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 $\eta = (\eta_1^T, \cdots, \eta_p^T)^T$ be a $(pqn)$-vector and $\Theta = (\Theta_1, \cdots, \Theta_p)$ be an $n \times (pqn)$ matrix. With matrix notation, the linear FAR criterion minimizes the following objective function

\begin{equation}
Q(\eta) = \frac{1}{2n} \|Y - \Theta \eta\|^2 + \sum_{j=1}^{p} \rho_{\lambda_n} \left( \frac{1}{\sqrt{n}} \|\Theta_j \eta_j\| \right).
\end{equation}

Define the $(q_n s_n)$-dimensional hypercube

\begin{equation}
N = \{ \eta \in \mathbb{R}^{pqn} : \eta_{2\mathbb{N}} = 0, \|\eta - \eta_0\|_{\infty} \leq \sqrt{c_0 q_n^{-1/2} n^{-\alpha}} \},
\end{equation}

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

**Lemma 1.1.** Define the event $E_1 = \{ \|\Theta_M^{T,\ast} \rho_{\ast}\|_{\infty} \leq n\lambda_n/2 \}$. Assume that $\lambda_n n^\alpha q_n \sqrt{s_n} \to 0$ with $\alpha$ defined in Condition 2(B), then under Condition 2 and conditional on event $E_1$, there exists a vector $\eta \in N$ such that $\eta_{2\mathbb{N}}$ is a solution to the following nonlinear equations

\begin{equation}
- \frac{1}{n} \Theta_{2\mathbb{N}} (Y - \Theta_{2\mathbb{N}} \eta_{2\mathbb{N}}) + v_{2\mathbb{N}}(\eta) = 0,
\end{equation}

where $v_{2\mathbb{N}}(\eta)$ is a vector obtained by stacking $v_k(\eta) = \rho'_{\lambda_n} \left( \frac{1}{\sqrt{n}} \|\Theta_k \eta_k\| \right) \frac{1}{\sqrt{n}} \|\Theta_k \eta_k\|$, $k \in 2\mathbb{N}$, one underneath another.

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

\begin{equation}
\frac{1}{\sqrt{n}} \max_{k \in 2\mathbb{N}} \|\Theta_k (\tilde{\eta}_k - \eta_{0,k})\| \leq c_0^{-1/2} \max_{k \in 2\mathbb{N}} \|\tilde{\eta}_k - \eta_{0,k}\| \\
\leq c_0^{-1/2} \sqrt{q_n} \max_{k \in 2\mathbb{N}} \|\tilde{\eta}_k - \eta_{0,k}\|_{\infty} \leq n^{-\alpha}.
\end{equation}

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

\begin{equation}
\|\Theta_k \tilde{\eta}_k\| \geq \|\Theta_k \eta_{0,k}\| - \|\Theta_k (\tilde{\eta}_k - \eta_{0,k})\| \geq \|\Theta_k \eta_{0,k}\| - n^{1/2 - \alpha} > \sqrt{n a_n}/2.
\end{equation}
Thus, by Condition 2(A), for any $k \in \mathcal{M}_0$, $\rho'_n \left( \frac{1}{\sqrt{n}} \| \Theta_k \tilde{\eta} \| \right) \leq \rho'_n \left( a_n/2 \right)$. Hence, by the definition of $v$ and Condition 2(D) we obtain that for any $\tilde{\eta} \in \mathcal{N}$,

$$\|v_{\mathcal{M}_0}(\tilde{\eta}_k)\|_{\infty} \leq \max_{k \in \mathcal{M}_0} \rho'_n \left( \frac{1}{\sqrt{n}} \| \Theta_k \tilde{\eta} \| \right) \max_{k \in \mathcal{M}_0} \frac{1}{\sqrt{n}} \| \Theta_k^T \Theta_k \tilde{\eta} \| \leq \rho'_n \left( a_n/2 \right) \frac{1}{\sqrt{c_0}}.$$

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

$$\| (\Theta^T_{\mathcal{M}_0} \Theta_{\mathcal{M}_0})^{-1} \|_{\infty} \leq s_n q_n \lambda_{\max} \left( (\Theta^T_{\mathcal{M}_0} \Theta_{\mathcal{M}_0})^{-1} \right) \leq c_0^{-1} n^{-1} \sqrt{s_n q_n}.$$

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

$$n \| (\Theta^T_{\mathcal{M}_0} \Theta_{\mathcal{M}_0})^{-1} v_{\mathcal{M}_0}(\tilde{\eta}_k) \|_{\infty} \leq n \| (\Theta^T_{\mathcal{M}_0} \Theta_{\mathcal{M}_0})^{-1} \|_{\infty} \| v_{\mathcal{M}_0}(\tilde{\eta}_k) \|_{\infty} \leq o \left( n^{-\alpha} q_n^{-1/2} \right).$$

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

$$\| (\Theta^T_{\mathcal{M}_0} \Theta_{\mathcal{M}_0})^{-1} \Theta^T_{\mathcal{M}_0} \epsilon^* \|_{\infty} \leq \| (\Theta^T_{\mathcal{M}_0} \Theta_{\mathcal{M}_0})^{-1} \|_{\infty} \| \Theta^T_{\mathcal{M}_0} \epsilon^* \|_{\infty} \leq o \left( n^{-\alpha} q_n^{-1/2} \right).$$

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

$$\| (\Theta^T_{\mathcal{M}_0} \Theta_{\mathcal{M}_0})^{-1} (n v_{\mathcal{M}_0}(\tilde{\eta}_k) - \Theta^T_{\mathcal{M}_0} \epsilon^*) \|_{\infty} \leq o \left( q_n^{-1/2} n^{-\alpha} \right).$$

Define the vector-valued continuous function $g : R^{s_n q_n} \to R^{s_n q_n}$ by $g(x) = \eta_{0,\mathcal{M}_0} - (\Theta^T_{\mathcal{M}_0} \Theta_{\mathcal{M}_0})^{-1} \left( n v_{\mathcal{M}_0}(x) - \Theta^T_{\mathcal{M}_0} \epsilon^* \right)$, where $x = (x_1^T, \ldots, x_n^T)^T$ with $x_k \in R^{d_n}$ for $k = 1, \ldots, s_n$, and $v_{\mathcal{M}_0}(x)$ is a vector obtained by stacking the vectors $v_k(x_k) = \rho'_n \left( \frac{1}{\sqrt{n}} \| \Theta_k x_k \| \right) \frac{1}{\sqrt{n}} \Theta_k^T \Theta_k x_k$, $k = 1, \ldots, s_n$ one underneath another. Then for any $x \in \mathcal{N}$, by (7) we have

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

for large enough $n$. The above inequality indicates that $g(\mathcal{N}) \subset \mathcal{N}$. Since $g(x)$ 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 $\mathcal{E}_2 = \{ \| \Theta^T_{\mathcal{M}_0} \epsilon^* \|_{\infty} \leq n \lambda_n/2 \}$. 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 $\alpha$ 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 $\tilde{\eta}$ of $Q(\eta)$ (1) such that $\tilde{\eta} \in \mathcal{N}$. 

\[ \square \]
Proof. Since $\lambda_n$ satisfying conditions in Lemma 1.2 also satisfies conditions in Lemma 1.1, by Lemma 1.1, we know that there exists a vector $\hat{\eta} \in \mathcal{N}$ such that $\hat{\eta}_{M_0}$ is a solution to (2). We next show that under some additional conditions, $\hat{\eta}$ is a local minimizer of $Q(\eta)$ in the original $R^{pM_0}$ space.

We first constraint the objective function $Q(\eta)$ to the $(q_n,s_n)$-dimensional subspace $\mathcal{N}$ defined in (2). We will show that under Condition 2 and conditional on $\mathcal{E}_1 \cap \mathcal{E}_2$, $Q(\eta)$ is strictly convex around $\hat{\eta}$. Then this together with Lemma 1.1 entails that the critical value $\hat{\eta}_{M_0}$ minimizes $Q(\eta)$ in the subspace $\mathcal{N}$.

We proceed to prove the strict convexity of $Q(\eta)$ in $\mathcal{N}$. Define $h(\eta) = \sum_{j=1}^p \rho_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_j\eta_j\|)$, which is a function in $R^{pM_0}$. Note that for each $k \in M_0$,

$$
\frac{\partial^2}{\partial \eta_k^2} h(\eta) = \Theta_k^T \Theta_k + \rho''_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_k\eta_k\|) \frac{\rho'_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_k\eta_k\|)}{\sqrt{n}} \Theta_k^T \Theta_k \eta_k - \rho''_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_k\eta_k\|) \frac{\rho'_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_k\eta_k\|)}{\sqrt{n}^3} \eta_k^T \Theta_k^T \Theta_k \eta_k.
$$

Since $\hat{\eta} \in \mathcal{N}$, similar to (5) we can show that $\|\Theta_k \hat{\eta}_k\| \geq \|\Theta_k \eta_{k,0}\| - \|\Theta_k(\hat{\eta}_k - \eta_{k,0})\| > \sqrt{n}a_n / 2$ for any $k \in M_0$ and large enough $n$. Thus it follows from Condition 2 (A), (B) and (C) that

$$
0 < \frac{\rho'_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_k\eta_k\|)}{\|\Theta_k \hat{\eta}_k\|/\sqrt{n}} \leq \frac{\rho'_{\lambda_n}(a_n/2)}{a_n/2} = o(1),
$$

$$
\rho''_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_k\hat{\eta}_k\|) = o(1),
$$

where the $o(\cdot)$ terms are uniformly over all $k \in M_0$. 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), $\rho''_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_k\hat{\eta}_k\|) < 0$ and $\rho'_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_k\hat{\eta}_k\|) > 0$.

These together with (8) and Condition 2(D) entail that uniformly over all $k \in M_0$,

$$
\Lambda_{\min}(\frac{\partial^2}{\partial \eta_k^2} h(\eta)) \geq \Lambda_{\min}(\Theta_k^T \Theta_k) \frac{\rho''_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_k\eta_k\|)}{\sqrt{n}} \|\Theta_k \hat{\eta}_k\| + \Lambda_{\max}(\Theta_k^T \Theta_k \eta_k^T \Theta_k \Theta_k) \left(\frac{\rho''_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_k\eta_k\|)^2}{\sqrt{n}} - \frac{\rho'_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_k\eta_k\|)}{\sqrt{n}^3} \|\Theta_k \eta_k\|\right)
$$

$$
\geq \Lambda_{\max}(\frac{1}{n} \Theta_k^T \Theta_k)(\rho''_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_k\eta_k\|) - \frac{\rho'_{\lambda_n}(\frac{1}{\sqrt{n}}\|\Theta_k\eta_k\|)}{\|\Theta_k \eta_k\|/\sqrt{n}}) = o(1),
$$

FUNCTIONAL ADDITIVE REGRESSION
where for the second inequality we used the fact that
\[
\Lambda_{\max}(\Theta_k^T \Theta_k \hat{\eta} \hat{\eta}_k^T \Theta_k^T \Theta_k) = \Lambda_{\max}(\hat{\eta}_k^T \Theta_k^T \Theta_k \hat{\eta} \hat{\eta}_k) \leq \Lambda_{\max}(\Theta_k^T \Theta_k)\|\Theta_k \hat{\eta}_k\|^2.
\]
Let \( H \) be a block diagonal matrix with block matrices \( \frac{\partial^2}{\partial \eta_k^2} h(\hat{\eta}), \ k \in \mathcal{M}_0. \)
Then it is easy to see that the Hessian matrix \( \frac{\partial^2}{\partial \eta_{2\mathcal{M}_0}^2} Q(\hat{\eta}) = n^{-1} \Theta_{2\mathcal{M}_0}^T \Theta_{2\mathcal{M}_0} + H. \)
Thus, it follows from the above inequality (9) that
\[
(10) \quad \Lambda_{\min}(\frac{\partial^2}{\partial \eta_{2\mathcal{M}_0}^2} Q(\hat{\eta})) \geq \frac{1}{n} \Lambda_{\min}(\Theta_{2\mathcal{M}_0}^T \Theta_{2\mathcal{M}_0}) + \min_{k \in \mathcal{M}_0} \Lambda_{\min}(\frac{\partial^2}{\partial \eta_k^2} h(\hat{\eta})) \geq c_0 - o(1).
\]
Therefore, for large enough \( n \), restricted on the space \( \mathcal{N} \), the function \( Q(\eta) \) is strictly convex around \( \hat{\eta} \) and thus has a unique minimizer in a ball \( \mathcal{N}_1 \subset \mathcal{N} \) centered at \( \hat{\eta} \). Since by Lemma 1.1 \( \hat{\eta} \) is a critical point, \( \hat{\eta} \) is indeed this strict local minimizer in \( \mathcal{N}_1 \).

We next show that \( \hat{\eta} \) is also a local minimizer in the original \( \mathcal{R}^{pmn} \)-dimensional space. We will first show that for \( \hat{\eta}_{2\mathcal{M}_0} \) defined in Lemma 1.1, conditional on \( \mathcal{E}_1 \cap \mathcal{E}_2 \),
\[
(11) \quad \max_{j \in \mathcal{M}_0} \{v_j^T (\Theta_j^T \Theta_j)^{-1} v_j\}^{1/2} = \max_{j \in \mathcal{M}_0} \|\Theta_j (\Theta_j^T \Theta_j)^{-1} v_j\| < n^{-1/2} \rho'_{\lambda_n}(0+), \forall j \in \mathcal{M}_0,
\]
where
\[
v_j = n^{-1} \Theta_j^T (Y - \Theta_{2\mathcal{M}_0} \hat{\eta}_{2\mathcal{M}_0}) = n^{-1} \Theta_j^T \Theta_{2\mathcal{M}_0} (\eta_{0,2\mathcal{M}_0} - \hat{\eta}_{2\mathcal{M}_0}) + n^{-1} \Theta_j^T \epsilon^*.
\]
By Lemma 1.1, we have \( \eta_{0,2\mathcal{M}_0} = \hat{\eta}_{2\mathcal{M}_0} = (\Theta_{2\mathcal{M}_0}^T \Theta_{2\mathcal{M}_0})^{-1} (n\nu_{2\mathcal{M}_0} - \Theta_{2\mathcal{M}_0}^T \epsilon^*). \) Plugging this into \( v_j \), we obtain that for \( j \in \mathcal{M}_0, v_j = \Theta_j \Theta_{2\mathcal{M}_0} (\Theta_{2\mathcal{M}_0}^T \Theta_{2\mathcal{M}_0})^{-1} \nu_{2\mathcal{M}_0} + n^{-1}[\Theta_j - \Theta_j \Theta_{2\mathcal{M}_0} (\Theta_{2\mathcal{M}_0}^T \Theta_{2\mathcal{M}_0})^{-1} \Theta_j^T] \epsilon^*. \) Therefore,
\[
(12) \quad \{v_j^T (\Theta_j^T \Theta_j)^{-1} v_j\}^{1/2} = \|\Theta_j (\Theta_j^T \Theta_j)^{-1} v_j\| \leq I_{1,j} + I_{2,j},
\]
where
\[
I_{1,j} = \|\Theta_j (\Theta_j^T \Theta_j)^{-1} \Theta_j^T \Theta_{2\mathcal{M}_0} (\Theta_{2\mathcal{M}_0}^T \Theta_{2\mathcal{M}_0})^{-1} \nu_{2\mathcal{M}_0}\|,
I_{2,j} = n^{-1} \|\Theta_j (\Theta_j^T \Theta_j)^{-1} \Theta_j^T (I - \Theta_{2\mathcal{M}_0} (\Theta_{2\mathcal{M}_0}^T \Theta_{2\mathcal{M}_0})^{-1} \Theta_j^T) \epsilon^*\|.
\]
By (6), Condition 2(B) and Condition 2(D), conditional on \( \mathcal{E}_1 \cap \mathcal{E}_2 \), we have
\[
I_{1,j} \leq \|\nu_{2\mathcal{M}_0}\|_{\infty} \|\Theta_j (\Theta_j^T \Theta_j)^{-1} \Theta_j^T \Theta_{2\mathcal{M}_0} (\Theta_{2\mathcal{M}_0}^T \Theta_{2\mathcal{M}_0})^{-1}\|_{\infty,2} < \frac{1}{2\sqrt{n}} \rho'_{\lambda_n}(0+),
I_{2,j} \leq n^{-1} \|\Theta_j (\Theta_j^T \Theta_j)^{-1} \Theta_j^T (I - \Theta_{2\mathcal{M}_0} (\Theta_{2\mathcal{M}_0}^T \Theta_{2\mathcal{M}_0})^{-1} \Theta_j^T) \epsilon\|
+ n^{-1} \|\Theta_j (\Theta_j^T \Theta_j)^{-1} \Theta_j^T (I - \Theta_{2\mathcal{M}_0} (\Theta_{2\mathcal{M}_0}^T \Theta_{2\mathcal{M}_0})^{-1} \Theta_j^T) \epsilon\| \equiv I_{2,1,j} + I_{2,2,j},
\]
where the inequality for $I_{1,j}$ is uniformly over all $j \in \mathcal{M}_0$. Since both $\Theta_j(\Theta_j^T \Theta_j)^{-1} \Theta_j^T$ and $(I - \Theta_{2n_0} (\Theta_{2n_0}^T \Theta_{2n_0})^{-1} \Theta_{2n_0}^T)$ are projection matrices and $\varepsilon$ 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]),

$$P(\max_{j \in \mathcal{M}_0} I_{2,1,j} > n^{-1} \sqrt{q_n + C \log p})$$

$$= P(\max_{j \in \mathcal{M}_0} n^2 I_{2,1,j}^2 > (q_n + C \log p)) \leq C(p - s_n) \exp(-C \log p) \to 0,$$

where $C$ is a large enough generic positive constant. Thus, $\max_{j \in \mathcal{M}_0} I_{2,1,j} = o_p(n^{-1}(q_n^{1/2} + \sqrt{\log p}))$. Now by Condition 1 and assumption that $q_n^{-2} s_n = o(\lambda_n),\, \|e\|_\infty = o(\lambda_n).$ Thus, $\|e\|_2 = o(n^{1/2} \lambda_n)$. This together with $\Theta_j(\Theta_j^T \Theta_j)^{-1} \Theta_j^T$ and $(I - \Theta_{2n_0} (\Theta_{2n_0}^T \Theta_{2n_0})^{-1} \Theta_{2n_0}^T)$ being projection matrix ensures that uniformly over all $j \in \mathcal{M}_0$,

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

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

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

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

Let $\mathcal{B} = \{\eta \in \mathbb{R}^{p_n} : \eta_{2n_0} = 0\}$ be a subspace in $\mathbb{R}^{p_n}$. Take a sufficiently small ball $\mathcal{N}_2$ in $\mathbb{R}^{p_n}$ centered at $\hat{\eta}$ 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 $\hat{\eta} \in \mathcal{N}_2$, appropriately shrink the radius of the ball $\mathcal{N}_2$ gives that there exists a $\delta \in (0, \infty)$ such that for any $\eta \in \mathcal{N}_2$,

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

Fix an arbitrary $\eta_1 = (\eta_{1,1}^T, \cdots, \eta_{1,p'}^T)^T \in \mathcal{N}_2 \cap \mathcal{N}_1^c$, we next show that $Q(\eta_1) > Q(\hat{\eta})$. Let $\eta_2 = (\eta_{2,1}^T, \cdots, \eta_{2,p'}^T)^T$ be the projection of $\eta_1$ onto $\mathcal{B}$. Then it follows from the definitions of $\mathcal{N}_1, \mathcal{N}_2, \mathcal{B}$ and $\hat{\eta}$ that $Q(\eta_2) > Q(\hat{\eta})$. Thus we only need to show $Q(\eta_1) \geq Q(\eta_2)$.

Note that $Q(\eta_1) - Q(\eta_2) = \nabla Q(\eta_3)(\eta_1 - \eta_2) = \sum_{j \in \mathcal{M}_0} \eta_{1,j}^T \frac{\partial Q(\eta_3)}{\partial \eta_{1,j}}$, where $\eta_3$ is a vector on the segment connecting $\eta_1$ and $\eta_2$. Since $\eta_{2k} = 0$ for any $k \in \mathcal{M}_0$, there exits a constant $0 < \gamma < 1$ such that $\eta_{3k} = \gamma \eta_{1k}, k \in \mathcal{M}_0$. Then by the definitions of $\mathcal{B}, \mathcal{N}_1, \mathcal{N}_2$, we know that $\eta_3 \in \mathcal{N}_2$. Shrink the
ball $N_2$ such that for any $\eta \in N_2$, $\|\Theta_k \eta_k\| = \|\Theta_k (\eta_k - \hat{\eta}_k)\| \leq \sqrt{n}\delta$, $k \in \mathbb{M}_0$. Since $\eta_3 \in N_2$, we have $\|\Theta_k \eta_{3k}\| \leq \sqrt{n}\delta$ and thus $\rho'_\lambda (\sqrt{n}\|\Theta_k \eta_{3k}\|) \geq \rho'_\lambda (\delta)$ for $k \in \mathbb{M}_0$. Therefore,

$$Q(\eta_1) - Q(\eta_2) = \nabla Q(\eta_3)(\eta_1 - \eta_2) = \sum_{j \in \mathbb{M}_0} \eta_{1j}^T \frac{\partial Q(\eta_3)}{\partial \eta_j}$$

$$= \sum_{j \in \mathbb{M}_0} \eta_{1j}^T \left( -\frac{1}{n} \Theta_j^T (Y - \Theta \eta_3) + \frac{\rho'_\lambda_n (\frac{1}{\sqrt{n}}\|\Theta_j \eta_{3j}\|)}{\sqrt{n}\|\Theta_j \eta_{3j}\|} \Theta_j^T \Theta_j \eta_{3j} \right)$$

$$\geq -\frac{1}{n} \sum_{j \in \mathbb{M}_0} \eta_{1j}^T \Theta_j^T (Y - \Theta \eta_3) + \frac{1}{\sqrt{n}\gamma} \rho'_\lambda_n (\delta) \sum_{j \in \mathbb{M}_0} \|\Theta_j \eta_{3j}\| \equiv \|I_3 + I_4\|.$$ 

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

$$|I_3| \leq \frac{1}{n} \sum_{j \in \mathbb{M}_0} \|\Theta_j \eta_{1j}\| \|\Theta_j (\Theta_j^T \Theta_j)^{-1} \Theta_j^T (Y - \Theta \eta_3)\|$$

$$= \frac{1}{n\gamma} \sum_{j \in \mathbb{M}_0} \|\Theta_j \eta_{3j}\| \|\Theta_j (\Theta_j^T \Theta_j)^{-1} \Theta_j^T (Y - \Theta \eta_3)\| \leq I_4.$$ 

Thus, $Q(\eta_1) \geq Q(\eta_2)$, which together with $Q(\eta_3) > Q(\eta)$ ensures that $\hat{\eta}$ is also a strict local minimizer in the original $R^{pn}$ dimensional space. The proof is completed. \(\square\)

### 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

$$P(\mathcal{E}_1 \cap \mathcal{E}_2) = 1 - P(\|\Theta^T \varepsilon^*\|_{\infty} \geq n\lambda_n/2)$$

$$\geq 1 - P(\|\Theta^T \varepsilon\|_{\infty} \geq n\lambda_n/2 - \|\Theta^T e\|_{\infty}).$$

By the assumption that $s_n q_n^{-2} = o(\lambda_n)$, it is easy to derive that $\|e\|_{\infty} = o(\lambda_n)$. Since each column of $\Theta$ has $\ell_2$ norm $\sqrt{n}$, it follows that $\|\Theta\|_{1} \leq n$. Thus, by Cauchy-Schwartz inequality, $\|\Theta^T e\|_{\infty} \leq \|\Theta\|_{1} \|e\|_{\infty} \leq o(n\lambda_n)$). This follows that

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

for large enough $n$.

Now we consider $\|\Theta^T \varepsilon\|_{\infty}$. Let $\xi = (\xi_1, \cdots, \xi_p)^T = \Theta^T \varepsilon$, 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 ℓ₂ norm √n, we have $d_i^2 = 1$ for $1 \leq i \leq q_n p$. Hence, by Bonferroni’s inequality and the assumption $n \lambda^2_n (\log(pq_n))^{-1} \rightarrow \infty$ we further obtain

$$P(\|\Theta^T\varepsilon\|_\infty > n\lambda_n/4) \leq \sum_{i=1}^{q_n p} P(|\xi_i| > n\lambda_n/4)$$

$$\leq \frac{4\sigma pq_n}{\sqrt{2\pi n\lambda_n}} \exp\left(-n\lambda^2_n/(32\sigma^2)\right) \rightarrow 0.$$

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

**Proof of Theorem 2**

**Proof.** Let $\hat{\v}_2(\eta) = v_{2n}(\hat{\eta})$ and $v_{0,2n0} = v_{2n0}(\eta_0)$ with the function $v_{2n0}(\cdot)$ defined in Lemma 1.1, $\hat{\v}_{2n0}$ the solution to (3), and $\eta_0$ the true regression coefficient vector. Since $\eta_{2n0}$ is a solution to (3), for any vector $c \in \mathbb{R}^{nq_n}$ satisfying $c^T c = 1$, we have the following decomposition

$$c^T [(\Theta_{2n0}^T \Theta_{2n0})^{-1/2} (\hat{\v}_{2n0} - \eta_{0,2n0}) + n(\Theta_{2n0}^T \Theta_{2n0})^{-1/2} v_{0,2n0}]$$

$$= c^T (\Theta_{2n0}^T \Theta_{2n0})^{-1/2} \Theta_{2n0}^T \hat{\v}_{2n0}^T \varepsilon + c^T (\Theta_{2n0}^T \Theta_{2n0})^{-1/2} \Theta_{2n0}^T \theta$$

$$+ n c^T (\Theta_{2n0}^T \Theta_{2n0})^{-1/2} (\hat{\v}_{2n0} - v_{0,2n0}) \equiv I_1 + I_2 + I_3.$$

It is easy to see

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

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

$$|I_2| \leq \|c^T (\Theta_{2n0}^T \Theta_{2n0})^{-1/2} \Theta_{2n0}^T \varepsilon\| = \|\varepsilon\| = o(1).$$

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

$$|I_3| \leq \|\sqrt{n}c^T (\Theta_{2n0}^T \Theta_{2n0})^{-1/2} \sqrt{n}(\hat{\v}_{2n0} - v_{0,2n0})\|$$

$$\leq c_0^{-1/2} \|\sqrt{n}(\hat{\v}_{2n0} - v_{0,2n0})\|.$$

Define $g(\eta_k) = \frac{1}{\sqrt{n}} \rho_k^T (\frac{1}{\sqrt{n}} \|\phi_k \eta_k\|) \bar{\Theta}_{k \eta_k}^T \eta_k$. Then by definitions of $\hat{\v}_{2n0}$ and $v_{0,2n0}$,

$$\hat{\v}_k - v_{0,k} = g(\hat{\eta}_k) - g(\eta_{0,k}) = \frac{\partial}{\partial \eta_k} g(\eta_k)(\hat{\eta}_k - \eta_{0,k}).$$
with \( \hat{\eta}_k \) lying on the segment connecting \( \eta_{0,k} \) and \( \hat{\eta}_k \). Thus, \( \hat{\eta} = (\hat{\eta}_1^T, \ldots, \hat{\eta}_p^T)^T \in \mathcal{N} \). It has been proved in (5) that \( \|\Theta_k \eta_k\| \geq \sqrt{n}a_n/2 \) for any \( \eta \in \mathcal{N} \). Note that for any \( \eta = (\eta_1^T, \ldots, \eta_p^T)^T \in \mathcal{N} \), and any \( k \in \mathcal{M}_0 \),

\[
\frac{\partial}{\partial \eta_k} g(\eta_k) = \rho''_{\lambda n}(\frac{1}{\sqrt{n}}\|\Theta_k \eta_k\|) \frac{\Theta_k^T \Theta_k \eta_k \eta_k^T \Theta_k \Theta_k}{n\|\Theta_k \eta_k\|^2} + \rho'_{\lambda n}(\frac{1}{\sqrt{n}}\|\Theta_k \eta_k\|) \left\{ \frac{\Theta_k^T \Theta_k}{\|\Theta_k \eta_k\|} - \frac{\Theta_k^T \Theta_k \eta_k \eta_k^T \Theta_k \Theta_k}{\|\Theta_k \eta_k\|^3} \right\}.
\]

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

\[
c_0^{-1} \left( -O\left(\frac{1}{\sqrt{n}}\right) - \frac{2\rho'_{\lambda n}(\frac{a_n}{\sqrt{n}})}{a_n} \right) \leq \Lambda_{\min} \left( \frac{\partial}{\partial \eta_k} g(\eta_k) \right) \leq \Lambda_{\max} \left( \frac{\partial}{\partial \eta_k} g(\eta_k) \right) \leq c_0^{-1} \frac{2\rho'_{\lambda n}(\frac{a_n}{\sqrt{n}})}{a_n}.
\]

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

\[
\|\hat{v}_{0,0} - v_{0,0}\| \leq c_0^{-1} \left( O\left(\frac{1}{\sqrt{n}}\right) + \frac{2\rho'_{\lambda n}(\frac{a_n}{\sqrt{n}})}{a_n} \right) \left\{ \sum_{k \in \mathcal{M}_0} \|\hat{\eta}_k - \eta_{0,k}\|^2 \right\}^{1/2} \leq c_0^{-3/2} \left( O\left(\frac{1}{\sqrt{n}}\right) + o\left(\frac{a_n}{\sqrt{n}}\right) \right) O_p(s_n^{1/2}n^{-\alpha}) = O_p(n^{-1/2}),
\]

So it follows that \( \sqrt{n}\|\hat{v}_{0,0} - v_{0,0}\| = O_p(1) \). Combing this with (17) yields \( I_3 \overset{p}{\rightarrow} 0 \). This together with (14) –(16) completes the proof. \( \square \)

2. Proof of Lemma 1. Observe that

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

\[
\sum_{j \in \mathcal{M}_0} P \left( \frac{(\varepsilon, \hat{f}_j - f_j^*)_n}{r_n + \|\hat{f}_j - f_j^*\|_n} > C_1 r_n \right) + \sum_{j \in \mathcal{M}_0} P \left( (\varepsilon, \hat{f}_j - f_j^*)_n > C_1 r_n \|\hat{f}_j - f_j^*\|_n \right).
\]

Consider an index \( j \in \mathcal{M}_0 \), and note that \( f_j^* \equiv 0 \). We have,

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

where \( \mathcal{F}_j(\delta) \) is defined for every positive \( \delta \) 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

\begin{equation}
\int_0^\delta H^{1/2}(u, \mathcal{F}_j(\delta), \|n\|)du \lesssim q_n^{1/2}\delta,
\end{equation}

which, by a maximal inequality for weighted sums of subgaussian variables, e.g. Corollary 8.3 of [2], implies $P(\sup_{f \in \mathcal{F}_j(1)}(\varepsilon, f)_n > C_1 r_n) \lesssim \exp(-c_3^2 C_1^2 nr_n^2)$ for some universal constants $C_1$ and $c_2$. Moreover, $c_2$ depends only on the distribution of the $\varepsilon_i$'s, and the bound holds for all $j$ and $n$, provided $C_1$ is above a certain universal threshold. Hence,

\begin{equation}
\sum_{j \in \mathcal{M}_0} P\left((\varepsilon, \hat{f}_j - f_j^*)_n > C_1 r_n \|\hat{f}_j - f_j^*\|_n\right) \lesssim p_n \exp\left(-c_3^2 C_1^2 n r_n^2\right).
\end{equation}

Now consider an index $j \in \mathcal{M}_0$. We will apply a peeling argument and intersect the set $A = \{(\varepsilon, \hat{f}_j - f_j^*)_n > C_1 r_n^2 + C_1 r_n \|\hat{f}_j - f_j^*\|_n\}$ with the sets $B_0 = \{|\|\hat{f}_j - f_j^*\|_n \leq r_n\}$, $B_s = \{2^{s-1} r_n < \|\hat{f}_j - f_j^*\|_n \leq 2^s r_n\}$, where $s = 1, 2, ..., S$, and $B_{S+1} = \{\tau/2 < \|\hat{f}_j - f_j^*\|_n\}$. Here $\tau$ is the constant from Condition 4(B) and $S = \lfloor \log_2(\tau r_n^{-1}) \rfloor$, which guarantees $\tau/2 \leq 2^S r_n \leq \tau$. Note that there exists a universal constant $\tilde{C}$, such that $\|f_j^*\|_n \leq \tilde{C}$ for all $j$ and $n$. Take $\tilde{c} = 1 + 2\tilde{C}/\tau$. On the event $B_{S+1}$, we have $\|\hat{f}_j\|_n/\|\hat{f}_j - f_j^*\|_n \leq \tilde{c}$ and $\|f_j^*\|_n/\|\hat{f}_j - f_j^*\|_n \leq \tilde{c}$ for all $j$ and $n$. Note that $P(A) \leq \sum_{s=0}^{S+1} P(AB_s)$, and, consequently,

\begin{equation}
P(A) \leq P\left( \sup_{g \in \mathcal{G}_j(r_n)} (\varepsilon, g)_n > C_1 r_n^2 \right) + \sum_{s=1}^{S} P\left( \sup_{g \in \mathcal{G}_j(2^s r_n)} (\varepsilon, g)_n > C_1 (2^{s-1} r_n) r_n \right) + P\left( \sup_{g \in \mathcal{G}_j(\tilde{c})} (\varepsilon, g)_n > C_1 r_n \right),
\end{equation}

where $\mathcal{G}_j(\delta) = \{g = f - f_j^*, \|g\|_n \leq \delta, f \in \mathcal{F}_j\}$ and $\mathcal{G}_j(\tilde{c}) = \mathcal{F}_j(\tilde{c}) \cap \mathcal{F}_j(\tilde{c})$. 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 $\delta \leq \tau$. Using Corollary 8.3 of [2] again we derive $P(\sup_{g \in \mathcal{G}_j(\delta)} (\varepsilon, g)_n > C_1 (\delta/2) r_n) \lesssim \exp(-c_3^2 C_1^2 n r_n^2)$, where $c_3$ is half the constant $c_2$, introduced earlier, provided $C_1$ is above a certain universal threshold. Thus,

\begin{equation}
P\left( \sup_{g \in \mathcal{G}_j(r_n)} (\varepsilon, g)_n > C_1 r_n^2 \right) + \sum_{s=1}^{S} P\left( \sup_{g \in \mathcal{G}_j(2^s r_n)} (\varepsilon, g)_n > C_1 (2^{s-1} r_n^2) \right) \lesssim \log n \exp(-c_3^2 C_1^2 n r_n^2).
\end{equation}
Similar arguments lead to $P(\sup_{\tilde{g}\in\tilde{G}} (\tilde{\epsilon}, \tilde{g})_n > C_1 r_n) \lesssim \exp(-c_3^2 C_1^2 n r_n^2)$, where $c_3 = c_2/(2\tilde{c})$. Consequently, $P(A) \lesssim \log n \exp(-c_3^2 C_1^2 n r_n^2)$, where $c_5 = \min(c_3, c_4)$. It follows from bounds (19) and (21) that

$$
P \left( (\tilde{\epsilon}, \hat{f} - f^*)_n > C_1 s_n r_n^2 + C_1 r_n \sum_{j=1}^{p_n} \|\hat{f}_j - f_j^*\|_n \right) \lesssim p_n \log n \exp(-c_3^2 C_1^2 n r_n^2),$$

provided $C_1$ is above a universal threshold. The right-hand side of the above bound tends to zero by the rate of growth on the $d_n$, provided $C_1^2 > 2c_3^{-2}$.

3. Proof of inequality (20). For each given $j$ and $\eta_j$, we will write $H_{\eta_j,j}(\cdot)$ for the $d_n$-dimensional row vector valued function $h_{\eta_j,j}(\eta_j^T \cdot)$. Note that $\|H_{\eta_2,j}\xi_2 - H_{\eta_1,j}\xi_1\|_n \leq \|H_{\eta_2,j}(\xi_2 - \xi_1)\|_n + \|H_{\eta_2,j}\xi_1 - H_{\eta_1,j}\xi_1\|_n$. Thus,

$$
H(u, F_j(\delta), ||\cdot||_n) \lesssim H_1(u/2) + H_2(u/2),
$$

where $\exp[H_1(u)]$ is the size of the grid of $\xi_1$ values, for which $\|H_{\eta_2,j}(\xi_2 - \xi_1)\|_n \leq u$ can be guaranteed for all $\xi_2$ and $\eta_2$ with $\|\eta_2\| = 1$ by choosing the appropriate grid point, while $\exp[H_2(u)]$ is the size of the grid of $\eta_1$ values, for which $\|H_{\eta_2,j}\xi_1 - H_{\eta_1,j}\xi_1\|_n \leq u$ can be ensured all $\xi_1$ and $\eta_2$ with $\|\eta_2\| = 1$.

First consider $H_1$. Note the general inequalities $d_n^{-1/2} ||\xi|| \lesssim \|H_{\eta,j}\xi\|_n \lesssim d_n^{-1/2} ||\xi||$, 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 $h_{\eta_2,j}(\eta_j^T \cdot) = h_{\eta_1,j}(a + b\eta_j^T \cdot)$, where $\max(|a|, |b - 1|) \lesssim \max_i ||\eta_2 - \eta_1||_i \theta_i$. Let $g = h_{\eta_1,j}\xi_1$, 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,

$$
\|H_{\eta_2,j}\xi_1 - H_{\eta_1,j}\xi_1\|_n = ||g(a + b\eta_j^T \cdot) - g(\eta_j^T \cdot)||_n \lesssim d_n^{-3/2} \delta \max_{i \leq n} ||\eta_2 - \eta_1||_i \theta_i.
$$

Write $\Delta_k$ for the $k$-th element of $\eta_2 - \eta_1$ 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

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

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

$$
\|H_{\eta_2,j}\xi_1 - H_{\eta_1,j}\xi_1\|_n \lesssim d_n^{3/2} \delta q_n^{1/2} \max_{k \leq d_n} |\Delta_k| k^{-2}.
$$
Construct the \( \eta_1 \) 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 \( \eta_2 \) and \( \xi_1 \), we can find a grid point \( \eta_1 \) 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

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

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

\[
\int_0^\delta H^{1/2}(u, F_j(\delta), \| \cdot \|_{n}) du \leq \int_0^\delta [H_1^{1/2}(u/2) + H_2^{1/2}(u/2)] du \lesssim q_n^{1/2} \left( \delta + \delta \int_0^1 \log^{1/2}(1/v) dv \right) \lesssim q_n^{1/2} \delta.
\]

References.