n,0}+\eta_n\delta) - Q_n(\theta_{n,0})\} \ge -\frac{1}{n}\{l_n(\theta_{n,0}+\eta_n\delta) - l_n(\theta_{n,0})\} - \eta_n\left(\sqrt{q_n}\xi_n\right)C$$

$$= -\frac{1}{n}\nabla^T_{\theta_n} l_n(\theta_{n,0})\left(\eta_n\delta\right) + \frac{1}{2}\left(\eta_n\delta\right)^T\{I(\theta_{n,0}) + o_p(1)\}\left(\eta_n\delta\right) - \eta_n^2 C \equiv \widetilde{A}_1 + \widetilde{A}_2 + \widetilde{A}_3$$

where

$$\widetilde{A}_1 = -\frac{1}{n}\nabla^T_{\theta_n} l_n(\theta_{n,0})\left(\eta_n\delta\right) \text{ and } \left|\widetilde{A}_1\right| \le n^{-1/2}\eta_n O_p(\sqrt{q_n})C = O_p(\eta_n^2)C,$$

$$\widetilde{A}_2 = \frac{1}{2}\left(\eta_n\delta\right)^T\{I(\theta_{n,0}) + o_p(1)\}\left(\eta_n\delta\right) = \frac{1}{2}\left(\eta_n\delta_a\right)^T\{\mathcal{I}_{an}(\beta_{a0},\gamma_{a0}) + o_p(1)\}\left(\eta_n\delta_a\right), \quad \widetilde{A}_3 = \eta_n^2 C,$$

and δ_a is the sub-vector of δ correspond to non-zero (β_{a0},γ_{a0}). Similarly, $\widetilde{A}_2$ dominates $\widetilde{A}_1$ and $\widetilde{A}_3$, and is positive since $\mathcal{I}_{an}(\beta_{a0},\gamma_{a0})$ is positive definite. Therefore, for any given $\epsilon > 0$, there exists a large enough constant C such that

$$P\left\{\inf_{\|\delta\|\leq C} Q_n(\theta_{n,0}+\eta_n\delta) > Q_n(\theta_{n,0})\right\} \geq 1-\epsilon.$$

This implies that with probability at least $1-\epsilon$, there exists a local minimizer in the ball $B_n(C)$ such that $\left\|\hat{\theta}_n - \theta_{n,0}\right\| = O_p(\eta_n)$.

We now show $P(\hat{\beta}_{bn}=0) \to 1$. It is sufficient to show that for any $j \in \{j : \beta_{n,j} \in \beta_{bn}\}$,

$$\frac{\partial Q_n(\hat{\theta}_n)}{\partial \beta_{n,j}} > 0 \; for \; 0 < \hat{\beta}_{n,j} < \epsilon_n \tag{4}$$

$$\frac{\partial Q_n(\hat{\theta}_n)}{\partial \beta_{n,j}} < 0 \; for \; -\epsilon_n < \hat{\beta}_{n,j} < 0 \tag{5}$$

with probability tending to 1, where $\epsilon_n = Cn^{-1/2}$ and $C > 0$ is any constant. Notice

$$\frac{\partial Q_n(\hat{\theta}_n)}{\partial \beta_{n,j}} = -\frac{\partial l_n(\hat{\theta}_n)}{\partial \beta_{n,j}} + \lambda_{j,n}^{\beta_n}\mathrm{sign}(\beta_{n,j})$$

$$= -\frac{\partial l_n(\theta_{n,0})}{\partial \beta_{n,j}} - \sum_{k=1}^{q_n} \frac{\partial^2 l_n(\theta_{n,0})}{\partial \beta_{n,j}\partial \theta_{n,k}}\left(\hat{\theta}_{n,k} - \theta_{n,0k}\right)$$

$$-\sum_{k=1}^{q_n}\sum_{l=1}^{q_n} \frac{\partial^3 l_n(\tilde{\theta})}{\partial \beta_{n,j}\partial \theta_{n,k}\partial \theta_{n,l}}\left(\hat{\theta}_{n,k} - \theta_{n,0k}\right)\left(\hat{\theta}_{n,l} - \theta_{n,0l}\right) + \lambda_{j,n}^{\beta_n}\mathrm{sign}(\beta_{n,j}),$$

where $\tilde{\theta}_n$ lies between $\hat{\theta}_n$ and $\theta_{n,0}$. By the regularity conditions, and notice $\left\|\hat{\theta}_n - \theta_{n,0}\right\| = O_p(\sqrt{q_n/n})$,

$$\frac{\partial Q_n(\hat{\theta}_n)}{\partial \beta_{n,j}} = \sqrt{nq_n}\left\{O_p(1) + \sqrt{\frac{n}{q_n}}\lambda_{j,n}^{\beta_n} sgn(\hat{\beta}_{n,j})\right\}.$$

As $\sqrt{n/q_n}\lambda_{j,n}^{\beta_n} \to \infty$ for $j \in \{j : \beta_{n,j} \in \beta_{bn}\}$, the sign of $\partial Q_n(\hat{\theta}_n)/\partial \beta_{n,j}$ is the same as $\mathrm{sign}(\hat{\beta}_{n,j})$. Therefore,

$$P\left[\frac{\partial Q_n(\hat{\theta}_n)}{\partial \beta_{n,j}} > 0 \text{ for } 0 < \hat{\beta}_{n,j} < \epsilon_n\right] \to 1 \; as \; n \to \infty$$

and (4) holds with probability tending to 1. Parallel to this, one can show (5) holds with probability tending to 1.

Similar argument can be used to prove $P(\hat{\gamma}_{n,j,j'} = 0) \to 1$ as $n \to \infty$, for $\hat{\gamma}_{n,j,j'} \in \hat{\gamma}_{bn}$, thus $P(\hat{\gamma}_{bn} = 0) \to 1$.

For $\hat{\gamma}_{n,j,j'} \in \hat{\gamma}_{cn}$, without loss of generality, assume that $\beta_{n,0j} = 0$. Notice that $\hat{\beta}_{n,j} = 0$ implies $\hat{\gamma}_{n,j,j'} = 0$, because if $\hat{\gamma}_{n,j,j'} \neq 0$, then the value of the loss function does not change but the value of the penalty function increases. Therefore, $P(\hat{\gamma}_{n,j,j'} = 0) \to 1$ follows since we have already shown $P(\hat{\beta}_{n,j} = 0) \to 1$.

Thus, Part (i) of Theorem 2 is proved.

Now, we prove the asymptotic normality. Denote $\theta_{an} = (\beta_{an}^T, \gamma_{an}^T)^T$, then

$$\sqrt{n}\Omega_n I_{an}^{1/2}(\theta_{an,0})\left(\hat{\theta}_{an} - \theta_{an,0}\right) = \sqrt{n}\Omega_n I_{an}^{-1/2}(\theta_{an,0}) I_{an}(\theta_{an,0})\left(\hat{\theta}_{an} - \theta_{an,0}\right)$$

$$= \sqrt{n}\Omega_n \mathcal{I}_{an}^{-1/2}(\beta_{a0}, \gamma_{a0})\left\{\frac{1}{n}\nabla l_n(\theta_{an,0}) + o_p(n^{-1/2})\right\}$$

$$= \frac{1}{\sqrt{n}}\Omega_n \mathcal{I}_{an}^{-1/2}(\beta_{a0}, \gamma_{a0})\sum_{i=1}^{n}\left[\nabla l_n(\theta_{an,0})\right] + o_p(1) \equiv \sum_{i=1}^{n} Y_{ni} + o_p(1),$$

where $Y_{ni} = n^{-1/2}\Omega_n \mathcal{I}_{an}^{-1/2}(\beta_{a0}, \gamma_{a0})\sum_{i=1}^{n}\left[\nabla l_n(\theta_{an,0})\right]$.

We now show that with probability tending to 1, $\sum_{i=1}^{n} Y_{ni} + o_p(1) \to_d N(0, \Sigma)$:

(i) We first show $\mathcal{I}_{an}(\beta_{a0}, \gamma_{a0})\left(\hat{\theta}_{an} - \theta_{an,0}\right) = n^{-1}\nabla l_n(\theta_{an,0}) + o_p(n^{-1/2})$. With probability tending to 1,

$$0 = \nabla_{\theta_{an}} Q_n(\hat{\theta}_{an}) = -\frac{1}{n}\nabla_{\theta_{an}} l_n(\hat{\theta}_{an}) + \nabla_{\theta_{an}}\left\{\sum_{\{j:\beta_{n,0j}\in\beta_{a0}\}} \lambda_{j,n}^{\beta_n}\hat{\beta}_{n,0j} + \sum_{\{(j,j'):\gamma_{n,0j,j'}\in\gamma_{a0}\}} \lambda_{j,j',n}^{\gamma_n}\hat{\gamma}_{n,0j,j'}\right\}.$$

Taking Taylor Expansion at $\theta_{an} = \theta_{a0}$, we have

$$0 = -\nabla_{\theta_{an}} l_n(\theta_{a0}) - \left[\nabla_{\theta_{an}}^2 l_n(\theta_{a0})\right]\left(\hat{\theta}_{an} - \theta_{a0}\right) - \frac{1}{2}\left(\hat{\theta}_{an} - \theta_{a0}\right)^T\left[\nabla_{\theta_{an}}^2\left(\nabla_{\theta_{an}} l_n(\theta_{a0})\right)\right]\left(\hat{\theta}_{an} - \theta_{a0}\right)$$

$$+n\nabla_{\theta_{an}}\left\{\sum_{\{j:\beta_{n,0j}\in\beta_{a0}\}} \lambda_{j,n}^{\beta_n}\beta_{n,0j} + \sum_{\{(j,j'):\gamma_{n,0j,j'}\in\gamma_{a0}\}} \lambda_{j,j',n}^{\gamma_n}\gamma_{n,0j,j'}\right\}.$$

Thus,

$$\mathcal{I}_{an}^{-1/2}(\beta_{a0}, \gamma_{a0})\left(\hat{\theta}_{an} - \theta_{a0}\right) = -\frac{1}{n}\nabla_{\theta_{an}}^2 l_n(\theta_{a0})\left(\hat{\theta}_{an} - \theta_{a0}\right) + \left\{\mathcal{I}_{an}^{-1/2}(\beta_{a0}, \gamma_{a0}) + \frac{1}{n}\nabla_{\theta_{an}}^2 l_n(\theta_{a0})\right\}\left(\hat{\theta}_{an} - \theta_{a0}\right)$$

$$= \frac{1}{n}\nabla_{\theta_{an}} l_n(\theta_{a0}) + \frac{1}{2n}\left(\hat{\theta}_{an} - \theta_{a0}\right)^T\left[\nabla_{\theta_{an}}^2\left(\nabla_{\theta_{an}} l_n(\theta_{a0})\right)\right]\left(\hat{\theta}_{an} - \theta_{a0}\right)$$

$$-\nabla_{\theta_{an}}\left\{\sum_{\{j:\beta_{n,0j}\in\beta_{a0}\}}\lambda_{j,n}^{\beta_n}\beta_{n,0j}+\sum_{\{(j,j'):\gamma_{n,0j,j'}\in\gamma_{a0}\}}\lambda_{j,j',n}^{\gamma_n}\gamma_{n,0j,j'}\right\}$$

$$+\left\{\mathcal{I}_{an}^{-1/2}(\beta_{a0},\gamma_{a0})+\frac{1}{n}\nabla_{\theta_{an}}^2 l_n(\theta_{a0})\right\}\left(\hat{\theta}_{an}-\theta_{a0}\right).$$

Therefore, it is sufficient to show that

$$\frac{1}{2n}\left(\hat{\theta}_{an}-\theta_{a0}\right)^T\left[\nabla_{\theta_{an}}^2\left(\nabla_{\theta_{an}} l_n(\theta_{a0})\right)\right]\left(\hat{\theta}_{an}-\theta_{a0}\right)$$

$$-\nabla_{\theta_{an}}\left\{\sum_{\{j:\beta_{n,0j}\in\beta_{a0}\}}\lambda_{j,n}^{\beta_n}\beta_{n,0j}+\sum_{\{(j,j'):\gamma_{n,0j,j'}\in\gamma_{a0}\}}\lambda_{j,j',n}^{\gamma_n}\gamma_{n,0j,j'}\right\}$$

$$+\left\{\mathcal{I}_{an}(\beta_{a0},\gamma_{a0})+\frac{1}{n}\nabla_{\theta_{an}}^2 l_n(\theta_{a0})\right\}\left(\hat{\theta}_{an}-\theta_{a0}\right)=o_p(n^{-1/2}).$$

Denote the three terms in the above equation as D_1, D_2, and D_3. First, by Cauchy-Schwarz inequality,

$$\|D_1\|^2\leq\frac{1}{4n^2}\left\|\nabla_{\theta_{an}}^2\left(\nabla_{\theta_{an}} l_n(\theta_{an,0})\right)\right\|^2\left\|\hat{\theta}_{an}-\theta_{a0}\right\|^4$$

$$=\frac{1}{4n^2}\sum_{\{(j,k,l):\theta_{n,j},\theta_{n,k},\theta_{n,l}\in\theta_{an}\}}n^2 O_p(1)O_p(\frac{q_n^2}{n})=O_p(q_n^5/n^2)=o_p(1/n)$$

Secondly, because $\xi_n=o(1/\sqrt{nq_n})$,

$$\|D_2\|^2=\left\|\left(\lambda_{1,n}^{\beta_n}\mathrm{sign}(\beta_{n,01}),\ldots,\lambda_{p_{n-1},p_n,n}^{\gamma_n}\mathrm{sign}(\gamma_{n,0(p_{n-1},p_n)})\right)^T\right\|^2$$

$$\leq|\theta_{an}|\,\xi_n^2=|\theta_{an}|\,o(1/nq_n)=o_p(1/n)$$

Third, it can be shown that

$$\|D_3\|^2\leq\left\|\mathcal{I}_{an}(\beta_{a0},\gamma_{a0})+\frac{1}{n}\nabla_{\theta_{an}}^2 l_n(\theta_{a0})\right\|^2\left\|\hat{\theta}_{an}-\theta_{a0}\right\|^2$$

$$=o_p(1/q_n^2)O_p(q_n/n)=o_p(1/nq_n)=o_p(1/n)$$

Therefore, $D_1+D_2+D_3=o_p(n^{-1/2})$.

Next, we show $\sum_{i=1}^n Y_{ni}+o_p(1)\longrightarrow_d N(0,\Sigma)$. It is sufficient to show that Y_{ni}, $i=1,\ldots,n$ satisfies the conditions for Lindeberg-Feller central limit theorem. For any given $\epsilon>0$, by

Cauchy-Schwarz inequality,

$$\sum_{i=1}^{n} E\left[\|Y_{ni}\|^2 I\left\{\|Y_{ni}\| > \epsilon\right\}\right] = nE\left[\|Y_{ni}\|^2 I\left\{\|Y_{ni}\| > \epsilon\right\}\right] \leq nD_4^{1/2} D_5^{1/2}$$

where $D_4 = \left[E\left\|Y_{ni}\right\|^4\right]$ and $D_5 = E\left\{I\left(\|Y_{ni}\| > \epsilon\right)\right\}$. Note

$$D_4 = \frac{1}{n^2} E\left\|\Omega_n \mathcal{I}_{an}^{-1/2}(\beta_{a0}, \gamma_{a0}) \nabla_{\theta_{an}} l_n(\theta_{a0})\right\|^4$$

$$\leq \frac{1}{n^2}\left\|\Omega_n^T \Omega_n\right\|^2 \left\|I_{an}(\theta_{a0})\right\|^{-2} E\left\|\nabla_{\theta_{an}}^T l_n(\theta_{a0}) \nabla_{\theta_{an}} l_n(\theta_{a0})\right\|^2$$

$$= \frac{1}{n^2}\lambda_{max}^2(\Omega_n^T\Omega_n)\lambda_{max}^2\left\{I_{an}^{-1}(\theta_{a0})\right\} O(|\theta_{an}|^2) = O(q_n^2/n^2).$$

By Markov inequality,

$$D_5 = E\left\{I\left(\|Y_{ni}\| > \epsilon\right)\right\} = P\left(\|Y_{n1}\| > \epsilon\right) \leq \frac{E\left\|Y_{n1}\right\|^2}{\epsilon^2} = O(q_n/n).$$

Therefore,

$$\sum_{i=1}^{n} E\left[\|Y_{ni}\|^2 1\{\|Y_{ni}\| > \epsilon\}\right] \leq nO(q_n/n)O(\sqrt{q_n/n}) = o(1),$$

and part (ii) of Theorem 2 follows.