Robust Variable Selection in High-dimensional Nonparametric Additive model

Suneel Babu Chatla and Abhijit Mandal

The University of Texas at El Paso, Texas

Supplementary Material

The supplementary material contains the proofs of Lemma 1 and Theorems 1 and 2. In the following lemma, we derive the maximal inequality for the product terms involving spline function and error. The proof requires results involving the Generalized Bernstein-Orlicz (GBO) norm. For completeness, we define it here. The function $\Psi_{\alpha,L}$ is defined in [5], with its inverse expressed as

$$\Psi_{\alpha,L}^{-1} := \sqrt{\log(1+t)} + L(\log(1+t))^{1/\alpha} \text{ for all } t \geq 0,$$

for fixed $\alpha > 0$ and $L \geq 0$. The GBO norm of a random variable X is then

$$\|X\|_{\Psi_{\alpha,L}} := \inf\{\eta > 0 \ : \ \mathbb{E}[\Psi_{\alpha,L}(|X|/\eta)] \leq 1\}.$$

Lemma 1. *Suppose assumptions (A1)–(A4) hold. Let*

$$T_{kj} = n^{-1/2} m_n^{1/2} \sum_{i=1}^{n} B_{kj}(X_{ij})\epsilon_i, \qquad 1 \leq k \leq m_n, 1 \leq j \leq p,$$

and $T_n = \max_{1\leq j\leq p, 1\leq k\leq m_n} |T_{kj}|$. *Let* $\alpha^* = \max\{0, \frac{\alpha-1}{\alpha}\}$. *When* $m_n \log(pm_n)/n \to 0$ *as* $n \to \infty$, *we obtain the following bound:*

$$\mathbb{E}[T_n] = \left(\sqrt{\log(pm_n)} + n^{-1/2+\alpha^*} m_n^{1/2-\alpha^*} (\log(pm_n))^{1/\alpha}\right) O(1).$$

Proof. From Remark A.1 in [5], conditional on X_{ij}'s, we obtain that

$$\begin{aligned} &\mathbb{E}\left[\max_{1\leq j\leq p, 1\leq k\leq m_n} |T_{kj}| \mid \{X_{ij}, 1 \leq j \leq p, 1 \leq i \leq n\}\right] \\ &\quad \leq n^{-1/2} m_n^{1/2} \max_{1\leq j\leq p, 1\leq k\leq m_n} \|T_{kj}\|_{\Psi_{\alpha,L}} C_\alpha \left\{\sqrt{\log(pm_n)} + L(\log(pm_n))^{1/\alpha}\right\}, \end{aligned} \tag{S0.1}$$

for some constant C_α depending on α. Similarly, the application of Theorem 3.1 in [5] yields that

$$\|T_{kj}\|_{\Psi_{\alpha,L}} \leq 2eC_1(\alpha)\|b\|_2, \tag{S0.2}$$

where $C_1(\alpha)$ is some constant involving α for which the explicit expression is given in Theorem 3.1 of [5], $\|b\|_2 = \left(\sum_{i=1}^{n} B_{kj}^2(X_{ij})\|\epsilon_i\|_{\psi_\alpha}^2\right)^{1/2}$, and

$$L(\alpha) = \frac{4^{1/\alpha}}{\sqrt{2}\|b\|_2} \times \begin{cases} \|b\|_\infty, & \text{if } \alpha < 1, \\ 4e\|b\|_\varpi / C_1(\alpha), & \text{if } \alpha \geq 1, \end{cases}$$

with ϖ is the Holder conjugate satisfying $1/\alpha + 1/\varpi = 1$. Therefore, combination of (S0.1) and (S0.2) yields

$$
\begin{aligned}
&\mathbb{E}\left[\max_{1\le j\le p,1\le k\le m_n} |T_{kj}| \mid \{X_{ij}, 1 \le j \le p, 1 \le i \le n\}\right] \\
&\quad\le 2eC_\alpha C_1(\alpha) n^{-1/2} m_n^{1/2} \max_{1\le j\le p,1\le k\le m_n} \|b\|_2 \sqrt{\log(pm_n)} \\
&\quad\quad + n^{-1/2} m_n^{1/2} \begin{cases} D_1(\alpha) \max_{1\le j\le p,1\le k\le m_n} \|b\|_\infty (\log(pm_n))^{1/\alpha} & \text{if } \alpha < 1, \\ D_2(\alpha) \max_{1\le j\le p,1\le k\le m_n} \|b\|_\varpi (\log(pm_n))^{1/\alpha} & \text{if } \alpha \ge 1, \end{cases}
\end{aligned} \tag{S0.3}
$$

where $D_1(\alpha) = \sqrt{2}eC_1(\alpha)C_\alpha 4^{1/\alpha}$ and $D_2(\alpha) = \sqrt{2}C_\alpha e^2 4^{1/\alpha+1}$.

Let $s_{n1} = \max_{1\le j\le p,1\le k\le m_n} \|b\|_2$, $s_{n2} = \max_{1\le j\le p,1\le k\le m_n} \|b\|_\infty$, and $s_{n3} = \max_{1\le j\le p,1\le k\le m_n} \|b\|_\varpi$. It follows that

$$
\begin{aligned}
&\mathbb{E}\left[\max_{1\le j\le p,1\le k\le m_n} |T_{kj}|\right] \\
&\quad\le 2eC_\alpha C_1(\alpha) n^{-1/2} m_n^{1/2} \mathbb{E}\left(s_{n1}\right) \sqrt{\log(pm_n)} \\
&\quad\quad + n^{-1/2} m_n^{1/2} \begin{cases} D_1(\alpha) \mathbb{E}\left(s_{n2}\right) (\log(pm_n))^{1/\alpha} & \text{if } \alpha < 1, \\ D_2(\alpha) \mathbb{E}\left(s_{n3}\right) (\log(pm_n))^{1/\alpha} & \text{if } \alpha \ge 1. \end{cases}
\end{aligned} \tag{S0.4}
$$

We now derive the upper bound for the $\mathbb{E}(s_{n3})$. From Theorem 4.2 of Chapter 5 in [1], we obtain that

$$
\sum_{i=1}^n \mathbb{E}\left\{B_{jk}^\varpi(X_{ij}) - \mathbb{E}B_{jk}^\varpi(X_{ij})\right\}^2 \le C_3 n m_n^{-1}, \qquad |B_{jk}(X_{ij})| \le 2, \tag{S0.5}
$$

and

$$
\max_{1\le j\le p,1\le k\le m_n} \sum_{i=1}^n \mathbb{E}B_{jk}^\varpi(X_{ij}) \le C_4 n m_n^{-1}. \tag{S0.6}
$$

By Lemma A.1 of [3] and (S0.5) we obtain

$$
\begin{aligned}
&\mathbb{E}\left(\max_{1\le j\le p,1\le k\le m_n} \left|\sum_{i=1}^n \mathbb{E}\left\{B_{jk}^\varpi(X_{ij}) - \mathbb{E}B_{jk}^\varpi(X_{ij})\right\}\right|\right) \\
&\quad\le \sqrt{2C_3 n m_n^{-1} \log(2pm_n)} + 2^\varpi \log(2pm_n).
\end{aligned} \tag{S0.7}
$$

Now, an application of Triangle inequality using (S0.7) and (S0.6) gives

$$
\mathbb{E}(s_{n3}^\varpi) \le \sqrt{2C_3 n m_n^{-1} \log(2pm_n)} + 2^\varpi \log(2pm_n) + C_4 n m_n^{-1}.
$$

Therefore, using Lyapunov's inequality, we obtain that, for $1 < \varpi < \infty$,

$$
\mathbb{E}(s_{n3}) \le \left(\sqrt{2C_3 n m_n^{-1} \log(2pm_n)} + 2^\varpi \log(2pm_n) + C_4 n m_n^{-1}\right)^{1/\varpi}. \tag{S0.8}
$$

Letting $\varpi = 2$ in (S0.8) yields the corresponding bound for $\mathbb{E}(s_{n1})$. By (S0.5), we can obtain that $\mathbb{E}(s_{n2}) \leq C_5$. Consequently,

$$\begin{aligned}
\mathbb{E}\left[T_n\right] \leq\ & 2eC_\alpha C_1(\alpha) n^{-1/2} m_n^{1/2} \Big(\sqrt{2C_3 n m_n^{-1} \log(2pm_n)} \\
& + 4\log(2pm_n) + C_4 n m_n^{-1} \Big)^{1/2} \sqrt{\log(pm_n)} + n^{-1/2} m_n^{1/2} \\
& \times \begin{cases} C_5 D_1(\alpha)(\log(pm_n))^{1/\alpha} & \text{if } \alpha < 1, \\ D_2(\alpha) \Big(\sqrt{2C_3 n m_n^{-1} \log(2pm_n)} & \\ +2^{\varpi} \log(2pm_n) + C_4 n m_n^{-1} \Big)^{1/\varpi} (\log(pm_n))^{1/\alpha} & \text{if } \alpha \geq 1. \end{cases}
\end{aligned} \tag{S0.9}$$

Let $\alpha^* = \max\{0, \frac{\alpha-1}{\alpha}\}$. When $m_n \log(pm_n)/n \to 0$ as $n \to \infty$, the above bound simplifies to the following bound:

$$\mathbb{E}\left[T_n\right] = \left(\sqrt{\log(pm_n)} + n^{-1/2+\alpha^*} m_n^{1/2-\alpha^*} (\log(pm_n))^{1/\alpha} \right) O(1).$$

□

Proof of Theorem 1. Note that the oracle estimator $\widehat{\beta}^0 = (\sqrt{m_n}\widehat{\mu}_0^0, \widehat{\beta}_1^0, \ldots, \widehat{\beta}_q^0)^T \in \mathbb{R}^{qm_n+1}$, minimizes the DPD loss function

$$\ell_n(\beta) := \frac{1}{n} \sum_{i=1}^n V_i(\beta; \sigma^2, \nu), \tag{S0.10}$$

where $V_i(\cdot)$ is defined in (7). We remark that penalization is not required for the oracle model as the true components are known. Using Taylor expansion, we have

$$\left.\frac{\partial \ell_n(\beta)}{\partial \beta}\right|_{\beta=\widehat{\beta}^0} = \frac{\partial \ell_n(\beta)}{\partial \beta} + \left.\frac{\partial^2 \ell_n(\beta)}{\partial \beta \partial \beta^T}\right|_{\beta=\overline{\beta}} (\widehat{\beta}^0 - \beta),$$

where $\overline{\beta} = t\widehat{\beta}^0 + (1-t)\beta$, $t \in [0,1]$. Therefore,

$$\widehat{\beta}^0 - \beta = -\left(\left.\frac{\partial^2 \ell_n(\beta)}{\partial \beta \partial \beta^T}\right|_{\beta=\overline{\beta}} \right)^{-1} \frac{\partial \ell_n(\beta)}{\partial \beta}.$$

Let $Z_i^0 = (1/\sqrt{m_n}, B_{11}(X_{i1}), \ldots, B_{m_n q}(X_{iq}))^T$ be the corresponding spline basis for the first q variables. By straightforward calculations, we have

$$\frac{\partial \ell_n(\beta)}{\partial \beta} = -\frac{1+\nu}{n\sigma^2} \sum_{i=1}^n f_i^\nu \times (Y_i - Z_i^{0^T}\beta) Z_i^0, \tag{S0.11}$$

and

$$\left.\frac{\partial^2 \ell_n(\beta)}{\partial \beta \partial \beta^T}\right|_{\beta=\overline{\beta}} = -\frac{1+\nu}{n\sigma^2} \sum_{i=1}^n \left\{ \frac{\nu}{\sigma^2} f_i^\nu \times (Y_i - Z_i^{0^T}\overline{\beta})^2 Z_i^0 Z_i^{0^T} - f_i^\nu Z_i^0 Z_i^{0^T} \right\}. \tag{S0.12}$$

First, we find the bound for (S0.11). Let $\delta_i = \sum_{j=1}^{q} g_j(X_i) - g_{nj}(X_i)$. Observe that

$$\left\| -\frac{1+\nu}{n\sigma^2} \sum_{i=1}^{n} f_i^{\nu} \times (Y_i - Z_i^{0^T}\beta) Z_i^0 \right\| = \left\| -\frac{1+\nu}{n\sigma^2} \sum_{i=1}^{n} f_i^{\nu} \times (\delta_i + \epsilon_i) Z_i^0 \right\|$$
$$\leq \left\| -\frac{1+\nu}{n\sigma^2} \sum_{i=1}^{n} f_i^{\nu} \delta_i Z_i^0 \right\| + \left\| -\frac{1+\nu}{n\sigma^2} \sum_{i=1}^{n} f_i^{\nu} \epsilon_i Z_i^0 \right\|.$$

From Fact 1 in (19) and Lemma 1 in [4], we have $\max_i \delta_i \leq C_{12} q m_n^{-d}$ for some constant C_{12}. Using the fact that $\max_i f_i^{\nu} \leq 1$, we obtain

$$\left\| -\frac{1+\nu}{n\sigma^2} \sum_{i=1}^{n} f_i^{\nu} (Y_i - Z_i^{0^T}\beta^0) Z_i^0 \right\|$$
$$\leq C_{12} q m_n^{-d} \frac{1+\nu}{\sigma^2} \left\| \frac{1}{n} \sum_{i=1}^{n} Z_i^0 \right\| + \frac{1+\nu}{\sigma^2} \left\| \frac{1}{n} \sum_{i=1}^{n} \epsilon_i Z_i^0 \right\|$$
$$= C_{12} q m_n^{-d-1/2} \frac{1+\nu}{\sigma^2} + \frac{1+\nu}{\sigma^2} \left\| \frac{1}{n} \sum_{i=1}^{n} \epsilon_i Z_i^0 \right\| \tag{S0.13}$$

where the second step follows because of $n^{-1} \sum_{i=1}^{n} B_{kj}(X_{ij}) = 0$, for $1 \leq k \leq m_n, 1 \leq j \leq q$. Consider

$$\left\| \frac{1}{n} \sum_{i=1}^{n} \epsilon_i Z_i^0 \right\|^2$$
$$= \sum_{j=1}^{q} \sum_{k=1}^{m_n} \left(\frac{1}{n} \sum_{i=1}^{n} \epsilon_i B_{kj}(X_{ij}) \right)^2 + \left(\frac{1}{n\sqrt{m_n}} \sum_{i=1}^{n} \epsilon_i \right)^2$$
$$\leq \max_{1 \leq j \leq q, 1 \leq k \leq m_n} \frac{q}{n} \left(\frac{m_n^{1/2}}{n^{1/2}} \sum_{i=1}^{n} \epsilon_i B_{kj}(X_{ij}) \right)^2 + \left(\frac{1}{n\sqrt{m_n}} \sum_{i=1}^{n} \epsilon_i \right)^2$$
$$= O_p(1)(q/n) \left(\log(q m_n) + n^{-1+2\alpha^*} m_n^{1-2\alpha^*} (\log(q m_n))^{2/\alpha} \right) + O_p(1/n m_n), \tag{S0.14}$$

where the first term in the last step follows from Lemma 1 and the second term follows from Condition (A4). Combination of (S0.11), (S0.13), and (S0.14) yields

$$\left\| \frac{\partial \ell_n(\beta)}{\partial \beta} \right\|$$
$$= O\left(m_n^{-d-1/2}\right) + O_p\left(\sqrt{ \frac{\log(q m_n) + n^{-1+2\alpha^*} m_n^{1-2\alpha^*} (\log(q m_n))^{2/\alpha}}{n} + \frac{1}{n m_n} } \right). \tag{S0.15}$$

We now show that,

$$\left\| \left(-\frac{1+\nu}{n\sigma^2} \sum_{i=1}^{n} \left\{ \frac{\nu}{\sigma^2} f_i^{\nu} \times (Y_i - Z_i^{0^T}\overline{\beta})^2 Z_i^0 Z_i^{0^T} - f_i^{\nu} Z_i^0 Z_i^{0^T} \right\} \right)^{-1} \right\| = O\left(m_n\right). \tag{S0.16}$$

From (S0.12), we write

$$\sum_{i=1}^{n} \frac{1+\nu}{n} \left\{ \frac{1}{\sigma^2} f_i^{\nu} \times \left[1 - \frac{\nu(Y_i - Z_i^{0^T}\overline{\beta})^2}{\sigma^2} \right] \right\} Z_i^0 Z_i^{0^T}.$$

First, observe that $\left[1 - \frac{\nu(Y_i - Z_i^{0^T}\overline{\beta})^2}{\sigma^2}\right] \leq \exp\{-\frac{\nu(Y_i - Z_i^{0^T}\overline{\beta})^2}{2\sigma^2}\} \leq 1$ for $i = 1, \ldots, n$. Consequently $f_i^{\nu} \times \left[1 - \frac{\nu(Y_i - Z_i^{0^T}\overline{\beta})^2}{\sigma^2}\right] \leq 1/\sigma\sqrt{2\pi}$. Further, since $\exp\{\frac{\nu(Y_i - Z_i^{0^T}\overline{\beta})^2}{2\sigma^2}\} \geq \left[1 - \frac{\nu(Y_i - Z_i^{0^T}\overline{\beta})^2}{\sigma^2}\right]$, we have the following lower bound

$$f_i^{\nu} \left[1 - \frac{\nu(Y_i - Z_i^{0^T}\overline{\beta})^2}{\sigma^2} \right] \geq \left[1 - \frac{\nu(Y_i - Z_i^{0^T}\overline{\beta})^2}{\sigma^2} \right]^2 / \sigma\sqrt{2\pi},$$

which is always positive and takes zero when $(Y_i - Z_i^{0^T}\overline{\beta})^2 = \sigma^2/\nu$. Together, we have,

$$1/\sigma\sqrt{2\pi} \geq f_i^{\nu} \times \left[1 - \frac{\nu(Y_i - Z_i^{0^T}\overline{\beta})^2}{\sigma^2} \right] \geq \left[1 - \frac{\nu(Y_i - Z_i^{0^T}\overline{\beta})^2}{\sigma^2} \right]^2 / \sigma\sqrt{2\pi}, \tag{S0.17}$$

which concludes that the weights are positive and bounded. Therefore, by Lemma 3 in [4], we prove (S0.16). Consequently, the result (21) follows from (S0.15) and (S0.16). □

Proof of Theorem 2. The main idea of the proof is similar to Theorem 3 in [2]. However, the details are more involved due to the group penalty and the basis functions. Without loss of generality, we assume the first q components, g_j, $j = 1, \ldots, q$, are nonzero. Let $\beta = (\beta^{(1)^T}, \beta^{(2)^T})^T \in \mathbb{R}^{pm_n+1}$ with $\beta^{(1)} = (\mu, \beta_1^T, \ldots, \beta_q^T)^T$ and $\beta^{(2)} = 0$.

Step1: Consistency in the $q-$dimensional space:

Let $Z_i^{(1)} = (1, B_{11}(X_{i1}), \ldots, B_{m_n q}(X_{iq}))^T$ and $Z_i^{(2)} = (B_{1q+1}(X_{iq+1}), \ldots, B_{m_n p}(X_{ip}))^T$ be the basis functions corresponding to first q nonzero functions (including intercept) and $p - q$ zero functions, respectively. First, we constrain the likelihood L_ν to $qm_n + 1$ dimensional subspace as the following:

$$\overline{L}_\nu(\delta) = -\ell_n(\delta) - \sum_{j=1}^{q} P_\lambda(\|\delta_j\|_2), \tag{S0.18}$$

where $\delta = (1, \delta_1^T, \ldots, \delta_q^T)^T$ with $\delta_j = (\delta_{1j}, \ldots, \delta_{m_n j})^T$ and ℓ_n is defined in (S0.10). Note that we take negative signs in $\overline{L}_\nu(\delta)$ so that we now maximize the likelihood instead of minimizing L_ν in (6). We show that there exists a local maximizer $\widehat{\beta}^{(1)}$ of $\overline{L}_\nu(\delta)$ such that $\|\widehat{\beta}^{(1)} - \beta^{(1)}\| = O_p(\gamma_n)$. Define an event

$$H_n = \left\{ \max_{\delta \in \partial N_\tau} \overline{L}_\nu(\delta) < \overline{L}_\nu(\beta^{(1)}) \right\}, \tag{S0.19}$$

where ∂N_τ is the boundary of the closed set

$$N_\tau = \{\delta \in \mathbb{R}^{qm_n+1} : \|\delta - \beta^{(1)}\| \leq \gamma_n \tau\},$$

and $\tau \in (0, \infty)$ and rate γ_n. Note that on event, H_n there exists a local maximizer $\widehat{\beta}^{(1)}$ of $\overline{L}_\nu(\delta)$ in N_τ. Therefore, it is sufficient to show that $P(H_n)$ approaches to 1 as $n \to \infty$.

By Taylor expansion, for any $\delta \in N_\tau$, we have

$$\overline{L}_\nu(\delta) - \overline{L}_\nu(\beta^{(1)}) = (\delta - \beta^{(1)})^T V - \frac{1}{2}(\delta - \beta^{(1)})^T D (\delta - \beta^{(1)}), \tag{S0.20}$$

where

$$V := \nabla \overline{L}_\nu(\beta^{(1)}) = \frac{1+\nu}{n\sigma^2} \sum_{i=1}^n f_i^\nu \cdot (Y_i - Z_i^{(1)^T} \beta^{(1)}) Z_i^{(1)} - \overline{P}_\lambda(\beta^{(1)}),$$

with $\overline{P}_\lambda(\beta^{(1)}) = (P'_\lambda(\|\beta_1\|_2)(D_1\beta_1)^T/\|\beta_1\|_2, \ldots, P'_\lambda(\|\beta_q\|_2)(D_q\beta_q)^T/\|\beta_q\|_2)^T$ where P'_λ is the derivative of the penalty function, and

$$\begin{aligned} D := -\nabla^2 \overline{L}_\nu(\beta^*) = &\sum_{i=1}^n \frac{1+\nu}{n} \left\{ \frac{1}{\sigma^2} f_i^\nu \left[1 - \frac{\nu(Y_i - Z_i^{(1)^T}\beta^*)^2}{\sigma^2} \right] \right\} Z_i^{(1)} Z_i^{(1)^T} \\ &+ \operatorname{diag}\left\{ \frac{P'_\lambda(\|\beta_j^*\|_2)}{\|\beta_j^*\|_2} D_j + \left[\frac{P''_\lambda(\|\beta_j^*\|_2)}{\|\beta_j^*\|_2^2} - \frac{P'_\lambda(\|\beta_j^*\|_2)}{\|\beta_j^*\|_2^3} \right] D_j \beta_j^* \beta_j^{*^T} D_j \right\}, \end{aligned} \tag{S0.21}$$

where β^* is on the line segment joining $\beta^{(1)}$ and δ. We note that the matrix D_j is positive definite, and its eigenvalues are of order m_n^{-1}. For any $\delta \in \partial N_\tau$, we have $\|\delta - \beta^{(1)}\| = \gamma_n \tau$ and $\beta^* \in N_\tau$. By doing calculations analogous to (S0.16), we obtain from (S0.21) that

$$\lambda_{\min}(D) \geq C_{21}(m_n^{-1} - \lambda\kappa_0 m_n^{-2}). \tag{S0.22}$$

Thus by (S0.20), we have

$$\max_{\delta \in \partial N_\tau} \overline{L}_\nu(\delta) - \overline{L}_\nu(\beta^{(1)}) \leq \gamma_n \tau [\|V\| - C_{21}\gamma_n\tau(m_n^{-1} - \lambda\kappa_0 m_n^{-2})],$$

which, together with Markov's inequality, gives

$$\begin{aligned} P(H_n) &\geq P\left(\|V\|^2 < C_{21}^2 \gamma_n^2 \tau^2 (m_n^{-1} - \lambda\kappa_0 m_n^{-2})^2\right) \\ &\geq 1 - \frac{\mathbb{E}\|V\|^2}{C_{21}^2\gamma_n^2\tau^2(m_n^{-1} - \lambda\kappa_0 m_n^{-2})^2}. \end{aligned}$$

Calculations analogous to (S0.15) in Theorem 1 yield that

$$\begin{aligned} \mathbb{E}\|V\|^2 &\leq \mathbb{E}\left\| \frac{1+\nu}{n\sigma^2} \sum_{i=1}^n f_i^\nu \cdot (Y_i - Z_i^{(1)^T}\beta^{(1)}) Z_i^{(1)} \right\|^2 + \|\overline{P}_\lambda(\beta^{(1)})\|^2 \\ &\leq O\left(m_n^{-2d-1} + \frac{\log(pm_n)}{n} + \frac{m_n^{1-2\alpha^*}(\log(pm_n))^{2/\alpha}}{n^{2-2\alpha^*}} + \frac{1}{nm_n} \right) + O(q\lambda^2 m_n^{-2}). \end{aligned}$$

Consequently,

$$P(H_n) \geq 1 - \frac{O\left(m_n^{-2d+1} + \frac{m_n^2 \log(pm_n)}{n} + \frac{m_n^{3-2\alpha^*}(\log(pm_n))^{2/\alpha}}{n^{2-2\alpha^*}} + \frac{m_n}{n} + q\lambda^2 \right)}{C_{21}^2\gamma_n^2\tau^2(1 - \lambda\kappa_0 m_n^{-1})^2}.$$

By choosing $\gamma_n^2 = m_n^{-2d+1} + \frac{m_n^2 \log(pm_n)}{n} + \frac{m_n^{3-2\alpha^*}(\log(pm_n))^{2/\alpha}}{n^{2-2\alpha^*}} + \frac{m_n}{n} + q\lambda^2$ and based on $\lambda\kappa_0 = o(1)$ (condition (A5)), we have

$$P(H_n) \geq 1 - o(\tau^{-2}).$$

It proves

$$\sum_{j=1}^{q} \|\widehat{\beta}_j - \beta_j\|^2 = O_p\left(\frac{m_n^2 \log(pm_n)}{n} + \frac{m_n^{3-2\alpha^*}(\log(pm_n))^{2/\alpha}}{n^{2-2\alpha^*}} + \frac{m_n}{n}\right) + O\left(m_n^{-2d+1} + q\lambda^2\right).$$

Step 2: Sparsity:

From Step 1 we have $\widehat{\beta}^{(1)} \in \mathbb{R}^{qm_n+1}$ is a local maximizer of $\overline{L}_\nu$ on N_τ. We now prove that $\widehat{\beta} = (\widehat{\beta}^{(1)^T}, 0^T)^T$ is indeed a maximizer of $-L_\nu$ on the space $\mathbb{R}^{pm_n+1}$. Let $\xi = (1, \xi_1^T, \ldots, \xi_p^T)^T = \sum_{i=1}^n f_i^\nu \cdot (Y_i - Z_i^T\beta)Z_i$. Let $\widehat{\beta}_{S_0} = \widehat{\beta}^{(1)}$ and $\widehat{\beta}_{S_0^c} = \widehat{\beta}^{(2)} = 0$. Consider the event,

$$\mathcal{E}_2 = \left\{\|\xi_{S_0^c}\|_\infty \leq u_n\right\}.$$

consider

$$\begin{aligned}
\|\xi_{S_0^c}\|_\infty &= \|\sum_{i=1}^n f_i^\nu \cdot (Y_i - Z_i^T\beta)Z_i^{(2)}\|_\infty \\
&\leq n^{1/2}m_n^{-1/2} \cdot \max_{1\leq j\leq p, 1\leq k\leq m_n} |n^{-1/2}m_n^{1/2}\sum_{i=1}^n (Y_i - Z_i^T\beta)B_{kj}(X_{ij})| \\
&= n^{1/2}m_n^{-1/2} \cdot T_n,
\end{aligned}$$

where $T_n = \max_{1\leq j\leq p, 1\leq k\leq m_n} |n^{-1/2}m_n^{1/2}\sum_{i=1}^n (Y_i - Z_i^T\beta)B_{kj}(X_{ij})|$. We have,

$$\begin{aligned}
P(\mathcal{E}_2) &\geq P\left(T_n \leq \frac{u_n}{(n/m_n)^{1/2}}\right) \\
&\geq 1 - \frac{\mathbb{E}T_n}{u_n}(n/m_n)^{1/2} \\
&\geq 1 - \frac{\left(\sqrt{\log(pm_n)} + n^{-1/2+\alpha^*}m_n^{1/2-\alpha^*}(\log(pm_n))^{1/\alpha}\right)(n/m_n)^{1/2}}{u_n},
\end{aligned}$$

which follows from Lemma 2 in [4] after ignoring the small spline approximation error. Choosing $u_n = n/m_n$, we obtain

$$P(\mathcal{E}_2) \geq 1 - \left(\sqrt{\frac{m_n \log(pm_n)}{n}} + \frac{m_n^{1-\alpha^*}(\log(pm_n))^{1/\alpha}}{n^{1-\alpha^*}}\right) \to 1,$$

under the assumption that $m_n \log(pm_n)/n \to 0$ as $n \to \infty$. Following Theorem 2 in [2], it is sufficient to show that,

$$\begin{aligned}
\|W\|_\infty &:= n^{-1}\| \sum_{i=1}^{n} f_i^{\nu} \cdot (Y_i - Z_i^T(-\beta + \widehat{\beta} + \beta))Z_i^{(2)}\|_\infty \\
&\leq n^{-1}\left[\|\xi_{S_0^c}\|_\infty + \| \sum_{i=1}^{n} f_i^{\nu} \cdot (Z_i^T(\widehat{\beta} - \beta))Z_i^{(2)}\|_\infty\right] \\
&\leq o(1) + n^{-1}\| \sum_{i=1}^{n} Z_i^T(\widehat{\beta} - \beta)Z_i^{(2)}\| \\
&\leq o(1) + n^{-1}\| \sum_{i=1}^{n} Z_i^{(1)^T}(\widehat{\beta}^{(1)} - \beta^{(1)})Z_i^{(2)}\| \\
&= o(1) + O(m_n^{-1})\|(\widehat{\beta}^{(1)} - \beta^{(1)})\| \\
&= o(1),
\end{aligned}$$

which concludes the proof. □

References

[1] DeVore RA, Lorentz GG (1993) Constructive approximation, vol 303. Springer Science & Business Media

[2] Fan J, Lv J (2011) Nonconcave penalized likelihood with np-dimensionality. IEEE Transactions on Information Theory 57(8):5467–5484

[3] van de Geer SA (2008) High-dimensional generalized linear models and the lasso. Annals of Statistics 36(2):614–645, DOI 10.1214/009053607000000929, URL https://doi.org/10.1214/009053607000000929

[4] Huang J, Horowitz JL, Wei F (2010) Variable selection in nonparametric additive models. Annals of Statistics 38(4):2282–2313

[5] Kuchibhotla AK, Chakrabortty A (2022) Moving beyond sub-gaussianity in high-dimensional statistics: Applications in covariance estimation and linear regression. Information and Inference: A Journal of the IMA 11(4):1389–1456