Submitted to the Annals of Statistics

SUPPLEMENTARY MATERIAL FOR: FUNCTIONAL ADDITIVE REGRESSION

By Yingying Fan and Gareth M. James and Peter Radchenko

University of Southern California

1. Proofs of Theorems 1 and 2. Let $\boldsymbol{\eta} = (\boldsymbol{\eta}_1^T, \cdots, \boldsymbol{\eta}_p^T)^T$ be a (pq_n)-vector and $\Theta = (\Theta_1, \cdots, \Theta_p)$ be an $n \times (pq_n)$ matrix. With matrix notation, the linear FAR criterion minimizes the following objective function

$$Q(\boldsymbol{\eta}) = \frac{1}{2n}\|\mathbf{Y} - \Theta\boldsymbol{\eta}\|^2 + \sum_{j=1}^{p} \rho_{\lambda_n}\left(\frac{1}{\sqrt{n}}\|\Theta_j \boldsymbol{\eta}_j\|\right). \tag{1}$$

Define the $(q_n s_n)$-dimensional hypercube

$$\mathcal{N} = \{\boldsymbol{\eta} \in R^{pq_n} : \boldsymbol{\eta}_{\mathfrak{M}_0^c} = \mathbf{0},\ \|\boldsymbol{\eta} - \boldsymbol{\eta}_0\|_\infty \le \sqrt{c_0} q_n^{-1/2} n^{-\alpha}\}, \tag{2}$$

where $\|\cdot\|_\infty$ stands for the infinity norm of a vector.

Lemma 1.1. *Define the event $\mathcal{E}_1 = \{\|\Theta_{\mathfrak{M}_0}^T \boldsymbol{\varepsilon}^*\|_\infty \le n\lambda_n/2\}$. Assume that $\lambda_n n^\alpha q_n \sqrt{s_n} \to 0$ with α defined in Condition 2(B), then under Condition 2 and conditional on event $\mathcal{E}_1$, there exits a vector $\boldsymbol{\eta} \in \mathcal{N}$ such that $\boldsymbol{\eta}_{\mathfrak{M}_0}$ is a solution to the following nonlinear equations*

$$-\frac{1}{n}\Theta_{\mathfrak{M}_0}^T(\mathbf{Y} - \Theta_{\mathfrak{M}_0}\boldsymbol{\eta}_{\mathfrak{M}_0}) + \mathbf{v}_{\mathfrak{M}_0}(\boldsymbol{\eta}) = 0, \tag{3}$$

where $\mathbf{v}_{\mathfrak{M}_0}(\boldsymbol{\eta})$ is a vector obtained by stacking $\mathbf{v}_k(\boldsymbol{\eta}) = \rho'_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_k \eta_k\|)\frac{1}{\sqrt{n}}\frac{\Theta_k^T\Theta_k\boldsymbol{\eta}_k}{\|\Theta_k\boldsymbol{\eta}_k\|}$, $k \in \mathfrak{M}_0$ one underneath another.

Proof. For any $\tilde{\boldsymbol{\eta}} = (\tilde{\boldsymbol{\eta}}_1^T, \tilde{\boldsymbol{\eta}}_2^T, \cdots, \tilde{\boldsymbol{\eta}}_p^T)^T \in \mathcal{N}$, by Condition 2(D) we have

$$\begin{aligned}\frac{1}{\sqrt{n}}\max_{k\in\mathfrak{M}_0}\|\Theta_k(\tilde{\boldsymbol{\eta}}_k - \boldsymbol{\eta}_{0,k})\| &\le c_0^{-1/2}\max_{k\in\mathfrak{M}_0}\|\tilde{\boldsymbol{\eta}}_k - \boldsymbol{\eta}_{0,k}\| \\ &\le c_0^{-1/2}\sqrt{q_n}\max_{k\in\mathfrak{M}_0}\|\tilde{\boldsymbol{\eta}}_k - \boldsymbol{\eta}_{0,k}\|_\infty \le n^{-\alpha}.\end{aligned} \tag{4}$$

This together with triangular inequality and Condition 2(B) entails that for n large enough,

$$\|\Theta_k\tilde{\boldsymbol{\eta}}_k\| \ge \|\Theta_k\boldsymbol{\eta}_{0,k}\| - \|\Theta_k(\tilde{\boldsymbol{\eta}}_k - \boldsymbol{\eta}_{0,k})\| \ge \|\Theta_k\boldsymbol{\eta}_{0,k}\| - n^{\frac{1}{2}-\alpha} > \sqrt{n}a_n/2. \tag{5}$$

Thus, by Condition 2(A), for any $k \in \mathfrak{M}_0$, $\rho'_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_k\tilde{\boldsymbol{\eta}}_k\|) \leq \rho'_{\lambda_n}(a_n/2)$. Hence, by the definition of $\mathbf{v}$ and Condition 2(D) we obtain that for any $\tilde{\boldsymbol{\eta}} \in \mathcal{N}$,

(6)
$$\|\mathbf{v}_{\mathfrak{M}_0}(\tilde{\boldsymbol{\eta}}_k)\|_\infty \leq \max_{k\in\mathfrak{M}_0} \rho'_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_k\tilde{\boldsymbol{\eta}}_k\|) \max_{k\in\mathfrak{M}_0} \frac{1}{\sqrt{n}} \frac{\|\Theta_k^T\Theta_k\tilde{\boldsymbol{\eta}}_k\|}{\|\Theta_k\tilde{\boldsymbol{\eta}}_k\|} \leq \frac{\rho'_{\lambda_n}(a_n/2)}{\sqrt{c_0}}.$$

Since $\frac{1}{n}\Theta_{\mathfrak{M}_0}^T\Theta_{\mathfrak{M}_0}$ has bounded eigenvalues, it follows from matrix norm calculations that

$$\|(\Theta_{\mathfrak{M}_0}^T\Theta_{\mathfrak{M}_0})^{-1}\|_\infty \leq \sqrt{s_nq_n}\Lambda_{\max}\Big((\Theta_{\mathfrak{M}_0}^T\Theta_{\mathfrak{M}_0})^{-1}\Big) \leq c_0^{-1}n^{-1}\sqrt{s_nq_n}.$$

Combining the above inequality with Cauchy-Schwartz inequality, Condition 2(C) and (6) yields

$$\begin{aligned} &n\|(\Theta_{\mathfrak{M}_0}^T\Theta_{\mathfrak{M}_0})^{-1}\mathbf{v}_{\mathfrak{M}_0}(\tilde{\boldsymbol{\eta}}_k)\|_\infty \\ &\leq n\|(\Theta_{\mathfrak{M}_0}^T\Theta_{\mathfrak{M}_0})^{-1}\|_\infty\|\mathbf{v}_{\mathfrak{M}_0}(\tilde{\boldsymbol{\eta}}_k)\|_\infty \leq o\big(n^{-\alpha}q_n^{-1/2}\big). \end{aligned}$$

Similarly, since $\lambda_n n^\alpha q_n\sqrt{s_n} \to 0$, conditional on the event $\mathcal{E}_1$ we have

$$\|(\Theta_{\mathfrak{M}_0}^T\Theta_{\mathfrak{M}_0})^{-1}\Theta_{\mathfrak{M}_0}^T\boldsymbol{\varepsilon}^*\|_\infty \leq \|(\Theta_{\mathfrak{M}_0}^T\Theta_{\mathfrak{M}_0})^{-1}\|_\infty\|\Theta_{\mathfrak{M}_0}^T\boldsymbol{\varepsilon}^*\|_\infty \leq o(n^{-\alpha}q_n^{-1/2}).$$

Combing the above two inequalities and by Cauchy-Schwartz inequality we obtain for large enough n,

$$\|(\Theta_{\mathfrak{M}_0}^T\Theta_{\mathfrak{M}_0})^{-1}(n\mathbf{v}_{\mathfrak{M}_0}(\tilde{\boldsymbol{\eta}}_k) - \Theta_{\mathfrak{M}_0}^T\boldsymbol{\varepsilon}^*)\|_\infty \leq o(q_n^{-1/2}n^{-\alpha}). \tag{7}$$

Define the vector-valued continuous function $\mathbf{g} : R^{s_nq_n} \to R^{s_nq_n}$ by $\mathbf{g}(\mathbf{x}) = \boldsymbol{\eta}_{0,\mathfrak{M}_0} - (\Theta_{\mathfrak{M}_0}^T\Theta_{\mathfrak{M}_0})^{-1}(n\mathbf{v}_{\mathfrak{M}_0}(\mathbf{x}) - \Theta_{\mathfrak{M}_0}^T\boldsymbol{\varepsilon}^*)$, where $\mathbf{x} = (\mathbf{x}_1^T, \cdots, \mathbf{x}_{s_n}^T)^T$ with $\mathbf{x}_k \in R^{q_n}$ for $k = 1, \cdots, s_n$, and $\mathbf{v}_{\mathfrak{M}_0}(\mathbf{x})$ is a vector obtained by stacking the vectors $\mathbf{v}_k(\mathbf{x}_k) = \rho'_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_k\mathbf{x}_k\|)\frac{1}{\sqrt{n}}\frac{\Theta_k^T\Theta_k\mathbf{x}_k}{\|\Theta_k\mathbf{x}_k\|}$, $k = 1, \cdots, s_n$ one underneath another. Then for any $\mathbf{x} \in \mathcal{N}$, by (7) we have

$$\|\mathbf{g}(\mathbf{x}) - \boldsymbol{\eta}_{0,\mathfrak{M}_0}\|_\infty \leq \sqrt{c_0}q_n^{-1/2}n^{-\alpha}$$

for large enough n. The above inequality indicates that $\mathbf{g}(\mathcal{N}) \subset \mathcal{N}$. Since $\mathbf{g}(\mathbf{x})$ is a continuous function on the convex, compact hypercube $\mathcal{N}$, applying Brouwer's fixed point theorem shows that (3) indeed has a solution in $\mathcal{N}$. $\square$

LEMMA 1.2. *Define* $\mathcal{E}_2 = \{\|\Theta_{\mathfrak{M}_0^c}^T\boldsymbol{\varepsilon}^*\|_\infty \leq n\lambda_n/2\}$. *Assume* $q_n^{-2}s_n = o(\lambda_n)$, $q_n + \log p = O(n\lambda_n^2)$, *and* $\lambda_n n^\alpha q_n\sqrt{s_n} \to 0$ *with* α *defined in Condition 2(B). Then under Condition 2 and conditional on the event* $\mathcal{E}_1 \cap \mathcal{E}_2$, *there exists a local minimizer* $\widehat{\boldsymbol{\eta}}$ *of* $Q(\boldsymbol{\eta})$ *(1) such that* $\widehat{\boldsymbol{\eta}} \in \mathcal{N}$.

PROOF. Since λ_n satisfying conditions in Lemma 1.2 also satisfies conditions in Lemma 1.1, by Lemma 1.1, we know that there exists a vector $\widehat{\boldsymbol{\eta}} \in \mathcal{N}$ such that $\widehat{\boldsymbol{\eta}}_{\mathfrak{M}_0}$ is a solution to (2). We next show that under some additional conditions, $\widehat{\boldsymbol{\eta}}$ is a local minimizer of $Q(\boldsymbol{\eta})$ in the original R^{pq_n} space.

We first constraint the objective function $Q(\boldsymbol{\eta})$ to the $(q_n s_n)$-dimensional subspace $\mathcal{N}$ defined in (2). We will show that under Condition 2 and conditional on $\mathcal{E}_1 \cap \mathcal{E}_2$, $Q(\boldsymbol{\eta})$ is strictly convex around $\widehat{\boldsymbol{\eta}}$. Then this together with Lemma 1.1 entails that the critical value $\widehat{\boldsymbol{\eta}}_{\mathfrak{M}_0}$ minimizes $Q(\boldsymbol{\eta})$ in the subspace $\mathcal{N}$.

We proceed to prove the strict convexity of $Q(\boldsymbol{\eta})$ in $\mathcal{N}$. Define $h(\boldsymbol{\eta}) = \sum_{j=1}^{p} \rho_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_j \boldsymbol{\eta}_j\|)$, which is a function in $\mathbf{R}^{pq_n}$. Note that for each $k \in \mathfrak{M}_0$,

$$
\begin{aligned}
\frac{\partial^2}{\partial \boldsymbol{\eta}_k^2} h(\widehat{\boldsymbol{\eta}}) &= \Theta_k^T \Theta_k \frac{\rho'_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_k \widehat{\boldsymbol{\eta}}_k\|)}{\sqrt{n}\|\Theta_k \widehat{\boldsymbol{\eta}}_k\|} \\
&\quad + \Theta_k^T \Theta_k \widehat{\boldsymbol{\eta}}_k \widehat{\boldsymbol{\eta}}_k^T \Theta_k^T \Theta_k \Big(\frac{\rho''_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_k \widehat{\boldsymbol{\eta}}_k\|)}{n\|\Theta_k \widehat{\boldsymbol{\eta}}_k\|^2} - \frac{\rho'_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_k \widehat{\boldsymbol{\eta}}_k\|)}{\sqrt{n}\|\Theta_k \widehat{\boldsymbol{\eta}}_k\|^3} \Big).
\end{aligned} \tag{8}
$$

Since $\widehat{\boldsymbol{\eta}} \in \mathcal{N}$, similar to (5) we can show that $\|\Theta_k \widehat{\boldsymbol{\eta}}_k\| \geq \|\Theta_k \boldsymbol{\eta}_{k,0}\| - \|\Theta_k(\widehat{\boldsymbol{\eta}}_k - \boldsymbol{\eta}_{k,0})\| > \sqrt{n} a_n/2$ for any $k \in \mathfrak{M}_0$ and large enough n. Thus it follows from Condition 2 (A), (B) and (C) that

$$
\begin{aligned}
&0 < \frac{\rho'_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_k \widehat{\boldsymbol{\eta}}_k\|)}{\|\Theta_k \widehat{\boldsymbol{\eta}}_k\|/\sqrt{n}} \leq \frac{\rho'_{\lambda_n}(a_n/2)}{a_n/2} = o(1), \\
&\rho''_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_k \widehat{\boldsymbol{\eta}}_k\|)) = o(1),
\end{aligned}
$$

where the $o(\cdot)$ terms are uniformly over all $k \in \mathfrak{M}_0$. By linear algebra, for any matrices A, B and C satisfying $A = B + C$, we have $\Lambda_{\min}(A) \geq \Lambda_{\min}(B) + \Lambda_{\min}(C)$. By Condition 2(A), $\rho''_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_k \widehat{\boldsymbol{\eta}}_k\|) < 0$ and $\rho'_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_k \widehat{\boldsymbol{\eta}}_k\|) > 0$. These together with (8) and Condition 2(D) entail that uniformly over all $k \in \mathfrak{M}_0$,

$$
\begin{aligned}
\Lambda_{\min}\Big(\frac{\partial^2}{\partial \boldsymbol{\eta}_k^2} h(\widehat{\boldsymbol{\eta}})\Big) &\geq \Lambda_{\min}\big(\Theta_k^T \Theta_k\big) \frac{\rho'_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_k \widehat{\boldsymbol{\eta}}_k\|)}{\sqrt{n}\|\Theta_k \widehat{\boldsymbol{\eta}}_k\|} \\
&\quad + \Lambda_{\max}\big(\Theta_k^T \Theta_k \widehat{\boldsymbol{\eta}}_k \widehat{\boldsymbol{\eta}}_k^T \Theta_k^T \Theta_k\big) \Big(\frac{\rho''_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_k \widehat{\boldsymbol{\eta}}_k\|)}{n\|\Theta_k \widehat{\boldsymbol{\eta}}_k\|^2} - \frac{\rho'_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_k \widehat{\boldsymbol{\eta}}_k\|)}{\sqrt{n}\|\Theta_k \widehat{\boldsymbol{\eta}}_k\|^3} \Big) \\
&\geq \Lambda_{\max}\Big(\frac{1}{n}\Theta_k^T \Theta_k\Big) \Big(\rho''_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_k \widehat{\boldsymbol{\eta}}_k\|) - \frac{\rho'_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_k \widehat{\boldsymbol{\eta}}_k\|)}{\|\Theta_k \widehat{\boldsymbol{\eta}}_k\|/\sqrt{n}} \Big) = o(1), \qquad (9)
\end{aligned}
$$

where for the second inequality we used the fact that

$$\Lambda_{\max}\big(\Theta_k^T\Theta_k\widehat{\boldsymbol{\eta}}_k\widehat{\boldsymbol{\eta}}_k^T\Theta_k^T\Theta_k\big) = \Lambda_{\max}\big(\widehat{\boldsymbol{\eta}}_k^T\Theta_k^T\Theta_k\Theta_k^T\Theta_k\widehat{\boldsymbol{\eta}}_k\big) \le \Lambda_{\max}(\Theta_k^T\Theta_k)\|\Theta_k\widehat{\boldsymbol{\eta}}_k\|^2.$$

Let H be a block diagonal matrix with block matrices $\frac{\partial^2}{\partial\boldsymbol{\eta}_k^2}h(\widehat{\boldsymbol{\eta}})$, $k \in \mathfrak{M}_0$. Then it is easy to see that the Hessian matrix $\frac{\partial^2}{\partial\boldsymbol{\eta}_{\mathfrak{M}_0}^2}Q(\widehat{\boldsymbol{\eta}}) = n^{-1}\Theta_{\mathfrak{M}_0}^T\Theta_{\mathfrak{M}_0}+H$. Thus, it follows from the above inequality (9) that

$$\Lambda_{\min}\Big(\frac{\partial^2}{\partial\boldsymbol{\eta}_{\mathfrak{M}_0}^2}Q(\widehat{\boldsymbol{\eta}})\Big) \ge \frac{1}{n}\Lambda_{\min}(\Theta_{\mathfrak{M}_0}^T\Theta_{\mathfrak{M}_0}) + \min_{k\in\mathfrak{M}_0}\Lambda_{\min}\big(\frac{\partial^2}{\partial\boldsymbol{\eta}_k^2}h(\widehat{\boldsymbol{\eta}})\big) \ge c_0 - o(1). \tag{10}$$

Therefore, for large enough n, restricted on the space $\mathcal{N}$, the function $Q(\boldsymbol{\eta})$ is strictly convex around $\widehat{\boldsymbol{\eta}}$ and thus has a unique minimizer in a ball $\mathcal{N}_1 \subset \mathcal{N}$ centered at $\widehat{\boldsymbol{\eta}}$. Since by Lemma 1.1 $\widehat{\boldsymbol{\eta}}$ is a critical point, $\widehat{\boldsymbol{\eta}}$ is indeed this strict local minimizer in $\mathcal{N}_1$.

We next show that $\widehat{\boldsymbol{\eta}}$ is also a local minimizer in the original R^{pq_n}-dimensional space. We will first show that for $\widehat{\boldsymbol{\eta}}_{\mathfrak{M}_0}$ defined in Lemma 1.1, conditional on $\mathcal{E}_1 \cap \mathcal{E}_2$,

$$\max_{j\in\mathfrak{M}_0^c}\{\hat{\mathbf{v}}_j^T(\Theta_j^T\Theta_j)^{-1}\hat{\mathbf{v}}_j\}^{1/2} = \max_{j\in\mathfrak{M}_0^c}\|\Theta_j(\Theta_j^T\Theta_j)^{-1}\hat{\mathbf{v}}_j\| < n^{-1/2}\rho'_{\lambda_n}(0+), \forall j \in \mathfrak{M}_0^c, \tag{11}$$

where

$$\hat{\mathbf{v}}_j = n^{-1}\Theta_j^T(\mathbf{Y} - \Theta_{\mathfrak{M}_0}\widehat{\boldsymbol{\eta}}_{\mathfrak{M}_0}) = n^{-1}\Theta_j^T\Theta_{\mathfrak{M}_0}(\boldsymbol{\eta}_{0,\mathfrak{M}_0} - \widehat{\boldsymbol{\eta}}_{\mathfrak{M}_0}) + n^{-1}\Theta_j^T\boldsymbol{\varepsilon}^*.$$

By Lemma 1.1, we have $\boldsymbol{\eta}_{0,\mathfrak{M}_0} - \widehat{\boldsymbol{\eta}}_{\mathfrak{M}_0} = (\Theta_{\mathfrak{M}_0}^T\Theta_{\mathfrak{M}_0})^{-1}(n\mathbf{v}_{\mathfrak{M}_0} - \Theta_{\mathfrak{M}_0}^T\boldsymbol{\varepsilon}^*)$. Plugging this into $\hat{\mathbf{v}}_j$, we obtain that for $j \in \mathfrak{M}_0^c$, $\hat{\mathbf{v}}_j = \Theta_j^T\Theta_{\mathfrak{M}_0}(\Theta_{\mathfrak{M}_0}^T\Theta_{\mathfrak{M}_0})^{-1}\mathbf{v}_{\mathfrak{M}_0} + n^{-1}[\Theta_j - \Theta_j^T\Theta_{\mathfrak{M}_0}(\Theta_{\mathfrak{M}_0}^T\Theta_{\mathfrak{M}_0})^{-1}\Theta_{\mathfrak{M}_0}^T]\boldsymbol{\varepsilon}^*$. Therefore,

$$\{\hat{\mathbf{v}}_j^T(\Theta_j^T\Theta_j)^{-1}\hat{\mathbf{v}}_j\}^{1/2} = \|\Theta_j(\Theta_j^T\Theta_j)^{-1}\hat{\mathbf{v}}_j\| \le I_{1,j} + I_{2,j}, \tag{12}$$

where

$$\begin{aligned} I_{1,j} &= \|\Theta_j(\Theta_j^T\Theta_j)^{-1}\Theta_j^T\Theta_{\mathfrak{M}_0}(\Theta_{\mathfrak{M}_0}^T\Theta_{\mathfrak{M}_0})^{-1}\mathbf{v}_{\mathfrak{M}_0}\|, \\ I_{2,j} &= n^{-1}\|\Theta_j(\Theta_j^T\Theta_j)^{-1}\Theta_j^T\big(\mathbf{I} - \Theta_{\mathfrak{M}_0}(\Theta_{\mathfrak{M}_0}^T\Theta_{\mathfrak{M}_0})^{-1}\Theta_{\mathfrak{M}_0}^T\big)\boldsymbol{\varepsilon}^*\|. \end{aligned}$$

By (6), Condition 2(B) and Condition 2(D), conditional on $\mathcal{E}_1 \cap \mathcal{E}_2$, we have

$$\begin{aligned} I_{1,j} &\le \|\mathbf{v}_{\mathfrak{M}_0}\|_\infty\|\Theta_j(\Theta_j^T\Theta_j)^{-1}\Theta_j^T\Theta_{\mathfrak{M}_0}(\Theta_{\mathfrak{M}_0}^T\Theta_{\mathfrak{M}_0})^{-1}\|_{\infty,2} < \frac{1}{2\sqrt{n}}\rho'_{\lambda_n}(0+), \\ I_{2,j} &\le n^{-1}\|\Theta_j(\Theta_j^T\Theta_j)^{-1}\Theta_j^T\big(\mathbf{I} - \Theta_{\mathfrak{M}_0}(\Theta_{\mathfrak{M}_0}^T\Theta_{\mathfrak{M}_0})^{-1}\Theta_{\mathfrak{M}_0}^T\big)\boldsymbol{\varepsilon}\| \\ &\quad + n^{-1}\|\Theta_j(\Theta_j^T\Theta_j)^{-1}\Theta_j^T\big(\mathbf{I} - \Theta_{\mathfrak{M}_0}(\Theta_{\mathfrak{M}_0}^T\Theta_{\mathfrak{M}_0})^{-1}\Theta_{\mathfrak{M}_0}^T\big)\mathbf{e}\| \equiv I_{2,1,j} + I_{2,2,j}, \end{aligned}$$

where the inequality for $I_{1,j}$ is uniformly over all $j \in \mathfrak{M}_0$. Since both $\Theta_j(\Theta_j^T\Theta_j)^{-1}\Theta_j^T$ and $(\mathbf{I} - \Theta_{\mathfrak{M}_0}(\Theta_{\mathfrak{M}_0}^T\Theta_{\mathfrak{M}_0})^{-1}\Theta_{\mathfrak{M}_0}^T)$ are projection matrices and $\boldsymbol{\varepsilon}$ is a n-vector of Gaussian random variables, it follows that $n^2 I_{2,1,j}^2$ is a Chi-square random variable with degrees of freedom at most q_n. Thus, by Chi-square tail probability inequality (see [1]),

$$
\begin{aligned}
&P(\max_{j\in\mathfrak{M}_0^c} I_{2,1,j} > n^{-1}\sqrt{q_n + C\log p}) \\
&= P\big(\max_{j\in\mathfrak{M}_0^c} n^2 I_{2,1,j}^2 > (q_n + C\log p)\big) \le C(p - s_n)\exp(-C\log p) \to 0,
\end{aligned}
$$

where C is a large enough generic positive constant. Thus, $\max_{j\in\mathfrak{M}_0^c} I_{2,1,j} = o_p(n^{-1}(q_n^{1/2} + \sqrt{\log p}))$. Now by Condition 1 and assumption that $q_n^{-2}s_n = o(\lambda_n)$, it is easy to derive that $\|\mathbf{e}\|_\infty = o(\lambda_n)$. Thus, $\|\mathbf{e}\|_2 = o(n^{1/2}\lambda_n)$. This together with $\Theta_j(\Theta_j^T\Theta_j)^{-1}\Theta_j^T$ and $(\mathbf{I} - \Theta_{\mathfrak{M}_0}(\Theta_{\mathfrak{M}_0}^T\Theta_{\mathfrak{M}_0})^{-1}\Theta_{\mathfrak{M}_0}^T)$ being projection matrix ensures that uniformly over all $j \in \mathfrak{M}_0^c$,

$$
I_{2,2,j} \le n^{-1}\|\mathbf{e}\|_2 = o(n^{-1/2}\lambda_n).
$$

Since it is assumed in the theorem that $q_n + \log p = O(n\lambda_n^2)$, combining the above results on $I_{2,1,j}$ and $I_{2,2,j}$ yields

$$
\max_{j\in\mathfrak{M}_0^c} I_{2,j} = o_p(n^{-1}(q_n^{1/2} + \sqrt{\log(p)})) = o_p(\lambda_n/\sqrt{n}) < \rho'_{\lambda_n}(0+)/(2\sqrt{n}).
$$

In summary, the results on I_1 and I_2 show that inequality (11) holds.

Let $\mathcal{B} = \{\boldsymbol{\eta} \in R^{q_n p} : \boldsymbol{\eta}_{\mathfrak{M}_0^c} = 0\}$ be a subspace in R^{pq_n}. Take a sufficiently small ball $\mathcal{N}_2$ in R^{pq_n} centered at $\widehat{\boldsymbol{\eta}}$ such that $\mathcal{N}_2 \cap \mathcal{B} \subset \mathcal{N}_1$. Since $\rho'_{\lambda_n}(t)$ is a continuous decreasing function and (11) holds for $\widehat{\boldsymbol{\eta}} \in \mathcal{N}_2$, appropriately shrink the radius of the ball $\mathcal{N}_2$ gives that there exists a $\delta \in (0,\infty)$ such that for any $\boldsymbol{\eta} \in \mathcal{N}_2$,

$$
\max_{j\in\mathfrak{M}_0} \|\Theta_j(\Theta_j^T\Theta_j)^{-1}\Theta_j^T(\mathbf{Y} - \Theta\boldsymbol{\eta})\| < n^{1/2}\rho'_{\lambda_n}(\delta). \tag{13}
$$

Fix an arbitrary $\boldsymbol{\eta}_1 = (\boldsymbol{\eta}_{1,1}^T, \cdots, \boldsymbol{\eta}_{1p}^T)^T \in \mathcal{N}_2 \cap \mathcal{N}_1^c$, we next show that $Q(\boldsymbol{\eta}_1) > Q(\widehat{\boldsymbol{\eta}})$. Let $\boldsymbol{\eta}_2 = (\boldsymbol{\eta}_{2,1}^T, \cdots, \boldsymbol{\eta}_{2p}^T)^T$ be the projection of $\boldsymbol{\eta}_1$ onto $\mathcal{B}$. Then it follows from the definitions of $\mathcal{N}_1$, $\mathcal{N}_2$, $\mathcal{B}$ and $\widehat{\boldsymbol{\eta}}$ that $Q(\boldsymbol{\eta}_2) > Q(\widehat{\boldsymbol{\eta}})$. Thus we only need to show $Q(\boldsymbol{\eta}_1) \ge Q(\boldsymbol{\eta}_2)$.

Note that $Q(\boldsymbol{\eta}_1) - Q(\boldsymbol{\eta}_2) = \nabla Q(\boldsymbol{\eta}_3)(\boldsymbol{\eta}_1 - \boldsymbol{\eta}_2) = \sum_{j\in\mathfrak{M}_0^c} \boldsymbol{\eta}_{1j}^T \frac{\partial Q(\boldsymbol{\eta}_3)}{\partial \boldsymbol{\eta}_j}$, where $\boldsymbol{\eta}_3$ is a vector on the segment connecting $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$. Since $\boldsymbol{\eta}_{2k} = 0$ for any $k \in \mathfrak{M}_0^c$, there exits a constant $0 < \gamma < 1$ such that $\boldsymbol{\eta}_{3k} = \gamma\boldsymbol{\eta}_{1k}$, $k \in \mathfrak{M}_0^c$. Then by the definitions of $\mathcal{B}$, $\mathcal{N}_1$, $\mathcal{N}_2$, we know that $\boldsymbol{\eta}_3 \in \mathcal{N}_2$. Shrink the ball $\mathcal{N}_2$ such that for any $\boldsymbol{\eta} \in \mathcal{N}_2$, $\|\Theta_k \boldsymbol{\eta}_k\| = \|\Theta_k(\boldsymbol{\eta}_k - \widehat{\boldsymbol{\eta}}_k)\| \leq \sqrt{n}\delta, k \in \mathfrak{M}_0^c$. Since $\boldsymbol{\eta}_3 \in \mathcal{N}_2$, we have $\|\Theta_k \boldsymbol{\eta}_{3k}\| \leq \sqrt{n}\delta$ and thus $\rho'_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_k \boldsymbol{\eta}_{3k}\|) \geq \rho'_{\lambda_n}(\delta)$ for $k \in \mathfrak{M}_0^c$. Therefore,

$$\begin{aligned}
Q(\boldsymbol{\eta}_1) - Q(\boldsymbol{\eta}_2) &= \nabla Q(\boldsymbol{\eta}_3)(\boldsymbol{\eta}_1 - \boldsymbol{\eta}_2) = \sum_{j \in \mathfrak{M}_0^c} \boldsymbol{\eta}_{1j}^T \frac{\partial Q(\boldsymbol{\eta}_3)}{\partial \boldsymbol{\eta}_j} \\
&= \sum_{j \in \mathfrak{M}_0^c} \boldsymbol{\eta}_{1j}^T \Big(-\frac{1}{n}\Theta_j^T(\mathbf{Y} - \Theta\boldsymbol{\eta}_3) + \frac{\rho'_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_j \boldsymbol{\eta}_{3j}\|)}{\sqrt{n}\|\Theta_j \boldsymbol{\eta}_{3j}\|}\Theta_j^T\Theta_j\boldsymbol{\eta}_{3j}\Big) \\
&\geq -\frac{1}{n}\sum_{j \in \mathfrak{M}_0^c} \boldsymbol{\eta}_{1j}^T\Theta_j^T(\mathbf{Y} - \Theta\boldsymbol{\eta}_3) + \frac{1}{\sqrt{n}\gamma}\rho'_{\lambda_n}(\delta)\sum_{j \in \mathfrak{M}_0^c}\|\Theta_j\boldsymbol{\eta}_{3j}\| \equiv I_3 + I_4.
\end{aligned}$$

Next note that by Cauchy-Schwartz inequality and (13),

$$\begin{aligned}
|I_3| &\leq \frac{1}{n}\sum_{j \in \mathfrak{M}_0^c}\|\Theta_j\boldsymbol{\eta}_{1j}\|\|\Theta_j(\Theta_j^T\Theta_j)^{-1}\Theta_j^T(\mathbf{Y} - \Theta\boldsymbol{\eta}_3)\| \\
&= \frac{1}{n\gamma}\sum_{j \in \mathfrak{M}_0^c}\|\Theta_j\boldsymbol{\eta}_{3j}\|\|\Theta_j(\Theta_j^T\Theta_j)^{-1}\Theta_j^T(\mathbf{Y} - \Theta\boldsymbol{\eta}_3)\| \leq I_4.
\end{aligned}$$

Thus, $Q(\boldsymbol{\eta}_1) \geq Q(\boldsymbol{\eta}_2)$, which together with $Q(\boldsymbol{\eta}_2) > Q(\widehat{\boldsymbol{\eta}})$ ensures that $\widehat{\boldsymbol{\eta}}$ is also a strict local minimizer in the original R^{pq_n} dimensional space. The proof is completed. □

Proof of Theorem 1

PROOF. We only need to show that $P(\mathcal{E}_1 \cap \mathcal{E}_2) \to 1$. Then Theorem 1 follows easily from Lemmas 1.1 and 1.2. To this end, note that

$$\begin{aligned}
P(\mathcal{E}_1 \cap \mathcal{E}_2) &= 1 - P(\|\Theta^T\boldsymbol{\varepsilon}^*\|_\infty \geq n\lambda_n/2) \\
&\geq 1 - P(\|\Theta^T\boldsymbol{\varepsilon}\|_\infty \geq n\lambda_n/2 - \|\Theta^T\mathbf{e}\|_\infty).
\end{aligned}$$

By the assumption that $s_n q_n^{-2} = o(\lambda_n)$, it is easy to derive that $\|\mathbf{e}\|_\infty = o(\lambda_n)$. Since each column of Θ has ℓ_2 norm $\sqrt{n}$, it follows that $\|\Theta\|_1 \leq n$. Thus, by Cauchy-Schwartz inequality, $\|\Theta^T\mathbf{e}\|_\infty \leq \|\Theta\|_1\|\mathbf{e}\|_\infty \leq o(n\lambda_n)$. This follows that

$$\|\Theta^T\mathbf{e}\|_\infty \leq n\lambda_n/4$$

for large enough n.

Now we consider $\|\Theta^T\boldsymbol{\varepsilon}\|_\infty$. Let $\boldsymbol{\xi} = (\xi_1, \cdots, \xi_{pq})^T = \Theta^T\boldsymbol{\varepsilon}$, then $\xi_i \sim N(0, n\sigma^2 d_i^2)$ with d_i^2 the i-th diagonal of matrix $n^{-1}\Theta^T\Theta$. Since each column of Θ has ℓ_2 norm $\sqrt{n}$, we have $d_i^2 = 1$ for $1 \le i \le q_n p$. Hence, by Bonferroni's inequality and the assumption $n\lambda_n^2(\log(pq_n))^{-1} \to \infty$ we further obtain

$$P(\|\Theta^T \boldsymbol{\varepsilon}\|_\infty > n\lambda_n/4) \le \sum_{i=1}^{q_n p} P(|\xi_i| > n\lambda_n/4)$$
$$\le \frac{4\sigma p q_n}{\sqrt{2\pi n}\lambda_n \sigma} \exp\big(-n\lambda_n^2/(32\sigma^2)\big) \to 0.$$

Combining the above two results we have completed the proof of Theorem 1. □

Proof of Theorem 2

Proof. Let $\hat{\mathbf{v}}_{\mathfrak{M}_0} = \mathbf{v}_{\mathfrak{M}_0}(\widehat{\boldsymbol{\eta}})$ and $\mathbf{v}_{0,\mathfrak{M}_0} = \mathbf{v}_{\mathfrak{M}_0}(\boldsymbol{\eta}_0)$ with the function $\mathbf{v}_{\mathfrak{M}_0}(\cdot)$ defined in Lemma 1.1, $\widehat{\boldsymbol{\eta}}_{\mathfrak{M}_0}$ the solution to (3), and $\boldsymbol{\eta}_0$ the true regression coefficient vector. Since $\widehat{\boldsymbol{\eta}}_{\mathfrak{M}_0}$ is a solution to (3), for any vector $\mathbf{c} \in \mathbf{R}^{s_n q_n}$ satisfying $\mathbf{c}^T\mathbf{c} = 1$, we have the following decomposition

$$\begin{aligned}(14)\qquad &\mathbf{c}^T\big[(\Theta_{\mathfrak{M}_0}^T\Theta_{\mathfrak{M}_0})^{1/2}(\widehat{\boldsymbol{\eta}}_{\mathfrak{M}_0} - \boldsymbol{\eta}_{0,\mathfrak{M}_0}) + n(\Theta_{\mathfrak{M}_0}^T\Theta_{\mathfrak{M}_0})^{-1/2}\mathbf{v}_{0,\mathfrak{M}_0}\big]\\ =&\mathbf{c}^T(\Theta_{\mathfrak{M}_0}^T\Theta_{\mathfrak{M}_0})^{-1/2}\Theta_{\mathfrak{M}_0}^T\boldsymbol{\varepsilon} + \mathbf{c}^T(\Theta_{\mathfrak{M}_0}^T\Theta_{\mathfrak{M}_0})^{-1/2}\Theta_{\mathfrak{M}_0}^T\mathbf{e}\\ &+ n\mathbf{c}^T(\Theta_{\mathfrak{M}_0}^T\Theta_{\mathfrak{M}_0})^{-1/2}(\hat{\mathbf{v}}_{\mathfrak{M}_0} - \mathbf{v}_{0,\mathfrak{M}_0}) \equiv I_1 + I_2 + I_3.\end{aligned}$$

It is easy to see

$$(15)\qquad I_1 \sim N(0,\sigma^2).$$

As for I_2, note that similar to Theorem 1 we can prove that $\|\mathbf{e}\|_\infty = o(n^{-1/2})$. Thus, $\|\mathbf{e}\| = o(1)$. So we can derive

$$(16)\qquad |I_2| \le \|\mathbf{c}^T(\Theta_{\mathfrak{M}_0}^T\Theta_{\mathfrak{M}_0})^{-1/2}\Theta_{\mathfrak{M}_0}^T\|\|\mathbf{e}\| = \|\mathbf{e}\| = o(1).$$

Now let us consider I_3. By Cauchy-Schwartz inequality we obtain

$$\begin{aligned}(17)\qquad |I_3| &\le \|\sqrt{n}\mathbf{c}^T(\Theta_{\mathfrak{M}_0}^T\Theta_{\mathfrak{M}_0})^{-1/2}\|\|\sqrt{n}(\hat{\mathbf{v}}_{\mathfrak{M}_0} - \mathbf{v}_{0,\mathfrak{M}_0})\|\\ &\le c_0^{-1/2}\|\sqrt{n}(\hat{\mathbf{v}}_{\mathfrak{M}_0} - \mathbf{v}_{0,\mathfrak{M}_0})\|.\end{aligned}$$

Define $g(\boldsymbol{\eta}_k) = \frac{1}{\sqrt{n}}\rho'_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_k\boldsymbol{\eta}_k\|)\frac{\Theta_k^T\Theta_k\boldsymbol{\eta}_k}{\|\Theta_k\boldsymbol{\eta}_k\|}$. Then by definitions of $\hat{\mathbf{v}}_{\mathfrak{M}_0}$ and $\mathbf{v}_{0,\mathfrak{M}_0}$,

$$(18)\qquad \hat{\mathbf{v}}_k - \mathbf{v}_{0,k} = g(\widehat{\boldsymbol{\eta}}_k) - g(\boldsymbol{\eta}_{0,k}) = \frac{\partial}{\partial\boldsymbol{\eta}_k}g(\tilde{\boldsymbol{\eta}}_k)(\widehat{\boldsymbol{\eta}}_k - \boldsymbol{\eta}_{0,k})$$

with $\tilde{\boldsymbol{\eta}}_k$ lying on the segment connecting $\boldsymbol{\eta}_{0,k}$ and $\widehat{\boldsymbol{\eta}}_k$. Thus, $\tilde{\boldsymbol{\eta}} = (\tilde{\boldsymbol{\eta}}_1^T, \cdots, \tilde{\boldsymbol{\eta}}_p^T)^T \in \mathcal{N}$. It has been proved in (5) that $\|\Theta_k \boldsymbol{\eta}_k\| \geq \sqrt{n} a_n/2$ for any $\boldsymbol{\eta} \in \mathcal{N}$. Note that for any $\boldsymbol{\eta} = (\boldsymbol{\eta}_1^T, \cdots, \boldsymbol{\eta}_p^T)^T \in \mathcal{N}$, and any $k \in \mathfrak{M}_0$,

$$\frac{\partial}{\partial \boldsymbol{\eta}_k} g(\boldsymbol{\eta}_k) = \rho''_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_k \boldsymbol{\eta}_k\|) \frac{\Theta_k^T \Theta_k \boldsymbol{\eta}_k \boldsymbol{\eta}_k^T \Theta_k^T \Theta_k}{n \|\Theta_k \boldsymbol{\eta}_k\|^2} + \frac{\rho'_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_k \boldsymbol{\eta}_k\|)}{\sqrt{n}} \left\{ \frac{\Theta_k^T \Theta_k}{\|\Theta_k \boldsymbol{\eta}_k\|} - \frac{\Theta_k^T \Theta_k \boldsymbol{\eta}_k \boldsymbol{\eta}_k^T \Theta_k^T \Theta_k}{\|\Theta_k \boldsymbol{\eta}_k\|^3} \right\}.$$

Using similar arguments to (9) and by Condition 2(A) and the assumption $\sup_{t \geq \frac{a_n}{2}} \rho''_{\lambda_n}(t) = O(n^{-1/2})$, we have for any $k \in \mathfrak{M}_0$,

$$c_0^{-1}\Big(-O(\frac{1}{\sqrt{n}}) - \frac{2\rho'_{\lambda_n}(\frac{a_n}{2})}{a_n}\Big) \leq \Lambda_{\min}(\frac{\partial}{\partial \boldsymbol{\eta}_k} g(\boldsymbol{\eta}_k)) \leq \Lambda_{\max}(\frac{\partial}{\partial \boldsymbol{\eta}_k} g(\boldsymbol{\eta}_k)) \leq c_0^{-1} \frac{2\rho'_{\lambda_n}(\frac{a_n}{2})}{a_n}.$$

This together with (18), Theorem 1, and the theorem assumptions ensures that

$$\begin{aligned} \|\hat{\mathbf{v}}_{\mathfrak{M}_0} - \mathbf{v}_{0,\mathfrak{M}_0}\| &\leq c_0^{-1} \left(O(\frac{1}{\sqrt{n}}) + \frac{2\rho'_{\lambda_n}(\frac{a_n}{2})}{a_n} \right) \Big\{ \sum_{k \in \mathfrak{M}_0} \|\widehat{\boldsymbol{\eta}}_k - \boldsymbol{\eta}_{0,k}\|^2 \Big\}^{1/2} \\ &\leq c_0^{-3/2} \left(O(\frac{1}{\sqrt{n}}) + o\big(n^{\alpha - \frac{1}{2}} s_n^{-1/2}\big) \right) O_p(s_n^{1/2} n^{-\alpha}) = o_p(n^{-1/2}), \end{aligned}$$

So it follows that $\sqrt{n}\|\hat{\mathbf{v}}_{\mathfrak{M}_0} - \mathbf{v}_{0,\mathfrak{M}_0}\| = o_p(1)$. Combing this with (17) yields $I_3 \xrightarrow{\mathrm{P}} 0$. This together with (14) –(16) completes the proof. ☐

2. Proof of Lemma 1. Observe that

$$P\left((\varepsilon, \widehat{f} - f^*)_n > C_1 s_n r_n^2 + C_1 r_n \sum_{j=1}^{p_n} \|\widehat{f}_j - f_j^*\|_n \right) \leq \tag{19}$$

$$\sum_{j \in \mathfrak{M}_0} P\left(\frac{(\varepsilon, \widehat{f}_j - f_j^*)_n}{r_n + \|\widehat{f}_j - f_j^*\|_n} > C_1 r_n \right) + \sum_{j \in \mathfrak{M}_0^c} P\left((\varepsilon, \widehat{f}_j - f_j^*)_n > C_1 r_n \|\widehat{f}_j - f_j^*\|_n \right).$$

Consider an index $j \in \mathfrak{M}_0^c$, and note that $f_j^* \equiv 0$. We have,

$$P\left((\varepsilon, \widehat{f}_j - f_j^*)_n > C_1 r_n \|\widehat{f}_j - f_j^*\|_n \right) \leq P\left(\sup_{f \in \mathcal{F}_j(1)} (\varepsilon, f)_n > C_1 r_n \right),$$

where $\mathcal{F}_j(\delta)$ is defined for every positive δ as $\{f \in \mathcal{F}_j^0, \|f\|_n \leq \delta\}$. Given a pseudo-metric space $(\mathcal{X}, d)$, we will use $N(u, \mathcal{X}, d)$ to denote the smallest number N, such that N balls of d-radius u can cover $\mathcal{X}$. We will also write $H(u,\mathcal{X},d)$ for $\log N(u,\mathcal{X},d)$. In Appendix 3 we demonstrate that

$$\int_0^\delta H^{1/2}(u,\mathcal{F}_j(\delta),||\cdot||_n)du \lesssim q_n^{1/2}\delta, \tag{20}$$

which, by a maximal inequality for weighted sums of subgaussian variables, e.g. Corollary 8.3 of [2], implies $P(\sup_{f\in\mathcal{F}_j(1)}(\varepsilon,f)_n > C_1 r_n) \lesssim \exp(-c_2^2C_1^2nr_n^2)$ for some universal constants C_1 and c_2. Moreover, c_2 depends only on the distribution of the ε_i's, and the bound holds for all j and n, provided C_1 is above a certain universal threshold. Hence,

$$\sum_{j\in\mathfrak{M}_0^c} P\left((\varepsilon,\widehat{f}_j-f_j^*)_n > C_1r_n\|\widehat{f}_j-f_j^*\|_n\right) \lesssim p_n\exp\left(-c_2^2C_1^2nr_n^2\right). \tag{21}$$

Now consider an index $j\in\mathfrak{M}_0$. We will apply a peeling argument and intersect the set $A=\{(\varepsilon,\widehat{f}_j-f_j^*)_n > C_1r_n^2+C_1r_n\|\widehat{f}_j-f_j^*\|_n\}$ with the sets $B_0=\{\|\widehat{f}_j-f_j^*\|_n\le r_n\}$, $B_s=\{2^{s-1}r_n<\|\widehat{f}_j-f_j^*\|_n\le 2^s r_n\}$, where $s=1,2,...,S$, and $B_{S+1}=\{\tau/2<\|\widehat{f}_j-f_j^*\|_n\}$. Here τ is the constant from Condition 4(B) and $S=\lfloor\log_2(\tau r_n^{-1})\rfloor$, which guarantees $\tau/2\le 2^S r_n\le\tau$. Note that there exists a universal constant $\tilde{C}$, such that $\|f_j^*\|_n\le\tilde{C}$ for all j and n. Take $\tilde{c}=1+2\tilde{C}/\tau$. On the event B_{S+1}, we have $\|\widehat{f}_j\|_n/\|\widehat{f}_j-f_j^*\|_n\le\tilde{c}$ and $\|f_j^*\|_n/\|\widehat{f}_j-f_j^*\|_n\le\tilde{c}$ for all j and n. Note that $P(A)\le\sum_{s=0}^{S+1}P(AB_s)$, and, consequently,

$$\begin{aligned} P(A) &\le P\Big(\sup_{g\in\mathcal{G}_j(r_n)}(\varepsilon,g)_n > C_1r_n^2\Big)+\sum_{s=1}^S P\Big(\sup_{g\in\mathcal{G}_j(2^sr_n)}(\varepsilon,g)_n > C_1(2^{s-1}r_n)r_n\Big)\\ &\quad +P\Big(\sup_{\tilde{g}\in\tilde{\mathcal{G}}_j(\tilde{c})}(\varepsilon,\tilde{g})_n > C_1r_n\Big),\end{aligned}$$

where $\mathcal{G}_j(\delta)=\{g=f-f_j^*,\ \|g\|_n\le\delta,\ f\in\mathcal{F}_j^0\}$ and $\tilde{\mathcal{G}}_j(\tilde{c})=\mathcal{F}_j(\tilde{c})\ominus\mathcal{F}_j(\tilde{c})$. Arguing as in Appendix 3, while taking advantage of Condition 4(B), we can derive $\int_0^\delta H^{1/2}(u,\mathcal{G}_j(\delta),||\cdot||_n)du\lesssim q_n^{1/2}\delta$, for $\delta\le\tau$. Using Corollary 8.3 of [2] again we derive $P(\sup_{g\in\mathcal{G}_j(\delta)}(\varepsilon,g)_n > C_1(\delta/2)r_n)\lesssim\exp(-c_3^2C_1^2nr_n^2)$, where c_3 is half the constant c_2, introduced earlier, provided C_1 is above a certain universal threshold. Thus,

$$\begin{aligned} P\big(\sup_{g\in\mathcal{G}_j(r_n)}(\varepsilon,g)_n > C_1r_n^2\big)+\sum_{s=1}^S P\big(\sup_{g\in\mathcal{G}_j(2^sr_n)}(\varepsilon,g)_n > C_12^{s-1}r_n^2\big)&\\ \lesssim \log n\exp(-c_3^2C_1^2nr_n^2).&\end{aligned}$$

Similar arguments lead to $P(\sup_{\tilde{g}\in\tilde{\mathcal{G}}_j(\tilde{c})}(\varepsilon,\tilde{g})_n > C_1 r_n) \lesssim \exp(-c_4^2 C_1^2 n r_n^2)$, where $c_4 = c_2/(2\tilde{c})$. Consequently, $P(A) \lesssim \log n \exp(-c_5^2 C_1^2 n r_n^2)$, where $c_5 = \min(c_3, c_4)$. It follows from bounds (19) and (21) that

$$P\left((\varepsilon, \widehat{f} - f^*)_n > C_1 s_n r_n^2 + C_1 r_n \sum_{j=1}^{p_n} \|\widehat{f}_j - f_j^*\|_n\right) \lesssim p_n \log n \exp(-c_5^2 C_1^2 n r_n^2),$$

provided C_1 is above a universal threshold. The right-hand side of the above bound tends to zero by the assumption on the rate of growth for d_n, provided $C_1^2 > 2c_5^{-2}$.

3. Proof of inequality (20). For each given j and $\boldsymbol{\eta}_j$, we will write $H_{\boldsymbol{\eta}_j,j}(\cdot)$ for the d_n-dimensional row vector valued function $\mathbf{h}_{\boldsymbol{\eta}_j,j}(\boldsymbol{\eta}_j^T\cdot)$. Note that $||H_{\boldsymbol{\eta}_2,j}\boldsymbol{\xi}_2 - H_{\boldsymbol{\eta}_1,j}\boldsymbol{\xi}_1||_n \leq ||H_{\boldsymbol{\eta}_2,j}(\boldsymbol{\xi}_2 - \boldsymbol{\xi}_1)||_n + ||H_{\boldsymbol{\eta}_2,j}\boldsymbol{\xi}_1 - H_{\boldsymbol{\eta}_1,j}\boldsymbol{\xi}_1||_n$. Thus,

$$H(u, \mathcal{F}_j(\delta), ||\cdot||_n) \lesssim H_1(u/2) + H_2(u/2), \tag{22}$$

where $\exp[H_1(u)]$ is the size of the grid of $\boldsymbol{\xi}_1$ values, for which $||H_{\boldsymbol{\eta}_2,j}(\boldsymbol{\xi}_2 - \boldsymbol{\xi}_1)||_n \leq u$ can be guaranteed for all $\boldsymbol{\xi}_2$ and $\boldsymbol{\eta}_2$ with $\|\boldsymbol{\eta}_2\| = 1$ by choosing the appropriate grid point, while $\exp[H_2(u)]$ is the size of the grid of $\boldsymbol{\eta}_1$ values, for which $||H_{\boldsymbol{\eta}_2,j}\boldsymbol{\xi}_1 - H_{\boldsymbol{\eta}_1,j}\boldsymbol{\xi}_1||_n \leq u$ can be ensured all $\boldsymbol{\xi}_1$ and $\boldsymbol{\eta}_2$ with $\|\boldsymbol{\eta}_2\| = 1$.

First consider H_1. Note the general inequalities $d_n^{-1/2}\|\boldsymbol{\xi}\| \lesssim \|H_{\boldsymbol{\eta},j}\boldsymbol{\xi}\|_n \lesssim d_n^{-1/2}\|\boldsymbol{\xi}\|$, which follow from Condition 3(E) and Lemma 6.1 in [3]. Using these bounds, Corollary 2.6 of [2] implies $H_1(u/2) \lesssim d_n[1 + \log(\delta/u)]$.

Now consider H_2. Note that $\mathbf{h}_{\boldsymbol{\eta}_2}(\boldsymbol{\eta}_2^T\cdot) = \mathbf{h}_{\boldsymbol{\eta}_1}(a + b\boldsymbol{\eta}_2^T\cdot)$, where $\max(|a|, |b - 1|) \lesssim \max_i |(\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1)^T\boldsymbol{\theta}_i|$. Let $g = \mathbf{h}_{\boldsymbol{\eta}_1}\boldsymbol{\xi}_1$, and note that $|g(z_2) - g(z_1)| \lesssim d_n^{3/2}\delta|z_2 - z_1|$ by the properties of the cubic B-spline derivatives. Consequently,

$$||H_{\boldsymbol{\eta}_2,j}\boldsymbol{\xi}_1 - H_{\boldsymbol{\eta}_1,j}\boldsymbol{\xi}_1||_n = ||g(a + b\boldsymbol{\eta}_2^T\cdot) - g(\boldsymbol{\eta}_1^T\cdot)||_n \lesssim d_n^{3/2}\delta \max_{i\leq n}|(\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1)^T\boldsymbol{\theta}_i|. \tag{23}$$

Write Δ_k for the k-th element of $\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1$ and note that the right-hand side of the above inequality is written as $d_n^{3/2}\delta \max_{i\leq n}|\sum_{k=1}^{q_n}\Delta_k\theta_{ik}|$. Observe that

$$\max_{i\leq n}|\sum_{k=1}^{q_n}\Delta_k\theta_{ik}| \leq \max_{i\leq n}\Big(\sum_{k=1}^{q_n}\Delta_k^2 k^{-4}\Big)^{1/2}\Big(\sum_{k=1}^{q_n}\theta_{ik}^2 k^4\Big)^{1/2} \lesssim \Big(\sum_{k=1}^{q_n}\Delta_k^2 k^{-4}\Big)^{1/2},$$

where the last inequality holds by Condition 3(A). It follows from (23) that

$$||H_{\boldsymbol{\eta}_2,j}\boldsymbol{\xi}_1 - H_{\boldsymbol{\eta}_1,j}\boldsymbol{\xi}_1||_n \lesssim d_n^{3/2}\delta q_n^{1/2}\max_{k\leq d_n}|\Delta_k|k^{-2}. \tag{24}$$

Construct the $\boldsymbol{\eta}_1$ grid by selecting the locations for the k-th coordinate from a uniform grid with step u on $[0, d_n^{3/2}\delta q_n^{1/2}k^{-2}]$. Then, for each $\boldsymbol{\eta}_2$ and $\boldsymbol{\xi}_1$, we can find a grid point $\boldsymbol{\eta}_1$ for which the right-hand side of (24) is bounded by u. The total number of the corresponding grid points is bounded by a constant factor of

$$\prod_{k=1}^{q_n}(\delta d_n^{3/2}q_n^{1/2}k^{-2}/u) \lesssim (4\delta e^2/u)^{q_n}, \tag{25}$$

where the last inequality follows from Stirling's formula and $d_n \lesssim q_n$. Hence, $H_2(u/2) \lesssim q_n[1+\log(\delta/u)]$, and

$$\begin{aligned}\int_0^\delta H^{1/2}(u, \mathcal{F}_j(\delta), ||\cdot||_n)du &\leq \int_0^\delta [H_1^{1/2}(u/2) + H_2^{1/2}(u/2)]du \\ &\lesssim q_n^{1/2}\left(\delta + \delta\int_0^1 \log^{1/2}(1/v)dv\right) \lesssim q_n^{1/2}\delta.\end{aligned}$$

References.

[1] Inglot, T. (2010). Inequalities for quantiles of the chi-square distribution. *Probability and Mathematical Statistics* **30**, 339–351.

[2] van de Geer, S. (1999). *Empirical Processes in M-Estimation.* Cambridge University Press.

[3] Zhou, S., Shen, X. and Wolfe, D. (1998). Local asymptotics for regression splines and confidence regions. *The Annals of Statistics* **26**, 1760–1782.

Data Sciences and Operations Department
Marshall School of Business
University of Southern California
Los Angeles, CA 90089
USA
E-mail: fanyingy@marshall.usc.edu
gareth@marshall.usc.edu
radchenk@marshall.usc.edu